Chapter 5

Simple random variables

Printed: February 19, 2019
This section is based on [Billingsley, Section 5].

Definition 5.1. A random variable $X$ is a simple random variable if it has a finite range.

If the range $X(\Omega)$ of $X$ is $\{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}$ (distinct real numbers), then

\begin{equation}
X = \sum_{j=1}^{n} x_j I_{A_j},
\end{equation}

where $A_j = X^{-1}(\{x_j\}) \in \mathcal{F}$. Note that if $x_j$ are distinct then $A_j$ are disjoint, and that $\bigcup_{j=1}^{n} A_j = \Omega$.

Theorem 5.1. Let $X_1, \ldots, X_n$ be simple random variables. A simple random variable $Y$ is $\sigma(X_1, \ldots, X_n)$-measurable if and only if there exists $f : \mathbb{R}^n \to \mathbb{R}$ such that $Y = f(X_1, \ldots, X_n)$.

Proof. If $Y = f(X_1, \ldots, X_n)$ then $Y^{-1}(\{y\}) = \{\omega : (X_1(\omega), \ldots, X_n(\omega)) \in f^{-1}(\{y\})\}$. Of course, $f^{-1}(\{y\})$ could be a non-measurable set. But its intersection with a finite set $F_1 \times F_2 \times \cdots \times F_n$ is measurable. So $Y^{-1}(\{y\})$ is an inverse image of a measurable set in a measurable mapping $(X_1, \ldots, X_n) : \Omega \to \mathbb{R}^n$ (compare Exercise 4.20).

Suppose now that $Y$ is $\sigma(X_1, \ldots, X_n)$. Denote by $y_1, \ldots, y_r$ its distinct values. Then there exists a set $U_i \subset \mathbb{R}^n$ such that

$\{\omega : Y(\omega) = y_i\} = \{\omega : (X_1(\omega), \ldots, X_n(\omega)) \in U_i\}$

Take $f = \sum y_j I_{U_j}$. (The sets $U_j$ are not disjoint, but their intersections with the range of $(X_1, \ldots, X_n)$ are disjoint.) \hfill \Box

The importance of simple random variables lies in their usefulness for approximations.

Theorem 5.2. If $X : \Omega \to [0, \infty)$ is a nonnegative random variable then there exist a sequence of simple random variables $X_1 \leq X_2 \leq X_3 \leq \cdots \leq X_n \leq \cdots$ such that $X(\omega) = \lim_{n \to \infty} X_n(\omega)$. 61
Remark 5.3. It is easy to produce good approximations on the sets of large probability,
\[ \sum_{k=1}^{n} \frac{k-1}{n} I_{\frac{k-1}{n} \leq X < \frac{k}{n}} \to X, \]
or discrete uniform approximations from below on entire \( \Omega \). For the latter, take
\[ X_n = \sum_{k=1}^{\infty} \frac{k-1}{n} I_{\frac{k-1}{2n} \leq X < \frac{k}{2n}}. \]
Then \( X - 1/n \leq X_n \leq X \).
For the proof, we want to use only finite number of values, work on all \( \Omega \), and be sure that the approximation is also increasing so that \( X_n \uparrow X \).

Proof. \(^1\) We find an appropriate function \( \varphi_n(x) \) and take as \( X_n \) the value of \( \varphi_n(X) \). Here is one such function:
\[ X_n := n I_{X \geq n/2^n} + \sum_{k=1}^{2^n} \frac{k-1}{2^n} I_{\frac{k-1}{2n} \leq X < \frac{k}{2n}} \uparrow X. \]
See Fig 1. \( \Box \)

1. Expected value
A simple random variable of the form (5.1) is assigned expected value
\[ \mathbb{E}[X] = \sum_{j=1}^{n} x_j P(A_j) \]
Remark 5.4. A special case of (5.3) is \( \mathbb{E}[I_A] = P(A) \).
In particular, if \( \mathbb{E}[I_A] = 0 \) then \( P(A) = 0 \).

Remark 5.5. Note that if \( \Omega = [0,1] \) and \( A_j \) are intervals, then \( E(X) = \int_0^1 X(\omega) d\omega \), defined as the Riemann integral.

Example 5.1. Suppose \( X \) is Binomial \( Bin(n,p) \), see Example 1.5 on page 13. Then \( x_j = j \) and \( P(A_j) = \binom{n}{j} p^j (1-p)^{n-j} \) with \( j = 0,1,\ldots,n \), so
\[
\mathbb{E}[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!} p^k (1-p)^{n-k} = np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\
= np (p+1-p)^{n-1} = np. \]

---

\(^1\)This proof is repeated at the beginning of next chapter!
It is clear that if $X$ is simple and $f : \mathbb{R} \to \mathbb{R}$ is an arbitrary function then $Y = f(X)$ is simple, and that
\[(5.4) \quad \mathbb{E}[Y] = \sum_j f(x_j)P(A_j)\]
In particular, the moments of $X$ are
\[m_k = \mathbb{E}[X^k] = \sum_j x_j^k P(A_j).\]

Proposition 5.6. For simple random variables the expected value has the following properties:
- **linearity:**
  \[(5.5) \quad \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]\]
  and more generally, $\mathbb{E}[aX + Y] = a\mathbb{E}[X] + \mathbb{E}[Y]$.
- **monotonicity:** if $X \leq Y$, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

(5.6) \quad |\mathbb{E}[X]| \leq \mathbb{E}|X|.

This implies $|\mathbb{E}[X - Y]| \leq \mathbb{E}|X - Y|$.

