1 Sequences of events and their limits

1.1 Monotone sequences of events

Sequences of events arise naturally when a probabilistic experiment is repeated many times. For example, if a coin is flipped consecutively, the "event" [3]

$$A = \big\{\text{'heads' never seen}\big\}$$

is just the intersection, $A = \cap_{n\ge1} A_n$, of the events

$$A_n = \big\{\text{'heads' not seen in the first } n \text{ tosses}\big\}\,.$$

This simple remark leads to the following important observations: 1) taking countable operations is not that exotic in probabilistic models, and thus any reasonable theory should deal with σ-fields; b) the event A is in some sense the limit of the sequence $(A_n)_{n\ge1}$, so understanding limits of sequences of sets (events) might be useful.

In general, finding a limit of a sequence of sets is not easy and we will not do this here. [4] Instead, we will mostly consider monotone sequences of events.

Definition 1.1. A sequence $(A_n)_{n\ge1}$ of events is *increasing* if $A_n \subset A_{n+1}$ for all $n \ge 1$. It is *decreasing* if $A_n \supset A_{n+1}$ for all $n \ge 1$.

Example 1.2. *If $(A_n)_{n\ge1}$ is a sequence of arbitrary events, then the sequence $(B_n)_{n\ge1}$ with $B_n = \cup_{k=1}^n A_k$ is increasing, whereas the sequence $(C_n)_{n\ge1}$ with $C_n = \cap_{k=1}^n A_k$ is decreasing.*

The following result shows that the probability measure is continuous along monotone sequences of events.

Lemma 1.3. *If $(A_n)_{n\ge1}$ is increasing with $A = \lim_n A_n = \cup_{n\ge1} A_n$, then*

$$\mathsf{P}(A) = \mathsf{P}\big(\lim_{n\to\infty} A_n\big) = \lim_{n\to\infty} \mathsf{P}(A_n)\,.$$

If $(A_n)_{n\ge1}$ is a decreasing sequence with $A = \lim_n A_n = \cap_{n\ge1} A_n$, then

$$\mathsf{P}(A) = \mathsf{P}\big(\lim_{n\to\infty} A_n\big) = \lim_{n\to\infty} \mathsf{P}(A_n)\,.$$

Remark 1.3.1. If $(A_n)_{n\ge1}$ is not a monotone sequence of events, the claim of the lemma is not necessarily true (find a counterexample!).

Proof. Let $(A_n)_{n\ge1}$ be increasing with $A = \cup_{n\ge1} A_n$. Denote $C_1 = A_1$ and, for $n \ge 2$, put $C_n = A_n \setminus A_{n-1} = A_n \cap A_{n-1}^{\mathsf{c}}$. We then have (**why?**) [5]

$$A_n = \bigcup_{k=1}^n A_k = \bigcup_{k=1}^n C_k\,, \qquad \bigcup_{k=1}^\infty A_k = \bigcup_{k=1}^\infty C_k\,.$$

[3] *Apriori* we do not know that A is an event, ie, can be assigned probability to!

[4] The corresponding theory is the subject of 'pure' courses such as set theory or (real) analysis/measure theory; if interested, have a look at problems E26–E28 and/or get in touch!

[5] Decompositions in the form $A_n = \bigcup_{k=1}^n \big(A_k \setminus (\cup_{m=1}^{k-1} A_m)\big)$ are often called telescopic; they are analogous to those in sequential Bayes formulae.

Since the events in $(C_k)_{k\ge1}$ are mutually incompatible, the σ-additivity property P3 of the probability measure gives

$$\mathsf{P}(A) = \mathsf{P}\Big(\bigcup_{k\ge1} A_k\Big) = \mathsf{P}\Big(\bigcup_{k\ge1} C_k\Big) = \sum_{k\ge1} \mathsf{P}\big(C_k\big) \le 1$$

and therefore

$$0 \le \mathsf{P}(A) - \mathsf{P}(A_n) = \mathsf{P}\big(A \setminus A_n\big) = \mathsf{P}\Big(\bigcup_{k>n} C_k\Big) = \sum_{k>n} \mathsf{P}(C_k) \to 0$$

as $n \to \infty$, as a tail sum of a convergent series $\sum_{k\ge1} \mathsf{P}\big(C_k\big)$.

