Chapter 4

Product spaces and independence

SECTION 1 introduces independence as a property that justifies some sort of factorization of probabilities or expectations. A key factorization Theorem is stated, with proof deferred to the next Section, as motivation for the measure theoretic approach. The Theorem is illustrated by a derivation of a simple form of the strong law of large numbers, under an assumption of bounded fourth moments.

SECTION 2 formally defines independence as a property of sigma-fields. The key Theorem from Section 1 is used as motivation for the introduction of a few standard techniques for dealing with independence. Product sigma-fields are defined.

SECTION 3 describes a method for constructing measures on product spaces, starting from a family of kernels.

SECTION 4 specializes the results from Section 3 to define product measures. The Tonelli and Fubini theorems are deduced. Several important applications are presented.

*SECTION *5 discusses some difficulties encountered in extending the results of Sections 3 and 4 when the measures are not sigma-finite.*

SECTION 6 introduces a blocking technique to refine the proof of the strong law of large numbers from Section 1, to get a version that requires only a second moment condition.

*SECTION *7 introduces a truncation technique to further refine the proof of the strong law of large numbers, to get a version that requires only a first moment condition for identically distributed summands.*

*SECTION *8 discusses the construction of probability measures on products of countably many spaces.*

1. Independence

Much classical probability theory, such as the laws of large numbers and central limit theorems, rests on assumptions of independence, which justify factorizations for probabilities of intersections of events or expectations for products of random variables.

An elementary treatment usually starts from the definition of independence for events. Two events A and B are said to be independent if $\mathbb{P}(AB) = (\mathbb{P}A)(\mathbb{P}B)$; three events A, B, and C, are said to be independent if not only $\mathbb{P}(ABC) = (\mathbb{P}A)(\mathbb{P}B)(\mathbb{P}C)$ but also $\mathbb{P}(AB) = (\mathbb{P}A)(\mathbb{P}B)$ and $\mathbb{P}(AC) = (\mathbb{P}A)(\mathbb{P}C)$ and $\mathbb{P}(BC) = (\mathbb{P}B)(\mathbb{P}C)$. And so on. There are similar definitions for independence of random variables, in terms of joint distribution functions or joint densities. The definitions have two

things in common: they all assert some type of factorization; and they do not lend themselves to elementary derivation of desirable facts about independence. The measure theoretic approach, by contrast, simplifies the study of independence by eliminating unnecessary duplications of definitions, replacing them by a single concept of independence for sigma-fields, from which useful consequences are easily deduced. For example, the following key assertion is impossible to derive by elementary means, but requires only routine effort (see Section 2) to establish by measure theoretic arguments.

<1> **Theorem.** *Let $Z_1, \ldots, Z_n$ be independent random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. If $f \in \mathcal{M}^+(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ and $g \in \mathcal{M}^+(\mathbb{R}^{n-k}, \mathcal{B}(\mathbb{R}^{n-k}))$ then $f(Z_1, \ldots, Z_k)$ and $g(Z_{k+1}, \ldots, Z_n)$ are independent random variables, and*

$$\mathbb{P} f(Z_1, \ldots, Z_k) g(Z_{k+1}, \ldots, Z_n) = \mathbb{P} f(Z_1, \ldots, Z_k) \mathbb{P} g(Z_{k+1}, \ldots, Z_n).$$

<2> **Corollary.** *The same conclusion (independendence and factorization) holds for Borel measurable functions f and g taking both positive and negative values if both $f(Z_1, \ldots, Z_k)$ and $g(Z_{k+1}, \ldots, Z_n)$ are integrable.*

As you will see at the end of Section 2, the Corollary follows easily from addition and subtraction of analogous results for the functions $f^{\pm}$ and $g^{\pm}$. Problem [10] shows that the result also extends to cases where some of the integrals are infinite, provided $\infty - \infty$ problems are ruled out.

The best way for you to understand the worth of Theorem <1> and its Corollary is to see it used. At the risk of interrupting the flow of ideas, I will digress slightly to present an instructive application.

