Simple random variables

This section is based on [Billingsley, Section 5]. 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

\[ X = \sum_{j=1}^{n} x_j I_{A_j}, \]

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 4.15).

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.)

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 $X(\omega) = \lim_{n \to \infty} X_n(\omega)$. 

55
It is easy to produce good approximations on sets of large probability,

\[ \sum_{k=1}^{n^2} \frac{k}{n} I_{\frac{k-1}{n} \leq X < \frac{k}{n}} \to X, \]

or discrete uniform approximations on entire \( \Omega \). For the latter, take

\[ X_n = \sum_{k=1}^{\infty} \frac{k}{n} I_{\frac{k-1}{n} \leq X < \frac{k}{n}}. \]

Then \( |X_n - X| \leq 1/n \).

For the proof, we want to be sure that the approximation is also increasing so that \( X_n \uparrow X \).

**Proof.** We find an appropriate function \( \varphi_n(x) \) and take as \( X_n \) the value of \( \varphi_n(X) \). Here is one such function:

(5.2) \[ X_n := nI_{X \geq n/2^n} + \sum_{k=1}^{n^{2^n}} \frac{k-1}{2^n} I_{\frac{k-1}{2^n} \leq X < \frac{k}{2^n}} \uparrow X. \]

See Fig 1. \( \square \)

![Figure 1](image-url). Diadic approximations \( \varphi_n(x) \uparrow x \) from (5.2). Drawn \( \varphi_1 \) (dashed) and \( \varphi_2 \) (dotted).
1. Expected value

A simple random variable of the form (5.1) is assigned expected value

\[ E[X] = \sum_{j=1}^{n} x_j P(A_j) \]

Remark 5.1. 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.

Remark 5.2. A special case of (5.3) is

\[ E(I_A) = P(A) \]

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

\[ E(Y) = \sum_{j} f(x_j) P(A_j) \]

In particular, the moments of \( X \) are

\[ m_k = E(X^k) \]

Proposition 5.3. For simple random variables we have:

- linearity:

\[ E(X + Y) = E(X) + E(Y) \]

and more generally, \( E(aX + Y) = aE(X) + E(Y) \).

- monotonicity: if \( X \leq Y \), then \( E(X) \leq E(Y) \).

\[ |E(X)| \leq E|X| \]

This implies \( |E(X - Y)| \leq E|X - Y| \).

Proof. If \( X = \sum x_j I_{A_j} \) and \( Y = \sum y_k I_{B_k} \) then \( X + Y = \sum x_j y_k I_{A_j \cap B_k} \). 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).

\[ \square \]

Theorem 5.4. If \( X_n \overset{P}{\to} X \) and \( \{X_n\} \) is uniformly bounded, then \( E(X_n) = \lim_{n \to \infty} E(X_n) \).

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

\[ |X(\omega) - X_n(\omega)| \leq 2K I_{A_n} + \varepsilon I_{A_n^c} \]

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

\[ \square \]
Example 5.1. 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 5.4, doesn’t it?

Remark 5.3. Suppose \( X \geq 0 \) is arbitrary, and \( X_n \to X \) are simple random variables from Theorem 5.1. Then \( \lim_{n \to \infty} E(X_n) \) exists, perhaps as \( \infty \). Furthermore, if \( X \) is simple, then by Theorem 5.4, \( \lim_{n \to \infty} E(X_n) \) 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 \( X_n \to X \) will give the same answer!)

It is tempting to compute by this technique an answer that we know from somewhere else.

Definition 5.1. The variance of a simple random variable \( X \) is

\[
\text{Var}(X) = E(X - m)^2 = E(X^2) - m^2
\]

where \( m = E(X) \).

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

2. Expected values and independence

If \( X_1, \ldots, X_n \) are independent then

\[
E(X_1X_2 \cdots X_n) = E(X_1)E(X_2) \cdots E(X_n)
\]

It is enough to verify this for two independent random variables. If \( X = \sum x_j I_{A_j} \) and \( Y = \sum y_k I_{B_k} \) 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 \( E(X_1X_2 \cdots X_n) = E(X_1)E(X_2) \cdots E(X_n) \) and inductively we can pull one factor at a time. In particular, if \( X_1, \ldots, X_n \) are independent then

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

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 \( \text{Var}(X + Y) = E(X + Y)^2 = E(X^2) + E(Y^2) + 2E(XY) = E(X^2) + E(Y^2) = \text{Var}(X) + \text{Var}(Y) \).

Definition 5.2. A moment generating function is \( M(t) = E \exp(tX) \).

