2.2. INDEPENDENCE AND STATIONARITY

For a fixed event F in a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with $\mathbf{P}(F) > 0$, we define the *conditional probability* measure *given* F by

$$\mathbf{P}_F(E) := \frac{\mathbf{P}(E \cap F)}{\mathbf{P}(F)}, \quad E \in \mathcal{F}. \tag{2.1}$$

Now suppose that for some event $E \in \mathcal{F}$ we have $\mathbf{P}_F(E) = \mathbf{P}(E)$, i.e., that knowledge about the occurrence (or not) of the event F does not change the probability assigned by the measure $\mathbf{P}$ to E. Or equivalently, that

$$\mathbf{P}(E \cap F) = \mathbf{P}(E)\,\mathbf{P}(F), \tag{2.2}$$

a relation which is symmetric in E and F and unambiguous, even when the probabilities $\mathbf{P}(E)$ or $\mathbf{P}(F)$ vanish.

We say that the two events E, F are **independent,** if (2.2) holds. It is interesting to check that, if this is the case, then E, F^c (and E^c, F as well as E^c, F^c), are also independent.

For instance, if $\Omega \to \Omega$ is measure preserving and weakly mixing (Exercise 1.9.6), then (2.2) holds for every $F \in \mathcal{F}$ and every $T-$invariant set $E \in \mathcal{F}$ (i.e., for which $T^{-1}E = E$ holds mod. $\mathbf{P}$).

2.1 DEFINITION: INDEPENDENT EVENTS, RANDOM VARIABLES.

(i) The events in an arbitrary family $\mathcal{E} = \{E_\alpha\}_{\alpha \in A}$ are said to be *independent,* if

$$\mathbf{P}\left(\bigcap_{j=1}^{n} E_{\alpha_j}\right) = \prod_{j=1}^{n} \mathbf{P}(E_{\alpha_j})$$

holds for any $n \in \mathbf{N}$ and any $\{\alpha_1, \cdots, \alpha_n\} \subseteq A$.

(ii) The random variables in a family $\{X_\alpha\}_{\alpha \in A}$ are said to be *independent,* if the events $\{X_\alpha^{-1}(B_\alpha)\}_{\alpha \in A}$ are independent for any family of Borel subsets $\{B_\alpha\}_{\alpha \in A} \subseteq \mathcal{B}(\mathbf{R})$.

The condition (ii) is equivalent to $\mu_{X_{\alpha_1}, \cdots, X_{\alpha_n}} = \otimes_{j=1}^{n} \mu_{X_{\alpha_j}}$, or to the condition

$$F_{X_{\alpha_1}, \cdots, X_{\alpha_n}}(x_1, \cdots, x_n) = F_{X_{\alpha_1}}(x_1) \cdots F_{X_{\alpha_n}}(x_n), \qquad \forall\, (x_1, \cdots, x_n) \in \mathbf{R}^n$$

for any $n \in \mathbf{N}$ and indices $\{\alpha_1, \cdots, \alpha_n\} \subseteq A^n$.

(iii) Two collections $\mathcal{G}$ and $\mathcal{H}$ of events are called *independent*, if for every $G \in \mathcal{G}$ and $H \in \mathcal{H}$ we have $\mathbf{P}(G \cap H) = \mathbf{P}(G) \cdot \mathbf{P}(H)$.

More generally, suppose $\mathcal{E}^{(1)}, \mathcal{E}^{(2)}, \cdots$ are collections of events; we say that *these collections are independent*, if for any $n \in \mathbf{N}$, any distinct integers $i_1, \cdots, i_n$, and any events $E_j \in \mathcal{E}^{(j)}$, $j \in \{i_1, \cdots, i_n\}$ we have

$$\mathbf{P}\left(E_{i_1} \cap \cdots \cap E_{i_n}\right) = \prod_{k=1}^{n} \mathbf{P}(E_{i_k}).$$

It is easily verified that if $\{E_\alpha\}_{\alpha \in A}$ are independent events, then $\{F_\alpha\}_{\alpha \in A}$ are also independent, where F_α can be either E_α or E_α^c.

Similarly, if $\{X_\alpha\}_{\alpha \in A}$ are independent random variables and $\{f_\alpha\}_{\alpha \in A}$ are Borel-measurable functions, then $\{f_\alpha(X_\alpha)\}_{\alpha \in A}$ are also independent. Indeed, if $\{\alpha_j\}_{j=1}^n \subseteq A$ is any set of n indices, and $\{B_j\}_{j=1}^n$ are Borel subsets in $\mathbf{R}$, then

$$\mathbf{P}\left[\bigcap_{j=1}^{n} \{f_{\alpha_j}(X_{\alpha_j}) \in B_j\}\right] = \mathbf{P}\left[\bigcap_{j=1}^{n} \{X_{\alpha_j} \in f_{\alpha_j}^{-1}(B_j)\}\right]$$
$$= \prod_{j=1}^{n} \mathbf{P}\left(X_{\alpha_j} \in f_{\alpha_j}^{-1}(B_j)\right) = \prod_{j=1}^{n} \mathbf{P}\left[f_{\alpha_j}(X_{\alpha_j}) \in A_j\right].$$

2.1 Lemma. *Suppose that $\mathcal{G}$ and $\mathcal{H}$ are sub-σ-algebras of $\mathcal{F}$; and that $\mathcal{I}$ and $\mathcal{J}$ are π-systems (Exercise 1.3.8) such that $\mathcal{G} = \sigma(\mathcal{I})$, $\mathcal{H} = \sigma(\mathcal{J})$.*

Then $\mathcal{G}$ and $\mathcal{H}$ are independent, if and only if $\mathcal{I}$ and $\mathcal{J}$ are independent.

Proof: Suppose $\mathcal{I}$ and $\mathcal{J}$ are independent; for any given $I \in \mathcal{I}$, the set-functions

$$H \mapsto \mathbf{P}(I \cap H), \qquad H \mapsto \mathbf{P}(I) \cdot \mathbf{P}(H)$$

agree on $\mathcal{J}$, are measures on $(\Omega, \mathcal{H})$, and have the same total mass $\mathbf{P}(I)$. By Exercise 1.3.8 they agree on $\mathcal{H} = \sigma(\mathcal{J})$, that is: $\mathbf{P}(I \cap H) = \mathbf{P}(I) \cdot \mathbf{P}(H)$ for every $H \in \mathcal{H}$.

Thus, for any given $H \in \mathcal{H}$, the set-functions

$$G \mapsto \mathbf{P}(G \cap H), \qquad G \mapsto \mathbf{P}(G) \cdot \mathbf{P}(H)$$

agree on $\mathcal{I}$; they are measures on $(\Omega, \mathcal{G})$ and have the same total mass $\mathbf{P}(H)$. By the same token as above, they they agree on $\mathcal{G} = \sigma(\mathcal{I})$, to wit: $\mathbf{P}(G \cap H) = \mathbf{P}(G) \cdot \mathbf{P}(H)$ for every $G \in \mathcal{G}$. $\diamond$

2.1 Exercise: If $\{X_\alpha\}_{\alpha \in A}$ is a family of independent random variables, then the $\sigma-$algebras generated by disjoint subfamilies are independent.