The proof of the ***strong law of large numbers*** (often referrred to by means of the acronym SLLN) illustrates well the use of Corollary <2>. Actually, several slightly different results answer to the name SLLN. A law of large numbers asserts convergence of averages to expectations, in some sense. The word "strong" specifies almost sure convergence. The various SLLN's differ in the assumptions made about the individual summands. The most common form invoked in statistical applications goes as follows.

<3> **Theorem.** *(Kolmogorov) Let $X_1, X_2, \ldots$ be independent, integrable random variables, each with the same distribution and common expectation μ. Then the average $(X_1 + \ldots + X_n)/n$ converges almost surely to μ.*

> REMARK. If $\mathbb{P}|X_1| = \infty$ then $(X_1 + \ldots + X_n)/n$ cannot converge almost surely to a finite limit (Problem [21]). Moreover Kolmogorov's zero-one law (Example <12>) implies that it cannot even converge to a finite limit at each point of a set with strictly positive probability. If only one of $\mathbb{P}X_1^{\pm}$ is infinite, the average still converges almost surely to $\mathbb{P}X_1$ (Problem [20]).

A complete proof of this form of the SLLN is quite a challenge. The classical proof (a modified version of which appears in Sections 6 and 7) combines a number of tricks that are more easily understood if introduced as separate ideas and not just rolled into one monolithic argument. The basic idea is not too hard to grasp when we have bounded fourth moments; it involves little more than an application of Corollary <2> and an appeal to the Borel-Cantelli lemma from Section 2.6.

For theoretical purposes, for summands that need not all have the same distribution, it is cleaner to work with the centered variables $X_i - \mathbb{P}X_i$, which is equivalent to an assumption that all variables have zero expected values.

<4> **Theorem.** *Let $X_1, X_2, \ldots$ be independent random variables with $\mathbb{P}X_i = 0$ for every i and $\sup_i \mathbb{P}X_i^4 < \infty$. Then $(X_1 + \ldots + X_n)/n \to 0$ almost surely.*

Proof. Define $S_n = X_1 + \ldots + X_n$. It is good enough to show, for each $\epsilon > 0$, that

<5> $$\sum_{n=1}^{\infty} \mathbb{P}\left\{\frac{|S_n|}{n} > \epsilon\right\} < \infty.$$

Do you remember why? If not, you should refer to Section 2.6 for a detailed explanation of the Borel-Cantelli argument: the series $\sum_n \{|S_n|/n > \epsilon\}$ must converge almost surely, which implies that $\limsup |S_n/n| \le \epsilon$ almost surely, from which the conclusion $\limsup |S_n/n| = 0$ follows after a casting out of a sequence of negligible sets.

Bound the nth term of the sum in <5> by $(n\epsilon)^{-4}\mathbb{P}(X_1 + \ldots + X_n)^4$. Expand the fourth power.

$$\begin{aligned}(X_1 + \ldots + X_n)^4 = \quad & X_1^4 + \ldots + X_n^4 && \boxed{1}\\ & + (\text{lots of terms like } X_1^3 X_2) && \boxed{2}\\ & + \binom{n}{2} \text{ terms like } 6X_1^2X_2^2 && \boxed{3}\\ & + (\text{lots of terms like } X_1^2 X_2 X_3) && \boxed{4}\\ & + (\text{lots of terms like } X_1X_2X_3X_4) && \boxed{5}\end{aligned}$$

The contributions to $\mathbb{P}(X_1 + \ldots + X_n)^4$ from the five groups of terms are:

- 1 $\sum_{i \le n} \mathbb{P}X_i^4 \le nM$, where $M = \sup_i \mathbb{P}X_i^4$;
- 2 zero, because $\mathbb{P}(X_1^3 X_2) = (\mathbb{P}X_1^3)(\mathbb{P}X_2) = 0$;
- 3 less than $12\binom{n}{2}M$, because $\mathbb{P}(X_1^2X_2^2) \le \mathbb{P}X_1^4 + \mathbb{P}X_2^4 \le 2M$;
- 4 zero, because $\mathbb{P}(X_1^2X_2X_3) = (\mathbb{P}X_1^2X_2)(\mathbb{P}X_3) = 0$;
- 5 zero, because $\mathbb{P}(X_1X_2X_3X_4) = (\mathbb{P}X_1X_2X_3)(\mathbb{P}X_4) = 0$.

