Part I

Theory

1

Basic probability concepts

1.1 Random experiments and probabilities

An experiment is random if its outcome cannot be predicted with certainty. A simple example is the throwing of a die. This experiment can result in any of six unpredictable outcomes 1, 2, 3, 4, 5, 6 which we list in what is usually called a *sample space* $\Omega = \{1, 2, 3, 4, 5, 6\} \stackrel{\triangle}{=} \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$. Another example is the amount of yearly rainfall in each of the next 10 years in Auckland. Each outcome here is an ordered set containing ten nonnegative real numbers (a vector in $\mathbb{R}_+^{10}$); however, one has to wait 10 years before observing the outcome ω.

Another example is the following.

Let X_t be the water level of a dam at time t. If we are interested in the behavior of X_t during an interval of time $[t_0, t_1]$ say, then it is necessary to consider simultaneously an uncountable family of X_ts, that is,

$$\Omega = \{0 \leq X_t < \infty, \quad t_0 \leq t \leq t_1\}.$$

The "smallest" observable outcome ω of an experiment is called *simple*.

The set $\{1\}$ containing 1 resulting from a throw of a die is simple. The outcome "odd number" is not simple and it occurs if and only if the throw results in any of the three simple outcomes 1, 3, 5. If the throw results in a 5, say, then the same throw results also in "a number larger than 3" or "odd number." Sets containing outcomes are called *events*. The events "odd number" and "a number larger than 3" are not mutually exclusive, that is, both can happen simultaneously, so that we can define the event "odd number *and* a number larger than 3."

The event "odd number *and* even number" is clearly impossible or empty. It is called the *impossible event* and is denoted, in analogy with the empty set in set theory, by $\emptyset$. The event "odd number *or* even number" occurs no matter what is the event ω. It is Ω itself and is called the *certain event*.

In fact possible events of the experiment can be combined naturally using the set operations *union, intersection*, and *complementation*. This leads to the concept of field or algebra (σ-field (sigma-field) or σ-algebra, respectively) which is of fundamental importance in the theory of probability.

A nonempty class $\mathcal{F}$ of subsets of a nonempty set Ω is called a *field* or *algebra* if

1. $\Omega \in \mathcal{F}$,
2. $\mathcal{F}$ is closed under finite unions (or finite intersections),
3. $\mathcal{F}$ is closed under complementation.

It is a σ*-field* or (σ*-algebra*) if the stronger condition

2.′ $\mathcal{F}$ is closed under countable unions (or countable intersections)

holds.

If $\{\mathcal{F}\}$ is a σ-field the pair $(\Omega, \mathcal{F})$ is called a *measurable space*. The sets $B \in \mathcal{F}$ are called *events* and are said to be *measurable sets*.

For instance, the collection of finite unions of the half open intervals $(a, b]$, $(-\infty < a < b \leq +\infty)$ in $\mathbb{R}$ plus the empty set is a field but not a σ-field because it is not closed under infinite countable unions. The open interval $(0, 1) = \bigcup_{n=1}^{\infty}(0, 1 - 1/n]$ is not in this collection despite the fact it contains *each* interval $(0, 1 - 1/n]$. Neither does it contain the singletons $\{x\}$, even though $\{x\} = \bigcap_{n=1}^{\infty}(x - 1/n, x]$ and it does not contain many other useful sets. This suggests that the notion of σ-field is indeed needed. There exists a minimal σ-field denoted $\mathcal{B}(\mathbb{R})$ containing all half open intervals $(a, b]$. This is the *Borel* σ-field on the real line and it is the smallest σ-field containing the collection of open intervals and hence all intervals. It contains also:

