6. Independence

Heuristically, two objects are independent if information concerning one of them does not contribute to one's knowledge about the other. The correct way to formally codify this notion in a manner amenable to proving theorems is in terms of a sort of multiplication rule.

- Two events A and B are independent if $P(A \cap B) = P(A)P(B)$.
- Two random variables X and Y are independent if $P(X \in E, Y \in F) = P(X \in E)P(Y \in F)$ for all $E, F \in \mathcal{B}$. (That is, if the events $\{X \in E\}$ and $\{Y \in F\}$ are independent.)
- Two sub-σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$ are independent if for all $A \in \mathcal{F}_1$, $B \in \mathcal{F}_2$, the events A and B are independent.

Observe that if $A \in \mathcal{F}$ has $P(A) = 0$ or $P(A) = 1$, then A is independent of every $B \in \mathcal{F}$.
This also implies that if X is a.s. constant, then X is independent of every $Y \in \mathcal{F}$.

An infinite collection of objects (sub-σ-algebras, random variables, events) is said to be independent if every finite subcollection is independent, where

- Events $A_1, ..., A_n \in \mathcal{F}$ are independent if for any $I \subseteq [n]$, we have
$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i).$$
- Random variables $X_1, ..., X_n \in \mathcal{F}$ are independent if for any choice of $E_i \in \mathcal{B}_i$, $i = 1, ..., n$, we have
$$P(X_1 \in E_1, ..., X_n \in E_n) = \prod_{i=1}^{n} P(X_i \in E_i).$$
- sub-σ-algebras $\mathcal{F}_1, ..., \mathcal{F}_n$ are independent if for any choice of $A_i \in \mathcal{F}_i$, $i = 1, ..., n$, we have
$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i).$$

Note that σ-algebras and random variables are implicitly subject to the same subcollection constraint as events since special cases of the definition include taking $A_i = \Omega$, $E_i = \mathbb{R}$ for $i \in I^C$.

For any $n \in \mathbb{N}$, it is possible to construct families of objects which are not independent, but every subcollection of size $m \leq n$ satisfies the applicable multiplication rule. For example, just because a collection of events $A_1, ..., A_n$ satisfies $P(A_i \cap A_j) = P(A_i)P(A_j)$ for all $i \neq j$ (called *pairwise independence*), it is not necessarily the case that $A_1, ..., A_n$ is an independent collection of events.
(Flip two fair coins and let $A = \{$1st coin heads$\}$, $B = \{$2nd coin heads$\}$, $C = \{$both coins same$\}$.)

One can show that independence of events is a special case of independence of random variables (via indicator functions), which in turn is a special case of independence of σ-algebras (via the σ-algebras the random variables generate). We will take as our running definition of independence, the further generalization:

Definition. Given a probability space $(\Omega, \mathcal{F}, P)$, collections of events $\mathcal{A}_1, ..., \mathcal{A}_n \subseteq \mathcal{F}$ are independent if for all $I \subseteq [n]$,
$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$$
whenever $A_i \in \mathcal{A}_i$ for each $i \in I$.
An infinite collection of subsets of $\mathcal{F}$ is independent if every finite subcollection is.

Observe that if $\mathcal{A}_1, ..., \mathcal{A}_n$ is independent and we set $\overline{\mathcal{A}_i} = \mathcal{A}_i \cup \{\Omega\}$, then $\overline{\mathcal{A}_1}, ..., \overline{\mathcal{A}_n}$ is independent as well. In this case, the independence criterion reduces to $P\left(\bigcap_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$ for any choice of $A_i \in \overline{\mathcal{A}_i}$.

Sufficient Conditions for Independence.

The preceding definitions often require us to check a huge number of cases to determine whether a given collection of objects is independent. The following results are useful for simplifying this task.

Theorem 6.1. *Suppose that $\mathcal{A}_1, ..., \mathcal{A}_n$ are independent collections of events. If each $\mathcal{A}_i$ is a π-system, then the sub-σ-algebras $\sigma(\mathcal{A}_1), ..., \sigma(\mathcal{A}_n)$ are independent.*

Proof. Because $\sigma(\mathcal{A}_i) = \sigma(\overline{\mathcal{A}_i})$ where $\overline{\mathcal{A}_i} = \mathcal{A}_i \cup \{\Omega\}$, we can assume without loss of generality that $\Omega \in \mathcal{A}_i$ for all i so that we need only consider intersections/products over $[n]$.

