4

Conditional Probabilities and Independence

4.1 Conditional Probabilities

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space, and let $A, B \in \mathcal{F}$ be two events. We assume that $\mathrm{P}(B) > 0$.

Definition 4.1. *The conditional probability of A given B is*

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \bigcap B)}{\mathrm{P}(B)} \ .$$

While the conditional probability depends on both A and B, this dependence has a very different nature for the two sets. As a function of A the conditional probability has the usual properties of a probability measure:

1. $\mathrm{P}(A|B) \geq 0$.
2. $\mathrm{P}(\Omega|B) = 1$.
3. For a finite or infinite sequence of disjoint events A_i with $A = \bigcup_i A_i$ we have
$$\mathrm{P}(A|B) = \sum_i \mathrm{P}(A_i|B) \ .$$

As a function of B, the conditional probability satisfies the so-called formula of total probability. Let $\{B_1, B_2, \ldots\}$ be a finite or countable partition of the space Ω, that is $B_i \bigcap B_j = \emptyset$ for $i \neq j$ and $\bigcup_i B_i = \Omega$. We also assume that $\mathrm{P}(B_i) > 0$ for every i. Take $A \in \mathcal{F}$. Then

$$\mathrm{P}(A) = \sum_i \mathrm{P}(A \bigcap B_i) = \sum_i \mathrm{P}(A|B_i)\mathrm{P}(B_i) \qquad (4.1)$$

is called the formula of total probability. This formula is reminiscent of multiple integrals written as iterated integrals. The conditional probability plays the role of the inner integral and the summation over i is the analog of the outer integral.

L. Koralov and Y.G. Sinai, *Theory of Probability and Random Processes*, Universitext, DOI 10.1007/978-3-540-68829-7_4,

In mathematical statistics the events B_i are sometimes called hypotheses, and probabilities $\mathrm{P}(B_i)$ are called prior probabilities (i.e., given pre-experiment). We assume that as a result of the trial an event A occurred. We wish, on the basis of this, to draw conclusions regarding which of the hypotheses B_i is most likely. The estimation is done by calculating the probabilities $\mathrm{P}(B_k|A)$ which are sometimes called posterior (post-experiment) probabilities. Thus

$$\mathrm{P}(B_k|A) = \frac{\mathrm{P}(B_k \bigcap A)}{\mathrm{P}(A)} = \frac{\mathrm{P}(A|B_k)\mathrm{P}(B_k)}{\sum_i \mathrm{P}(B_i)\mathrm{P}(A|B_i)} \ .$$

This relation is called Bayes' formula.

4.2 Independence of Events, σ-Algebras, and Random Variables

Definition 4.2. *Two events A_1 and A_2 are called independent if*

$$\mathrm{P}(A_1 \bigcap A_2) = \mathrm{P}(A_1)\mathrm{P}(A_2) \ .$$

The events $\emptyset$ and Ω are independent of any event.

Lemma 4.3. *If (A_1, A_2) is a pair of independent events, then $(\overline{A}_1, A_2)$, $(A_1, \overline{A}_2)$, and $(\overline{A}_1, \overline{A}_2)$, where $\overline{A}_j = \Omega \backslash A_j$, $j = 1, 2$, are also pairs of independent events.*

Proof. If A_1 and A_2 are independent, then

$$\mathrm{P}(\overline{A}_1 \bigcap A_2) = \mathrm{P}((\Omega \backslash A_1) \bigcap A_2)) =$$
$$\mathrm{P}(A_2) - \mathrm{P}(A_1 \bigcap A_2) = \mathrm{P}(A_2) - \mathrm{P}(A_1)\mathrm{P}(A_2) = \tag{4.2}$$
$$(1 - \mathrm{P}(A_1))\mathrm{P}(A_2) = \mathrm{P}(\overline{A}_1)\mathrm{P}(A_2).$$

Therefore, $\overline{A}_1$ and A_2 are independent. By interchanging A_1 and A_2 in the above argument, we obtain that A_1 and $\overline{A}_2$ are independent. Finally, $\overline{A}_1$ and $\overline{A}_2$ are independent since we can replace A_2 by $\overline{A}_2$ in (4.2). □

The notion of pair-wise independence introduced above is easily generalized to the notion of independence of any finite number of events.