**Proof.** If $X = \sum_j x_j I_{A_j}$ and $Y = \sum_k y_k I_{B_k}$ then $X + Y = \sum_{j,k} (x_j + y_k) I_{A_j \cap B_k}$. Thus $E(X + Y) = \sum_{j,k} (x_j + y_k) P(A_j \cap B_k) = \sum_j x_j \sum_k P(A_j \cap B_k) + \sum_k y_k \sum_j P(A_j \cap B_k)$. This gives linearity (5.5).

Expected value also preserves order: if $X \geq 0$ for all $\omega$ then $E(X) \geq 0$. Thus if $X \leq Y$ (i.e. $Y - X \geq 0$) then $E(X) \leq E(Y)$.

Since $X \leq |X|$, this gives $E(X) \leq E|X|$. Since $-X$ satisfies this, too, we get (5.6).

**Definition 5.2.** We will say that $\{X_n\}$ is uniformly bounded if there exists a real number $K$ such that $|X_n(\omega)| \leq K$ for all $\omega \in \Omega$ and all $n$.

We say that $\{X_n\}$ is uniformly bounded with probability one if there exists a real number $K$ such that $P(|X_n(\omega)| \leq K) = 1$ for all $n$.

We say that $\{X_n\}$ is stochastically bounded, if for every $\varepsilon > 0$ there exists $K \in \mathbb{R}$ such that $P(|X_n| > K) < \varepsilon$ for all $n$.

**Example 5.2.** Here are some examples that illustrate the definition

(i) A sequence $X_n = \pm 1$ is of course uniformly bounded.

(ii) If $P(A_n) = 0$ and $A_n \neq \emptyset$ and $B \in \mathcal{F}$ then random variables $X_n = nI_{A_n} + I_{B \cap A_n} - I_{A_n \cap B^c}$ are uniformly bounded with probability one, but not uniformly bounded, as the values of $X_n$ are generically $\{-1, 1, n\}$.

(iii) A sequence of, say independent, Poisson random variables $X_n$ with the same distribution is stochastically bounded but not bounded.

**Theorem 5.7.** Suppose that $X_n, X$ are simple random variables such that $X_n \xrightarrow{P} X$. If $\{X_n\}$ is uniformly bounded, then $\mathbb{E}[X] = \lim_{n \to \infty} \mathbb{E}[X_n]$. 

5. Simple random variables

Proof. Suppose $|X_n| \leq K$. Since $X$ is simple, we can increase $K$ to ensure also $|X| \leq K$.

If $A_n = \{ \omega : |X - X_n| \geq \varepsilon \}$ then

$$|X(\omega) - X_n(\omega)| \leq 2KI_{A_n} + \varepsilon I_{A_n^c}$$

Thus $E|X - X_n| \leq 2KP(|X_n - X| \geq \varepsilon) + \varepsilon \to \varepsilon$. Inequality (5.6) ends the proof. \qed

Example 5.3. Suppose $P(X_n = 0) = (n - 1)/n$ and $P(X_n = (-1)^n/n) = 1/n$. Then $X_n \xrightarrow{P} 0$ but $E(X_n) = (-1)^n$ does not converge. This contradicts Theorem 5.7, doesn’t it?

Remark 5.8. Suppose $X \geq 0$ is arbitrary, and $X_n \uparrow X$ are simple random variables from Theorem 5.1. Then $E(X_n) \leq E(X_{n+1})$ so $\lim_{n \to \infty} E(X_n)$ exists, perhaps as $\infty$. Furthermore, if $X$ is simple, then by Theorem 5.7, $\lim_{n \to \infty} E(X_n)$ is just $E(X)$. This suggests that we can try to define $E(X)$ by this limit. (It would be nice to know that any other sequence $X_n \uparrow X$ will give the same answer!)

It is tempting to compute by this technique an answer that we know from somewhere else. But anything more complicated than Exercises 5.5 or 5.6 seems to require too much work.

Definition 5.3. The variance of a simple random variable $X$ is

$$(5.7) \quad \text{Var}(X) = E(X - m)^2 = E(X^2) - m^2$$

where $m = E(X)$.

The mean and variance of a linear transformation $Y = aX + B$ of $X$ are $E(Y) = aE(X) + b$, $\text{Var}(Y) = a^2 \text{Var}(X)$.

Somewhat more generally,

$$\text{Var}(aX + bY + c) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X,Y)$$

where, trivially,

$$\text{Cov}(X,Y) = \frac{1}{2} \left( \text{Var}(X + Y) - \text{Var}(X) - \text{Var}(Y) \right)$$

This gives

$$\text{Cov}(X,Y) = E((X - m_X)(Y - m_Y)) = E(XY) - EXEY$$

2. Expected values and independence

If $X_1, \ldots, X_n$ are independent then

$$(5.8) \quad E(X_1X_2 \ldots X_n) = E(X_1)E(X_2) \ldots E(X_n)$$

It is enough to verify this for two independent random variables. If $X = \sum_j x_j I_{A_j}$ and $Y = \sum_k y_k I_{B_k}$ then $XY = \sum_{j,k} x_j y_k I_{A_j \cap B_k}$. Thus $E(XY) = \sum_{j,k} x_j y_k P(A_j \cap B_k) = \sum_j x_j P(A_j) \sum_k y_k P(B_k)$. Thus $E(X_1X_2 \ldots X_n) = E(X_1)E(X_2) \ldots E(X_n)$ and inductively we can pull one factor at a time.

For independent $X_1, \ldots, X_n$ we have

$$\text{Var}(X_1 + \cdots + X_n) = \text{Var}(X_1) + \cdots + \text{Var}(X_n)$$

Again, we verify this for the sum of two independent variables $X, Y$. Replacing $X$ by $X - m$ if needed, without loss of generality we may assume $E(X) = E(Y) = 0$. Then $\text{Var}(X + Y) = E(X + Y)^2 = E(X^2 + Y^2 + 2XY) = E(X^2) + E(Y^2) = \text{Var}(X) + \text{Var}(Y)$. 