2.2 Exercise: (a) Show by example, that three events can be pairwise-independent, but not independent in the sense of Definition 2.1.(i).
(b) Any two square-integrable random variables that are independent, are also uncorrelated. However, show by example that two random variables can be uncorrelated, *without* being independent.
(c) Show that for random variables $X_1, \cdots, X_n$ in $\mathbf{L}^2$, we have

$$\mathrm{Var}(\textstyle\sum_{j=1}^n X_j) = \sum_{j=1}^n \mathrm{Var}(X_j) + 2\sum_{j=1}^n \sum_{i=j+1}^n \mathrm{Cov}(X_i, X_j)\,.$$

(d) Observe also, that if the $X_j = \chi_{A_j}$, $j = 1, \cdots, n$ are indicators, then

$$\mathrm{Var}(\textstyle\sum_{j=1}^n X_j) \le \mathbf{E}(\sum_{j=1}^n X_j) + 2\sum_{j=1}^n \sum_{i=j+1}^n \mathrm{Cov}(X_i, X_j)\,.$$

2.1 REMARK: In the context and with the notation of Example 1.6.2, let us take the $n^{th}-$coördinate mapping $X_n(\omega) \equiv \omega_n$ on the canonical space $(\Omega, \mathcal{F}) = (\mathbf{R}^{\mathbf{N}}, \sigma(\mathcal{C}^*))$, for each $n \in \mathbf{N}$. This way we create a *sequence of random variables* $X_1, X_2, \cdots$ *with prescribed probability distribution functions* $F_1, F_2, \cdots$, respectively, under the probability measure $\mathbf{P}$ of (1.6.16).

If all these distributions are the same $F_n \equiv F$, $\forall\, n \in \mathbf{N}$, then we say that the sequence $X_1, X_2, \cdots$ consists of **independent, identically distributed (I.I.D.) random variables, under P**.

2.1 EXAMPLE : TOSSING A COIN. Let us place ourselves on the space $(\Omega, \mathcal{F}) = (\{0,1\}^{\mathbf{N}}, \sigma(\mathcal{C}^*))$, consisting of sequences $\omega = (\omega_1, \omega_2, \cdots)$ with $\omega_j = 0$ or 1 for every $j \in \mathbf{N}$. Such an ω can be visualized as an infinite sequence of tosses of a coin, with the outcome "heads" (success) represented by 1, and the outcome "tails" (failure) represented by 0. Suppose also that we assign the **Bernoulli distribution**

$$\mathbf{P}[X_j = 1] = p\,, \quad \mathbf{P}[X_j = 0] = 1 - p =: q\,, \quad \text{for some } \ p \in (0,1) \tag{2.3}$$

to each of the coördinate mappings $X_j(\omega) = \omega_j$, $j \in \mathbf{N}$. Indeed, according to Example 1.6.2 and Remark 2.1 (see also Example 2.2 below), there exists a probability measure $\mathbf{P}$ on the space $(\Omega, \mathcal{F})$ under which the random variables $X_1, X_2, \cdots$ are independent, with common Bernoulli distribution $\mathbf{P} \circ X_j^{-1} \equiv \mathbf{P} \circ X_1^{-1}$ as in (2.3), $\forall\, j \in \mathbf{N}$. This corresponds to the intuitive notion that "different tosses of the coin are independent".

Consider now the number of successes ("heads") $S_n(\omega) = \sum_{j=1}^n X_j(\omega)$ obtained during the first n tosses in the realization $\omega \in \Omega$ of our experiment. It is not hard to see that this random variable has the **Binomial disribution**

$$\mathbf{P}[S_n = k] \;=\; \frac{n!}{k!(n-k)!}\; p^k (1-p)^{n-k}\,, \qquad k = 0, \cdots, n\,. \tag{2.4}$$

It is also intuitively plausible that, as the number of tosses n becomes large, the proportion of successes

$$\overline{X}_n(\omega) := \frac{S_n(\omega)}{n} \quad \text{should converge, in some sense, to the number } \; p\,, \tag{2.5}$$

the probability of success on each individual trial of our coin-tossing experiment. A statement of the form (2.5) is a prototypical *Law of Large Numbers;* it will be justified in Exercise 2.3 and Example 2.5 below. We discuss this issue more thoroughly in the next section.

2.1 THEOREM : BOREL-CANTELLI LEMMATA. *For a sequence* $\{E_n\}_{n\in\mathbf{N}} \subseteq \mathcal{F}$ *of measurable sets in a measure space* $(\Omega, \mathcal{F}, \mathbf{P})$*, we have*

(i) $\mathbf{P}(E_n,\ \text{i.o.}) = \mathbf{P}(\limsup_n E_n) = 0\,,$ if $\sum_{n\in\mathbf{N}} \mathbf{P}(E_n) < \infty\,.$

(ii) *If* $\mathbf{P}(\Omega) = 1\,,$ $\sum_{n\in\mathbf{N}} \mathbf{P}(E_n) = \infty\,,$ *and the events* $\{E_n\}_{n\in\mathbf{N}}$ *are independent, then*

$$\mathbf{P}(E_n,\ i.o.) \;=\; \mathbf{P}\left(\limsup_n \, E_n\right) \;=\; 1$$

or equivalently $\sum_{n\in\mathbf{N}} \chi_{E_n} = \infty$ *a.e.*

In other words: if the measures of the sets decrease "very rapidly to zero", we cannot expect to "see too many of these sets"!

And if the probabilities of *independent* events "do not decrease too rapidly to zero", we can expect to "see these events realized quite often".

PROOF : (i) Recall that $\{E_n,\,\text{i.o.}\} := \limsup_n E_n = \bigcap_{k\in\mathbf{N}} F_k\,,$ where $F_k := \cup_{n\geq k} E_n$. Clearly, $\{F_k\}_{k\in\mathbf{N}}$ is a decreasing sequence and F_1 has finite measure, by assumption: $\mu(F_1) \leq \sum_{n\in\mathbf{N}} \mu(E_n) < \infty\,.$ Then the continuity-from-above property (1.2.15) implies $\mathbf{P}(\limsup_n E_n) = \mathbf{P}\left(\bigcap_{k\in\mathbf{N}} F_k\right) = \lim_{k\to\infty}\mathbf{P}(F_k) \leq \lim_{k\to\infty} \sum_{n\geq k} \mathbf{P}(E_n) = 0$. This proves (i).

(ii) On the other hand, for any $1 \leq k < m\,,$ we have

$$1 - \mathbf{P}\left(\bigcup_{n=k}^{m} E_n\right) = \mathbf{P}\left(\bigcap_{n=k}^{m} E_n^c\right) = \prod_{n=k}^{m} \mathbf{P}(E_n^c) = \prod_{n=k}^{m} (1 - \mathbf{P}(E_n)) \leq \exp\left(-\sum_{n=k}^{m} \mathbf{P}(E_n)\right),$$

where we have used independence and the elementary inequality $1-x \leq e^{-x}$ for $0 \leq x \leq 1$. Since $\sum_{n\in\mathbf{N}} \mathbf{P}(E_n) = \infty$, we find by letting $m \to \infty$ that $1-\mathbf{P}(F_k) = 1-\mathbf{P}\left(\bigcup_{n\geq k} E_n\right) = 0\,,$ for every $k \in \mathbf{N}$. But then the finiteness of the measure gives $\mathbf{P}(\limsup_n E_n) = \lim_{k\to\infty}\mathbf{P}(F_k) = 1\,,$ and (ii) is proved.