1. all singletons $\{x\}$ since $\{x\} = \bigcap_{n=1}^{\infty}\left(x - \frac{1}{n}, x + \frac{1}{n}\right)$,
2. the set Q of all rational numbers because it is a countable union: $Q = \bigcup_{r \in Q}\{r\}$,
3. the complement of Q, which is the set of all irrational numbers,
4. all open sets since any open set $\mathcal{O} = \bigcup_n I_n$, where $\{I_n\}$ are disjoint intervals. To see this recall that since $\mathcal{O}$ is open, then for any $x \in \mathcal{O}$ there exits a maximal interval I_x containing x and contained in $\mathcal{O}$ and $I_x = \mathcal{O}$ if $\mathcal{O}$ is itself an interval. If $\mathcal{O}$ is not an interval then there is a collection of disjoint maximal intervals contained in $\mathcal{O}$, one for each $x \in \mathcal{O}$. Moreover, each of these intervals contains a rational number because of the density of Q. Let $\{r_n : n = 1, 2, \dots\}$ be an enumeration of these rationals. Consequently, there is only at most a countable number of these intervals $I_1, I_2, \dots$. Therefore, since each of these intervals is contained in $\mathcal{O}$, their union $\bigcup_n I_n \subset \mathcal{O}$. Conversely, for each $x \in \mathcal{O}$ there exits a maximal interval $I_{n(x)}$ containing x and contained in $\bigcup_n I_n$, that is, $\mathcal{O} \subset \bigcup_n I_n$. Consequently $\mathcal{O} = \bigcup_n I_n$.

Sets in $\mathcal{B}(\mathbb{R})$ are called *Borel sets*. Note that a topological space, unlike a measure space, is not closed under complementation. A word of caution here: even σ-fields are not in general closed under uncountable unions.

The largest possible σ-field on any set Ω is the *power class* 2^{Ω} containing all the subsets of Ω. However this σ-field is in general "too big" to be of any use in probability theory. At the other extreme we have the smallest σ-field consisting of Ω and the empty set $\emptyset$.

Given any collection C of subsets of Ω, the σ-field generated by C, denoted by $\sigma\{C\}$, is made up of the class of all countable unions, all countable intersections and all complements of the subsets in C and all countable unions, intersections and complements of these sets, and so on. For instance, if C contains one subset, F say, then $\sigma\{F\}$ consists of the subset

F itself, its complement $\bar{F}$ (also denoted F^c), their union $F \cup \bar{F}$ (which is always Ω) and their intersection $F \cap \bar{F}$ (which is always $\emptyset$).

The σ-field $\sigma\{C\}$ generated by a class of subsets C contains by definition C itself (as a subset); however, there are other σ-fields also containing C, one of them being 2^Ω (the largest one). The point here is that $\sigma\{C\}$ is the *smallest* σ-field containing C. In the set theory context "smallest" means that $\sigma\{C\}$ is in the intersection of all the σ-fields containing C. In summary:

$$C \subset \sigma\{C\} \subset \{\text{any } \sigma\text{-field containing } C\}.$$

It is left as an exercise to show that any σ-field is either finite or uncountably infinite.

Fields, or σ-fields, are convenient mathematical objects that express how much we know about the outcome ω of a random experiment. For instance, if $\Omega = \{1, 2, 3, 4, 5, 6\}$ we may not be able to observe ω but we may observe a "larger" event like "odd number"$= \{(1, 3, 5)\}$, so that our "observed" σ-field is smaller than the one generated by Ω. In fact it is equal to $\{(1, 3, 5), (2, 4, 6), \Omega, \emptyset\}$, which does not contain events like $\{(1, 3)\}$ or $\{6\}$.

When the sample space Ω is finite, it is enough to represent information through partitions of Ω into *atoms*, which are the smallest observable events. Since a field is just a collection of finite unions and complements of these atoms, it represents the same information as the partition. This is not true on infinite sample spaces as partitions and fields are not big enough to represent information in all practical situations.

Suppose that when the experiment of throwing a die is performed, an indirect observer of the outcome ω can only learn that the event $\{1, 2\}$ did or did not occur. So for this observer the (smallest) decidable events, or atoms, are in the field

$$\mathcal{F}_1 = \sigma\{\{1, 2\}, \{3, 4, 5, 6\}\} = \{\emptyset, \{1, 2, 3, 4, 5, 6\}, \{1, 2\}, \{3, 4, 5, 6\}\}.$$

Another observer with a better access to information might be able to observe the richer field

$$\mathcal{F}_2 = \sigma\{\{1, 2\}, \{3, 4\}, \{5, 6\}\},$$

which contains more atoms. The point here is that, given a set of outcomes Ω, it is possible to define many fields, or σ-fields, ranging from the coarsest (containing only Ω and the empty set $\emptyset$), to the finest (containing all the subsets of Ω).