Notice that if \( X, Y \) are independent then \( M_{X+Y}(t) = M_X(t)M_Y(t) \). In particular, one can check that the moment generating functions behave consistently with the facts stated in Proposition 4.9 (i) and (ii).

Example 5.2. If \( X \) is \( \text{Bin}(n, p) \), then its moment generating function is \( (1 + p(e^t - 1))^n \).

2.1. Tail integration formula. If \( X \geq 0 \) then

\[
E(X) = \int_0^\infty P(X > x)dx = \int_0^\infty P(X \geq x)dx
\]

Proof. For simple random variables this is just a picture. \( \square \)
3. Inequalities

3.1. Markov inequality. Markov’s inequality for non-negative $X$ is

\[(5.10)\]
\[P(X \geq \alpha) \leq \frac{1}{\alpha}E(X)\]

This follows from (5.9), as $E(X) \geq \int_{0}^{\alpha} P(X \geq x)dx \geq \int_{0}^{\alpha} P(X \geq \alpha)dx$.

This implies Chebyshev’s inequality

\[(5.11)\]
\[P(|X - m| \geq \alpha \leq \frac{\text{Var}(X)}{\alpha^2}).\]

Exercise 5.5 is another application of (5.10).

3.1.1. Jensen, Hölder. Recall that $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ is a convex function\(^1\) 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

\[(5.12)\]
\[\varphi(E(X)) \leq E(\varphi(X))\]

Similarly, if $\varphi : \mathbb{R}^{d} \rightarrow \mathbb{R}$ is convex, then

\[\varphi(EX_{1}, EX_{2}, \ldots, EX_{d}) \leq \varphi(X_{1}, X_{2}, \ldots, X_{d})\]

Special cases are $|E(X)| \leq E|X|$, $E(X)^{2} \leq E(X^{2})$, $\exp(E(X)) \leq E(\exp X)$, $E \ln X \leq \ln E(X)$.

In particular, $E(|X|) \leq \sqrt{E(X^{2})}$. More generally, we have Lyapunov’s inequality: if $\alpha \leq \beta$ then

\[E^{1/\alpha}(|X|^{\alpha}) \leq E^{1/\beta}(|X|^{\beta})\]

Indeed, with $p = \beta/\alpha \geq 1$ function $\varphi(x) = |x|^{p}$ is convex\(^2\). 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

\[(5.13)\]
\[|E(X Y)| \leq \sqrt{E(X^{2})}\sqrt{E(Y^{2})}\]

**Proof #1.** The simplest proof is to consider the quadratic polynomial in variable $t$ defined by $p(t) = E(X + t Y)^{2}$. (Without loss of generality we may assume that $E(Y^{2}) \neq 0$. Since $p(t) \geq 0$ and $p(t) = E(X^{2}) + 2tE(X Y) + t^{2}E(Y^{2})$ we have $(E(X Y))^{2} \leq E(X^{2})E(Y^{2})$.

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

---

\(^1\)A sufficient condition is $\varphi''(x) \geq 0$.

**Proof.** For $x < y$ the difference quotient $(\varphi(y) - \varphi(x))/(y - x) = \varphi'(u)$. Since $\varphi'' > 0$ we have $\varphi'(x) < \varphi'(u) < \varphi'(y)$. This implies that

\[
\frac{\varphi(at + b(1 - t)) - \varphi(a)}{(b - a)(1 - t)} < \varphi'(at + b(1 - t)) < \frac{\varphi(b) - \varphi(at + b(1 - t))}{(b - a)t}
\]

Thus

\[
\frac{\varphi(at + b(1 - t)) - \varphi(a)}{1 - t} < \frac{\varphi(b) - \varphi(at + b(1 - t))}{t}
\]

which is convex.

\(^2\)If $\varphi''(x) = p(p - 1)x^{p-2} > 0$ for $x > 0$ and $p > 1$. 

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

$$E(XY) = \sum_{i,j} x_i y_j P(A_i \cap B_j) \leq \sum_{i,j} \frac{1}{2} x_i^2 P(A_i \cap B_j) + \sum_{i,j} \frac{1}{2} y_j^2 P(A_i \cap B_j) = 1$$

□

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.