Definition 4.4. *The events $A_1, \ldots, A_n$ are called independent if for any $1 \leq k \leq n$ and any $1 \leq i_1 < \ldots < i_k \leq n$*

$$\mathrm{P}(A_{i_1} \bigcap \ldots \bigcap A_{i_k}) = \mathrm{P}(A_{i_1}) \ldots \mathrm{P}(A_{i_k}) \ .$$

For $n \geq 3$ the pair-wise independence of events A_i and A_j for all $1 \leq i < j \leq n$ does not imply that the events $A_1, \ldots, A_n$ are independent (see Problem 5).

Consider now a collection of σ-algebras $\mathcal{F}_1, \ldots, \mathcal{F}_n$, each of which is a σ-subalgebra of $\mathcal{F}$.

Definition 4.5. *The σ-algebras $\mathcal{F}_1, \ldots, \mathcal{F}_n$ are called independent if for any $A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}_n$ the events $A_1, \ldots, A_n$ are independent.*

Take a sequence of random variables $\xi_1, \ldots, \xi_n$. Each random variable ξ_i generates the σ-algebra $\mathcal{F}_i$, where the elements of $\mathcal{F}_i$ have the form $C = \{\omega : \xi_i(\omega) \in A\}$ for some Borel set $A \subseteq \mathbb{R}$. It is easy to check that the collection of such sets is indeed a σ-algebra, since the collection of Borel subsets of $\mathbb{R}$ is a σ-algebra.

Definition 4.6. *Random variables $\xi_1, \ldots, \xi_n$ are called independent if the σ-algebras $\mathcal{F}_1, \ldots, \mathcal{F}_n$ they generate are independent.*

Finally, we can generalize the notion of independence to arbitrary families of events, σ-algebras, and random variables.

Definition 4.7. *A family of events, σ-algebras, or random variables is called independent if any finite sub-family is independent.*

We shall now prove that the expectation of a product of independent random variables is equal to the product of expectations. The converse is, in general, not true (see Problem 6).

Theorem 4.8. *If ξ and η are independent random variables with finite expectations, then the expectation of the product is also finite and $\mathrm{E}(\xi\eta) = \mathrm{E}\xi\mathrm{E}\eta$.*

Proof. Let ξ_1 and ξ_2 be the positive and negative parts, respectively, of the random variable ξ, as defined above. Similarly, let η_1 and η_2 be the positive and negative parts of η. It is sufficient to prove that $\mathrm{E}(\xi_i\eta_j) = \mathrm{E}\xi_i\mathrm{E}\eta_j$, $i, j = 1, 2$. We shall prove that $\mathrm{E}(\xi_1\eta_1) = \mathrm{E}\xi_1\mathrm{E}\eta_1$, the other cases being completely similar. Define $f_n(\omega)$ and $g_n(\omega)$ by the relations

$$f_n(\omega) = k2^{-n} \quad \text{if} \quad k2^{-n} \leq \xi_1(\omega) < (k+1)2^{-n} \ ,$$

$$g_n(\omega) = k2^{-n} \quad \text{if} \quad k2^{-n} \leq \eta_1(\omega) < (k+1)2^{-n} \ .$$

Thus f_n and g_n are two sequences of simple random variables which monotonically approximate from below the variables ξ_1 and η_1 respectively. Also, the sequence of simple random variables $f_n g_n$ monotonically approximates the random variable $\xi_1\eta_1$ from below. Therefore,

$$\mathrm{E}\xi_1 = \lim_{n\to\infty} \mathrm{E}f_n, \quad \mathrm{E}\eta_1 = \lim_{n\to\infty} \mathrm{E}g_n, \quad \mathrm{E}\xi_1\eta_1 = \lim_{n\to\infty} \mathrm{E}f_n g_n \ .$$