2.2 THEOREM: (i) *If the random variables* $X_1, \cdots, X_n$ *are integrable and independent, then* $\prod_{j=1}^n X_j \in \mathbf{L}^1$ *and we have*

$$\mathbf{E}\left(\prod_{j=1}^n X_j\right) = \prod_{j=1}^n \mathbf{E}(X_j)\,.$$

(ii) *If the random variables* $X_1, \cdots, X_n$ *are square-integrable and pairwise-independent, then they are also pairwise-uncorrelated and we have:* $\mathrm{Var}\Big(\sum_{j=1}^n X_j\Big) = \sum_{j=1}^n \mathrm{Var}(X_j)$.

Proof: With $f(x_1, \cdots, x_n) = |x_1 \cdots x_n|$ we have, from Tonelli's theorem:

$$\mathbf{E}\left(\prod_{j=1}^n |X_j|\right) = \int_{\mathbf{R}^n} f \; d\left(\bigotimes_{j=1}^n \mu_{X_j}\right) = \prod_{j=1}^n \left(\int_{\mathbf{R}} |x_j|\, d\mu_{X_j}(x_j)\right) = \prod_{j=1}^n \mathbf{E}(|X_j|) < \infty\,,$$

so that $\prod_{j=1}^n X_j \in \mathbf{L}^1$; now apply Fubini's theorem (same argument, with absolute values removed).

For part (ii), observe that the random variables $\xi_j := X_j - \mathbf{E}(X_j),\ j = 1, \cdots, n$ have zero expectation and are pairwise independent, thus $\mathbf{E}(\xi_j \xi_k) = \mathbf{E}(\xi_j)\,\mathbf{E}(\xi_k) = 0$ for $j \neq k$; therefore, the variables $X_1, \ldots, X_n$ are pairwise-uncorrelated, and

$$\mathrm{Var}\left(\sum_{j=1}^n X_j\right) = \mathbf{E}\left(\sum_{j=1}^n \xi_j\right)^2 = \sum_{j=1}^n \mathbf{E}(\xi_j^2) + 2\sum_{j=1}^n \sum_{k=j+1}^n \mathbf{E}(\xi_j \xi_k) = \sum_{j=1}^n \mathrm{Var}(\xi_j)\,. \qquad \diamond$$

A: INSTANCES OF INDEPENDENCE

Here are a few examples of situations, where independence arises quite naturally and, sometimes, unexpectedly.

2.2 EXAMPLE : RADEMACHER FUNCTIONS. It is well known that every number $\omega \in [0, 1)$ has a binary expansion

$$\omega = \frac{\varepsilon_1}{2} + \frac{\varepsilon_2}{2^2} + \cdots + \frac{\varepsilon_n}{2^n} + \cdots$$

where each ε is either 0 or 1. In order to ensure the uniqueness of this expansion, we postulate that only expansions with infinitely many digits "0" are to be used; for instance, let us agree to write $\frac{3}{4}$ as $\frac{1}{2}+\frac{1}{2^2}+\frac{0}{2^3}+\frac{0}{2^4}+\cdots$ rather than $\frac{1}{2}+\frac{0}{2^2}+\frac{1}{2^3}+\frac{1}{2^4}+\cdots$. With this recipe the digits $\varepsilon \in \{0,1\}$ become functions of ω, so let us write more appropriately

$$\omega \;=\; \frac{\varepsilon_1(\omega)}{2}+\frac{\varepsilon_2(\omega)}{2^2}+\cdots+\frac{\varepsilon_n(\omega)}{2^n}+\cdots$$

or equivalently

$$\underbrace{1-2\omega \;=\; \sum_{k\in\mathbf{N}} \frac{r_k(\omega)}{2^k}}\,, \qquad \text{where} \quad r_k(\omega) \;:=\; 1-2\,\varepsilon_k(\omega)\,, \qquad k\in\mathbf{N}$$

are the so-called *Rademacher functions.*

For instance, $\varepsilon_1(\omega)=0$ for $0\le\omega<\frac{1}{2}$ and $\varepsilon_1(\omega)=1$ for $\frac{1}{2}\le\omega<1$; similarly, $\varepsilon_2(\omega)=0$ for $0\le\omega<\frac{1}{4}$ or $\frac{1}{2}\le\omega<\frac{3}{4}$, and $\varepsilon_2(\omega)=1$ for $\frac{1}{4}\le\omega<\frac{1}{2}$ or $\frac{3}{4}\le\omega<1$; and so on. (Plot the first four functions in this sequence!)

Now let us do the simple trigonometry

$$\sin x \;=\; 2\sin\left(\frac{x}{2}\right)\cos\left(\frac{x}{2}\right) \;=\; 2^2\sin\left(\frac{x}{4}\right)\cos\left(\frac{x}{4}\right)\cos\left(\frac{x}{2}\right)$$

$$=\;\cdots\;=\;2^n\sin\left(\frac{x}{2^n}\right)\prod_{k=1}^{n}\cos\left(\frac{x}{2^k}\right)$$

and note that $\lim_{n\to\infty} 2^n\sin(x2^{-n})=x$, to obtain the *generalized Vieta formula*

$$\underbrace{\frac{\sin x}{x} \;=\; \prod_{k\in\mathbf{N}}\cos\left(\frac{x}{2^k}\right)}\,.$$

But observe

$$\int_0^1 e^{\,ix(1-2\omega)}\,d\omega \;=\; \frac{\sin x}{x}\,, \qquad \int_0^1 \exp\left\{ix\,\frac{r_k(\omega)}{2^k}\right\}d\omega \;=\; \cos\left(\frac{x}{2^k}\right)\,,$$

so that generalized Vieta formula can be written as

$$\int_0^1 \exp\left\{ix\sum_{k\in\mathbf{N}}\frac{r_k(\omega)}{2^k}\right\}d\omega \;=\; \frac{\sin x}{x} \;=\; \prod_{k\in\mathbf{N}}\cos\left(\frac{x}{2^k}\right) \;=\; \prod_{k\in\mathbf{N}}\int_0^1 \exp\left\{ix\,\frac{r_k(\omega)}{2^k}\right\}d\omega\,.$$

In particular, we get the *Rademacher formula*

$$\underbrace{\int_0^1 \prod_{k\in\mathbf{N}}\exp\left\{ix\,\frac{r_k(\omega)}{2^k}\right\}d\omega \;=\; \prod_{k\in\mathbf{N}}\int_0^1 \exp\left\{ix\,\frac{r_k(\omega)}{2^k}\right\}d\omega}\,,$$

where an integral of products is expressed as a product of integrals, much like the situation of Theorem 2.2.