2.1. Tail integration formula. If $X \geq 0$ then

\begin{equation}
E(X) = \int_0^\infty P(X > x) \, dx = \int_0^\infty P(X \geq x) \, dx
\end{equation}

Proof. For simple random variables this is just a picture. □

Remark 5.9. Remark 5.8 suggested a definition of $E(X)$ that was hard to use in specific examples. Here is another approach. The function $x \mapsto P(X \geq x)$ is decreasing, so it is Riemann-integrable, \( \int_0^T P(X \geq x) \, dx \) exists. Furthermore, the integral is an increasing (or at least non-decreasing) function of its upper limit $T$, so for $X \geq 0$ it is tempting to define $\mathbb{E}[X]$ as the (possibly infinite) improper Riemann integral (5.9).

3. Inequalities

3.1. Markov inequality. Markov’s inequality says the following.

Proposition 5.10. For non-negative simple r.v. $X$ and $\alpha > 0$, we have

\begin{equation}
P(X \geq \alpha) \leq \frac{1}{\alpha} E(X)
\end{equation}

Proof. This follows from (5.9), as $E(X) \geq \int_0^\alpha P(X \geq x) \, dx \geq \int_0^\alpha P(X \geq \alpha) \, dx$. □

From Remark 5.4 we see that one can have $\mathbb{E}[|X|] = 0$ even if $X = I_A$. But we have the following.

Corollary 5.11. If $\mathbb{E}[|X|] = 0$ then $P(X = 0) = 1$.

Proof. By (5.10) applied to non-negative random variable $|X|$, for every $\alpha > 0$, we have $0 \leq P(|X| > \alpha) \leq P(X \geq \alpha) = 0$. So $P(|X| > 0) = P(\bigcup_n |X| > 1/n) = \lim n \to \infty P(|X| > 1/n) = 0$.

This implies Chebyshev’s inequality.

Proposition 5.12. For $\alpha > 0$,

\begin{equation}
P(|X - m| \geq \alpha) \leq \frac{\text{Var}(X)}{\alpha^2}.
\end{equation}

Corollary 5.13. If $X$ is a simple random variable such that $\text{Var}(X) = 0$ and $\mathbb{E}(X) = m$ then $X = m$ with probability one.

Proof. Write $X = \sum x_j I_{A_j}$ and let $\alpha = \min \{|x_j - m| : x_j - m \neq 0\}$ be the smallest. Note that $\alpha > 0$. By (5.11),

\[
P(X - m \neq 0) = \sum_{j: x_j - m \neq 0} P(A_j) = P(|X - m| \geq \alpha) = 0
\]

Exercise 5.9 is another application of (5.10).
3.1.1. Jensen, Hölder. Recall that \( \varphi : \mathbb{R} \to \mathbb{R} \) is a convex function\(^2\) if \( \varphi(px + (1-p)y) \leq p\varphi(x) + (1-p)\varphi(y) \). Inductively, \( \varphi(\sum_j x_j p_j) \leq \sum_j \varphi(x_j)p_j \). This gives Jensen’s inequality
\[
\varphi(E(X)) \leq E(\varphi(X)) \tag{5.12}
\]
Similarly, if \( \varphi : \mathbb{R}^d \to \mathbb{R} \) is convex, then
\[
\varphi(\mathbb{E}X_1, \mathbb{E}X_2, \ldots, \mathbb{E}X_d)) \leq \mathbb{E}(\varphi(X_1, X_2, \ldots, X_d))
\]
Special cases are \( |E(X)| \leq E|X|, E(X)^2 \leq E(X^2), \) \( \exp(E(X)) \leq E(\exp X), E\ln X \geq \ln E(X) \).
In particular, \( E(|X|) \leq \sqrt{E(X^2)} \). More generally, we have Lyapunov’s inequality: if \( \alpha \leq \beta \) then
\[
E^{1/\alpha}(|X|^\alpha) \leq E^{1/\beta}(|X|^\beta) \tag{5.13}
\]
Indeed, with \( p = \beta/\alpha \geq 1 \) function \( \varphi(x) = |x|^p \) is convex\(^3\). Write \( |X|^\beta = (|X|^\alpha)^p = \varphi(|X|^\alpha) \). Then by Jensen’s inequality,
\[
(E(|X|^\alpha))^{\beta/\alpha} \leq E|X|^\beta
\]
Another important inequality is Cauchy-Schwarz inequality
\[
|E(XY)| \leq \sqrt{E(X^2)E(Y^2)} \tag{5.14}
\]
**Proof #1.** The simplest proof is to consider the quadratic polynomial in variable \( t \) defined by \( p(t) = E(Y + tX)^2 \). (Without loss of generality we may assume that \( E(Y^2) \neq 0 \).) Since \( p(t) \geq 0 \) and \( p(t) = E(Y^2) + 2tE(XY) + t^2E(X^2) \) we have \( (E(XY))^2 \leq E(X^2)E(Y^2) \), as the quadratic polynomial \( p(t) \) cannot have two real roots. \( \square \)

**Proof #2.** Here is a proof using Jensen’s inequality: The function \( (x, y) \to -\sqrt{x}\sqrt{y} \) is convex on \( [0, \infty) \times [0, \infty) \). We apply Jensen’s inequality to non-negative random variables \( X^2 \) and \( Y^2 \).
\[
\mathbb{E}\sqrt{X^2}\sqrt{Y^2} \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)} \tag*{□}
\]

**Proof #3.** Here is a proof based on the elementary inequality \( ab \leq a^2/2 + b^2/2 \).

By homogeneity we may assume that \( E(X^2) = E(Y^2) = 1 \). Then we apply the elementary inequality with \( a = x_i\sqrt{P(A_i \cap B_j)} \) and \( b = y_j\sqrt{P(A_i \cap B_j)} \). We get
\[
E(XY) = \sum_{i,j} x_i y_j P(A_i \cap B_j) \leq \sum_{i,j} \frac{1}{2} x_i^2 P(A_i \cap B_j) + \sum_{i,j} \frac{1}{2} y_j^2 P(A_i \cap B_j) = 1
\]
\( \square \)

\(^2\)A sufficient condition is \( \varphi''(x) \geq 0 \).