Since the limit of a product is the product of the limits, it remains to show that $\mathrm{E}f_n g_n = \mathrm{E}f_n\mathrm{E}g_n$ for any n, . Let A_k^n be the event $\{k2^{-n} \leq \xi_1 < (k+1)2^{-n}\}$ and B_k^n be the event $\{k2^{-n} \leq \eta_1 < (k+1)2^{-n}\}$. Note that for any k_1 and

k_2 the events $A^n_{k_1}$ and $B^n_{k_2}$ are independent due to the independence of the random variables ξ and η. We write

$$\mathrm{E}f_n g_n = \sum_{k_1,k_2} k_1 k_2 2^{-2n} \mathrm{P}(A^n_{k_1} \bigcap B^n_{k_2}) =$$

$$\sum_{k_1} k_1 2^{-n} \mathrm{P}(A^n_{k_1}) \sum_{k_2} k_2 2^{-n} \mathrm{P}(B^n_{k_2}) = \mathrm{E}f_n \mathrm{E}g_n \ ,$$

which completes the proof of the theorem. □

Consider the space Ω corresponding to the homogeneous sequence of n independent trials, $\omega = (\omega_1, \ldots, \omega_n)$, and let $\xi_i(\omega) = \omega_i$.

Lemma 4.9. *The sequence $\xi_1, \ldots, \xi_n$ is a sequence of identically distributed independent random variables.*

Proof. Each random variable ξ_i takes values in a space X with a σ-algebra $\mathcal{G}$, and the probabilities of the events $\{\omega : \xi_i(\omega) \in A\}$, $A \in \mathcal{G}$, are equal to the probability of A in the space X. Thus they are the same for different i if A is fixed, which means that ξ_i are identically distributed. Their independence follows from the definition of the sequence of independent trials. □

4.3 π-Systems and Independence

The following notions of a π-system and of a Dynkin system are very useful when proving independence of functions and σ-algebras.

Definition 4.10. *A collection $\mathcal{K}$ of subsets of Ω is said to be a π-system if it contains the empty set and is closed under the operation of taking the intersection of two sets, that is*

1. *$\emptyset \in \mathcal{K}$.*
2. *$A, B \in \mathcal{K}$ implies that $A \bigcap B \in \mathcal{K}$.*

Definition 4.11. *A collection $\mathcal{G}$ of subsets of Ω is called a Dynkin system if it contains Ω and is closed under the operations of taking complements and finite and countable non-intersecting unions, that is*

1. *$\Omega \in \mathcal{G}$.*
2. *$A \in \mathcal{G}$ implies that $\Omega \backslash A \in \mathcal{G}$.*
3. *$A_1, A_2, \ldots \in \mathcal{G}$ and $A_n \bigcap A_m = \emptyset$ for $n \neq m$ imply that $\bigcup_n A_n \in \mathcal{G}$.*

Note that an intersection of Dynkin systems is again a Dynkin system. Therefore, it makes sense to talk about the smallest Dynkin system containing a given collection of sets $\mathcal{K}$—namely, it is the intersection of all the Dynkin systems that contain all the elements of $\mathcal{K}$.

Lemma 4.12. *Let $\mathcal{K}$ be a π-system and let $\mathcal{G}$ be the smallest Dynkin system such that $\mathcal{K} \subseteq \mathcal{G}$. Then $\mathcal{G} = \sigma(\mathcal{K})$.*

