7.1 Agenda

- Almost sure events.
- Markov’s inequality.
- Some properties of a nonnegative random variable with finite expectation.
- When is a random variable independent of itself?

7.2 Almost Sure Events

Suppose that \( x \in \mathbb{R} \) is a real number, and it is known that \( x \geq 0 \). As we know, the interpretation of this condition is that \( x \) is a number that is greater than or equal to 0. However, suppose now that \( f: \mathbb{R} \rightarrow \mathbb{R} \) is a real-valued function defined on \( \mathbb{R} \). Borrowing from the example of real numbers, we may want to make statements such as “\( f \geq 0 \)” or “\( f \geq g \)”, where \( g: \mathbb{R} \rightarrow \mathbb{R} \) is another real-valued function. How do we interpret such statements about functions? Typically, such statements about functions mean what is most intuitively deducible, i.e., the statement “\( f \geq 0 \)” is used to express the fact that \( f(x) \geq 0 \) for every \( x \in \mathbb{R} \). Similarly, the statement “\( f \geq g \)” is used to express the fact that \( f(x) \geq g(x) \) for every \( x \in \mathbb{R} \).

In the above, notice that we imposed the condition \( f(x) \geq 0 \) or \( f(x) \geq g(x) \) to hold for every \( x \in \mathbb{R} \), which may be stringent in certain cases. This raises the following question: can we relax the aforementioned conditions to not hold for every \( x \in \mathbb{R} \) and still retain the usefulness of our statements? In other words, we seek to know if we can come up with a definition such as “\( f \geq 0 \)” means that \( f(x) \geq 0 \) only for those \( x \in A \)”, or “\( f \geq g \)” means that \( f(x) \geq g(x) \) only for those \( x \in A \)” where \( A \subseteq \mathbb{R} \) is a rich and informative set.

In the preceding definitions, it may so be the case that \( f(x) < 0 \) or \( f(x) < g(x) \) for \( x \in A^c \), but we do not care about this since as per the definitions, \( A \) is a rich and informative set, which means that \( A^c \) is a poor and uninformative set. While, in general, we may not be able to come up with meaningful definitions such as the ones above, one instance where we can do so is in probability theory. This motivates the following definitions.

**Definition 7.2.1** (Sure events). Given a probability space \((\Omega, \mathcal{F}, P)\), an event \( A \in \mathcal{F} \) is called a sure event if \( A = \Omega \).

**Definition 7.2.2** (Almost sure events). Given a probability space \((\Omega, \mathcal{F}, P)\), an event \( A \in \mathcal{F} \) is called an almost sure event if \( P(A) = 1 \).

**Remarks:**

1. Notice that \( \Omega \) is both a sure event and an almost sure event.
2. If $A$ is an almost sure event, i.e., $P(A) = 1$, this does not imply that $A = \Omega$.

3. Going back to our earlier examples on functions, if we now consider $f, g : \Omega \to \mathbb{R}$, then the first definition $f \geq 0$ (or $f \geq g$) implies that $f(\omega) \geq 0$ (or $f(\omega) \geq g(\omega)$) for all $\omega \in \Omega$, which implies that $\{\omega \in \Omega : f(\omega) \geq 0\} = \Omega$ (or $\{\omega \in \Omega : f(\omega) \geq g(\omega)\} = \Omega$) is a sure event. However, the relaxed definitions imply that the conditions in the definitions hold only for those $\omega \in A$ where $A \subseteq \Omega$ is a set which has probability 1. We do not care about what happens outside the set $A$.

4. In probability theory, for all purposes, an almost sure event is considered as a “certain event”.

5. “Almost surely” is often abbreviated as “a.s.”. Sometimes, the statement “with probability 1” is used interchangeably with almost surely.

We now present more examples.

**Example 7.2.3.** $\Omega = \{1, 2, 3, 4\}$, $\mathcal{F} = 2^\Omega$ (power set, or set of all subsets, of $\Omega$),

$$P(\{\omega\}) = \begin{cases} \frac{1}{2}, & \forall \omega \in \{1, 2\} \\ 0, & \forall \omega \in \{3, 4\}. \end{cases}$$

Let $X, Y$ be two $\mathcal{F}$-measurable random variables defined as follows:

$$X(\omega) = \begin{cases} 1, & \forall \omega \in \{1, 2\} \\ 0, & \forall \omega \in \{3, 4\}; \end{cases}$$

$$Y(\omega) = \begin{cases} 0, & \forall \omega \in \{1, 2\} \\ 5000, & \forall \omega \in \{3, 4\}. \end{cases}$$

Notice that $Y(\omega)$ takes a value much greater than $X(\omega)$ on the set $\{3, 4\}$. But notice that $P(\{3, 4\}) = 0$, and $P(\{1, 2\}) = 1$. Therefore on a set of probability 1, $X \geq Y$. In other words, the event $\{\omega : X(\omega) \geq Y(\omega)\}$ is an almost sure event. In shorthand, this is also written as $X(\omega) \geq Y(\omega)$ a.s..

