2.5. **Tail events.** Let $(X_n : n \in \mathbb{N})$ be a sequence of random variables. Define

$$\mathcal{T}_n = \sigma(X_{n+1}, X_{n+2}, \dots), \quad \mathcal{T} = \bigcap_n \mathcal{T}_n.$$

Then $\mathcal{T}$ is a σ-algebra, called the *tail σ-algebra* of $(X_n : n \in \mathbb{N})$. It contains the events which depend only on the limiting behaviour of the sequence.

Theorem 2.5.1 (Kolmogorov's zero-one law)**.** *Suppose that $(X_n : n \in \mathbb{N})$ is a sequence of independent random variables. Then the tail σ-algebra $\mathcal{T}$ of $(X_n : n \in \mathbb{N})$ contains only events of probability 0 or 1. Moreover, any $\mathcal{T}$-measurable random variable is almost surely constant.*

Proof. Set $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Then $\mathcal{F}_n$ is generated by the π-system of events

$$A = \{X_1 \le x_1, \dots, X_n \le x_n\}$$

whereas $\mathcal{T}_n$ is generated by the π-system of events

$$B = \{X_{n+1} \le x_{n+1}, \dots, X_{n+k} \le x_{n+k}\}, \quad k \in \mathbb{N}.$$

We have $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all such A and B, by independence. Hence $\mathcal{F}_n$ and $\mathcal{T}_n$ are independent, by Theorem 1.12.1. It follows that $\mathcal{F}_n$ and $\mathcal{T}$ are independent. Now $\bigcup_n \mathcal{F}_n$ is a π-system which generates the σ-algebra $\mathcal{F}_\infty = \sigma(X_n : n \in \mathbb{N})$. So by Theorem 1.12.1 again, $\mathcal{F}_\infty$ and $\mathcal{T}$ are independent. But $\mathcal{T} \subseteq \mathcal{F}_\infty$. So, if $A \in \mathcal{T}$,

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A)$$

so $\mathbb{P}(A) \in \{0, 1\}$.

Finally, if Y is any $\mathcal{T}$-measurable random variable, then $F_Y(y) = \mathbb{P}(Y \le y)$ takes values in $\{0, 1\}$, so $\mathbb{P}(Y = c) = 1$, where $c = \inf\{y : F_Y(y) = 1\}$. □

2.6. **Convergence in measure and convergence almost everywhere.**
Let $(E, \mathcal{E}, \mu)$ be a measure space. A set $A \in \mathcal{E}$ is sometimes defined by a property shared by its elements. If $\mu(A^c) = 0$, then we say that property holds *almost everywhere* (or *a.e.*). The alternative *almost surely* (or *a.s.*) is often used in a probabilistic context. Thus, for a sequence of measurable functions $(f_n : n \in \mathbb{N})$, we say *f_n converges to f a.e.* to mean that

$$\mu(\{x \in E : f_n(x) \not\to f(x)\}) = 0.$$

If, on the other hand, we have that

$$\mu(\{x \in E : |f_n(x) - f(x)| > \varepsilon\}) \to 0, \quad \text{for all } \varepsilon > 0,$$

then we say *f_n converges to f in measure* or, in a probabilistic context, *in probability.*

Theorem 2.6.1. *Let $(f_n : n \in \mathbb{N})$ be a sequence of measurable functions.*

(a) *Assume that $\mu(E) < \infty$. If $f_n \to 0$ a.e. then $f_n \to 0$ in measure.*
(b) *If $f_n \to 0$ in measure then $f_{n_k} \to 0$ a.e. for some subsequence (n_k).*

Proof. (a) Suppose $f_n \to 0$ a.e.. For each $\varepsilon > 0$,

$$\mu(|f_n| \leq \varepsilon) \geq \mu\left(\bigcap_{m\geq n}\{|f_m| \leq \varepsilon\}\right) \uparrow \mu(|f_n| \leq \varepsilon \text{ ev.}) \geq \mu(f_n \to 0) = \mu(E).$$

Hence $\mu(|f_n| > \varepsilon) \to 0$ and $f_n \to 0$ in measure.

(b) Suppose $f_n \to 0$ in measure, then we can find a subsequence (n_k) such that

$$\sum_k \mu(|f_{n_k}| > 1/k) < \infty.$$

So, by the first Borel–Cantelli lemma,

$$\mu(|f_{n_k}| > 1/k \text{ i.o.}) = 0$$

so $f_{n_k} \to 0$ a.e.. □

2.7. **Large values in sequences of IID random variables.** Consider a sequence $(X_n : n \in \mathbb{N})$ of independent random variables, all having the same distribution function F. Assume that $F(x) < 1$ for all $x \in \mathbb{R}$. Then, almost surely, the sequence $(X_n : n \in \mathbb{N})$ is unbounded above, so $\limsup_n X_n = \infty$. A way to describe the occurrence of large values in the sequence is to find a function $g : \mathbb{N} \to (0, \infty)$ such that, almost surely,

$$\limsup_n X_n/g(n) = 1.$$

We now show that $g(n) = \log n$ is the right choice when $F(x) = 1 - e^{-x}$. The same method adapts to other distributions.

Fix $\alpha > 0$ and consider the event $A_n = \{X_n \geq \alpha \log n\}$. Then $\mathbb{P}(A_n) = e^{-\alpha \log n} = n^{-a}$, so the series $\sum_n \mathbb{P}(A_n)$ converges if and only if $\alpha > 1$. By the Borel–Cantelli Lemmas, we deduce that, for all $\varepsilon > 0$,

$$\mathbb{P}(X_n/\log n \geq 1 \text{ i.o.}) = 1, \quad \mathbb{P}(X_n/\log n \geq 1 + \varepsilon \text{ i.o.}) = 0.$$

Hence, almost surely,

$$\limsup_n X_n/\log n = 1.$$