Proof. Since $\sigma(\mathcal{K})$ is a Dynkin system, we obtain $\mathcal{G} \subseteq \sigma(\mathcal{K})$. In order to prove the opposite inclusion, we first note that if a π-system is a Dynkin system, then it is also a σ-algebra. Therefore, it is sufficient to show that $\mathcal{G}$ is a π-system. Let $A \in \mathcal{G}$ and define

$$\mathcal{G}_A = \{B \in \mathcal{G} : A \bigcap B \in \mathcal{G}\}.$$

The collection of sets $\mathcal{G}_A$ obviously satisfies the first and the third conditions of Definition 4.11. It also satisfies the second condition since if $A, B \in \mathcal{G}$ and $A \bigcap B \in \mathcal{G}$, then $A \bigcap (\Omega \backslash B) = \Omega \backslash [(A \bigcap B) \bigcup (\Omega \backslash A)] \in \mathcal{G}$. Moreover, if $A \in \mathcal{K}$, then $\mathcal{K} \subseteq \mathcal{G}_A$. Thus, for $A \in \mathcal{K}$ we have $\mathcal{G}_A = \mathcal{G}$, which implies that if $A \in \mathcal{K}$, $B \in \mathcal{G}$, then $A \bigcap B \in \mathcal{G}$. This implies that $\mathcal{K} \subseteq \mathcal{G}_B$ and therefore $\mathcal{G}_B = \mathcal{G}$ for any $B \in \mathcal{G}$. Thus $\mathcal{G}$ is a π-system. □

Lemma 4.12 can be re-formulated as follows.

Lemma 4.13. *If a Dynkin system $\mathcal{G}$ contains a π-system $\mathcal{K}$, then it also contains the σ-algebra generated by $\mathcal{K}$, that is $\sigma(\mathcal{K}) \subseteq \mathcal{G}$.*

Let us consider two useful applications of this lemma.

Lemma 4.14. *If P_1 and P_2 are two probability measures which coincide on all elements of a π-system $\mathcal{K}$, then they coincide on the minimal σ-algebra which contains $\mathcal{K}$.*

Proof. Let $\mathcal{G}$ be the collection of sets A such that $\mathrm{P}_1(A) = \mathrm{P}_2(A)$. Then $\mathcal{G}$ is a Dynkin system, which contains $\mathcal{K}$. Consequently, $\sigma(\mathcal{K}) \subseteq \mathcal{G}$. □

In order to discuss sequences of independent random variables and the laws of large numbers, we shall need the following statement.

Lemma 4.15. *Let $\xi_1, \ldots, \xi_n$ be independent random variables, $m_1 + \ldots + m_k = n$ and $f_1, \ldots, f_k$ be measurable functions of $m_1, \ldots, m_k$ variables respectively. Then the random variables $\eta_1 = f_1(\xi_1, \ldots, \xi_{m_1})$, $\eta_2 = f_2(\xi_{m_1+1}, \ldots, \xi_{m_1+m_2})$,..., $\eta_k = f(\xi_{m_1+\ldots+m_{k-1}+1}, \ldots, \xi_n)$ are independent.*

Proof. We shall prove the lemma in the case $k = 2$ since the general case requires only trivial modifications. Consider the sets $A = A_1 \times \ldots \times A_{m_1}$ and $B = B_1 \times \ldots \times B_{m_2}$, where $A_1, \ldots, A_{m_1}, B_1, \ldots, B_{m_2}$ are Borel subsets of $\mathbb{R}$. We shall refer to such sets as rectangles. The collections of all rectangles in $\mathbb{R}^{m_1}$ and in $\mathbb{R}^{m_2}$ are π-systems. Note that by the assumptions of the lemma,

$$\begin{aligned} &\mathrm{P}((\xi_1, \ldots, \xi_{m_1}) \in A)\mathrm{P}((\xi_{m_1+1}, \ldots, \xi_{m_1+m_2}) \in B) = \\ &\mathrm{P}((\xi_1, \ldots, \xi_{m_1}) \in A, (\xi_{m_1+1}, \ldots, \xi_{m_1+m_2}) \in B). \end{aligned} \tag{4.3}$$

Fix a set $B = B_1 \times \ldots \times B_{m_2}$ and notice that the collection of all the measurable sets A that satisfy (4.3) is a Dynkin system containing all the rectangles in $\mathbb{R}^{m_1}$. Therefore, the relation (4.3) is valid for all sets A in the smallest σ-algebra containing all the rectangles, which is the Borel σ-algebra on $\mathbb{R}^{m_1}$. Now we can fix a Borel set A and, using the same arguments, demonstrate that (4.3) holds for any Borel set B.