Notice all the factorizations due to independence. Combining these bounds and equalities we get $\mathbb{P}\{|S_n|/n > \epsilon\} = O\left(n^{-2}\right)$, from which <5> follows. □

If you feel that Theorem <4> is good enough for 'practical purposes,' and that all the extra work to whittle a fourth moment assumption down to a first moment assumption is hardly worth the gain in generality, you might like to contemplate the following example. How natural, or restrictive, would it be if we were to assume finite fourth moments?

<6> **Example.** Let $\{P_\theta : \theta = 0, 1, \ldots, N\}$ be a finite family of distinct probability measures, defined by densities $\{p_\theta\}$ with respect to a measure μ. Suppose observations $X_1, X_2, \ldots$ are generated independently from P_0. The maximum likelihood estimator $\widehat{\theta}_n(\omega)$ is defined as the value that maximizes $L_n(\theta, \omega) := \prod_{i \le n} p_\theta(X_i(\omega))$. The SLLN will show that $\mathbb{P}\{\widehat{\theta}_n = 0 \text{ eventually}\} = 1$. That is, the maximum likelihood estimator eventually picks the true value of θ.

It will be enough to show, for each $\theta \neq 0$, that with probability one, $\log(L_n(\theta)/L_n(0)) < 0$ eventually. For fixed $\theta \neq 0$ define $\ell_i = \log(p_\theta(X_i)/p_0(X_i))$. By Jensen's inequality, with a strict inequality because $P_\theta \neq P_0$,

$$\begin{aligned}\mathbb{P}\ell_i = P_0^x \log\left(\frac{p_\theta(x)}{p_0(x)}\right) &< \log \mu^x\left(p_0(x)\frac{p_\theta(x)}{p_0(x)}\right)\\ &= \log \mu^x p_\theta(x)\{p_0(x) \neq 0\} \leq 0.\end{aligned}$$

By the SLLN (or its extension from Problem [20] if $\mathbb{P}\ell_i = -\infty$), for almost all ω there exists a finite $n_0(\omega,\theta)$ for which $0 > n^{-1}\sum_{i\leq n}\ell_i := n^{-1}\log(L_n(\theta)/L_n(0))$ when $n \geq n_0(\omega,\theta)$. When $n \geq \max_{\theta=1}^N n_0(\omega,\theta)$, we have $\max_{\theta=1}^N L_n(\theta) < L_n(0)$, in which case the maximizing $\widehat{\theta}_n$ prefers 0 to each $\theta \geq 1$. □

REMARK. Notice that the argument would not work if the index set were infinite. To handle such sets, one typically imposes compactness assumptions to reduce to the finite case, by means of a much-imitated method originally due to Wald (1949).

2. Independence of sigma-fields

Technically speaking, the best treatment of independence starts with the concept of independent sub-sigma-fields of $\mathcal{F}$, for a fixed probability space $(\Omega,\mathcal{F},\mathbb{P})$. This Section will develop the appropriate definitions and techniques for dealing with independence of sigma-fields, using the ideas needed for the proof of Theorem <2> as motivation.

<7> **Definition.** *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space. Sub-sigma-fields $\mathcal{G}_1,\ldots,\mathcal{G}_n$ of $\mathcal{F}$ are said to be independent if*

$$\mathbb{P}(G_1\ldots G_n) = (\mathbb{P}G_1)\ldots(\mathbb{P}G_n) \qquad \text{for all } G_i \in \mathcal{G}_i, \text{ for } i = 1,\ldots n.$$

An infinite collection of sub-sigma-fields $\{\mathcal{G}_i : i \in I\}$ is said to be independent if each finite subcollection is independent, that is, if $\mathbb{P}(\cap_{i\in S}G_i) = \prod_{i\in S}\mathbb{P}G_i$ for all finite subsets S of I, and all choices $G_i \in \mathcal{G}_i$ for each i in S.