A natural question is: what extra conditions will make a field into a σ-field? We have the following useful result.

A field is a σ-field if and only if it is closed under monotonic sequences of events, that is, it contains the limit of every monotonically increasing or decreasing sequence of events. (A sequence of events A_i, $i \in \mathbb{N}$, is monotonic increasing if $A_1 \subset A_2 \subset A_3 \dots$).

Let the index parameter t be either a nonnegative integer or a nonnegative real number.

To keep track, to record, and to benefit from the flow of information accumulating in time and to give a mathematical meaning to the notions of *past*, *present* and *future* the concept of *filtration* is introduced. This is done by equipping the measurable space $(\Omega, \mathcal{F})$ with a nondecreasing family $\{\mathcal{F}_t, t \geq 0\}$ of "observable" sub-σ-fields of $\mathcal{F}$ such that $\mathcal{F}_t \subset \mathcal{F}_{t'}$ whenever $t \leq t'$. That is, as time flows, our information structures or σ-fields are becoming finer and finer.

We define $\mathcal{F}_\infty = \sigma(\bigcup_{t\geq 0} \mathcal{F}_t) \triangleq \bigvee_{t\geq 0} \mathcal{F}_t$ where the symbol $\triangleq$ stands for "by definition."

Example 1.1.1 Let $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$. The σ-fields

$$\begin{aligned}
\mathcal{F}_0 &= \sigma\{\Omega, \emptyset\},\\
\mathcal{F}_1 &= \sigma\{\{\omega_1, \omega_2, \omega_3\}, \{\omega_4, \omega_5, \omega_6\}\},\\
\mathcal{F}_2 &= \sigma\{\{\omega_1, \omega_2\}, \{\omega_3\}, \{\omega_4, \omega_5, \omega_6\}\},\\
\mathcal{F}_3 &= \sigma\{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4, \omega_5, \omega_6\}\},
\end{aligned}$$

form a filtration since $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$. However, the σ-fields

$$\begin{aligned}
\mathcal{F}_0 &= \sigma\{\Omega, \emptyset\},\\
\mathcal{F}_1 &= \sigma\{\{\omega_1, \omega_2, \omega_3\}, \{\omega_4, \omega_5, \omega_6\}\},\\
\mathcal{F}_2 &= \sigma\{\{\omega_1, \omega_4\}, \{\omega_2, \omega_5\}, \{\omega_3, \omega_6\}\},\\
\mathcal{F}_3 &= \sigma\{\{\omega_1\}, \{\omega_2\}, \{\omega_3, \omega_4\}, \{\omega_5, \omega_6\}\},
\end{aligned}$$

do not form a filtration since, for instance, $\mathcal{F}_1 \not\subset \mathcal{F}_2$. □

Example 1.1.2 Suppose Ω is the unit interval (0, 1] and consider the following σ-fields:

$$\begin{aligned}
\mathcal{F}_0 &= \sigma\{\Omega, \emptyset\},\\
\mathcal{F}_1 &= \sigma\{(0, \tfrac{1}{2}], (\tfrac{1}{2}, \tfrac{3}{4}], (\tfrac{3}{4}, 1]\},\\
\mathcal{F}_2 &= \sigma\{(0, \tfrac{1}{4}], (\tfrac{1}{4}, \tfrac{1}{2}], (\tfrac{1}{2}, \tfrac{3}{4}], (\tfrac{3}{4}, 1]\},\\
\mathcal{F}_3 &= \sigma\{(0, \tfrac{1}{8}], (\tfrac{1}{8}, \tfrac{2}{8}], \dots, (\tfrac{7}{8}, 1]\}.
\end{aligned}$$

These form a filtration since $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$. □

When the time index $t \in \mathbb{R}^+$ we are led naturally to introduce the concepts of *right-continuity* and *left-continuity* of a filtration as a function of t.