It remains to apply (4.3) to $A = f_1^{-1}(\overline{A})$ and $B = f_2^{-1}(\overline{B})$, where $\overline{A}$ and $\overline{B}$ are arbitrary Borel subsets of $\mathbb{R}$. $\square$

4.4 Problems

1. Let P be the probability distribution of the sequence of n Bernoulli trials, $\omega = (\omega_1, \ldots, \omega_n)$, $\omega_i = 1$ or 0 with probabilities p and $1-p$. Find $\mathrm{P}(\omega_1 = 1|\omega_1 + \ldots + \omega_n = m)$.

2. Find the distribution function of a random variable ξ which takes positive values and satisfies $\mathrm{P}(\xi > x + y|\xi > x) = \mathrm{P}(\xi > y)$ for all $x, y > 0$.

3. Two coins are in a bag. One is symmetric, while the other is not—if tossed it lands heads up with probability equal to 0.6. One coin is randomly pulled out of the bag and tossed. It lands heads up. What is the probability that the same coin will land heads up if tossed again?

4. Suppose that each of the random variables ξ and η takes at most two values, a and b. Prove that ξ and η are independent if $\mathrm{E}(\xi\eta) = \mathrm{E}\xi E\eta$.

5. Give an example of three events A_1, A_2, and A_3 which are not independent, yet pair-wise independent.

6. Give an example of two random variables ξ and η which are not independent, yet $\mathrm{E}(\xi\eta) = \mathrm{E}\xi\mathrm{E}\eta$.

7. A random variable ξ has Gaussian distribution with mean zero and variance one, while a random variable η has the distribution with the density

$$p_\eta(t) = \begin{cases} te^{-\frac{t^2}{2}} & \text{if } t \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Find the distribution of $\zeta = \xi \cdot \eta$ assuming that ξ and η are independent.

8. Let ξ_1 and ξ_2 be two independent random variables with Gaussian distribution with mean zero and variance one. Prove that $\eta_1 = \xi_1^2 + \xi_2^2$ and $\eta_2 = \xi_1/\xi_2$ are independent.

9. Two editors were independently proof-reading the same manuscript. One found a misprints, the other found b misprints. Of those, c misprints were found by both of them. How would you estimate the total number of misprints in the manuscript?

10. Let ξ, η be independent Poisson distributed random variables with expectations λ_1 and λ_2 respectively. Find the distribution of $\zeta = \xi + \eta$.

11. Let ξ, η be independent random variables. Assume that ξ has the uniform distribution on $[0, 1]$, and η has the Poisson distribution with parameter λ. Find the distribution of $\zeta = \xi + \eta$.

12. Let $\xi_1, \xi_2, \ldots$ be independent identically distributed Gaussian random variables with mean zero and variance one. Let $\eta_1, \eta_2, \ldots$ be independent identically distributed exponential random variables with mean one. Prove that there is $n > 0$ such that

$$\mathrm{P}(\max(\eta_1, \ldots, \eta_n) \geq \max(\xi_1, \ldots, \xi_n)) > 0.99.$$

13. Suppose that $\mathcal{A}_1$ and $\mathcal{A}_2$ are independent algebras, that is any two sets $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$ are independent. Prove that the σ-algebras $\sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{A}_2)$ are also independent. (Hint: use Lemma 4.12.)

14. Let $\xi_1, \xi_2, \ldots$ be independent identically distributed random variables and N be an $\mathbb{N}$-valued random variable independent of ξ_i's. Show that if ξ_1 and N have finite expectation, then

$$E \sum_{i=1}^{N} \xi_i = E(N) E(\xi_1).$$