Theorem 5.5 (Minkowski’s inequality).
(5.14) $\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$

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

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

□

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

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

□

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

$$E(\sqrt[p]{X} + \sqrt[p]{Y})^p \leq (\sqrt[p]{E(X)} + \sqrt[p]{E(Y)})^p$$

We now replace $X, Y \geq 0$ by $|X|^p, |Y|^p$ □

The following generalization of Cauchy-Schwarz inequality is often useful
5. The law of large numbers

**Theorem 5.6** (Hölder’s inequality). Suppose $p, q > 1$ are conjugate exponents $1/p + 1/q = 1$. Then

\[ |E(XY)| \leq \|X\|_p \|Y\|_q \]

**Sketch of proof.** We apply Jensen’s inequality to convex function $-\sqrt[p]{x} \sqrt[q]{y}$, $x, y \geq 0$. We get

\[ E\left( \sqrt[p]{X} \sqrt[q]{Y} \right) \leq \sqrt[p]{E(X)} \sqrt[q]{E(Y)} \]

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 $a \alpha^{1/p} b^{1/q} \leq \alpha/p + \beta/q$. (Or, what is the same, for $0 < u < 1$ we have $f(u) = u^{1/q} \leq 1/p + u/q$ as by the mean value theorem $f(1) - f(u) = (1 - u)f'(\theta) = (1 - u)^{p-1}/q \geq (1 - u)^{p}/q$.

This gives $|ab| \leq (|a|^{p}/p + |b|^{q}/q)$, and we can now modify proof #3 of the Cauchy-Schwarz inequality. □

**Another proof of Minkowski’s inequality for $p > 1$.** We apply monotonicity, linearity, and Hölder inequalities:

\[ E(|X+Y|^p) = E(|X+Y|^{p-1}|X+Y|) \leq E(|X+Y|^{p-1}|X|+|Y|) = E(|X||X+Y|^{p-1})+E(|Y||X+Y|^{p-1}) \]

\[ \leq (E(|X|^p))^{1/p}(E|X+Y|^q(p-1))^{1/q} + ... \]

□

**Definition 5.3.** 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 5.7.** $1/nS_n \to m$ with 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,

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

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 3.6) $P(\frac{1}{n}|S_n| > \varepsilon \text{ i.o.}) = 0$. (See discussion of convergence with probability one in the proof of Proposition 4.11.) □
5. Simple random variables

Required Exercises

Exercise 5.1 (Statistics). Show that the number \( m = \mathbb{E}(X) \) minimizes the function\(^3\) \( x \mapsto f(x) = \mathbb{E}((X - x)^2) \).

Exercise 5.2. Suppose \( X \) has non-negative integers \( \{0, 1, 2, \ldots\} \) as values. Prove that \( \mathbb{E}(X) = \sum_{n=1}^{\infty} P(X \geq n) \).

Exercise 5.3. Suppose \( X \) is uniform \( U(0, 1) \) random variable, \( X_n \) is its approximation from the proof of Theorem 5.1. Compute \( \mathbb{E}(X_n) \) solely in terms of \( F \).

Exercise 5.4. 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 \( \mathbb{E}(X_n) \) solely in terms of \( F \).

Exercise 5.5. Prove that for any simple r.v. \( X \) (positive or not) and any real numbers \( a, t \) we have
\[
(5.16) \quad P(X > t) \leq e^{-at} \mathbb{E}\exp(aX)
\]

Exercise 5.6. We say that random variables are centered if their mean is zero. We say that random variables \( X, Y \) are uncorrelated if \( \mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) \). Show that if \( X_1, X_2, \ldots \) are (pairwise) uncorrelated, have the same variance \( \sigma^2 \), and centered then \( \frac{1}{n}S_n \to 0 \) in mean square.

Exercise 5.7. Show that \( L_p \)-convergence implies convergence in probability: if \( \|X_n - X\|_p \to 0 \) then \( X_n \overset{P}{\to} X \). (Thus Exercise 5.6 proves the so called weak law of large numbers: \( \frac{1}{n}S_n \overset{P}{\to} 0 \). Hint: the proof relies on a suitable application of (5.10).

Additional Exercises

Exercise 5.8. 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 5.9 (Computer Science). Suppose \( X_n \) is \( Bin(n, 1/2) \). Apply (5.16) to sample proportion \( \tilde{X} = X_n/n \) 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 \( 1/2 < p < 1 \).

Exercise 5.10. Complete details in the sketch of proof for Minkowski’s inequality.

Exercise 5.11. Complete details in the sketch of proof for Hölder’s inequality.

Exercise 5.12. Show that \( X_n \to X \) with probability 1 iff for every \( \varepsilon > 0 \) there exists \( n \) such that \( P(|X_k - X| < \varepsilon, n \leq k \leq m) \geq 1 - \varepsilon \) for all \( m > n \).

\(^3\) Quadratic Loss Function
**Exercise 5.13.** 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

\[(5.17) \quad \frac{1}{n} \sum_{j=1}^{n} X_j X_{j+1} \to 0\]

with probability 1.

*Hint:* Verify that $\Omega_0 = \{\omega : \frac{1}{n^2} \sum_{j=1}^{n^2} X_j X_{j+1} \to 0\}$ has probability 1. Then show that this implies convergence in (5.17) for every $\omega \in \Omega_0$.