A filtration $\{\mathcal{F}_t,\ t \geq 0\}$ is right-continuous if $\mathcal{F}_t$ contains events *immediately after* t, that is $\mathcal{F}_t = \bigcap_{\epsilon>0} \mathcal{F}_{t+\epsilon}$. We may also say that a filtration $\{\mathcal{F}_t,\ t \geq 0\}$ is right-continuous if new information at time t arrives precisely at time t and not an instant after t.

It is left-continuous if $\{\mathcal{F}_t\}$ contains events *strictly prior* to t, that is $\mathcal{F}_t = \bigvee_{s<t} \mathcal{F}_s$.

Probability measures

Given a measurable space $(\Omega, \mathcal{F})$ a *probability measure* P is a countably additive function defined on events in $\mathcal{F}$ with values in [0, 1]. More precisely:

A set function $P: \mathcal{F} \to [0, 1]$, where $\mathcal{F}$ is either a field or a σ-field, is called a probability measure if

1. $P(\Omega) = 1$;
2. If B_k is a countable sequence of pairwise disjoint events in $\mathcal{F}$, then $P(\bigcup B_k) = \sum P(B_k)$. This is termed σ-additivity of P.

Of course, (1) and (2) imply that $P(\emptyset) = 0$. Also, if A is an event such that $P(A) = 0$ and B is any event contained in A then $P(B) = 0$.

It is easily seen that if Ω is finite we need only specify P on atoms of $\mathcal{F}$.

The triple $(\Omega, \mathcal{F}, P)$ is called a *probability space*.

Nonempty events which are unlikely to occur and to which a zero probability is assigned are called *negligible events* or *null* events.

A σ-field $\mathcal{F}$ is *P-complete* if all subsets of null events are also events. Of course, their probability is zero.

A filtration is *complete* if $\mathcal{F}_0$ is complete, i.e. all the null events are known at the initial time.

The mathematical object $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, where the filtration $\{\mathcal{F}_t,\ t \geq 0\}$ is right-continuous and complete, is sometimes called a *stochastic basis* or a *filtered probability space* .

The filtration $\{\mathcal{F}_t, t \geq 0\}$ is said to satisfy the "usual conditions" if it is right-continuous and complete.

For monotonic sequences of events we have the following result on *continuity* of probability measures.

Theorem 1.1.3 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. If $\{A_n\}$ is an increasing sequence of events with limit A, then*

$$P(A_n) \uparrow P(A),$$

and if $\{B_n\}$ is a decreasing sequence of events with limit B, then

$$P(B_n) \downarrow P(B).$$

Proof To prove the first statement, visualize the sequence $\{A_n\}$ as a sequence of increasing concentric disks and then define the sequence of disjoint rings $\{R_n\}$ (except for R_1 which is the disk A_1):

$$R_1 = A_1,\ \ R_2 = A_2 - A_1, \dots, R_n = A_n - A_{n-1}.$$

Note that

$$A_k = \cup_{n=1}^{k} R_n,\ \ A = \cup_{n=1}^{\infty} A_n = \cup_{n=1}^{\infty} R_n,$$

so that by σ-additivity

$P(A) = \sum_{n=1}^{\infty} P(R_n) = \lim_k \sum_{n=1}^{k} P(R_n) = \lim_k P(\cup_{n=1}^{k} R_n) = \lim_k P(A_k)$.

The proof of the second statement follows by considering the sequence of complementary events $\{\bar{B}_n\}$ which is increasing with limit $\bar{B}$, so that

$$1 - P(A_n) \uparrow 1 - P(A) \Longrightarrow P(A_n) \downarrow P(A).$$

■

Example 1.1.4 Consider the experiment of tossing a fair coin infinitely many times and "observing" the outcomes of *all* tosses. Here each $\omega \in \Omega = (H, T)^{\infty}$ is a countably infinite sequence of "Heads" and "Tails". If we denote "Heads" and "Tails" by 0 and 1, each ω is a sequence of 0s and 1s and it can be shown that there are as many ωs as there are points in the interval [0, 1)!