Is this simply a coincidence, or is it perhaps symptomatic of some underlying "independence" structure? To make some headway, let us endow $\Omega = [0,1)$ with its $\sigma-$algebra of Borel sets $\mathcal{F} = \mathcal{B}([0,1))$ which measures ε_n for all $n \in \mathbf{N}$, and with Lebesgue measure λ. Fix an arbitrary sequence $\{d_j\}_{j\in\mathbf{N}}$ of 0's and 1's, and look at the sets

$$I_j := \{\, \omega \in \Omega \,|\, \varepsilon_j(\omega) = d_j \,\}, \qquad K_n := \bigcap_{j=1}^{n} I_j := \{\, \omega \in \Omega \,|\, \varepsilon_1(\omega) = d_1, \cdots, \varepsilon_n(\omega) = d_n \,\};$$

this latter is the set of numbers ω, in whose binary expansion the first n digits are $d_1, \cdots, d_n$ and the rest are arbitrary:

$$\omega = \frac{d_1}{2} + \frac{d_2}{2^2} + \cdots + \frac{d_n}{2^n} + \frac{\varepsilon_{n+1}(\omega)}{2^{n+1}} + \frac{\varepsilon_{n+2}(\omega)}{2^{n+2}} + \cdots .$$

Clearly, K_n is just an interval of length 2^{-n}, whereas each I_j is an interval of length $1/2$. In other words,

$$\lambda\big(\{\, \omega \in \Omega \,|\, \varepsilon_1(\omega) = d_1, \cdots, \varepsilon_n(\omega) = d_n \,\}\big) = \lambda(K_n) = 2^{-n} = \prod_{j=1}^{n}(1/2)$$

$$= \prod_{j=1}^{n} \lambda(I_j) = \prod_{j=1}^{n} \lambda\big(\{\, \omega \in \Omega \,|\, \varepsilon_j(\omega) = d_j \,\}\big).$$

This is true for every choice of sequence $\{d_j\}_{j\in\mathbf{N}}$ and integer n, so the measurable functions $\varepsilon_1, \varepsilon_2, \cdots$ are independent "coin-tosses", that is, $\mathbf{P}(\varepsilon_j = 0) = \mathbf{P}(\varepsilon_j = 1) = 1/2$. This means, in particular, that recourse to the Daniel-Kolmogorov theorem was not strictly necessary for the construction needed in Example 2.1.

But then the Rademacher functions $r_1, r_2, \cdots$ are also independent; thus so are the functions $\{e^{ix2^{-k}r_k}\}_{k\in\mathbf{N}}$. The Rademacher formula thus becomes a special case of Theorem 2.2. $\diamond$

2.3 EXAMPLE: INVERSIONS IN RANDOM PERMUTATIONS. Let Ω be the symmetric group of all $n!$ distinct permutations $\omega = (\omega_1, \cdots, \omega_n)$ of the integers $(1, \cdots, n)$, and let $\mathbf{P}$ assign probability $1/n!$ to each such permutation. For every $j \in \{1, \cdots, n\}$ and $\omega \in \Omega$, let $X_{nj}(\omega)$ be the number of *inversions* caused by j in ω: to wit, $X_{nj}(\omega) = k$ means that j precedes exactly k $(0 \le k \le j-1)$ of the integers $1, \cdots, j-1$ in the permutation ω. With this notation,

$$S_n(\omega) := \sum_{j=1}^{n} X_{nj}(\omega)$$

is the total number of inversions in the permutation ω. For instance, with $n = 5$ the permutation $\omega = (3, 2, 5, 1, 4)$ of $(1, 2, 3, 4, 5)$ has $X_{n1}(\omega) = 0$, $X_{n2}(\omega) = 1$, $X_{n3}(\omega) = 2$, $X_{n4}(\omega) = 0$ and $X_{n5}(\omega) = 2$, thus $S_n(\omega) = 5$.

We have then the following, rather remarkable, fact: *The random variables* X_{nj}, $j = 1, \cdots, n$ *are independent, and*

$$\mathbf{P}\big(X_{nj} = k\big) = \frac{1}{j} \qquad \textit{for} \;\; k = 0, \cdots, j-1 .$$

The following argument is from Chung (1974): Let us start by observing that the values $X_{n1}(\omega), \cdots, X_{nj}(\omega)$ are determined, as soon as the sites occupied by the integers $1, \cdots, j$ in the permutation ω are known ("allotted"); the sites occupied by the remaining integers do not matter. Given j arbitrary sites among n ordered slots, there are $j!\,(n-j)!$ permutations ω in which the integers $1, \cdots, n$ occupy these sites in some order. Among these permutations, there are $(j-1)!\,(n-j)!$ permutations in which the integer j occupies the $(j-k)^{th}$ site, with the order from left to right, for some given $k \in \{0, 1, \cdots, j-1\}$. With this site fixed, there are $(j-1)!$ ways in which the integers $1, \cdots, j-1$ may occupy the remaining "allotted" sites; and each such way corresponds to *exactly one* of the possible values that the vector $(X_{n1}, \cdots, X_{n,j-1})$ can take.

Let us fix some such value $(c_1, \cdots, c_{j-1})$, and consider all permutations ω in which

(a) the integers $1, \cdots, j$ occupy the "allotted" sites; and

(b) $X_{n1}(\omega) = c_1, \cdots, X_{n,j-1}(\omega) = c_{j-1}, \quad X_{nj}(\omega) = k$.

There are $(n-j)!$ such ω's; thus, the number of ω's that satisfy condition (b) is $\frac{n!}{j!\,(n-j)!} \cdot (n-j)! = n!/j!$.

Now sum up over $k \in \{0, 1, \cdots, j-1\}$ to find the number of ω's in which $X_{n1}(\omega) = c_1, \cdots, X_{n,j-1}(\omega) = c_{j-1}$, namely: $j\,(n!/j!) = n!/(j-1)!$. Therefore,

$$\frac{\mathbf{P}[\,\omega \in \Omega \,:\, X_{n1}(\omega) = c_1, \cdots, X_{n,j-1}(\omega) = c_{j-1}, \quad X_{nj}(\omega) = k\,]}{\mathbf{P}[\,\omega \in \Omega \,:\, X_{n1}(\omega) = c_1, \cdots, X_{n,j-1}(\omega) = c_{j-1}\,]} = \frac{\frac{n!}{j!}}{\frac{n!}{(j-1)!}} = \frac{1}{j},$$

proving the claim.

2.4 EXAMPLE: RANKS AND RECORDS. Suppose $X_1, X_2, \cdots$ are independent random variables with *common* distribution function $F(\cdot)$ which is *continuous*. Consider the event

$$A_k := \Big\{ X_k > \max_{1 \le j \le k-1} X_j \Big\}$$

that "a record is set on day $t = k$", the number $W_n := \sum_{k=1}^{n} \chi_{A_k}$ of records set by day $t = n$", as well as the random variable

$$R_n := 1 + \sum_{j=1}^{n-1} \chi_{\{X_n < X_j\}}$$

which stands for the relative rank of the random variable X_n among $X_1, \cdots, X_n$. Clearly, $A_n = \{R_n = 1\}$.