Let $A_2, ..., A_n$ be events with $A_i \in \mathcal{A}_i$, set $F = \bigcap_{i=2}^n A_i$, and set $\mathcal{L} = \{A \in \mathcal{F} : P(A \cap F) = P(A)P(F)\}$.

Since $P(\Omega \cap F) = P(F) = P(\Omega)P(F)$, we have that $\Omega \in \mathcal{L}$.

Now suppose that $A, B \in \mathcal{L}$ with $A \subseteq B$. Then

$$\begin{aligned} P\left((B \setminus A) \cap F\right) &= P\left((B \cap F) \setminus (A \cap F)\right) = P(B \cap F) - P(A \cap F) \\ &= P(B)P(F) - P(A)P(F) = \left(P(B) - P(A)\right)P(F) = P(B \setminus A)P(F), \end{aligned}$$

hence $(B \setminus A) \in \mathcal{L}$.

Finally, let $B_1, B_2, ... \in \mathcal{L}$ with $B_n \nearrow B$. Then $(B_n \cap F) \nearrow (B \cap F)$, so

$$P(B \cap F) = \lim_{n\to\infty} P(B_n \cap F) = \lim_{n\to\infty} P(B_n)P(F) = P(B)P(F),$$

so $B \in \mathcal{L}$ as well.

Therefore, $\mathcal{L}$ is a λ-system, so, since $\mathcal{A}_1$ is a π-system contained in $\mathcal{L}$ by assumption, the π-λ Theorem shows that $\sigma(\mathcal{A}_1) \subseteq \mathcal{L}$.

Because $A_2, ..., A_n$ were arbitrary members of $\mathcal{A}_2, ..., \mathcal{A}_n$, we conclude that $\sigma(\mathcal{A}_1), \mathcal{A}_2, ..., \mathcal{A}_n$ are independent.

Repeating this argument for $\mathcal{A}_2, \mathcal{A}_3, ..., \mathcal{A}_n, \sigma(\mathcal{A}_1)$ shows that $\sigma(\mathcal{A}_2), \mathcal{A}_3, ..., \mathcal{A}_n, \sigma(\mathcal{A}_1)$ are independent, and $n-2$ more iterations completes the proof. ☐

A useful corollary is given by

Corollary 6.1. *Random variables $X_1, ..., X_n$ are independent if*

$$P(X_1 \le x_1, ..., X_n \le x_n) = \prod_{i=1}^n P(X_i \le x_i) \text{ for all } x_1, ..., x_n \in \mathbb{R}.$$

Proof. Let $\mathcal{A}_i = \{\{X_i \le x\} : x \in \mathbb{R}\}$ for $i = 1, ..., n$.

Since $\{X_i \le x\} \cap \{X_i \le y\} = \{X_i \le x \wedge y\}$, the $\mathcal{A}_i' s$ are π-systems, so $\sigma(\mathcal{A}_1), ..., \sigma(\mathcal{A}_n)$ are independent by Theorem 6.1.

Because $\{(-\infty, x] : x \in \mathbb{R}\}$ generates $\mathcal{B}$, $\sigma(\mathcal{A}_i) = \sigma(X_i)$, and the result follows. ☐

Since the converse of Corollary 6.1 is true by definition, independence of random variables $X_1, ..., X_n$ is equivalent to the condition that their joint cdf factors as a product of the marginals cdfs.

One can prove analogous results for density and mass functions using the same basic ideas.

It is clear that if $X_1, ..., X_n$ are independent random variables and $f_1, ..., f_n : \mathbb{R} \to \mathbb{R}$ are measurable, then $f(X_1), ..., f(X_n)$ are independent random variables since for any choice of $B_i \in \mathcal{B}_i$,

$$\begin{aligned} P\left(f_1(X_i) \in B_1, ..., f_n(X_n) \in B_n\right) &= P\left(X_1 \in f_1^{-1}(B_1), ..., X_n \in f_n^{-1}(B_n)\right) \\ &= \prod_{i=1}^{n} P\left(X_i \in f_i^{-1}(B_i)\right) = \prod_{i=1}^{n} P\left(f_i(X_i) \in B_i\right). \end{aligned}$$

With the help of Theorem 6.1, we can prove the stronger result that functions of disjoint sets of independent random variables are independent.

Lemma 6.1. *Suppose $\mathcal{F}_{i,j}$, $1 \le i \le n$, $1 \le j \le m(i)$, are independent sub-σ-algebras and let $\mathcal{G}_i = \sigma\left(\bigcup_j \mathcal{F}_{i,j}\right)$. Then $\mathcal{G}_1, ..., \mathcal{G}_n$ are independent.*

Proof. Let $\mathcal{A}_i = \left\{\bigcap_j A_{i,j} : A_{i,j} \in \mathcal{F}_{i,j}\right\}$.

