Section 2

Random variables and their properties. Expectation.

Let \((\Omega, \mathcal{A}, \mathbb{P})\) be a probability space and \((S, \mathcal{B})\) be a measurable space where \(\mathcal{B}\) is a \(\sigma\)-algebra of subsets of \(S\). A \textit{random variable} \(X : \Omega \to S\) is a measurable function, i.e.
\[
B \in \mathcal{B} \implies X^{-1}(B) \in \mathcal{A}.
\]

When \(S = \mathbb{R}\) we will usually consider a \(\sigma\)-algebra \(\mathcal{B}\) of Borel measurable sets generated by sets \(\bigcup_{i \leq n} (a_i, b_i]\) (or, equivalently, generated by sets \((a_i, b_i]\) or by open sets).

\textbf{Lemma 3} \(X : \Omega \to \mathbb{R}\) is a random variable iff for all \(t \in \mathbb{R}\)
\[
\{X \leq t\} := \{\omega \in \Omega : X(\omega) \in (-\infty, t]\} \in \mathcal{A}.
\]

\textbf{Proof.} Only \(\Leftarrow\) direction requires proof. We will prove that
\[
\mathcal{D} = \{D \subseteq \mathbb{R} : X^{-1}(D) \in \mathcal{A}\}
\]
is a \(\sigma\)-algebra. Since sets \((-\infty, t] \in \mathcal{D}\) this will imply that \(\mathcal{B} \subseteq \mathcal{D}\). The result follows simply because taking pre-image preserves set operations. For example, if we consider a sequence \(D_i \in \mathcal{D}\) for \(i \geq 1\) then
\[
X^{-1}\left(\bigcup_{i \geq 1} D_i\right) = \bigcup_{i \geq 1} X^{-1}(D_i) \in \mathcal{A}
\]
because \(X^{-1}(D_i) \in \mathcal{A}\) and \(\mathcal{A}\) is a \(\sigma\)-algebra. Therefore, \(\bigcup_{i \geq 1} D_i \in \mathcal{D}\). Other properties can be checked similarly, so \(\mathcal{D}\) is a \(\sigma\)-algebra.

Let us define a measure \(\mathbb{P}_X\) on \(\mathcal{B}\) by \(\mathbb{P}_X = \mathbb{P} \circ X^{-1}\), i.e. for \(B \in \mathcal{B}\),
\[
\mathbb{P}_X(B) = \mathbb{P}(X \in B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P} \circ X^{-1}(B).
\]

\((S, \mathcal{B}, \mathbb{P}_X)\) is called the \textit{sample space} of a random variable \(X\) and \(\mathbb{P}_X\) is called the \textit{law} of \(X\). Clearly, on this space a random variable \(\xi : S \to S\) defined by the identity \(\xi(s) = s\) has the same law as \(X\).

When \(S = \mathbb{R}\), a function \(F(t) = \mathbb{P}(X \leq t)\) is called the cumulative distribution function (c.d.f.) of \(X\).

\textbf{Lemma 4} \(F\) is a c.d.f. of some r.v. \(X\) iff
1. \(0 \leq F(t) \leq 1\),
2. \(F\) is non-decreasing, right-continuous,
3. \( \lim_{t \to -\infty} F(t) = 0, \lim_{t \to +\infty} F(t) = 1. \)

**Proof.** The fact that any c.d.f. satisfies properties 1 - 3 is obvious. Let us show that \( F \) which satisfies properties 1 - 3 is a c.d.f. of some r.v. \( X \). Consider algebra \( A \) consisting of sets \( \bigcup_{i \leq n} (a_i, b_i] \) for disjoint intervals and for all \( n \geq 1 \). Let us define a function \( \mathbb{P} \) on \( A \) by

\[
\mathbb{P}\left(\bigcup_{i \leq n} (a_i, b_i]\right) = \sum_{i \leq n} (F(a_i) - F(b_i)).
\]

One can show that \( \mathbb{P} \) is countably additive on \( A \). Then, by Caratheodory extension Theorem 1, \( \mathbb{P} \) extends uniquely to a measure \( \mathbb{P} \) on \( \sigma(A) = B \) - Borel measurable sets. This means that \((\mathbb{R}, B, \mathbb{P})\) is a probability space and, clearly, random variable \( X : \mathbb{R} \to \mathbb{R} \) defined by \( X(x) = x \) has c.d.f. \( \mathbb{P}(X \leq t) = F(t) \). Below we will sometimes abuse the notations and let \( F \) denote both c.d.f. and probability measure \( \mathbb{P} \).