We claim that the events $\{A_k\}_{k\in\mathbf{N}}$ are *independent*, with $\mathbf{P}(A_k) = 1/k$, $k \in \mathbf{N}$. Similarly, the random variables $\{R_n\}_{n\in\mathbf{N}}$ are *also independent*, with

$$\mathbf{P}(R_n = \varrho) = \frac{1}{n}, \quad \varrho = 1, \cdots, n.$$

Indeed, because $F(\cdot)$ is continuous, we have $\mathbf{P}(X_1 = X_2) = 0$, and in fact we have $\mathbf{P}\big(\bigcup_{m\neq n}\{X_n = X_m\}\big) = 0$; consult Exercise 2.4(ii). (Nothing here depends on the particular form of $F(\cdot)$, as long as this distribution function is continuous.) For fixed $n \in \mathbf{N}$, list the variables $X_1, \cdots, X_n$ in decreasing order

$$\max_{1\le i\le n} X_i =: Y_1^{(n)} > Y_2^{(n)} > \cdots > Y_n^{(n)} := \min_{1\le i\le n} X_i$$

and define the random permutation $\pi^{(n)} = \big(\pi_1^{(n)} \cdots, \pi_n^{(n)}\big)$ of $(1, \cdots, n)$ as $\pi_i^{(n)} = r$ if $X_i = Y_r^{(n)}$; to wit, if the random variable X_i has relative rank r among $X_1, \cdots, X_n$. There are $n!$ permutations of $(1, \cdots, n)$, each of them corresponding to a particular ordering of the $X_1, \cdots, X_n$ (relative rankings $r_1, \cdots, r_n$). Each particular configuration $\{R_1 = \varrho_1, \cdots, R_n = \varrho_n\}$ determines uniquely an ordering of $X_1, \cdots, X_n$. Because of our assumptions, all such permutations are equally likely and we have

$$\mathbf{P}\big(\pi_1^{(n)} = r_1, \cdots, \pi_n^{(n)} = r_n\big) = \frac{1}{n!} = \mathbf{P}\big(R_1 = \varrho_1, \cdots, R_n = \varrho_n\big).$$

In particular,

$$\mathbf{P}(R_n = \varrho_n) = \sum_{\varrho_1, \cdots, \varrho_{n-1}} \mathbf{P}\big(R_1 = \varrho_1, \cdots, R_n = \varrho_n\big) = \sum_{\varrho_1, \cdots, \varrho_{n-1}} \frac{1}{n!} \;;$$

each ϱ_j in this sum ranges over j values, so the number of terms in the sum is given by $1 \cdot 2 \cdots (n-1) = (n-1)!$. Therefore $\mathbf{P}(R_n = \varrho_n) = (n-1)!/\,n! = 1/n$, and

$$\mathbf{P}\big(R_1 = \varrho_1, \cdots, R_n = \varrho_n\big) = \frac{1}{n!} = \mathbf{P}\big(R_1 = \varrho_1\big) \cdots \mathbf{P}\big(R_n = \varrho_n\big)$$

for all possible values of $\varrho_j \in \{1, \cdots, j\}$ and $j = 1 \cdots, n$. The independence of the events $\{A_n\}_{n\in\mathbf{N}}$ follows now from that of the random variables $\{R_n\}_{n\in\mathbf{N}}$, since $A_n = \{R_n = 1\}$; in particular, $\mathbf{P}(A_n) = \mathbf{P}(R_n = 1) = \mathbf{P}\big(\pi_n^{(n)} = 1\big) = 1/n$.

In particular, this allows us to compute $\mathbf{E}(W_n) = \sum_{k=1}^n (1/k) \sim \log n$ and $\mathrm{Var}(W_n) = \sum_{k=1}^n (k-1)/k^2 \sim \log n$, as $n \to \infty$. ◇

B: SOME ELEMENTARY LIMIT THEOREMS

Let us introduce now the notion of "convergence in probability". This is just the analogue of the notion of "convergence in measure" in Exercise 1.5.5, translated to the context of a probability space.

2.2 Definition : Convergence in Probability. We say that a sequence of random variables $\{\Xi_n\}_{n\in\mathbf{N}}$ *converges in probability* to the random variable Ξ (defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$), if for every $\varepsilon > 0$ we have $\lim_{n\to\infty} \mathbf{P}(|\Xi_n - \Xi| > \varepsilon) = 0$.

In terms of this notion we can already begin to explore some elementary limit theorems, as shown in the next exercise. We shall take up this subject in earnest in Sections 2.3-2.5 and 4.2.

2.3 EXERCISE : An Elementary Weak Law of Large Numbers. (i) Consider a sequence of pairwise-uncorrelated random variables $\{X_n\}_{n\in\mathbf{N}}$ in $\mathbf{L}^2$. Under the condition

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{j=1}^{n} \mathrm{Var}(X_j) = 0,$$

show that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \big(X_j - \mathbf{E}(X_j)\big) = 0, \qquad \text{in probability.}$$

(ii) In particular, if $X_1, X_2, \cdots$ are pairwise-uncorrelated random variables in $\mathbf{L}^2$ with the same expectation, then under the same condition as above, we have

$$\overline{X}_n := \frac{1}{n} \sum_{j=1}^{n} X_j \longrightarrow \mathbf{E}(X_1) \qquad \text{in probability, as} \quad n \to \infty.$$

This is the case, for instance, in the coin-tossing Example 2.1, where $\mathbf{E}(X_1) = p$ is the "probability of success" on each individual toss.

2.4 Exercise : Convolution. Let X, Y be independent random variables with distributions μ and ν, respectively, and $F(\cdot) = \mu((-\infty, \cdot])$, $G(\cdot) = \nu((-\infty, \cdot])$.

(i) Show that the probability distribution function $H(x) = \mathbf{P}(Z \leq x)$, $x \in \mathbf{R}$ of the sum $Z := X + Y$ is given by the *convolution*

$$H(x) \equiv (F * G)(x) := \int_{-\infty}^{\infty} F(x-y)\, dG(y) = \int_{-\infty}^{\infty} G(y-x)\, dF(x), \qquad x \in \mathbf{R}$$

of the distribution functions $F(\cdot)$ and $G(\cdot)$.

(ii) Show that $\mathbf{P}(X+Y=0) = \sum_y \mu(\{-y\})\,\nu(\{y\})$. Thus, if either F or G is continuous, we get $\mathbf{P}(X+Y=0)=0$.

(iii) If $F(x) = \int_{-\infty}^x f(u)\,du$ $G(x) = \int_{-\infty}^x g(u)\,du$ are both absolutely continuous with densities $f,\, g \in \mathbf{L}^+ \cap \mathbf{L}^1$, then $H \equiv F * G$ is also absolutely continuous, with density

$$h(x) = \int_{-\infty}^{\infty} f(x-y)\,g(y)\,dy = \int_{-\infty}^{\infty} g(y-x)\,f(x)\,dx = (f*g)(x)\,, \quad x \in \mathbf{R}$$

given by the convolution of the two densities $f(\cdot)$ and $g(\cdot)$, as in (1.6.6).

(iv) If X has Gamma $\Gamma(\lambda, r)$ distribution and Y has $\Gamma(\lambda, s)$ distribution, then $X+Y$ has $\Gamma(\lambda, r+s)$ distribution. In particular, if $X_1, \cdots, X_n$ are independent and exponentially distributed with the same parameter λ , then $X_1+\cdots+X_n$ has $\Gamma(\lambda, n)$ distribution.

(v) If X has normal $\mathcal{N}(m_1, \sigma_1^2)$ distribution and Y has $\mathcal{N}(m_2, \sigma_2^2)$ distribution, then $X+Y$ has $\mathcal{N}(m_1+m_2, \sigma_1^2+\sigma_2^2)$ distribution.