**Exercise 5.14.** 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, 60  
$\lambda$-system, 23  
$\pi$-system, 23  
$\sigma$-field, 16  
$\sigma$-field generated by $X$, 41

distribution of a random variable, 42

Bernoulli random variables, 46  
Binomial distribution, 17, 68  
bivariate cumulative distribution function, 27  
Bonferroni’s correction, 18  
Boole’s inequality, 18  
Borel $\sigma$-field, 41  
Borel sigma-field, 16

Cantelli’s inequality, 63  
cardinality, 9  
Cauchy distribution, 99  
Cauchy-Schwarz inequality, 59  
centered, 62  
Central Limit Theorem, 103  
characteristic function, 95  
characteristic function – continuity theorem, 99  
Characteristic functions – uniqueness, 98  
Characteristic functions – inversion formula, 98  
Chebyshev’s inequality, 58  
complex numbers, 94  
conjugate exponents, 60  
continuity condition, 14  
converge in $L_p$, 61  
converge in mean square, 61  
convergence in distribution, 48, 85  
converges in distribution, 109  
converges in probability, 46  
converges pointwise, 7  
converges uniformly, 7  
converges with probability one, 47  
convex function, 59  
copula, 52  
countable additivity, 14  
covariance matrix, 112

cumulative distribution function, 24, 43  
cyindrical sets, 29

DeMorgan’s law, 8  
density function, 26  
diadic interval, 117  
discrete random variable, 68  
discrete random variables, 45

equal in distribution, 43  
events, 13, 17  
extpected value, 56  
Exponential distribution, 69  
exponential distribution, 26

Fatou’s lemma, 67  
field, 13  
finitely-additive probability measure, 14  
Fubini’s Theorem, 74

Geometric distribution, 69  
Hölder’s inequality, 60, 70

inclusion-exclusion, 18  
independent $\sigma$-fields, 33  
independent events, 33  
independent identically distributed, 46  
independent random variables, 44  
indicator functions, 9  
induced measure, 42  
infinite number of tosses of a coin, 117  
integrable, 66  
intersection, 8

Jensen’s inequality, 59  
joint cumulative distribution function, 27  
joint distribution of random variables, 43

Kolmogorov’s maximal inequality, 79  
Kolmogorov’s one series theorem, 80  
Kolmogorov’s three series theorem, 80  
Kolmogorov’s two series theorem, 80  
Kolmogorov’s zero-one law, 79
Kolmogorov-Smirnov metric, 10, 11
Kronecker’s Lemma, 81

Levy distance, 91
law of $X$, 42
Lebesgue’s dominated convergence theorem, 67
Lebesgue’s dominated convergence theorem – used, 68, 78, 87, 100
Levy’s metric, 11
Levy’s theorem, 82
Lindeberg condition, 105
Lyapunov’s condition, 106
Lyapunov’s inequality, 59

marginal cumulative distribution functions, 27
Markov’s inequality, 58
maximal inequality, Etemadi’s, 82
maximal inequality, Kolmogorov’s, 79
measurable function, 41
measurable rectangle, 73
metric, 10
metric space, 10
Minkowski’s inequality, 60
Minkowski’s inequality, 70
moment generating function, 58, 72
Monotone Convergence Theorem, 67
multivariate normal, 111
multivariate normal distribution, 112
multivariate random variable, 41

negative binomial distribution, 18
normal distribution, 26
Poisson distribution, 18, 69
Polya’s distribution, 18
Portmanteau Theorem, 87
power set, 7
probability, 13
probability measure, 14
probability space, 13, 17
product measure, 74
quantile function, 44, 87

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

sample space, 13
Scheffe’s theorem, 85
section, 73
semi-algebra, 15
semi-ring, 15
sigma-field generated by $\mathcal{A}$, 16
simple random variable, 55
simple random variables, 45
Skorohod’s theorem, 87
Slutsky’s Theorem, 86
Standard normal density, 69
stochastic process with continuous trajectories, 43
stochastic processes, 42

stochastically bounded, 52
symmetric distribution, 83
tail $\sigma$-field, 34
Tail integration formula, 75
Taylor polynomials, 93
tight, 52
tight probability measure, 19
Tonelli’s theorem, 74
total variation metric, 11
truncation of r.v., 51

uncorrelated, 62
uniform continuous, 25
Uniform density, 69
uniform discrete, 25
uniform singular, 25
uniformly integrable, 68, 89
union, 8

variance, 58
Wasserstein distance, 11
weak law of large numbers, 62

zero-one law, 34, 79