**Alternative proof.** Consider a probability space \((\{0, 1\}, B, \lambda)\), where \( \lambda \) is the Lebesgue measure. Define r.v. \( X : [0, 1] \to \mathbb{R} \) by the quantile transformation

\[
X(t) = \inf\{x \in \mathbb{R}, F(x) \geq t\}.
\]

The c.d.f. of \( X \) is \( \lambda(t : X(t) \leq a) = F(a) \) since

\[
X(t) \leq a \iff \inf\{x : F(x) \geq t\} \leq a \iff \exists a_n \to a, F(a_n) \geq t \iff F(a) \geq t.
\]

![Figure 2.1: A random variable defined by quantile transformation.](image)

**Definition.** Given a probability space \((\Omega, A, \mathbb{P})\) and a r.v. \( X : \Omega \to S \) let \( \sigma(X) \) be a \( \sigma \)-algebra generated by a collection of sets \( \{X^{-1}(B) : B \in B\} \). Clearly, \( \sigma(X) \subseteq A \). Moreover, the above collection of sets is itself a \( \sigma \)-algebra. Indeed, consider a sequence \( A_i = X^{-1}(B_i) \) for some \( B_i \in B \). Then

\[
\bigcup_{i \geq 1} A_i = \bigcup_{i \geq 1} X^{-1}(B_i) = X^{-1}\left(\bigcup_{i \geq 1} B_i\right) = X^{-1}(B)
\]

where \( B \in \bigcup_{i \geq 1} B_i \in B \). \( \sigma(X) \) is called the \( \sigma \)-algebra generated by a r.v. \( X \).

![Figure 2.2: \( \sigma(X) \) generated by \( X \).](image)

**Example.** Consider a r.v. defined in figure 2.2. We have \( \mathbb{P}(X = 0) = \frac{1}{2} \), \( \mathbb{P}(X = 1) = \frac{1}{2} \) and

\[
\sigma(X) = \{\emptyset, [0, \frac{1}{2}], (\frac{1}{2}, 1], [0, 1]\}.
\]
Lemma 5 Consider a probability space \((\Omega, \mathcal{A}, \mathbb{P})\), a measurable space \((\mathcal{S}, \mathcal{B})\) and random variables \(X : \Omega \to \mathcal{S}\) and \(Y : \Omega \to \mathbb{R}\). Then the following are equivalent:

1. \(Y = g(X)\) for some (Borel) measurable function \(g : \mathcal{S} \to \mathbb{R}\).
2. \(Y : \Omega \to \mathbb{R}\) is measurable on \((\Omega, \sigma(X))\), i.e. with respect to the \(\sigma\)-algebra generated by \(X\).

Remark. It should be obvious from the proof that \(\mathbb{R}\) can be replaced by any separable metric space.

Proof. The fact that 1 implies 2 is obvious since for any Borel set \(B \subseteq \mathbb{R}\) the set \(B' := g^{-1}(B) \in \mathcal{B}\) and, therefore,

\[
\{Y = g(X) \in B\} = \{X \in g^{-1}(B) = B'\} = X^{-1}(B') \in \sigma(X).
\]

Let us show that 2 implies 1. For all integer \(n\) and \(k\) consider sets

\[
A_{n,k} = \left\{ \omega : Y(\omega) \in \left[ \frac{k}{2^n} , \frac{k+1}{2^n} \right) \right\} = Y^{-1}\left( \left[ \frac{k}{2^n}, \frac{k+1}{2^n} \right) \right).
\]

By 2, \(A_{n,k} \in \sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}\}\) and, therefore, \(A_{n,k} = X^{-1}(B_{n,k})\) for some \(B_{n,k} \in \mathcal{B}\). Let us consider a function

\[
g_n(X) = \sum_{k \in \mathbb{Z}} \frac{k}{2^n} I(X \in B_{n,k}).
\]

By construction, \(|Y - g_n(X)| \leq \frac{1}{4^n}\) since

\[
Y(\omega) \in \left[ \frac{k}{2^n} , \frac{k+1}{2^n} \right) \iff X(\omega) \in B_{n,k} \iff g_n(X(\omega)) = \frac{k}{2^n}.
\]

It is easy to see that \(g_n(x) \leq g_{n+1}(x)\) and, therefore, \(g(x) = \lim_{n \to \infty} g_n(x)\) is a measurable function on \((\mathcal{S}, \mathcal{B})\) and, clearly, \(Y = g(X)\).

\[\square\]

Discrete random variables.
A r.v. \(X : \Omega \to \mathcal{S}\) is called discrete if \(\mathbb{P}_X(\{S_i\}_{i \geq 1}) = 1\) for some sequence \(S_i \in \mathcal{S}\).