2.5 Exercise : DeMoivre & Laplace. In the coin-tossing Example 2.1, justify the Bernoulli distribution (2.4) for the random variable S_n (number of successes in n trials), and note that $\mathbf{E}(S_n) = np$, $\mathrm{Var}(S_n) = npq$. Then, with the help of the Stirling formula $n! \sim \sqrt{2\pi n}\, n^n\, e^{-n}$, derive the *DeMoivre-Laplace Limit Theorem*

$$\mathbf{P}\left[a \le \frac{S_n - np}{\sqrt{npq}} \le b\right] \longrightarrow \Phi(b) - \Phi(a) = \int_a^b \frac{e^{-x^2/2}}{\sqrt{2\pi}}\,dx\,, \qquad a < b \quad \text{in } \mathbf{R}\,. \tag{2.6}$$

(*Hint:* Start with the “local” form $\sqrt{npq}\,(n!/k_n!(n-k_n))!\,p^{k_n} q^{n-k_n} \longrightarrow e^{-x^2/2}/\sqrt{2\pi}$ of this result, where $k_n = x\sqrt{npq} + np$, and observe that this convergence is uniform over x in the bounded interval $[a,b]$.)

C: GAUSSIANS

We have used in Exercise 2.5 the notation

$$\Phi(x) := \int_{-\infty}^x \varphi(u)\,du\,, \qquad \varphi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}} \quad ; \qquad x \in \mathbf{R} \tag{2.7}$$

for the *standard normal*, or *Gaussian*, distribution $\Phi(\cdot)$ and density $\varphi(\cdot)$ functions, respectively (verify the properties of Definition 1.4.1). Translating by $m \in \mathbf{R}$ and dilating by $\sigma > 0$, we generate a two-parameter family of normal (Gaussian) distribution functions

$$F_{m,\sigma^2}(x) \equiv \Phi\left(\frac{x-m}{\sigma}\right) = \int_{-\infty}^x \frac{1}{\sigma}\,\varphi\left(\frac{u-m}{\sigma}\right)du\,, \quad x \in \mathbf{R} \tag{2.8}$$

indexed by their expectation $m = \int x \, dF_{m,\sigma^2}(x)$ and their variance $\sigma^2 = \int (x - m)^2 \, dF_{m,\sigma^2}(x)$. These distributions play a central rôle in Probability Theory and its applications, for instance in the study of the fundamental *Brownian Motion process.* The DeMoivre-Laplace result (2.6) is the prototypical *Central Limit Theorem,* a result that we shall state in great generality in section 2.4 and prove in section 3.3.

2.5 Example: Multivariate Normal Distribution. A random vector $X = (X_1, \cdots, X_d)'$ is said to have a multivariate normal distribution with *mean-vector* $m = (m_1, \cdots, m_d)'$ and symmetric, non-singular *covariance-matrix* $\Sigma = \{\Sigma_{ij}\}_{1 \le i,j \le d}$, if $\mathbf{P}[(X_1, \cdots, X_d) \in A] = \int_A f(x)\, dx$ for every $A \in \mathcal{B}(\mathbf{R}^d)$ with

$$f(x) \;=\; \Big(\big(2\pi\big)^d \, |\det(\Sigma)| \Big)^{-1/2} \cdot \exp \left[-\frac{1}{2} \, \langle (x-m), \, \Sigma^{-1}(x-m) \rangle \right] , \quad x \in \mathbf{R}^d .$$

The reader should verify that, in this case,

- each X_i has a (univariate) normal distribution with $\mathbf{E}(X_i) = m_i$, $\mathrm{Var}(X_i) = \Sigma_{ii}$;
- $\mathrm{Cov}(X_i, X_j) = \Sigma_{ij}$ for $i \neq j$;
- $(X_1, \cdots, X_d)$ are independent, if and only if they are pairwise-uncorrelated (that is, $\Sigma_{ij} = 0$ for $i \neq j$);
- the characteristic function $\varphi_X(\xi) = \mathbf{E}(e^{\, i \langle \xi, X \rangle})$ of Remark 1.1 is given as

$$\varphi(\xi) = e^{\sqrt{-1}\, \langle \xi, x \rangle - (1/2)\, \xi' \Sigma \xi} , \qquad \xi \in \mathbf{R}^d ;$$

and

- each linear combination $\lambda_1 X_1 + \cdots + \lambda_d X_d$ has a (univariate) normal distribution.

2.3 Definition: Gaussian Family. A family $\mathcal{X} = \{X_\alpha\}_{\alpha \in A}$ of random variables is called *Gaussian*, if for each $d \in \mathbf{N}$ and any $\{\alpha_1, \cdots, \alpha_d\} \subseteq A$ the random vector $\big(X_{\alpha_1}, \cdots, X_{\alpha_d}\big)'$ has a multivariate normal (Gaussian) distribution.

In this case we denote by $m(\alpha) := \mathbf{E}(X_\alpha)$, $\alpha \in A$ the expectation function, and by

$$\Sigma(\alpha, \beta) \;:=\; \mathbf{E} \left[\big(X_\alpha - m(\alpha) \big) \big(X_\beta - m(\beta) \big) \right] , \quad (\alpha, \beta) \in A^2$$

the variance/covariance function. These two functions characterize the finite-dimensional distributions of the family $\mathcal{X}$.

A Gaussian family is closed in $\mathbf{L}^2(\mathbf{P})$.

2.6 Exercise : Poisson Approximation of "Rare Events". Show that the Poisson distribution $e^{-\lambda}\,\lambda^k\,/\,k!\,,\ k\in\mathbf{N}_0$ provides an approximation to the Binomial distribution of (2.4) for n large and p small (small probability of "success") – in the sense that, if $\{p_n\}_{n\in\mathbf{N}}$ is a sequence of positive numbers that decreases to zero, then

$$\frac{n!}{k!(n-k)!}\,p_n^k(1-p_n)^{n-k}\;-\;\left(e^{-\lambda}\,\frac{\lambda^k}{k!}\right)\Big|_{\lambda=np_n}\quad\longrightarrow\quad 0\,,\qquad\forall\;k\in\mathbf{N}_0$$

as $n\to\infty$. (*Hint:* Start with the "easy" case $p_n=\lambda\,/\,n$, some $\lambda>0$.)

2.7 Exercise: Geometric Distribution. In the coin-tossing Example 2.1, consider the random "time of first success"

$$T(\omega):=\inf\{n\in\mathbf{N}\,|\,X_n(\omega)=1\}\,,\ \omega\in\Omega$$

(always with the understanding $\inf\emptyset\equiv\infty$), and show that the random variable T has the geometric distribution: $\mathbf{P}[T=k]=p\,(1-p)^{k-1}$, $k\in\mathbf{N}$.

2.5 Example: Multinomial Distribution. Consider the space $\Omega=\{1,\cdots,d\}^n$ of n-tuples $\omega=(\omega_1,\cdots,\omega_n)$ with $\omega_i\in\{1,\cdots,d\}$ for each $i=1,\cdots,n$. Such an ω can be visualized as representing the result of n repetitions of a random experiment with d possible outcomes (such as throwing a die, if $d=6$).