The definition neatly captures all the factorizations involved in the elementary definitions of independence for more than two events.

<8> **Example.** Let A, B, and C be events. They generate sigma-fields $\mathcal{A} = \{\emptyset, A, A^c, \Omega\}$, and $\mathcal{B} = \{\emptyset, B, B^c, \Omega\}$, and $\mathcal{C} = \{\emptyset, C, C^c, \Omega\}$. Independence of the three sigma-fields requires factorization for $4^3 = 64$ triples of events, amongst which are the four factorizations stated at the start of Section 1 as the elementary definition of independence for the three events A, B, and C. In fact, all 64 factorizations are consequences of those four. For example, any factorization where one of the factors is the empty set will reduce to the identity $0 = 0$. The factorization $\mathbb{P}(AB^cC) = (\mathbb{P}A)(\mathbb{P}B^c)(\mathbb{P}C)$ follows from $\mathbb{P}(AC) = (\mathbb{P}A)(\mathbb{P}C)$ and $\mathbb{P}(ABC) = (\mathbb{P}A)(\mathbb{P}B)(\mathbb{P}C)$, by subtraction. And so on. □

Generating class arguments, such as the π–λ Theorem from Section 2.10, make it easy to derive facts about independent sigma-fields. For example, Problem [8] uses such arguments in a routine way to establish the following result.

<9> **Theorem.** *Let $\mathcal{E}_1, \ldots, \mathcal{E}_n$ be classes of measurable sets, each class stable under finite intersections and containing the whole space Ω. If*

$$\mathbb{P}(E_1 E_2 \ldots E_n) = (\mathbb{P}E_1)(\mathbb{P}E_2)\ldots(\mathbb{P}E_n) \qquad \text{for all } E_i \in \mathcal{E}_i, \text{ for } i = 1, 2, \ldots, n,$$

then the sigma-fields $\sigma(\mathcal{E}_1), \sigma(\mathcal{E}_2), \ldots, \sigma(\mathcal{E}_n)$ are independent.

REMARK. The requirement that $\Omega \in \mathcal{E}_i$ for each i is just a sneaky way of getting factorizations for intersections of fewer than n sets.

<10> **Corollary.** *Let $\{\mathcal{E}_i : i \in I\}$ be classes of measurable sets, each stable under finite intersections. If $\mathbb{P}(\cap_{i\in S} E_i) = \prod_{i\in S} \mathbb{P}E_i$ for all finite subsets S of I, and all choices $E_i \in \mathcal{E}_i$ for each i in S, then the sigma-fields $\sigma(\mathcal{E}_i)$, for $i \in I$, are independent.*

Proof. Notice the alternative to requiring $\Omega \in \mathcal{E}_i$ for every i. Theorem <9> establishes independence for each finite subcollection. □

<11> **Corollary.** *Let $\{\mathcal{G}_i : i \in I\}$ be independent sigma-fields. If $\{I_j : j \in J\}$ are disjoint subsets of I, then the sigma-fields $\sigma\left(\cup_{i\in I_j} \mathcal{G}_i\right)$, for $j \in J$, are independent.*

Proof. Invoke Corollary <10> with $\mathcal{E}_j$ consisting of the collection of all finite intersections of sets chosen from $\cup_{i\in I_j} \mathcal{G}_i$. □

<12> **Example.** Let $\{\mathcal{G}_i : i \in \mathbb{N}\}$ be a sequence of independent sigma-fields. For each n let $\mathcal{H}_n$ denote the sigma-field generated by $\cup_{i>n}\mathcal{G}_i$. The ***tail sigma-field*** is defined as $\mathcal{H}_\infty := \cap_n \mathcal{H}_n$. Kolmogorov's ***zero-one law*** asserts that, for each H in $\mathcal{H}_\infty$, either $\mathbb{P}H = 0$ or $\mathbb{P}H = 1$. Equivalently, the sigma-field $\mathcal{H}_\infty$ is independent of itself, so that $\mathbb{P}(HH) = (\mathbb{P}H)(\mathbb{P}H)$ for every H in $\mathcal{H}_\infty$.