**Exercises**

1. In the above example, check that $\{\omega : X(\omega) = 1\}$ is an almost sure event.

2. In the above example, check that $\{\omega : Y(\omega) = 0\}$ is an almost sure event.

3. Consider the following probability space: for a positive integer $n \geq 1$, let

$$\Omega = \{1, \ldots, 2n\},$$

$$\mathcal{F} = 2^\Omega,$$

$$P(\{\omega\}) = \frac{1}{2n}, \quad \forall \omega \in \{1, \ldots, 2n\}.$$ 

Let $X, (Y_i)_{i \in \mathbb{N}}$ be a collection of $\mathcal{F}$-measurable random variables defined as follows:

$$X(\omega) = \omega \quad \forall \omega \in \Omega,$$

$$Y_i(\omega) = 2n - \omega - i \quad \forall \omega \in \Omega, \quad i \in \mathbb{N}. $$
(a) For \( n = 5 \) and \( i = 5 \), is the event 
\[
\{ \omega : X(\omega) \geq Y_i(\omega) \}
\]
an almost sure event?

(b) Find the smallest value of \( i \) (in terms of \( n \)) such that the event 
\[
\{ \omega : X(\omega) \geq Y_i(\omega) \}
\]
is an almost sure event.

### 7.3 Markov’s Inequality

**Theorem 7.3.1.** [Markov’s Inequality] Let \( \{ \Omega, \mathcal{F}, P \} \) be a probability space, and let \( X \) be an \( \mathcal{F} \)-measurable random variable. Furthermore, let \( X \geq 0 \) a.s. and \( E[X] < \infty \). Then,

\[
P(\{ \omega : X(\omega) \geq a \}) \leq \frac{E[X]}{a} \quad \forall a > 0.
\]

**Proof.** Let 
\[
A := \{ \omega : X(\omega) \geq a \}.
\]
Notice that
\[
1_A(\omega) + 1_{A^c}(\omega) = 1, \quad \forall \omega \in \Omega
\]
since \( 1_A(\omega) \) is 1 if \( \omega \in A \), and 0 if \( \omega \in A^c \). Similarly, \( 1_{A^c}(\omega) \) is 1 if \( \omega \in A^c \), and 0 if \( \omega \in A \). This further leads to

\[
X(\omega)(1_A(\omega) + 1_{A^c}(\omega)) = X(\omega), \quad \forall \omega \in \Omega. \quad (7.1)
\]

Now, since we are given that \( X \geq 0 \) a.s., this implies that there exists a set \( B \in \mathcal{F} \), \( B \subseteq \Omega \) satisfying \( P(B) = 1 \), for which the condition

\[
X(\omega) \geq 0, \quad \forall \omega \in B
\]
holds. Also, notice that
\[
1_{A^c}(\omega) \geq 0, \quad \forall \omega \in \Omega.
\]
Therefore, we get
\[
X(\omega)1_{A^c}(\omega) \geq 0, \quad \forall \omega \in B.
\]

By using the above inequality in (7.1), we have

\[
X(\omega) \geq X(\omega)(1_A(\omega)), \quad \forall \omega \in B. \quad (7.2)
\]

Now, notice that
\[
X(\omega)(1_A(\omega)) \geq \begin{cases} 
a, & \forall \omega \in A \\
0, & \forall \omega \in A^c.
\end{cases}
\]

The above equation can be compactly rewritten as follows:

\[
X(\omega)(1_A(\omega)) \geq a1_A(\omega), \quad \forall \omega \in \Omega.
\]
Since $B \subseteq \Omega$, the above equation in particular implies
\[ X(\omega)(1_A(\omega)) \geq a1_A(\omega), \quad \forall \omega \in B. \]
Using this in the right hand side of (7.2), we get
\[ X(\omega) \geq a(1_A(\omega)) \quad \forall \omega \in B. \]
Now, since $P(B) = 1$ this implies
\[ X(\omega) \geq a1_A(\omega) \quad a.s. \]
Taking expectation on both sides, we get $E[X] \geq aP(A)$, thus proving the result\(^1\).