**Proof.** For \( x < y \) the difference quotient \( (\varphi(y) - \varphi(x))/(y - x) = \varphi'(\xi) \). We apply this to \( x = a, y = at + b(1-t) \) and \( x = at + b(1-t), y = b \) and write \( \xi = as + b(1-s) \) as a linear combination of \( a, b \). Since \( \varphi'' > 0 \) we have \( \varphi'(\xi_1) < \varphi'(\xi_2) \) for \( \xi_1 < \xi_2 \).
\[
\frac{\varphi(at + b(1-t)) - \varphi(a)}{(b-a)(1-t)} = \varphi'(as + b(1-s)) < \varphi'(at + b(1-t)) < \varphi'(au + b(1-u)) \frac{\varphi(b) - \varphi(at + b(1-t))}{(b-a)t}
\]
Thus
\[
\frac{\varphi(at + b(1-t)) - \varphi(a)}{1-t} < \frac{\varphi(b) - \varphi(at + b(1-t))}{t}
\]
which is equivalent to the convexity condition. \( \square \)

\(^3\)\( f''(x) = p(p - 1)x^{p-2} > 0 \) for \( x > 0 \) and \( p > 1 \).
3.1.2. The correlation coefficient. The correlation coefficient between non-degenerate random variables $X, Y$ is a real number defined as

$$\rho = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Applying Cauchy-Schwartz inequality (5.14) to the product $X - m_X)(Y - m_Y)$ we see that $|\rho| \leq 1$.

**Proposition 5.14.** If $\rho^2 = 1$ then, with probability one, $Y$ is a linear function of $X$.

**Proof.** Without loss of generality, we may assume $E(X) = E(Y) = 0$. Inspecting proof #1 of Cauchy-Schwartz, we see that polynomial $p(t)$ there has a minimum at $t_0 = -E(XY)/E(X^2) = -\rho\sqrt{E(Y^2)/E(X^2)}$. Inserting this into the definition, we see that

$$E(Y + tX)^2 \geq (1 - \rho^2)E(Y^2),$$

and that at $t_0$ we have equality. In particular, if $\rho^2 = 1$ then $E(Y + t_0X)^2 = 0$. So by Corollary 5.13, $|P(Y = t_0X) = 1$.

\[\square\]

4. $L_p$-norms

For $p \geq 1$, define the $L_p$-norm of $X$ as

$$\|X\|_p = \sqrt[p]{E(|X|^p)}$$

Lyapunov’s inequality says that of $p_1 \leq p_2$ then $\|X\|_{p_1} \leq \|X\|_{p_2}$. In particular, $\|X\|_1 \leq \|X\|_2$.

The Cauchy-Schwarz inequality can be stated concisely as

$$|E(XY)| \leq \|X\|_2\|Y\|_2$$

It is clear that $\|\alpha X\|_p = |\alpha|\|X\|_p$ and that $\|X\|_p \geq 0$ is zero only if $X = 0$ (with probability one). What is less obvious is that this is indeed a norm in the vector space of all simple random variables. For this, we need a triangle inequality.

**Theorem 5.15** (Minkowski’s inequality).

(5.15) $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$

**Proof of Minkowski’s inequality for** $p = 1$. Using triangle inequality and monotonicity of expectation, we have

$$\|X + Y\|_1 = E|X + Y| \leq E(|X|) + E(|Y|) = \|X\|_1 + \|Y\|_1$$

\[\square\]

**Proof of Minkowski’s inequality for** $p = 2$.

$$\|X + Y\|_2^2 = E(X + Y)^2 = E(X^2) + E(Y^2) + 2E(XY) \leq \|X\|_2^2 + \|Y\|_2^2 + 2\|X\|_2\|Y\|_2 = (\|X\|_2 + \|Y\|_2)^2$$

\[\square\]
Sketch of proof for general $p \geq 1$. We will use the more general version of Jensen’s inequality: if $\varphi : [0, \infty) \times [0, \infty) \to \mathbb{R}$ is convex then and $X, Y \geq 0$ then $\varphi(E(X), E(Y)) \leq E(\varphi(X, Y))$.

We apply this to the convex function
\[
\varphi(x, y) = -(x^{1/p} + y^{1/p})^p, \quad x, y \geq 0
\]
We get
\[
E(\sqrt[p]{X} + \sqrt[q]{Y})^p \leq (\sqrt[p]{E(X)} + \sqrt[q]{E(X)})^p
\]
We now replace $X, Y \geq 0$ by $|X|^p, |Y|^p$.

The following generalization of Cauchy-Schwarz inequality is often useful

**Theorem 5.16** (Hölder’s inequality). Suppose $p, q > 1$ are conjugate exponents $1/p + 1/q = 1$. Then
\[
(5.16) \quad |E(XY)| \leq \|X\|_p \|Y\|_q
\]

**Sketch of proof.** We apply Jensen’s inequality to convex function $-\sqrt[p]{x}\sqrt[q]{y}$, $x, y \geq 0$. We get
\[
E\left(\sqrt[p]{X} \sqrt[q]{E(Y)}\right) \leq \sqrt[p]{E(X)} \sqrt[q]{E(Y)}
\]
We then replace $X, Y \geq 0$ by $|X|^p$ and $|Y|^q$ to get $|E(XY)| \leq E(|X||Y|) \leq \|X\|_p \|Y\|_q$.