For each finite n, Corollary <11> implies independence of $\mathcal{H}_n, \mathcal{G}_1, \ldots, \mathcal{G}_n$. From the fact that $\mathcal{H}_\infty \subseteq \mathcal{H}_n$ for every n, it then follows that each finite subcollection of $\{\mathcal{H}_\infty, \mathcal{G}_i : i \in \mathbb{N}\}$ is independent, and hence the whole collection of sigma-fields is independent. From Corollary <11> again, $\mathcal{H}_\infty$ and $\mathcal{F}_\infty := \sigma\left(\cup_{i\in\mathbb{N}}\mathcal{G}_i\right)$ are independent. To complete the argument, note that $\mathcal{F}_\infty \supseteq \mathcal{H}_\infty$. □

Random variables (or random vectors, or random elements of more general spaces) inherit their definition of independence from the sigma-fields they generate. Recall that if X is a map from Ω into a set $\mathcal{X}$, equipped with a sigma-field $\mathcal{A}$, then the sigma-field $\sigma(X)$ on Ω generated by X is defined as the smallest sigma-field $\mathcal{G}$ for which X is $\mathcal{G}\backslash\mathcal{A}$-measurable. It consists of all sets of the form $\{\omega \in \Omega : X(\omega) \in A\}$, with $A \in \mathcal{A}$.

REMARK. The extra generality gained by allowing maps into arbitrary measurable spaces will not be wasted; but in the first instance you could safely imagine each space to be the real line, ignoring the fact that the definition also covers independence of random vectors and independence of stochastic processes.

<13> **Definition.** *Measurable maps X_i, for $i \in I$, from Ω into measurable spaces $(\mathcal{X}_i, \mathcal{A}_i)$ are said to be independent if the sigma-fields that they generate are independent, that is, if*

<14>
$$\mathbb{P}\left(\bigcap_{i\in S}\{X_i \in A_i\}\right) = \prod_{i\in S}\mathbb{P}\{X_i \in A_i\},$$

for all finite subsets S of the index set I, and all choices of $A_i \in \mathcal{A}_i$ for $i \in S$.

Results about independent random variables are usually easy to deduce from the corresponding results about independent sigma-fields.

<15> **Example.** Real random variables X_1 and X_2 for which

$$\mathbb{P}\{X_1 \le x_1,\ X_2 \le x_2\} = \mathbb{P}\{X_1 \le x_1\}\mathbb{P}\{X_2 \le x_2\} \qquad \text{for all } x_1, x_2 \text{ in } \mathbb{R}$$

are independent, because the collections of sets $\mathcal{E}_i = \{\{X_i \le x\} : x \in \mathbb{R}\}$ are both stable under finite intersections, and $\sigma(X_i) = \sigma(\mathcal{E}_i)$. □

We now have the tools needed to establish Theorem <2>. Write X_1 for $f(Z_1, \ldots, Z_k)$ and X_2 for $g(Z_{k+1}, \ldots, Z_n)$. Write $\mathcal{G}_i$ for $\sigma(Z_i)$, the sigma-field generated by the random variable Z_i. From Corollary <11>, the sigma-fields $\mathcal{F}_1 := \sigma(\mathcal{G}_1 \cup \ldots \cup \mathcal{G}_k)$ and $\mathcal{F}_2 := \sigma(\mathcal{G}_{k+1} \cup \ldots \cup \mathcal{G}_n)$ are independent. If we can show that X_1 is $\mathcal{F}_1\backslash\mathcal{B}(\mathbb{R})$-measurable and X_2 is $\mathcal{F}_2\backslash\mathcal{B}(\mathbb{R})$-measurable, then their independence will follow: we will have the desired factorization for all sets of the form $\{X_1 \in A_1\}$ and $\{X_2 \in A_2\}$, for Borel sets A_1 and A_2.