Suppose we wish to estimate the probability of the event consisting of those ωs for which the proportion of heads converges to $1/2$. The so-called *Strong Law of Large Numbers* says that this probability is equal to one, i.e. the ωs for which the convergence to $1/2$ does not hold form a negligible set. However, this negligible set is rather huge, as can be imagined! □

Example 1.1.5 In Example 1.1.4 let $F_{n,S}$ be the collection of infinite sequences of Hs and Ts with some restriction S put on the first n tosses. For instance, if $n = 3$,

$$S = \{HHT\dots, HTH\dots, THH\dots\} \subset (H,T)^3,$$

$F_{3,S}$ is the collection of infinite sequences of Hs and Ts for which the first three entries contain exactly two Hs.

It is left as an exercise to show that the class
$\mathcal{F} = \{F_{n,S},\ S \subset (H,T)^n,\ n \in \mathbb{N}\}$ is a field. □

We now quote without proof from [4] the following result on extending a function P defined on sets in a field.

Theorem 1.1.6 *([4]) If P is a probability measure on a field $\mathcal{A}$, then it can be extended uniquely to the σ-field $\mathcal{F} = \sigma\{\mathcal{A}\}$ generated by $\mathcal{A}$, i.e. the restriction of the extension measure to the field $\mathcal{A}$ is P itself and by tradition they are both denoted by P.*

Let us return to the coin-tossing situation of Example 1.1.5.

Using the extension theorem (Theorem 1.1.6) one can construct a (unique) probability measure P called *product probability measure* on the space $((H,T)^\infty, \mathcal{F})$, starting from an initial probability $(p(H), p(T)) = (1/2, 1/2)$ by setting

$$P(F_{n,S}) = \sum_S \left(\frac{1}{2}\right)^n = (\text{number of infinite sequences in } S) \times \left(\frac{1}{2}\right)^n.$$

It is left as an exercise to show that P does not depend on the representations of sets in $\mathcal{F}$ and that it is countably additive. (See [4]).

An immediate generalization of the coin tossing experiment in Example 1.1.5 is to consider an infinite sequence of independent experiments, to which corresponds an infinite sequence of probability spaces $(\Omega_1, \mathcal{F}_1, P_1), (\Omega_2, \mathcal{F}_2, P_2), \dots$. We are interested in the space $\Omega^{(\infty)} = \Omega_1 \times \Omega_2 \times \dots$ of all infinite sequences $\omega = (\omega_1, \omega_2, \dots)$. Events of interest are again cylinder sets, i.e. infinite sequences with restrictions put on the first n outcomes. The collection of all these cylinders form a field which generates a σ-field $\mathcal{F}$, often denoted $\mathcal{F}_1 \otimes \mathcal{F}_2 \otimes \dots$. A probability measure P can be defined on cylinder sets then extended uniquely to $\mathcal{F}$ using the Extension Theorem 1.1.6.

In the coin-tossing experiment, an example of an event which is in $\mathcal{F}$ is the event F that a "Head" will occur. Clearly, $F = \bigcup_{k=1}^\infty F_k$, where F_k is the event that a "Head" occurs on the k-th trial and not before. Since each F_k is a cylinder set, $P(F_k)$ is well defined for each

$k \geq 1$. Moreover the F_ks are pairwise disjoint, hence

$$P(F) = \sum_{k=1}^{\infty} P(F_k) = \sum_{k=1}^{\infty} \frac{1}{2^k} = 1.$$

Note that this probability is still 1 regardless of the size of the probability of occurrence of a "Head", (as long as it is not 0).

Modeling with infinite sample spaces is not a mathematical fantasy. In many very simple minded problems infinite sequences of outcomes cannot be avoided. For example, "the first time a Head occurs" event cannot be described in a finite sample space model because the number of trials before it occurs cannot be bounded in advance.

In general, it is impossible to define a probability measure on all the subsets of an infinite sample space; that is, one cannot say any subset is an event. However, consider the following case.