A similar argument holds for decreasing sequences (**do this!**). □

Example 1.4. *A standard six-sided die is tossed repeatedly. Let N_1 denote the total number of ones observed. Assuming that the individual outcomes are independent, show that $\mathsf{P}(N_1 = \infty) = 1$.*

Solution. We show that $\mathsf{P}(N_1 < \infty) = 0$. First, notice that $\{N_1 < \infty\} = \cup_{n\ge1} B_n$ with $B_n = \big\{\text{no 'ones' after } n\text{th toss}\big\}$, so it is enough to show that $\mathsf{P}(B_n) = 0$ for all n. However, $B_n = \cap_{m>0} C_{n,m}$ with $C_{n,m} = \big\{\text{no 'one' on tosses } n+1, \dots n+m\big\}$ being a decreasing sequence, $C_{n,m} \supset C_{n,m+1}$ for all $m \ge 1$. Since $\mathsf{P}(C_{n,m}) = (5/6)^m \to 0$ as $m \to \infty$, Lemma 1.3.1 implies $\mathsf{P}(B_n) = \lim_{m\to\infty} \mathsf{P}(C_{n,m}) = 0$, as requested. □

Example 1.5. *Let X be a positive random variable with $\mathsf{P}(X < \infty) = 1$. For $k \ge 1$, denote $X_k = \frac{1}{k}X$. Show that the event $A(\varepsilon) \equiv \big\{|X_k| > \varepsilon$ finitely often$\big\}$ satisfies $\mathsf{P}\big(A(\varepsilon)\big) = 1$ for every $\varepsilon > 0$.*

Solution. Let $\Omega_0 = \{\omega \in \Omega : X(\omega) < \infty\}$ be the event 'X is finite'; by assumption, $\mathsf{P}(\Omega_0) = 1$. Consider the events $B_k = \{|X_k| > \varepsilon\} = \{\omega : |X(\omega)| > k\varepsilon\}$. Since the random variables X_k form a pointwise decreasing sequence, namely

$$^\forall\omega \in \Omega\,, \qquad X_k(\omega) \ge X_{k+1}(\omega) \quad \text{for all } k \ge 1\,,$$

the events B_k are decreasing (ie., $B_k \supset B_{k+1}$ for all $k \ge 1$) towards $\{X = \infty\}$, we deduce that $A(\varepsilon) = \big\{B_k \text{ finitely often}\big\} \equiv \Omega_0$. □

Remark 1.5.1. The previous argument shows that the event $\{\omega : X_k(\omega) \to 0\}$ coincides with $\cap_{\varepsilon>0} A(\varepsilon) \equiv \Omega_0$; in other words, the sequence of random variables X_k converges (to zero) with probability one (or almost surely), $\mathsf{P}(X_k \to 0) = 1$.

1.2 Borel-Cantelli lemma

Let $(A_k)_{k\ge1}$ be an infinite sequence of events from some probability space $\big(\Omega, \mathcal{F}, \mathsf{P}\big)$. One is often interested in finding out how many of the events A_n occur.[6] The event that infinitely many of the events A_n occur, written $\big\{A_n \text{ i.o.}\big\}$ or $\big\{A_n \text{ infinitely often}\big\}$, is

$$\big\{A_n \text{ i.o.}\big\} = \bigcap_{n\ge1} \bigcup_{k=n}^{\infty} A_k\,. \tag{1.1}$$

[6] Eg., some results in Number Theory about rational approximations of irrational numbers are formulated in a form similar to Lemma 1.6!

The next result is very important for applications. Its proof uses the intrinsic monotonicity structure of the definition (1.1).