Suppose also that we assign the common distribution $\mathbf{P}_i[X_i=k]=p_k>0$, $k=1,\cdots,d$ $(\sum_{k=1}^d p_k=1)$ to *each* of the coördinate mappings $X_i(\omega)=\omega_i$ $(1\le i\le n)$, and the probability measure $\mathbf{P}=\bigotimes_{i=1}^n\mathbf{P}_i$ to Ω itself. This corresponds to the intuitive notion, that different repetitions of the experiment are independent. If we count the occurrences $S_k(\omega)=\sum_{j=1}^n X_k(\omega_j)$ of the k^{th} outcome in these n repetitions of the experiment, then the random vector $(S_1,\cdots,S_d)$ has the *Multinomial Distribution*

$$\mathbf{P}[S_1=n_1,\cdots,S_d=n_d]\;=\;\frac{n!}{n_1!\cdots n_d!}\cdot p_1^{n_1}\cdots p_d^{n_d}\,;\qquad n_k\ge 0\quad\sum_{i=1}^d n_i=n\,.\qquad(2.9)$$

2.8 Exercise: Justify the form of the distribution (2.9). Show that each S_k has the Binomial distribution

$$\mathbf{P}[S_k=m]=\frac{n!}{m!(n-m)!}\,p_k^m(1-p_k)^{n-m}\,,\qquad m=0,\cdots,n$$

of (2.4); and that $\mathrm{Cov}(S_k,S_\ell)\;=\;-np_kp_\ell$, for $k\neq\ell$.

2.9 Exercise: For I.I.D. random variables $X_1, X_2, \cdots$ with $\sum_{j=1}^{\infty} \mathbf{P}[X_1 = j] = 1$, show that

$$\mathbf{P}\big[X_{n+1} > nX_n\,,\ \text{i.o.}\big] \;=\; 1 \qquad \Leftrightarrow \qquad \mathbf{E}\big(X_1\big) = \infty\,.$$

2.10 Exercise: Let $X_1, X_2, \cdots$ be random variables with

$$\mathbf{P}(X_n = n^2 - 1) \;=\; \frac{1}{n^2} \qquad \text{and} \qquad \mathbf{P}(X_n = -1) \;=\; 1 - \frac{1}{n^2}\;.$$

In particular, $\mathbf{E}(X_n) = 0$ for every $n \in \mathbf{N}$. Show that $\lim_{n\to\infty}\left(\frac{1}{n}\sum_{j=1}^{n} X_j\right) = -1$, a.e.

2.11 Exercise: If $X_1, X_2, \cdots$ are non-negative random variables with $\sum_n \mathbf{P}(X_n > n) < \infty$, show that

$$\limsup_{n\to\infty}\left(\frac{X_n}{n}\right) \le 1 \qquad \text{holds a.e.}$$

2.12 Exercise: Let $X_1, X_2, \cdots$ be independent random variables with common exponential distribution $\mathbf{P}(X_n > \xi) = e^{-\xi}$, $\xi \ge 0$. Show that

$$\limsup_{n\to\infty}\left(\frac{X_n}{\log n}\right) \;=\; 1\,, \qquad \lim_{n\to\infty}\frac{1}{\log n}\left(\max_{1\le k\le n} X_k\right) \;=\; 1\,, \qquad \text{a.e.}$$

E: THE TAIL (REMOTE) SIGMA-ALGEBRA

For any given sequence of random variables $X_1, X_2, \cdots$ let us denote by

- $\mathcal{F}_n := \sigma(X_1, \cdots, X_n)$ the smallest $\sigma-$algebra that measures the first $n \ge 1$ of them; by
- $\mathcal{T}^n := \sigma(X_{n+1}, X_{n+2}, \cdots)$ the smallest $\sigma-$algebra that measures all but the first $n \ge 0$ of them; and by
- $\mathcal{T} := \bigcap_{n\in\mathbf{N}_0} \mathcal{T}^n$ the *tail* or *remote* $\sigma-$algebra of this sequence.

Intuitively, the $\sigma-$algebra $\mathcal{T}$ contains all events whose occurrence is not affected by changing the values of finitely many terms in the sequence and leaving all others the same. For instance, $\{X_n \in B_n,\ \text{i.o.}\}$ belongs to $\mathcal{T}$, for any sequence $\{B_n\}_{n\in\mathbf{N}} \subset \mathcal{B}(\mathbf{R})$; so do the sets $\{\lim_n S_n \text{ exists}\}$ and $\{\limsup_n (S_n/c_n) > a\}$, where $S_n = \sum_{j=1}^{n} X_j$, $a \in \mathbf{R}$ and $\{c_n\}_n \subset (0,\infty)$ is a sequence that grows to infinity. But $\{\limsup_n S_n > 0\}$ does not belong to $\mathcal{T}$.

A celebrated result of Kolmogorov asserts that, for a sequence of *independent* random variables, the tail $\sigma-$algebra is trivial.

2.3 THEOREM: KOLMOGOROV's ZERO-ONE LAW. *If the random variables $X_1, X_2, \cdots$ are independent, then $\mathcal{T} = \{\emptyset, \Omega\}$ mod. $\mathbf{P}$, that is:*

$$\mathbf{P}(A) \,=\, 0 \quad \text{or} \quad 1\,, \qquad \text{for every} \quad A \in \mathcal{T}\,.$$

Proof: Let us take any $A \in \mathcal{T}$ with $\mathbf{P}(A) > 0$, if such a set exists (if not, there is nothing to prove); we shall try to show that $\mathbf{P}(A) = 1$. From Exercise 2.1 we know that $\mathcal{F}_n$ and $\mathcal{T}^n$ are independent for every $n \in \mathbf{N}$, so we have

$$\mathbf{P}(A \cap B) \,=\, \mathbf{P}(A) \cdot \mathbf{P}(B) \tag{2.10}$$

for every $B \in \mathcal{F}_n$, $n \in \mathbf{N}$, because $A \in \mathcal{T} \subseteq \mathcal{T}_n$. Thus (2.10) holds for every $A \in \mathcal{T}$, $B \in \cup_{n\in\mathbf{N}}\mathcal{F}_n$. Note that $\cup_{n\in\mathbf{N}}\mathcal{F}_n$ is closed under finite intersections: if $B_j \in \mathcal{F}_{n_j}$ for some $n_j \in \mathbf{N}$, $j = 1, 2$, then $B_1 \cap B_2 \in \mathcal{F}_n$ for $n := \max(n_1, n_2)$.

Now we define a new probability measure $\mathbf{P}_A(\cdot) = \mathbf{P}(A \cap \cdot)/\mathbf{P}(A)$ on $\mathcal{F}$ as in (2.1), and observe that the two probability measures $\mathbf{P}_A$ and $\mathbf{P}$ agree on the $\pi-$system $\cup_{n\in\mathbf{N}}\mathcal{F}_n$. Thus, by Exercise 1.3.8 these two measures agree on the σ-algebra $\sigma\left(\cup_{n\in\mathbf{N}}\mathcal{F}_n\right) = \sigma(X_1, X_2, \cdots) = \mathcal{T}^0$, which contains $\mathcal{T}$. But this means that we can write (2.10) with $B = A$, namely $\mathbf{P}(A) = \mathbf{P}(A \cap A) = \mathbf{P}(A) \cdot \mathbf{P}(A) = \big(\mathbf{P}(A)\big)^2$, and leads to $\mathbf{P}(A) = 1$. $\diamond$

F: STATIONARY RANDOM SEQUENCES

We say that a sequence of random variables $X_1, X_2, \cdots$ is *stationary*, if $(X_1, \cdots, X_n)$ and $(X_{k+1}, \cdots, X_{k+n})$ have the same distribution, for every $k \in \mathbf{N}$ and $n \in \mathbf{N}$.