Consider first the measurability property for X_1. Temporarily write $\mathbf{Z}$ for $(Z_1, \ldots, Z_k)$, a map from Ω into $\mathbb{R}^k$. We need to show that the set

$$\{X_1 \in A\} = \{\mathbf{Z} \in f^{-1}(A)\}$$

belongs to $\mathcal{F}_1$ for every A in $\mathcal{B}(\mathbb{R})$. The $\mathcal{B}(\mathbb{R}^k)\backslash\mathcal{B}(\mathbb{R})$-measurability of f ensures that $f^{-1}(A) \in \mathcal{B}(\mathbb{R}^k)$. We therefore need only show that $\{\mathbf{Z} \in B\} \in \mathcal{F}_1$ for every B in $\mathcal{B}(\mathbb{R}^k)$, that is, that the map $\mathbf{Z}$ is $\mathcal{F}_1\backslash\mathcal{B}(\mathbb{R}^k)$-measurable.

As with many measure theoretic problems, it is better to turn the question around and ask: For how extensive a class of sets B does $\{\mathbf{Z} \in B\}$ belong to $\mathcal{F}_1$? It is very easy to show that the class $\mathcal{B}_0$ of all such B is a sigma-field; so $\mathbf{Z}$ is an $\mathcal{F}_1\backslash\mathcal{B}_0$-measurable function. Moreover, for all choices of $D_i \in \mathcal{B}(\mathbb{R})$, the set

$$D := \{(z_1, \ldots, z_k) \in \mathbb{R}^k : z_i \in D_i \text{ for } i = 1, \ldots, k\}$$

belongs to $\mathcal{B}_0$ because $\{\mathbf{Z} \in D\} = \cap_i\{Z_i \in D_i\} \in \mathcal{F}_1$. As shown in Problem [6], the collection of all such D sets generates the Borel sigma-field $\mathcal{B}(\mathbb{R}^k)$. Thus $\mathcal{B}(\mathbb{R}^k) \subseteq \mathcal{B}_0$, and $\{\mathbf{Z} \in B\} \in \mathcal{F}_1$ for all $B \in \mathcal{B}(\mathbb{R}^k)$. It follows that X_1 is $\mathcal{F}_1\backslash\mathcal{B}(\mathbb{R})$-measurable. Similarly, X_2 is $\mathcal{F}_2\backslash\mathcal{B}(\mathbb{R})$-measurable. The random variables X_1 and X_2 are independent, as asserted by Theorem <2>.

The whole argument can be carried over to random elements of more general spaces if we work with the right sigma-fields.

<16> **Definition.** *Let $\mathcal{X}_1, \ldots, \mathcal{X}_n$ be sets equipped with sigma-fields $\mathcal{A}_1, \ldots, \mathcal{A}_n$. The set of all ordered n-tuples $(x_1, \ldots, x_n)$, with $x_i \in \mathcal{X}_i$ for each i is denoted by $\mathcal{X}_1 \times \ldots \times \mathcal{X}_n$ or $\mathsf{X}_{i \le n}\mathcal{X}_i$. It is called the **product** of the $\{\mathcal{X}_i\}$. A set of the form*

$$A_1 \times \ldots \times A_n = \{(x_1, \ldots, x_n) \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_n : x_i \in A_i \text{ for each } i\},$$

*with $A_i \in \mathcal{A}_i$ for each i, is called a **measurable rectangle**. The product sigma-field $\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n$ on $\mathcal{X}_1 \times \ldots \times \mathcal{X}_n$ is defined to be the sigma-field generated by all measurable rectangles.*

REMARK. Even if n equals 2 and $\mathcal{X}_1 = \mathcal{X}_2 = \mathbb{R}$, there is is no presumption that either A_1 or A_2 is an interval—a measurable rectangle might be composed of

many disjoint pieces. The symbol $\otimes$ in place of $\times$ is intended as a reminder that $\mathcal{A}_1 \otimes \mathcal{A}_2$ consists of more than the set of all measurable rectangles $A_1 \times A_2$.