Example 1.1.7 Suppose that Ω is countable and let $\mathcal{F}$ be the σ-field 2^{Ω}. Then it is not difficult to define a probability measure on $\mathcal{F}$. Choose P such that

$$0 \leq P(\{\omega\}) \leq 1 \text{ and } P(\{\Omega\}) = \sum_{\omega \in \Omega} P(\omega) = 1,$$

and for any $F \in \mathcal{F}$, define $P(F) = \sum_{\omega \in F} P(\omega)$.

Let $\{F_n\}_{n \in \mathbb{N}}$ be a sequence of disjoint sets in $\mathcal{F}$ and let $\omega_{n,}$ denote the simple events in F_n. Since we have an infinite series of nonnegative numbers,

$$P(\bigcup_n F_n) = \sum_{n,m} P(\omega_{n,m}) = \sum_n \sum_m P(\omega_{n,m}) = \sum_n P(F_n).$$

$\square$

1.2 Conditional probabilities and independence

Given a probability space $(\Omega, \mathcal{F}, P)$ and some event B with $P(B) \neq 0$, we define a new *posterior* probability measure as follows. If A is any event we define the probability of A given B as

$$P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(A \cap B)}{P(B)},$$

provided $P(B) > 0$. Otherwise $P(A \mid B)$ is left undefined.

What we mean by "given event B" is that *we know* that event B has occurred, that is we know that $\omega \in B$, so that we no longer assign the same probabilities given by P to events but assign new, or updated, probabilities given by the probability measure $P(. \mid B)$. Any event which is mutually exclusive with B has probability zero under $P(. \mid B)$ and the new probability space is now $(B, \mathcal{F} \cap B, P(. \mid B))$.

If our observation is limited to knowing whether event B has occurred or not we may as well define $P(. \mid \overline{B})$, where $\overline{B}$ is the complement of B within Ω. Prior to knowing where the outcome ω is we define the, now random, quantity:

$$P(. \mid B \text{ or } \overline{B})(\omega) = P(. \mid \sigma\{B\})(\omega) \triangleq P(. \mid B)I_B(\omega) + P(. \mid \overline{B})I_{\overline{B}}(\omega).$$

This definition extends in an obvious way to a σ-field $\mathcal{G}$ generated by a finite or countable partition $\{B_1, B_2, \dots\}$ of Ω and the random variable $P(. \mid \mathcal{G})(\omega)$ is called the *conditional probability given* $\mathcal{G}$. The random function $P(. \mid \mathcal{G})(\omega)$ whose values on the atoms B_i are ordinary conditional probabilities $P(. \mid B_i) = \dfrac{P(. \cap B_i)}{P(B_I)}$ is not defined if $P(B_i) = 0$. In this case we have a family of functions $P(. \mid \mathcal{G})(\omega)$, one for each possible arbitrary value assigned to the undefined $P(. \mid B_i)$. Usually, one *version* is chosen and different versions differ only on a set of probability 0.

Example 1.2.1 Phone calls arrive at a switchboard between 8:00 a.m. and 12:00 p.m. according to the following probability distribution:

1. $P(k \text{ calls within an interval of length } l) = e^{-l}\dfrac{l^k}{k!}$;
2. If I_1 and I_2 are disjoint intervals,

$$\begin{aligned} &P((k_1 \text{ calls within } I_1) \cap (k_2 \text{ calls within } I_2)) \\ &\quad = P(k_1 \text{ calls within } I_1) P(k_2 \text{ calls within } I_2), \end{aligned}$$

   that is, events occurring within disjoint time intervals are independent.

Suppose that the operator wants to know the probability that 0 calls arrive between 8:00 and 9:00 given that the total number of calls from 8:00 a.m. to 12:00 p.m., N_{8-12}, is known. From past experience, the operator assumes that this number is near 30 calls, say. Hence

$$\begin{aligned} &P(0 \text{ calls within } [8, 9) \mid 30 \text{ calls within } [8, 12]) \\ &= \frac{P((0 \text{ calls within } [8, 9)) \bigcap (30 \text{ calls within } [9, 12]))}{P(30 \text{ calls within } [8, 12])} \\ &= \frac{P(0 \text{ calls within } [8, 9)) P(30 \text{ calls within } [9, 12])}{P(30 \text{ calls within } [8, 12])} = \left(\frac{3}{4}\right)^{30}, \end{aligned}$$

which can be written as

$$P(0 \text{ calls within } [8, 9) \mid N_{8-12} = N) = \left(\frac{3}{4}\right)^{N}. \tag{1.2.1}$$

□

Remarks 1.2.2 Consider again Example 1.2.1.