**Other proofs.** The geometric mean is smaller than the arithmetic mean, so $\alpha^{1/p} \beta^{1/q} \leq \alpha/p + \beta/q$. Or, what is really the same, for $0 < u < 1$ we have $u^{1/q} \leq 1/p + u/q$ as by the mean value theorem applied to $f(u) = u^{1/q}$ we have $f(1) - f(u) = (1 - u)f'(\theta) = \frac{(1-u)^{1/p-1} \beta^{1/q} - 1}{q} \geq \frac{(1-u)^{1/p-1}}{q}$

This gives $|ab| \leq \left\langle \frac{|a|^p}{p} + \frac{|b|^q}{q} \right\rangle$, and we can now modify proof #3 of the Cauchy-Schwarz inequality.

**Another proof of Minkowski’s inequality for $p > 1$.** When $p > 1$ we have $q = p/(p-1)$. We apply monotonicity, linearity, and Hölder inequalities:
\[
E(|X+Y|^p) = E(|X+Y|^{p-1}|X+Y|) \leq E(|X+Y|^{p-1}|X|+|Y|) = E(|X||X+Y|^{p-1})+E(|Y||X+Y|^{p-1})
\]
\[
\leq (E(|X|^p))^{1/p}(E|X+Y|^q(p-1))^{1/q} + E(|Y|^p)^{1/p}(E|X+Y|^q(p-1))^{1/q} = (\|X\|_p + \|Y\|_p)E(|X+Y|^p)^{1/q}
\]
Here we used $q(p-1) = p$ and $1 - 1/q = 1/p$.

**Another proof of Lyapunov’s inequality (5.13).** Write $|X|^\alpha$ as product $1 \times |X|^\alpha$ and apply Hölder’s inequality with $p = \beta/\alpha, q = 1 - 1/p = (\beta - \alpha)/\beta$. We get $E|X|^\alpha \leq (E|X|^\beta)^{\alpha/\beta}$ which implies (5.13).

**Definition 5.4.** We say that random variables $X_n$ converges to $X$ converge in $L_p$, if $\|X_n - X\|_p \to 0$. When $p = 2$ we also say that $X_n$ converge in mean square.

It is clear that if $X_n \to X$ and $Y_n \to Y$ in $L_p$ then $X_n + Y_n \to X + Y$ in $L_p$. 
5. The law of large numbers

This is based on [Billingsley, Section 6]. Let $X_1, X_2, \ldots$ be a sequence of simple independent identically distributed random variables on some probability space $(\Omega, \mathcal{F}, P)$. Define $S_n = X_1 + \cdots + X_n$. Denote $m = E(X_n)$.

**Theorem 5.17.** $\frac{1}{n}S_n \to m$ with probability one.

**Convergence in probability.** In this proof we only show convergence in probability. Denote by $\sigma^2$ the variance of $X_1$. (Recall that simple random variables have all moments.) Then $\text{Var}(\frac{1}{n}S_n) = \frac{1}{n^2}\text{Var}(S_n) = \sigma^2/n \to 0$. Therefore, by Chebyshev’s inequality, for $\varepsilon > 0$,

$$P\left(\left|\frac{1}{n}S_n - m\right| > \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0$$

□

**Example 5.4** (The proof gives more!). We do not need independence: Suppose $\Omega = (0, 2\pi)$ with $P = \lambda/2\pi$. Let $X_k(\omega) = \cos(k\omega)$. Then $\frac{1}{n}S_n \to 0$ in probability.

**Example 5.5** (The proof gives more!). We do not need identical distributions: Suppose $X_n = a_n\gamma_n$ where $\gamma_n$ are independent mean zero variance 1. If $\{a_n\}$ is a bounded sequence then $\frac{1}{n}S_n \to 0$ in probability.

(Omitted in 2019) □

**Proof moved to Chapter 8**

**Convergence of a sub-sequence.** Here we show a specific subsequence, $\frac{1}{n}S_{n^2} \to m$ with probability one. Denote by $\sigma^2$ the variance of $X_1$. Then $\text{Var}(\frac{1}{n}S_n) = \frac{1}{n^2}\text{Var}(S_n) = \sigma^2/n \to 0$. Therefore, by Chebyshev’s inequality, for $\varepsilon > 0$,

$$\sum_n P\left|\frac{1}{n^2}S_{n^2} - m\right| > \varepsilon \leq \sum_n \frac{\sigma^2}{n^2\varepsilon^2} < \infty.$$ 

Therefore, by Borel-Cantelli lemma, $P\left(\left|\frac{1}{n^2}S_{n^2} - m\right| > \varepsilon \text{ i.o.}\right) = 0$, i.e. $\frac{1}{n^2}S_{n^2} \to m$ with probability one. This in fact implies the result by the following reasoning, in which without loss of generality we assume $m = 0$. Given $n \in \mathbb{N}$ take $k = k(n)$ such that $k^2 \leq n < (k+1)^2$. Then, since $|X_j| \leq M$, for every $\omega \in \Omega$ we have

$$S_{k^2} - M(n - k^2) \leq S_n \leq S_{k^2} + M(n - k^2)$$

So

$$\frac{k^2 S_{k^2}}{n} - M(1 - \frac{k^2}{n}) \leq \frac{S_n}{n} \leq \frac{k^2 S_{k^2}}{n} + M(1 - \frac{k^2}{n})$$