Remarks:

1. Markov’s inequality trivially holds if $E[X] = \infty$.
2. Also, notice that Markov’s inequality holds trivially for all constants $a$ satisfying $0 < a \leq E[X]$ since the right hand side of the inequality then becomes greater than or equal to 1, while the left hand side is a probability which always lies between 0 and 1.
3. Generally, when writing proofs in probability, you are not expected to specify the event (set) on which a condition holds almost surely. If a few condition for random variables hold almost surely, any new condition derived out of this conditions holds almost surely.

To make the above remark clear let us rewrite the proof of Markov’s inequality from (7.1) onwards.

A more compact proof of Markov’s Inequality: From (7.1),
\[ X(\omega)(1_A(\omega) + 1_{A^c}(\omega)) = X(\omega) \quad a.s. \]
Also, $X \geq 0 \quad a.s.$, and $1_{A^c}(\omega) \geq 0 \quad a.s.$ Therefore,
\[ X(\omega) \geq X(\omega)1_A(\omega) \quad a.s. \]
Also,
\[ X(\omega)1_A(\omega) \geq a1_A(\omega) \quad a.s. \]
Therefore,
\[ X(\omega) \geq a1_A(\omega) \quad a.s. \]
Taking expectations gives the result.
Of course, the proof above is conceptually the same as the more elaborate proof we carried out before. The idea here was to illustrate how one should compactly write such proofs.

7.4 Some Properties of Nonnegative Random Variables with Finite Expectation

Lemma 7.4.1. For $X \geq 0 \ a.s.$, if $E[X] = 0$, then
\[ P(\{\omega : X(\omega) = 0\}) = 1. \]
\(^1\)This holds since if $Y \geq Z \quad a.s.$, then $E[Y] \geq E[Z].$
Proof. It suffices to show that \(P\{\omega : X(\omega) > 0\})\). Towards this end, we note that for any \(\epsilon > 0\), by Markov’s inequality, we have

\[
P(\{\omega : X(\omega) \geq \epsilon\}) \leq \frac{E[X]}{\epsilon} = 0.
\]

But \(P(\{\omega : X(\omega) \geq \epsilon\}) \geq 0\), therefore it follows that for any \(\epsilon > 0\),

\[
P(\{\omega : X(\omega) \geq \epsilon\}) = 0.
\]

In particular, this implies that for all \(n \in \{1, 2, \ldots\}\), we have

\[
P\left(\left\{\omega : X(\omega) \geq \frac{1}{n}\right\}\right) = 0.
\]

By continuity of probability, we have

\[
P(\{\omega : X(\omega) > 0\}) = 0.
\]

Below is detailed clarification on the last line of the proof. For all \(n \in \{1, 2, \ldots\}\),

\[
P\left(\left\{\omega : X(\omega) \geq \frac{1}{n}\right\}\right) = 0.
\]

Therefore,

\[
\lim_{n \to \infty} P\left(\left\{\omega : X(\omega) \geq \frac{1}{n}\right\}\right) = 0.
\]

By continuity of probability,

\[
P\left(\lim_{n \to \infty} \left\{\omega : X(\omega) \geq \frac{1}{n}\right\}\right) = 0.
\]

Now since \(\{\omega : X(\omega) \geq \frac{1}{n}\}\) is an increasing sequence of sets,

\[
\lim_{n \to \infty} \left\{\omega : X(\omega) \geq \frac{1}{n}\right\} = \bigcup_{n=1}^{\infty} \left\{\omega : X(\omega) \geq \frac{1}{n}\right\}
\]

\[
= \{\omega : X(\omega) > 0\}.
\]

Remark: Check Section 5.4.1 of Tutorial 5 to see that Lemma 7.4.1 is used to prove equality in Cauchy-Schwartz inequality.