Absolutely continuous random variables.
On a measure space \((\mathcal{S}, \mathcal{B})\), a measure \(\mathbb{P}\) is called absolutely continuous w.r.t. a measure \(\lambda\) if

\[
\forall B \in \mathcal{B}, \lambda(B) = 0 \implies \mathbb{P}(B) = 0.
\]

The following is a well known result from measure theory.

Theorem 2 (Radon-Nikodym) If \(\mathbb{P}\) and \(\lambda\) are sigma-finite and \(\mathbb{P}\) is absolutely continuous w.r.t. \(\lambda\) then there exists a Radon-Nikodym derivative \(f \geq 0\) such that for all \(B \in \mathcal{B}\)

\[
\mathbb{P}(B) = \int_B f(s) d\lambda(s).
\]

\(f\) is uniquely defined up to a \(\lambda\)-null sets.

In a typical setting of \(\mathcal{S} = \mathbb{R}^k\), a probability measure \(\mathbb{P}\) and Lebesgue’s measure \(\lambda\), \(f\) is called the density of the distribution \(\mathbb{P}\).

\[\square\]

Independence.
Consider a probability space \((\Omega, \mathcal{C}, \mathbb{P})\) and two \(\sigma\)-algebras \(\mathcal{A}, \mathcal{B} \subseteq \mathcal{C}\). \(\mathcal{A}\) and \(\mathcal{B}\) are called independent if

\[
\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \quad \text{for all } A \in \mathcal{A}, B \in \mathcal{B}.
\]
σ-algebras \( A_i \subseteq C \) for \( i \leq n \) are independent if
\[
P(A_1 \cap \cdots \cap A_n) = \prod_{i \leq n} P(A_i) \quad \text{for all } A_i \in A_i.
\]

σ-algebras \( A_i \subseteq C \) for \( i \leq n \) are pairwise independent if
\[
P(A_i \cap A_j) = P(A_i)P(A_j) \quad \text{for all } A_i \in A_i, A_j \in A_j, i \neq j.
\]

Random variables \( X_i : \Omega \to S \) for \( i \leq n \) are (pairwise) independent if σ-algebras \( \sigma(X_i), i \leq n \) are (pairwise) independent which is just another convenient way to state the familiar
\[
P(X_1 \in B_1, \ldots, X_n \in B_n) = P(X_1 \in B_1) \times \cdots \times P(X_n \in B_n)
\]
for any events \( B_1, \ldots, B_n \in B \).

**Example.** Consider a regular tetrahedron die, Figure 2.3, with red, green and blue sides and a red-green-blue base. If we roll this die then indicators of different colors provide an example of pairwise independent r.v.s that are not independent since
\[
P(r) = P(b) = P(g) = \frac{1}{2} \quad \text{and} \quad P(rb) = P(rg) = P(bg) = \frac{1}{4}
\]
but
\[
P(rbg) = \frac{1}{4} \neq P(r)P(b)P(g) = \left(\frac{1}{2}\right)^3.
\]

![Figure 2.3: Pairwise independent but not independent r.v.s.](image)

Independence of σ-algebras can be checked on generating algebras:

**Lemma 6** If algebras \( A_i, i \leq n \) are independent then σ-algebras \( \sigma(A_i) \) are independent.

**Proof.** Obvious by Approximation Lemma 2.

**Lemma 7** Consider r.v.s \( X_i : \Omega \to \mathbb{R} \) on a probability space \( (\Omega, A, P) \).

1. \( X_i \)'s are independent iff
\[
P(X_1 \leq t_1, \ldots, X_n \leq t_n) = P(X_1 \leq t_1) \times \cdots \times P(X_n \leq t_n).
\]

2. If the laws of \( X_i \)'s have densities \( f_i(x) \) then \( X_i \)'s are independent iff a joint density exists and
\[
f(x_1, \ldots, x_n) = \prod f_i(x_i).
\]
Proof. 1 is obvious by Lemma 6 because (2.0.1) implies the same equality for intervals

\[ \mathbb{P}(X_1 \in (a_1, b_1], \ldots, X_n \in (a_n, b_n]) = \mathbb{P}(X_1 \in (a_1, b_1]) \times \ldots \times \mathbb{P}(X_n \in (a_n, b_n]) \]

and, therefore, for finite union of disjoint such intervals. To check this for intervals (for example, for \( n = 2 \)) we can write \( \mathbb{P}(a_1 < X_1 \leq b_1, a_2 < X_n \leq b_2) \) as