Now $k = k(n) \to \infty$ so $k^2/(k+1)^2 \to 1$. Since $k^2 \leq n < (k+1)^2$ we see that $k^2/n \to 1$, and the result follows by squeezing principle, as $\frac{S_{k^2}}{k^2} \to 0$ for all $\omega \in \Omega_0 \subset \Omega$ of probability one. □
Proof. Without loss of generality we can assume \( m = 0 \). (Replace \( X_n \) by \( X_n - m \).) We will use Borel-Cantelli lemma to verify that for every \( \varepsilon > 0 \), \( P(\frac{1}{n}|S_n| \geq \varepsilon \text{i.o.}) = 0 \). We use Markov’s inequality,

\[
P(\frac{1}{n}|S_n| \geq \varepsilon) \leq \frac{E[S_n^4]}{\varepsilon^4n^4}
\]

We note that

\[
E[S_n^4] = \sum_{j_1,j_2,j_3,j_4=1}^n E[X_{j_1}X_{j_2}X_{j_3}X_{j_4}] = nE(X_1^4) + 3n(n-1)(E[X_1^2])^2 \leq Cn^2
\]

Thus \( \sum_n P(\frac{1}{n}|S_n| \geq \varepsilon) < \infty \). By Borel-Cantelli (Theorem 3.8) \( P(\frac{1}{n}|S_n| > \varepsilon \text{i.o.}) = 0 \), ending the proof. (See discussion of convergence with probability one in the proof of Proposition 4.14.) \( \square \)

6. Large deviations

Suppose \( X_n \) is \( \text{Bin}(n,p) \) and let \( \hat{p}_n = X_n/n \) denote the sample proportion. Define \( I(x) = x \log x + (1-x) \log \frac{1-x}{1-p} \).

**Theorem 5.18.** For \( x > p \),

\[
P(\hat{p}_n > x) \leq e^{-nI(x)}
\]

**Proof.** We apply Markov inequality to non-negative (simple) random variable \( \exp(tX_n) \), and choose optimal \( t \).

\[
P(\hat{p}_n > x) = P(X_n > nx) = P(tX_n > ntx) = P(\exp(tX_n) > \exp(nxt)) \leq \frac{E[\exp(tX_n)]}{e^{nxt}} = e^{-nxt(1 + p(e^t - 1))^n} = e^{-n(xt - \log(1 + p(e^t - 1)))}
\]

Let's now choose \( t \) that maximizes \( f(t) = (xt - \log(1 + p(e^t - 1))) \). Setting \( f'(t) = 0 \) we get \( t_0 = \log \frac{1-p}{1-x}p \) and \( f(t_0) = I(x) \) as claimed.

The inequality is not true for \( x < p \), as then \( P(\hat{p}_n > x) \to 1 \) while the right hand side converges to 0.

**Remark 5.19.** When \( x < p \) we need to give a bound for \( P(\hat{p}_n < x) \). (Better non-asymptotic estimates are known.)

**Required Exercises**

**Exercise 5.1** (Statistics). Suppose that \( X \) is a simple random variable. Show that the number \( m = E(X) \) minimizes the function \( f(x) = E((X - x)^2) \).

Solution: Expanding the square, \( f(x) = EX^2 - 2xEX + x^2 = EX^2 - 2mx + x^2 \). So \( f'(x) = -2m + 2x = 0 \) and \( f'(x) = 0 \) for \( x = m \). Since \( f''(x) = 2 > 0 \), this is a minimum.

\[^4\text{Quadratic Loss Function}\]
Required Exercises

Exercise 5.2. Suppose that \( X_n \) is a sequence of random variables such that \( |X_n| \leq Y \) where \( Y \) has Poisson distribution. Show that \( \{X_n\} \) is stochastically bounded. (For definition, see Exercise 4.14.)

Solution:

\[
P(|X_n| > K) \leq P(Y > K)
\]

Since \( F_Y(K) = P(Y \leq K) \to 1 \) as \( K \to \infty \), for every \( \varepsilon > 0 \) we can find \( K \) such that \( P(Y > K) < \varepsilon \). (The fact that \( Y \) has Poisson distribution does not matter here!)

Exercise 5.3. Suppose \( X_n \) takes values \( \pm 1 \) with probability \( 1 - 1/(2n) \) and value \( n \) with probability \( 1/n \). Show that \( X_n \) is stochastically bounded. (For definition, see Exercise 4.14.)

Solution:

\[
P(|X_n| > K) = \begin{cases} 1/n & n > K \\ 0 & n \leq K \end{cases}
\]

So given \( \varepsilon > 0 \), choose \( 1/K < \varepsilon \). Then by the above formula, \( P(|X_n| > K) = 0 < \varepsilon \) if \( n < K \) and \( P(|X_n| > K) = 1/n < 1/K < \varepsilon \) if \( n > K \).

Exercise 5.4. Suppose that \( X \) is a simple random variable which has non-negative integers \( \{0, 1, 2, \ldots \} \) as values. Use the definition of \( \mathbb{E}(X) \), not (5.9), to prove that \( \mathbb{E}(X) = \sum_{n=1}^{\infty} P(X \geq n) \).

(Of course, this is a finite sum!). Other similar exercises are possible, and using (5.9) sometimes simplifies the solution, see Exercises 5.18, 5.20, 5.19 below.

Solution: Here is a (forbidden) solution that uses (5.9): 
\[
\mathbb{E}(X) = \int_0^\infty P(X > t)dt = \sum_{n=1}^{\infty} \int_{n-1}^{n} P(X > t)dt = \sum_{n=1}^{\infty} \int_{n-1}^{n} P(X \geq n)dt = \sum_{n=1}^{\infty} P(X \geq n)
\]

Exercise 5.5. Suppose \( X \) is uniform \( U(0, 1) \) random variable, \( X_n \) is its approximation from the proof of Theorem 5.1. Compute \( \mathbb{E}(X_n) \) and its limit as \( n \to \infty \).