Lemma 7.4.2. For \(X \geq 0\) a.s., if \(E[X] < \infty\), then \(^2\)

\[
P(\{\omega : X(\omega) = \infty\}) = 0.
\]

\(^2\)Notice for the event \(\{\omega : X(\omega) = \infty\}\) to be even valid, we need to define random variables to take values on extended real line \(\mathbb{R} = \mathbb{R} \cup \{-\infty, \infty\}\). In other words, we define \(X: \Omega \to \mathbb{R}\).
Proof. Notice that for all \( M \geq 0 \),

\[
\{ \omega : X(\omega) = \infty \} \subseteq \{ \omega : X(\omega) \geq M \}
\]

since \( X(\omega) = \infty \) implies (which in set theory means “subset of”) that \( X(\omega) \geq M \) for any choice of \( M \geq 0 \). Therefore, for all \( M \geq 0 \), by Markov’s inequality, we have

\[
P(\{ \omega : X(\omega) = \infty \}) \leq P(\{ \omega : X(\omega) \geq M \}) \leq \frac{E[X]}{M}.
\]

Since the above inequality holds for any \( M \), taking \( M \to \infty \), we have

\[
P(\{ \omega : X(\omega) = \infty \}) = 0.
\]

\[\square\]

7.5 When is a random variable independent of itself?

Let us first recap the definition of independence of two random variables.

**Definition 7.5.1.** Given a probability space \((\Omega, \mathcal{F}, P)\) and two \(\mathcal{F}\)-measurable random variables \(X\) and \(Y\), \(X\) is said to be independent of \(Y\) if and only if for all \(x, y \in \mathbb{R}\),

\[
P(\{ \omega : X(\omega) \leq x, Y(\omega) \leq y \}) = P(\{ \omega : X(\omega) \leq x \}) \cdot P(\{ \omega : Y(\omega) \leq y \}).
\]

We seek to investigate the conditions under which a random variable will be independent of itself. Towards this, we make the substitution \(Y = X\) and \(y = x\) in the above definition for independence to get

\[
P(\{ \omega : X(\omega) \leq x, X(\omega) \leq x \}) = P(\{ \omega : X(\omega) \leq x \}) \cdot P(\{ \omega : X(\omega) \leq x \})
\]

\[
\Rightarrow P(\{ \omega : X(\omega) \leq x \}) = (P(\{ \omega : X(\omega) \leq x \}))^2
\]

to be true for all \(x \in \mathbb{R}\). The above quadratic equation implies that for each \(x \in \mathbb{R}\), we have

\[
P(\{ \omega : X(\omega) \leq x \}) = 0 \text{ or } P(\{ \omega : X(\omega) \leq x \}) = 1.
\]

However, since \(F(x) := P(\{ \omega : X(\omega) \leq x \})\) is a nondecreasing function with the property that \(\lim_{x \to \infty} F(x) = 1\), it follows that there exists a constant \(c \in \mathbb{R}\) such that

\[
P(\{ \omega : X(\omega) \leq x \}) = \begin{cases} 0, & x < c \\ 1, & x \geq c. \end{cases}
\]

Therefore, if a random variable \(X\) is independent of itself, then its CDF \(F_X(\cdot)\) is as follows:

\[
F_X(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c \end{cases}
\]

for some \(c \in \mathbb{R}\).

In summary, if \(X\) is independent of itself, then \(X\) is a constant with probability 1 (or almost surely), i.e., \(P(X = c) = 1\) for some constant \(c \in \mathbb{R}\).
Exercises Assume \((\Omega, \mathcal{F}, P)\) is a given probability space. All random variables mentioned below are assumed to be \(\mathcal{F}\)-measurable. Expectations are with respect to \(P(\cdot)\).

1. Let \(\alpha \in \mathbb{R}\) be a constant, and let \(X\) and \(Y\) be two independent random variables satisfying the condition

\[
P(X + Y = \alpha) = 1.
\]

Then, show that \(X, Y\) must be constants (i.e., \(X\) and \(Y\) take constant values with probability 1).

Hint: Show that if \(X, Y\) are independent, then \(X\) and \(\alpha - Y\) are independent random variables.

2. Let \(X\) and \(Y\) have a joint pmf given by

\[
p_{X,Y}(j, k) = \frac{c(j+k) e^{j+k}}{j!k!}, \quad \forall j, k \geq 0,
\]

where \(c\) is a constant. Find \(c\), and find the pmf of \(X\). What is \(E[X]\)?

3. Let \(X\) and \(Y\) be independent random variables. Let \(U\) be a discrete random variable independent of both \(X\) and \(Y\) with the following pmf:

\[
P(U = 1) = P(U = -1) = \frac{1}{2}.
\]

Define two new random variables \(R\) and \(S\) as \(S = UX\) and \(R = UY\). Show that in general, \(S\) and \(R\) need not be independent, but \(S^2\) and \(R^2\) are always independent.

Above, the first exercise is from [1], second and third from [2].

References