✍ **Lemma 1.6** (Borel-Cantelli lemma). *Let $A = \bigcap_{n\ge1}\bigcup_{k=n}^{\infty} A_k$ be the event that infinitely many of the A_n occur. Then:*

a) *If $\sum_k \mathsf{P}(A_k) < \infty$, then $\mathsf{P}(A) = 0$, ie., with probability one only finitely many of the A_k occur.*

b) *If $\sum_k \mathsf{P}(A_k) = \infty$ and $A_1, A_2, \dots$ are independent events, then $\mathsf{P}(A) = 1$.*

Remark 1.6.1. The independence condition in part b) above cannot be relaxed. Otherwise, let $A_n \equiv E$ for all $n \ge 1$, where $E \in \mathcal{F}$ satisfies $0 < \mathsf{P}(E) < 1$ (and thus the events A_k are not independent). Then $A = E$ and $\mathsf{P}(A) = \mathsf{P}(E) \neq 1$.

Remark 1.6.2. An even more interesting counterexample to part b) without the independence property can be constructed as follows (do this!):
☞ *Let X be a uniform random variable on $(0,1)$, write $X \sim \mathcal{U}(0,1)$. For $n \ge 1$, consider the event $A_n = \bigl\{X < 1/n\bigr\}$. It is easy to see that $A = \bigl\{A_n \text{ i.o.}\bigr\} = \varnothing$, so that one can have $\sum_n \mathsf{P}(A_n) = \infty$ together with $\mathsf{P}(A) = \mathsf{P}(A_n \text{ i.o.}) = 0$.*

Example 1.7 (Infinite monkey theorem). *By the second Borel-Cantelli lemma, Lemma 1.6b), a monkey hitting keys at random on a typewriter keyboard for an infinite amount of time will almost surely (ie., with probability one) type any particular chosen text, such as the complete works of William Shakespeare (and, in fact, infinitely many copies of the chosen text).*
Idea of the argument. Suppose that the typewriter has 50 keys, and the word to be typed is 'banana'. The chance that the first letter typed is b is $1/50$, as is the chance that the second letter is a, and so on. These events are independent, so the chance of the first six letters matching 'banana' is $1/50^6$. For the same reason the chance that the next six letters match 'banana' is also $1/50^6$, and so on.
Now, the chance of not typing 'banana' in each block of six letters is $1-1/50^6$. Because each block is typed independently, the chance of not typing 'banana' in any of the first n blocks of six letters[7] is $\mathsf{p} = (1-1/50^6)^n$. If we were to count occurences of 'banana' that crossed blocks, p would approach zero even more quickly.[8] Finally, once the first copy of the word 'banana' appears, the process starts afresh independently of the past, so that the probability of obtaining the second copy of the word 'banana' within the same number of blocks is still p etc.; the result now follows from Lemma 1.6.
Of course, the same argument applies if the monkey were typing any other string of characters of finite length, eg., your favourite novel.[9]

[7] As n grows, p gets smaller. For $n = 10^6$, p is more than 99.99%, but for $n = 10^{10}$ the probability p is about 52.73% and for an $n = 10^{11}$ it is about 0.17%. As n goes to infinity, the probability p can be made as small as one likes.

[8] Using the theory of Markov chains, discussed later in the course, you should be able to show that the expected hitting time of the word 'banana' is exactly $50^6 \approx 1.5625 \cdot 10^{10}$.

[9] You can use the R script available from the course webpage to explore sequences of different length and/or different typewriters.

☞ **Remark 1.7.1.** By using an appropriate monotone approximation, one can deduce the result as in Example 1.4, without explicitly using the Borel-Cantelli lemma. Moreover, the same argument can be extended to the situations, when the probability p_n of typing 'banana' in the nth block of six letters varies with n, but remains uniformly positive, ie., $\mathsf{p}_n \ge \delta > 0$ for all $n \ge 1$. The true power of the lemma is seen in the situations when $\mathsf{p}_n \to 0$ slowly enough to have $\sum_n \mathsf{p}_n = \infty$ (provided the events in different blocks are independent).