Solution: \( X_n = k - 1/2^n \) when \( U \in [(k-1)/2^n, k/2^n) \) so \( \mathbb{E}(X_n) = \sum_{k=1}^{2^n} (k-1)/(2^n)^2 \).

It is simpler to denote \( m = 2^n \) and re-write this as \( \frac{1}{m^2} \sum_{k=1}^{m} (k-1) = \frac{m(m+1)}{2m^2} \to 1/2 \) as expected.

Exercise 5.6. Suppose \( X \) is Poisson random variable with parameter \( \lambda \) (see Example 1.6 on page 13), written as \( X = \sum_{k=0}^{\infty} kI_{A_k} \). Then as discrete monotone approximation we can take

\[
X_n = \sum_{k=0}^{n-1} kI_{A_k} + nI_{\bigcup_{k=n}^{\infty} A_k}.
\]

Compute \( \lim_{n \to \infty} \mathbb{E}(X_n) \).

Solution: Somewhat more generally, if \( X = \sum_{k=0}^{\infty} kI_{A_k} \) then \( X_n = \sum_{k=0}^{n-1} kI_{A_k} + nI_{\bigcup_{k=n}^{\infty} A_k} \) are simple random variables such that \( X_n \uparrow X \). So \( \mathbb{E}(X) = \lim_{n \to \infty} \left( n \sum_{k=n}^{\infty} p_k + \sum_{k=1}^{n-1} kp_k \right) \)

Exercise 5.7. For general \( X \geq 0 \), Remark 5.9 suggests a possible definition of the expected value via tail-integration formula.
(i) Suppose $X$ is uniform $U(0, 1)$ random variable. Use this approach to see what this approach gives for $E[X]$

(ii) Suppose $X$ is exponential random variable, see Example 2.4 on page 24. Use this approach to see what this approach gives for $E[X]$.

(iii) Suppose $X$ has CDF from Example 2.6 on page 25. Use this approach to see what this approach gives for $E[X]$.

Solution: Uniform: $F(x) = 1 - x$ so $E(X)$ should be $\int_0^1 (1 - t) dt = -\int_0^1 (1 - t)^2/2|_{t=0} = 1/2$
as expected

Exponential: With $F(x) = 1 - e^{-\lambda x}$ the tail integration formula says that $E(X)$ should be $\int_0^\infty e^{-\lambda x} dt = 1/\lambda$ as expected.

Exercise 5.8. Suppose $0 \leq X \leq 1$ has cumulative distribution function $F(x)$, and $X_n$ is its approximation from the proof of Theorem 5.1. Express $E(X_n)$ solely in terms of $F$.

Solution:

$$E(X_n) = \frac{2^n}{n^2} \sum_{k=1}^{n^2} \left( \lim_{x \uparrow \frac{k}{2^n}} F(x) - \lim_{x \uparrow \frac{k-1}{2^n}} F(x) \right)$$

Exercise 5.9. Prove that for any simple r.v. $X$ (positive or not) and any real numbers $a, t$ we have

$$P(X > t) \leq e^{-at} E \exp(aX)$$

Exercise 5.10. We say that random variables are centered if their mean is zero. We say that random variables $X, Y$ are uncorrelated if $E(XY) = E(X)E(Y)$. Show that if $X_1, X_2, \ldots$ are simple, (pairwise) uncorrelated, have the same variance $\sigma^2$, and centered then $\frac{1}{n} S_n \rightarrow 0$ in mean square.

Exercise 5.11. Show that $L_p$-convergence implies convergence in probability: if $\|X_n - X\|_p \rightarrow 0$ then $X_n \xrightarrow{P} X$. (Thus together with Exercise 5.10 this establishes the so called weak law of large numbers: $\frac{1}{n} S_n \xrightarrow{P} 0$. Hint: the proof relies on a suitable application of (5.10).

Exercise 5.12. Suppose $X_1, X_2, \ldots$ are independent uniformly bounded (say, $|X_n| \leq 17$ for all $n$) mean zero (simple) random variables. Prove that

$$\frac{1}{n^2} \sum_{j=1}^{n^2} X_j X_{j+1} \rightarrow 0$$

with probability one.

Hints: Use Borel-Cantelli Lemma to verify that $\Omega_0 = \{ \omega : \frac{1}{n^2} \sum_{j=1}^{n^2} X_j X_{j+1} \rightarrow 0 \}$ has probability one.
Additional Exercises

Exercise 5.13. Let $X, Y$ be simple random variables that (together) take values $0, 1, 2, \ldots, m$. Write

$$X = \sum_{j=0}^{m} jI_{A_j}, \quad Y = \sum_{j=0}^{m} jI_{B_j}.$$ 

Show that $\sigma(X, Y) = \sigma(A_0, A_1, \ldots, A_m, B_0, B_1, \ldots, B_m)$. Then describe $\sigma(Z)$ for $Z = X - Y$.

Exercise 5.14 (Computer Science). Suppose $X_n$ is $\text{Bin}(n, 1/2)$. Apply (5.17) to sample proportion $X = X_n/n$ choosing $a$ in that will minimize the right hand side. State the resulting inequality in terms of a bound for $\frac{1}{n} \log P(\frac{1}{n} X_n < p)$, where $p < 1/2$. (Compare Theorem 5.18.)

Exercise 5.15. Is Theorem 5.7 true if we replace assumption $X_n \xrightarrow{p} X$ by a weaker condition $X_n \overset{D}{\to} X$?

Exercise 5.16. Is Theorem 5.7 true if we replace boundedness assumption by a weaker condition that $\{X_n\}$ is bounded with probability one?

Exercise 5.17. Is Theorem 5.7 true if we replace boundedness assumption by a weaker condition that $\{X_n\}$ is bounded in probability?