If Z_i is an $\mathcal{F}\backslash\mathcal{A}_i$-measurable map from Ω into $\mathcal{X}_i$, for $i = 1, \ldots, n$, then the map $\omega \mapsto \mathbf{Z}(\omega) = (Z_1(\omega), \ldots, Z_n(\omega))$ from Ω into $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$ is $\mathcal{F}\backslash\mathcal{A}$-measurable, where $\mathcal{A}$ denotes the product sigma-field $\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n$. If f is an $\mathcal{A}\backslash\mathcal{B}(\mathbb{R})$-measurable real-valued function on $\mathcal{X}$ then $f(\mathbf{Z})$ is $\mathcal{F}\backslash\mathcal{B}(\mathbb{R})$-measurable.

The second assertion of Theorem <1> is now reduced to a factorization property for products of independent random variables, a result easily deduced from the defining factorization for independence of sigma-fields by means of the usual approximation arguments.

<17> **Lemma.** *Let X and Y be independent random variables. If either $X \geq 0$ and $Y \geq 0$, or both X and Y are integrable, then $\mathbb{P}(XY) = (\mathbb{P}X)(\mathbb{P}Y)$. The product XY is integrable if both X and Y are integrable.*

Proof. Consider first the case of nonnegative variables. Express X and Y as monotone increasing limits of simple random variables (as in Section 2.2), $X_n := 2^{-n}\sum_{1\leq i\leq 4^n}\{X \geq i/2^n\}$ and $Y_n := 2^{-n}\sum_{1\leq i\leq 4^n}\{Y \geq i/2^n\}$. Then, for each n,

$$\begin{aligned}
\mathbb{P}(X_nY_n) &= 4^{-n}\textstyle\sum_{i,j}\mathbb{P}\left(\{X \geq i/2^n\}\{Y \geq j/2^n\}\right) \\
&= 4^{-n}\textstyle\sum_{i,j}\left(\mathbb{P}\{X \geq i/2^n\}\right)\left(\mathbb{P}\{Y \geq j/2^n\}\right) \qquad \text{by independence} \\
&= \left(2^{-n}\textstyle\sum_i \mathbb{P}\{X \geq i/2^n\}\right)\left(2^{-n}\textstyle\sum_j \mathbb{P}\{Y \geq j/2^n\}\right) \\
&= (\mathbb{P}X_n)(\mathbb{P}Y_n).
\end{aligned}$$

Invoke Monotone Convergence twice in the passage to the limit to deduce $\mathbb{P}(XY) = (\mathbb{P}X)(\mathbb{P}Y)$.

For the case of integrable random variables, factorize expectations for products of positive and negative parts, $\mathbb{P}(X^{\pm}Y^{\pm}) = (\mathbb{P}X^{\pm})(\mathbb{P}Y^{\pm})$. Each of the four products represented by the right-hand side is finite. Complete the argument by splitting each term on the right-hand side of the decomposition

$$\mathbb{P}(XY) = \mathbb{P}(X^+Y^+) - \mathbb{P}(X^+Y^-) - \mathbb{P}(X^-Y^+) + \mathbb{P}(X^-Y^-)$$

into a product of expectations, then refactorize as $(\mathbb{P}X^+ - \mathbb{P}X^-)(\mathbb{P}Y^+ - \mathbb{P}Y^-)$. Integrability of XY follows from a similar decomposition for $\mathbb{P}|XY|$. □

3. Construction of measures on a product space

The probabilistic concepts of independence and conditioning are both closely related to the measure theoretic constructions for measures on product spaces. As you will see in Chapter 5, conditioning may be thought of as a inverse operation to a general construction whereby a measure on a product space is built from families of measures on the component spaces. For probability measures the components have the interpretation of distributions involved in a two-stage experiment. Product measures, and independence, correspond to the special case where the second stage of the experiment does not depend on the first stage. Many traditional facts about