Proof of Lemma 1.6. a) For every $n \ge 1$, let $B_n \stackrel{\text{def}}{=} \cup_{k=n}^{\infty} A_k$ be the event that at least one of A_k with $k \ge n$ occurs. Since $A \subset B_n$ for all $n \ge 1$, we have

$$\mathsf{P}(A) \le \mathsf{P}(B_n) \le \sum_{k=n}^{\infty} \mathsf{P}(A_k) \to 0$$

as $n \to \infty$, whenever $\sum_k \mathsf{P}(A_k) < \infty$.

b) The event $A^{\mathsf{c}} = \big\{ A_n \text{ occur finitely often} \big\}$ is related to the sequence

$$B_n^{\mathsf{c}} = \cap_{k=n}^{\infty} A_k^{\mathsf{c}} \equiv \big\{ \text{ none of } A_k,\ k \ge n, \text{ occurs} \big\}$$

via

$$A^{\mathsf{c}} = \bigcup_n \bigcap_{k=n}^{\infty} A_k^{\mathsf{c}} = \bigcup_n B_n^{\mathsf{c}} ,$$

so it is sufficient to show that $\mathsf{P}(B_n^{\mathsf{c}}) = 0$ for all $n \ge 1$. By independence and the elementary inequality $1 - x \le e^{-x}$ with $x \ge 0$, we get

$$\mathsf{P}\Big(\bigcap_{k=n}^{m} A_k^{\mathsf{c}} \Big) = \prod_{k=n}^{m} \mathsf{P}\big(A_k^{\mathsf{c}}\big) = \prod_{k=n}^{m} \Big(1 - \mathsf{P}\big(A_k\big) \Big) \le \exp\Big\{ - \sum_{k=n}^{m} \mathsf{P}\big(A_k\big) \Big\}$$

so that

$$\mathsf{P}\big(B_n^{\mathsf{c}}\big) = \lim_{m \to \infty} \mathsf{P}\Big(\bigcap_{k=n}^{m} A_k^{\mathsf{c}} \Big) \le \exp\Big\{ - \sum_{k=n}^{\infty} \mathsf{P}\big(A_k\big) \Big\} = 0 ,$$

as the sum diverges. □

Example 1.8. *A standard six-sided die is tossed repeatedly. Let N_k denote the total number of tosses when face k was observed. Assuming that the individual outcomes are independent, show that*

$$\mathsf{P}(N_1 = \infty) = \mathsf{P}(N_2 = \infty) = \mathsf{P}(N_1 = \infty, N_2 = \infty) = 1 .$$

Solution. Equalities $\mathsf{P}(N_1 = \infty) = \mathsf{P}(N_2 = \infty) = 1$ can be derived as in Example 1.4, so that the intersection event $\{N_1 = \infty, N_2 = \infty\}$ has probability one.

Alternatively, we derive the first equality from the Borel-Cantelli lemma. To this end, fix $k \in \{1, 2, \dots, 6\}$ and denote $A_n^k = \big\{ n\text{th toss shows } k \big\}$. For different n, the events A_n^k are independent and have the same probability $1/6$. Since $\sum_n \mathsf{P}(A_n^k) = \infty$, the Borel-Cantelli lemma implies that the event $\big\{ N_k = \infty \big\} \equiv \big\{ A_n^k \text{ infinitely often} \big\}$ has probability one. The remaining claims now follow as indicated above. □

Example 1.9. *A coin showing 'heads' with probability p is tossed repeatedly. With X_n denoting the result of the nth toss, let $C_n = \{X_n = \mathsf{T}, X_{n-1} = \mathsf{H}\}$. Show that $\mathsf{P}(C_n \text{ i.o.}) = 1$.*

Solution. We have $\{C_{2n} \text{ i.o.}\} \subset \{C_n \text{ i.o.}\}$, where $\mathsf{P}(C_{2n}) \equiv pq$ and C_{2n} are independent. The result follows from Lemma 1.6b) (or via monotone approximation). $\square$

The Borel-Cantelli lemma is often used, when one needs to describe long-term behaviour of sequences of random variables.