1. The events $F_i = \{\omega : N_{8-12}(\omega) = i\}$, $i = 0, 1, \dots$ form a partition of Ω and are atoms of the σ-field generated by observing only N_{8-12}, so we may write:

$$\begin{aligned} &P(0 \text{ calls within } [8, 9) \mid F_i,\ i \in \mathbb{N})(\omega) \\ &\quad = P(0 \text{ calls within } [8, 9) \mid \sigma\{F_i,\ i \in \mathbb{N}\})(\omega) \\ &\quad = \sum_i^{\infty} \left(\frac{3}{4}\right)^i I_{F_i}(\omega). \end{aligned}$$

2. Observe that since each event $F \in \sigma\{F_i,\ i \in \mathbb{N}\}$ is a union of some $F_{i_1}, F_{i_2}, \dots$, and since we know, at the end of the experiment, which F_j contains ω, then we know

whether or not ω lies in F, that is whether F or the complement of F has occurred. In this sense, $\sigma\{F_i,\ i \in \mathbb{N}\}$ is indeed *all* we can answer about the experiment from what we know. □

The likelihood of occurrence of any event A could be affected by the realization of B. Roughly speaking if the "proportion" of A within B is the same as the "proportion" of A within Ω then it is intuitively clear that $P(A \mid B) = P(A \mid \Omega) = P(A)$. Knowing that B has occurred does not change the prior probability $P(A)$. In that case we say that events A and B are *independent*. Therefore two events A and B are *independent* if and only if $P(A \cap B) = P(A)P(B)$.

Two σ-fields $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent if and only if $P(A_1 \cap A_2) = P(A_1)P(A_2)$ for all $A_1 \in \mathcal{F}_1$, $A_2 \in \mathcal{F}_2$.

If events A and B are independent so are $\sigma\{A\}$ and $\sigma\{B\}$ because the impossible event $\emptyset$ is independent of everything else including itself, and so is Ω. Also A and B^c, A^c and B, A^c and B^c are independent. We can say a bit more, if $P(E) = 0$ or $P(E) = 1$ then the event E is independent of any other event including E itself, which seems intuitively clear.

Mutually exclusive events with positive probabilities provide a good example of dependent events.

Example 1.2.3 In the die throwing experiment the σ-fields

$$\mathcal{F}_1 = \sigma\{\{1, 2\}, \{3, 4, 5, 6\}\},$$

and

$$\mathcal{F}_2 = \sigma\{\{1, 2\}, \{3, 4\}, \{5, 6\}\},$$

are not independent since if we know, for instance, that ω has landed in $\{5, 6\}$ (or equivalently $\{5, 6\}$ has occurred) in $\mathcal{F}_2$ then we also know that the event $\{3, 4, 5, 6\}$ in $\mathcal{F}_1$ has occurred. This fact can be checked by direct calculation using the definition. However, the σ-fields

$$\mathcal{F}_3 = \sigma\{\{1, 2, 3\}, \{4, 5, 6\}\},$$

and

$$\mathcal{F}_4 = \sigma\{\{1, 4\}, \{2, 5\}, \{3, 6\}\},$$

are independent. The occurrence of any event in any of $\mathcal{F}_3$ or $\mathcal{F}_4$ does not provide any nontrivial information about the occurrence of any (nontrivial) event in the other field. □

Another fundamental concept of probability theory is *conditional independence*. Events A and C are said to be conditionally independent given event B if $P(A \cap C \mid B) = P(A \mid B)P(C \mid B)$, $P(B) > 0$.

The following example shows that it is not always easy to decide, under a probability measure, if conditional independence holds or not between events.

Example 1.2.4 Consider the following two events:

A_1="person 1 is going to watch a football game next weekend,"
A_2="person 2, with no relation at all with person 1, is going to watch a football game next weekend."