Exercise 5.18 (*). Suppose that $X$ is a simple random variable which has non-negative integers $\{0, 1, 2, \ldots\}$ as values. Use (5.9), or some other means, to prove that

$$\mathbb{E}(X^2) = \sum_{n=1}^{\infty} (2n - 1)P(X \geq n)$$

(Of course, this is a finite sum!).

Solution: Let's use tail integration formula: $\mathbb{E}[X^2] = \int_{0}^{\infty} P(X^2 > t) dt = \int_{0}^{\infty} P(X > \sqrt{t}) dt = \int_{0}^{\infty} 2uP(X > u) du = \sum_{n=0}^{\infty} \int_{n+1}^{\infty} 2uP(X > u) du = \sum_{n=0}^{\infty} \int_{n+1}^{\infty} 2uP(X \geq n+1) du = \sum_{n=0}^{\infty} (2n+1)P(X \geq n+1) = \sum_{n=1}^{\infty} (2n-1)P(X \geq n)$.

Exercise 5.19 (*). Suppose that $X$ is a simple random variable which has non-negative integers $\{0, 1, 2, \ldots\}$ as values. Prove that

$$\mathbb{E}[2^X] = 1 + \sum_{n=0}^{\infty} 2^n P(X \geq n+1)$$

(Of course, this is a finite sum!).

Solution: Let's proceed with the right hand side: $\sum_{n=0}^{\infty} 2^n P(X \geq n+1) = \sum_{n=0}^{\infty} 2^n \sum_{k=n+1}^{\infty} P(X = k) = \sum_{k=1}^{\infty} P(X = k) \sum_{n=0}^{k-1} 2^n = \sum_{k=1}^{\infty} P(X = k)(2^k - 1) = \sum_{k=1}^{\infty} 2^k P(X = k) - \sum_{k=1}^{\infty} P(X = k) = \sum_{k=1}^{\infty} 2^k P(X = k) - (1 - p_0) = \sum_{k=0}^{\infty} 2^k P(X = k) - 1 = \mathbb{E}[X] - 1$. 

Additional Exercises

73
Exercise 5.20 (**). Suppose that \( X \) is a simple random variable which has positive integers \( \{1, 2, \ldots \} \) as values. Use (5.9), to prove that

\[
\mathbb{E} \left[ \frac{1}{X} \right] = 1 - \sum_{n=1}^{\infty} \frac{1}{n(n+1)} P(X \geq n + 1)
\]

(Of course, this is a finite sum!).

Solution: Here one can proceed either with tail integration and change of variable as in Exercise 5.20, or work with the right hand side as in Exercise 5.19.

First method requires noticing that \( P(1/X > t) = 0 \) when \( t > 1 \), and then working out \( \mathbb{E}[X] = \int_{0}^{\infty} P(1/X > t)dt = \int_{0}^{1} P(X < 1/t)dt = 1 - \int_{0}^{1} P(X > 1/t)dt = 1 + \int_{1}^{\infty} 1/u^2 P(X > u)du = 1 - \sum_{n=1}^{\infty} \int_{n}^{n+1} 1/u^2 P(X > u)du = 1 - \sum_{n=1}^{\infty} P(X \geq n + 1) \int_{n}^{n+1} 1/u^2 du... \)

The second method requires noticing the telescoping sum: \( \sum_{n=1}^{\infty} \frac{1}{n(n+1)} P(X \geq n + 1) = \sum_{n=1}^{\infty} (\frac{1}{n} - \frac{1}{n+1}) \sum_{k=1}^{\infty} P(X = k) = \sum_{k=2}^{\infty} P(X = k) \sum_{n=1}^{k-1} (\frac{1}{n} - \frac{1}{n+1}) = \sum_{k=2}^{\infty} P(X = k) (1 - \frac{1}{k}) = (1 - p_1) - \sum_{k=2}^{\infty} \frac{1}{k} P(X = k) \int_{n}^{n+1} 1/u^2 du... \)

Exercise 5.21. Complete details in the sketch of proof for Minkowski’s inequality.

Exercise 5.22. Complete details in the sketch of proof for Hölder’s inequality.

Exercise 5.23. Show that \( X_n \to X \) with probability 1 iff for every \( \varepsilon > 0 \) there exists \( n \) such that

\[ P(|X_k - X| < \varepsilon, n \leq k \leq m) \geq 1 - \varepsilon \text{ for all } m > n. \]

Exercise 5.24 (*). Suppose \( X_1, X_2, \ldots \) are independent uniformly bounded mean zero (simple) random variables. Prove that

\[
\frac{1}{n} \sum_{j=1}^{n} X_j X_{j+1} \to 0
\]

with probability 1.

Hint: Prove (5.18) first. Noting that every \( n \) can be put into an interval \( k^2 \leq n < (k+1)^2 \), use the uniform bound on \( X_j \) to show that if (5.18) hold for some \( \omega \in \Omega \), this implies convergence in (5.19) for the same \( \omega \).

Exercise 5.25. Suppose \( X \) has mean \( m \) and variance \( \sigma^2 \). For \( \alpha \geq 0 \), prove Cantelli’s inequality

\[
P(X - m \geq \alpha) \leq \frac{\sigma^2}{\sigma^2 + \alpha^2}
\]

Deduce that

\[
P(|X - m| \geq \alpha) \leq \frac{2\sigma^2}{\sigma^2 + \alpha^2}
\]

When is this better than Chebyshev’s inequality?

Hint: Assume \( m = 0 \). \( P(X \geq \alpha) \leq P((X + x)^2 \geq (\alpha + x)^2) \). Apply Markov’s inequality, minimize over \( x > 0 \).
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
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 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
Wasserstein distance, 11
weak convergence, 49
weak law of large numbers, 63
zero-one law, 36, 90