Example 1.10. *Let $(X_k)_{k\ge1}$ be i.i.d. random variables with common exponential distribution of mean $1/\lambda$, i.e., $\mathsf{P}(X_1 > x) = e^{-\lambda x}$ for all $x \ge 0$. One can show that X_n grows like $\frac{1}{\lambda}\log n$, more precisely, that*[10]

$$\mathsf{P}\Big(\limsup_{n\to\infty} \tfrac{X_n}{\log n} = \tfrac{1}{\lambda}\Big) = 1\,.$$

Solution. For $\varepsilon > 0$, denote

$$A_n^\varepsilon \stackrel{\mathsf{def}}{=} \Big\{\omega : X_n(\omega) > \tfrac{1+\varepsilon}{\lambda}\log n\Big\}\,, \qquad B_n^\varepsilon \stackrel{\mathsf{def}}{=} \Big\{\omega : X_n(\omega) > \tfrac{1-\varepsilon}{\lambda}\log n\Big\}\,.$$

We clearly have $\mathsf{P}(A_n^\varepsilon) = n^{-(1+\varepsilon)}$ and $\mathsf{P}(B_n^\varepsilon) = n^{-(1-\varepsilon)}$. Since $\sum_n \mathsf{P}(A_n^\varepsilon) < \infty$, by Lemma 1.6a) the event $\{A_n^\varepsilon \text{ infinitely often}\}$ has probability zero. Similarly, the events B_n^ε are independent and $\sum_n \mathsf{P}(B_n^\varepsilon) = \infty$, thus, by Lemma 1.6b), the event $\{B_n^\varepsilon \text{ infinitely often}\}$ has probability one. $\square$

Remark 1.10.1. (Records) A slightly more general version of the argument from Example 1.10 helps to control the limiting behaviour of records:[11]
Let $(X_k)_{k\ge1}$ be i.i.d. exponential r.v. with distribution $\mathsf{P}(X_k > x) = e^{-x}$, and let $M_n \stackrel{\mathsf{def}}{=} \max_{1\le k\le n} X_k$. Then $\mathsf{P}(M_n/(\log n) \to 1) = 1$, ie., the normalized maximum $M_n/(\log n)$ converges to one almost surely (as $n \to \infty$).

Example 1.11. *Let random variables $(X_n)_{n\ge1}$ be i.i.d. with $X_1 \sim \mathcal{U}[0,1]$. For $\alpha > 0$, we have $\mathsf{P}(X_n > 1 - n^{-\alpha}) = n^{-\alpha}$, so that $\mathsf{P}(X_n > 1 - n^{-\alpha} \text{ i.o. }) = 1$ iff $\alpha \le 1$. A similar analysis shows that*

$$\mathsf{P}\Big(X_n > 1 - \frac{1}{n(\log n)^\beta} \text{ i.o. }\Big) = \begin{cases} 1\,, & \beta \le 1\,,\\ 0\,, & \beta > 1\,.\end{cases}$$

Lemma 1.6 is one of the main methods of proving almost sure convergence:

Example 1.12. *If $(X_k)_{k\ge1}$ is a sequence of random variables such that for every $\varepsilon > 0$ the event $A(\varepsilon) \equiv \{|X_k| > \varepsilon \text{ finitely often}\}$ has probability one, then X_k is said to converge to zero with probability one, recall Remark 1.5.1. A simple example is with $X_k = \frac{1}{k}X$ for a variable $X \ge 0$ of finite mean, $\mathsf{E}X < \infty$. One then can show that $\sum_{k\ge1}\mathsf{P}(|X_k| > \varepsilon) = \sum_{k\ge1}\mathsf{P}(X > k\varepsilon) < \infty$, and thus the result follows from Lemma 1.6, see also Lemma 2.15 below.*

[10] Recall that for a real sequence $(a_n)_{n\ge1}$ one defines $\limsup\limits_{n\to\infty} a_n$ as the largest limiting point of the sequence $(a_n)_{n\ge1}$, equivalently, $\limsup\limits_{n\to\infty} a_n \equiv \lim\limits_{n\to\infty}\sup_{k\ge n} a_k$, see App. A below.

[11] Similar results hold for other distributions, see page E6 in the Problems Sheets.