Suppose that $T : \Omega \to \Omega$ is a *measure-preserving transformation* as in section 1.9; then starting from any given random variable $X : \Omega \to \mathbf{R}$ we can generate a stationary sequence

$$X_n(\omega) := X(T^{n-1}(\omega))\,, \qquad n \in \mathbf{N}\,, \tag{2.11}$$

with the understanding $T^0(\omega) = \omega$. Indeed, for any Borel subset B of $\mathbf{R}^n$ and with $E = \{\omega \in \Omega \,|\, (X_1(\omega), \cdots, X_n(\omega)) \in B\}$, we have

$$\mathbf{P}[\,(X_{k+1}, \cdots, X_{k+n}) \in B\,] \,=\, \mathbf{P}[\,\omega \in \Omega \,|\, T^k(\omega) \in E\,] \,=\, \mathbf{P}(E) \,=\, \mathbf{P}[\,(X_1, \cdots, X_n) \in B\,]\,.$$

This turns out to be not such a special case as it might seem. For suppose that $Y_1, Y_2, \cdots$ is *any* stationary sequence of random variables; then the Daniell-Kolmogorov Theorem 1.6.4 allows us to construct, on a canonical sequence space, a probability measure $\mathbf{P}$ under which the coördinate mappings $(X_n(\omega) = \omega_n\,,\ n \in \mathbf{N})$ have the same finite-dimensional distributions as $(Y_n\,,\ n \in \mathbf{N})$. Now set $X(\omega) := \omega_1 \equiv X_1(\omega)$, define the shift transformation $T(\omega_1, \omega_2, \cdots) = (\omega_2, \omega_3, \cdots)$, and observe that this T is measure-preserving and that (2.11) holds. In other words: *every stationary sequence of random variables can be cast in the form (2.11), for a suitable random variable* $X : \Omega \to \mathbf{R}$ *and a measure-preserving transformation* $T : \Omega \to \Omega$.

2.6 EXAMPLE: The I.I.D. Case. Suppose $X_1, X_2, \cdots$ is a sequence of independent and identically distributed random variables (independent copies of the random variable $X \equiv X_1$ as in (2.11)), say on the completion of the canonical space $\Omega = \mathbf{R}^{\mathbf{N}}$: $X_j(\omega) = \omega_j\,,\ j \in \mathbf{N}$. The shift transformation $T(\omega) = (\omega_2, \omega_3, \cdots)$ is then measure-preserving, and for any of its invariant sets $E \in \mathcal{I}$ we have $E = T^{-1}E = \{\omega \in \Omega \,|\, T(\omega) \in E\}$ modulo $\mathbf{P}$, therefore $E \in \sigma(X_2, X_3, \cdots)$. Iterating this, we obtain $E = T^{-n}E = \{\omega \in \Omega \,|\, T^n(\omega) \in E\}$ modulo $\mathbf{P}$, so $E \in \sigma(X_n, X_{n+1}, \cdots)$ for every $n \in \mathbf{N}$, and thus

$$E \in \bigcap_{n \in \mathbf{N}} \sigma(X_n, X_{n+1}, \cdots) = \mathcal{T}, \quad \text{the tail } \sigma\text{-algebra}: \quad \mathcal{I} \subseteq \mathcal{T}.$$

But Theorem 2.3 gives then $\mathbf{P}(E) = 0$ or 1 (the Kolmogorov zero-one law), so that $\mathcal{I}$ is trivial and the shift transformation T ergodic.

If, in addition, the random variable $X \equiv X_1$ is integrable (i.e., $\mathbf{E}(|X|) < \infty$), then from the Birkhoff pointwise ergodic Theorem 1.9.2 and its Corollary 1.9.1, we obtain the so-called **Strong Law of Large Numbers**

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} X_j(\omega) = \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} X\big(T^j(\omega)\big) = \mathbf{E}(X), \quad \text{for a.e. } \omega \in \Omega. \tag{2.12}$$

Theorem 1.9.2 also gives the mean convergence result

$$\lim_{n \to \infty} \mathbf{E}\,\big|\,\overline{X}_n - \mathbf{E}(X)\,\big| = 0, \qquad \text{where} \qquad \overline{X}_n(\omega) := \frac{1}{n} \sum_{j=1}^{n} X_j(\omega). \tag{2.13}$$

2.7 Example: The Bernoulli Shift. Take as probability space $\Omega = [0, 1)$ together with its Borel sets and Lebesgue measure, and consider the identity map $X(\omega) = \omega$ as well as the dyadic transformation $T(\omega) = 2\,\omega$, modulo 1. We saw in section 1.9 that this transformation is measure-preserving. Then $X_1(\omega) = X(\omega)$, $X_{n+1}(\omega) = T(X_n(\omega))$ for $n \in \mathbf{N}$ defines a stationary sequence.

This sequence is called the *Bernoulli shift*, because it admits the following representation: let $\xi_1, \xi_2, \cdots$ be independent binary (Bernoulli) random variables with common distribution $\mathbf{P}(\xi_j = 0) = \mathbf{P}(\xi_j = 1) = 1/2$, take

$$g(x) = \sum_{j \in \mathbf{N}} \frac{x_j}{2^j}, \qquad x \in \mathbf{R}^{\mathbf{N}}$$

and define a stationary sequence by $Y_k := g(\xi_k, \xi_{k+1}, \cdots)$ for $k \in \mathbf{N}$. Then $(X_1, X_2, \cdots)$ and $(Y_1, Y_2, \cdots)$ have the same finite-dimensional distributions.

2.8 Example: Autoregressive Sequence. A stationary sequence models the behavior of a random system which has "settled down" into a steady-state type of behavior. We expect such behavior in random dynamical systems that have been running already for a long time, so that transient (non-stationary) behavior has decayed to zero.

For instance, let $\{\Xi_n\}_{n\in\mathbf{Z}}$ be a sequence of I.I.D. random variables (the "noise" sequence) such that $\mathbf{E}(\Xi_n) = 0$, Var(Ξ_n)=1. Consider solving the equation

$$X_{n+1}^{(N)} \;=\; \alpha\, X_n^{(N)} + \Xi_{n+1}\,, \qquad n \ge -N\,, \quad X_{-N} = 0$$

where $\alpha \in (-1, 1)$. The solution is

$$X_n^{(N)} \;=\; \sum_{k\,=-N}^{n} \alpha^{n-k}\, \Xi_k\,, \qquad n \ge -N\,.$$

Now let N tend to infinity; this corresponds to the idea that on any fixed date n when the system is observed, it has already been running for a long time. Passing to this limit, we get the limiting sequence

$$X_n \;=\; \sum_{k\,=-\infty}^{n} \alpha^{n-k}\, \Xi_k \;=\; \sum_{k\,=0}^{\infty} \alpha^{k}\, \Xi_{n-k}\,, \qquad n \in \mathbf{Z}\,.$$

Exercise: Check that this sequence is $\{X_n\}_{n\in\mathbf{Z}}$ is stationary, and solves the "autoregressive equation" $X_{n+1} = \alpha X_n + \Xi_{n+1}$.