\[
\mathbb{P}(X_1 \leq b_1, X_2 \leq b_2) - \mathbb{P}(X_1 \leq a_1, X_2 \leq b_2) - \mathbb{P}(X_1 \leq b_1, X_2 \leq a_2) + \mathbb{P}(X_1 \leq a_1, X_2 \leq a_2) \\
= \mathbb{P}(X_1 \leq b_1)\mathbb{P}(X_2 \leq b_2) - \mathbb{P}(X_1 \leq a_1)\mathbb{P}(X_2 \leq b_2) - \mathbb{P}(X_1 \leq b_1)\mathbb{P}(X_2 \leq a_2) + \mathbb{P}(X_1 \leq a_1)\mathbb{P}(X_2 \leq a_2) \\
= (\mathbb{P}(X_1 \leq b_1) - \mathbb{P}(X_1 \leq a_1))(\mathbb{P}(X_2 \leq b_2) - \mathbb{P}(X_2 \leq a_2)) = \mathbb{P}(a_1 < X_1 \leq b_1)\mathbb{P}(a_2 < X_2 \leq b_2).
\]

To prove 2 we start with "\( \Leftarrow \)".

\[
\mathbb{P}(\cap\{X_i \in A_i\}) = \mathbb{P}(X \in A_1 \times \cdots \times A_n) = \int_{A_1 \times \cdots \times A_n} \prod f_i(x_i)dx \\
= \prod \int A_i f_i(x_i)dx, \text{ \{by Fubini’s Theorem\}} = \prod_{i \leq n} \mathbb{P}(X \in A_i).
\]

Next, we prove "\( \Rightarrow \)". First of all, by independence,

\[
\mathbb{P}(X \in A_1 \times \cdots \times A_n) = \prod \mathbb{P}(X_i \in A_i) = \int_{A_1 \times \cdots \times A_n} \prod f_i(x_i)dx.
\]

Therefore, the same equality holds for sets in algebra \( A \) that consists of finite unions of disjoint sets \( A_1 \times \cdots \times A_n \), i.e.

\[
\mathbb{P}(X \in B) = \int_B \prod f_i(x_i)dx \text{ for } B \in A.
\]

Both \( \mathbb{P}(X \in B) \), \( \int_B \prod f_i(x_i)dx \) are countably additive on \( A \) and finite,

\[
\mathbb{P}(\mathbb{R}^n) = \int_{\mathbb{R}^n} \prod f_i(x_i)dx = 1.
\]

By the Carathéodory extension Theorem 1, they extend uniquely to all Borel sets \( B = \sigma(A) \), so

\[
\mathbb{P}(B) = \int_B \prod f_i(x_i)dx \text{ for } B \in B.
\]

**Expectation.** If \( X : \Omega \to \mathbb{R} \) is a random variable on \( (\Omega, A, \mathbb{P}) \) then **expectation** of \( X \) is defined as

\[
\mathbb{E}X = \int_{\Omega} X(\omega)d\mathbb{P}(\omega).
\]

In other words, expectation is just another term for the integral with respect to a probability measure and, as a result, expectation has all the usual properties of the integrals. Let us emphasize some of them.

**Lemma 8.** 1. If \( F \) is the c.d.f. of \( X \) then for any measurable function \( g : \mathbb{R} \to \mathbb{R}, \)

\[
\mathbb{E}g(x) = \int_{\mathbb{R}} g(x)dF(x).
\]

2. If \( X \) is discrete, i.e. \( \mathbb{P}(X \in \{x_i\}_{i \geq 1}) = 1 \), then

\[
\mathbb{E}X = \sum_{i \geq 1} x_i \mathbb{P}(X = x_i).
\]
3. If $X : \Omega \to \mathbb{R}^k$ has a density $f(x)$ on $\mathbb{R}^k$ and $g : \mathbb{R}^k \to \mathbb{R}$ then

$$\mathbb{E}g(X) = \int g(x)f(x)dx.$$

**Proof.** All these properties follow by making a change of variables $x = X(\omega)$ or $\omega = X^{-1}(x)$, i.e.

$$\mathbb{E}g(X) = \int g(X(\omega))d\mathbb{P}(\omega) = \int g(x)d\mathbb{P} \circ X^{-1}(x) = \int g(x)d\mathbb{P}_X(x),$$

where $\mathbb{P}_X = \mathbb{P} \circ X^{-1}$ is the law of $X$. Another way to see this would be to start with indicator functions of sets $g(x) = \mathbb{I}(x \in B)$ for which

$$\mathbb{E}g(X) = \mathbb{P}(X \in B) = \mathbb{P}_X(B) = \int_{\mathbb{R}} \mathbb{I}(x \in B)d\mathbb{P}_X(x)$$

and, therefore, the same is true for simple step functions

$$g(x) = \sum_{i \geq n} w_i \mathbb{I}(x \in B_i)$$

for disjoint $B_i$. By approximation, this is true for any measurable functions.

$\square$