If $\bigcap_j A_{i,j}, \bigcap_j B_{i,j} \in \mathcal{A}_i$, then $\left(\bigcap_j A_{i,j}\right) \cap \left(\bigcap_j B_{i,j}\right) = \bigcap_j \left(A_{i,j} \cap B_{i,j}\right) \in \mathcal{A}_i$, thus $\mathcal{A}_i$ is a π-system, so $\sigma(\mathcal{A}_1), ..., \sigma(\mathcal{A}_n)$ are independent by Theorem 6.1.

Because $F \in \bigcup_j \mathcal{F}_{i,j}$ implies $F \in \mathcal{F}_{i,k}$ for some k and thus $F = \Omega \cap \cdots \cap \Omega \cap F \cap \Omega \cap \cdots \cap \Omega \in \mathcal{A}_i$, we have that $\bigcup_j \mathcal{F}_{i,j} \subseteq \mathcal{A}_i$, so $\mathcal{G}_i = \sigma\left(\bigcup_j \mathcal{F}_{i,j}\right) \subseteq \sigma(\mathcal{A}_i)$. Consequently, $\mathcal{G}_1, ..., \mathcal{G}_n$ are independent. □

Corollary 6.2. *If $X_{i,j}$, $1 \le i \le n$, $1 \le j \le m(i)$, are independent random variables and the functions $f_i : \mathbb{R}^{m(i)} \to \mathbb{R}$ are measurable, then $f_1(X_{1,1}, ..., X_{1,m(1)}), ..., f_n(X_{n,1}, ..., X_{n,m(n)})$ are independent.*

Proof. Let $\mathcal{F}_{i,j} = \sigma(X_{i,j})$. Since $f_i(X_{i,1}, ..., X_{i,m(i)})$ is measurable with respect to $\mathcal{G}_i = \sigma\left(\bigcup_j \mathcal{F}_{i,j}\right)$, the result follows from Lemma 6.1. □

Product Measure.

We now pause to recall the construction of product measures.

Proposition 6.1. *Given finite measure spaces* $(\Omega_1, \mathcal{F}_1, \mu_1)$ *and* $(\Omega_2, \mathcal{F}_2, \mu_2)$*, there exists a unique measure* $\mu_1 \times \mu_2$ *on* $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ *which satisfies* $(\mu_1 \times \mu_2)(A \times E) = \mu_1(A)\mu_2(E)$ *for all* $A \in \mathcal{F}_1,\ E \in \mathcal{F}_2$*.*

Proof.

Let $\mathcal{S} = \{A \times E : A \in \mathcal{F}_1, E \in \mathcal{F}_2\}$.

If $A_1 \times E_1, A_2 \times E_2 \in \mathcal{S}$, then $(A_1 \times E_1) \cap (A_2 \times E_2) = (A_1 \cap A_2) \times (E_1 \cap E_2)$ and $(A_1 \times E_1)^C = \left(A_1^C \times E_1\right) \sqcup \left(A_1 \times E_1^C\right) \sqcup \left(A_1^C \times E_1^C\right)$, hence $\mathcal{S}$ is a semialgebra.

Define $\nu : \mathcal{S} \to [0, \infty)$ by $\nu(A \times E) = \mu_1(A)\mu_2(E)$.

In light of the discussion in Section 3, the result will follow if we can show that for any countable disjoint union of sets $\{A_i \times E_i\}_{i \in I}$ in $\mathcal{S}$ such that $A \times E = \bigcup_{i \in I}(A_i \times E_i) \in \mathcal{S}$, we have $\nu\left(A \times E\right) = \sum_{i \in I} \nu(A_i \times E_i)$.

To see that this is so, observe that for all $(x, y) \in \Omega_1 \times \Omega_2$,

$$1_A(x)1_E(y) = 1_{A \times E}(x, y) = \sum_{i \in I} 1_{A_i \times E_i}(x, y) = \sum_{i \in I} 1_{A_i}(x)1_{E_i}(y).$$

Consequently,

$$\begin{aligned}\mu_1(A)1_E(y) &= \int_{\Omega_1} 1_A(x)1_E(y)d\mu_1(x) = \int_{\Omega_1} \sum_{i \in I} 1_{A_i}(x)1_{E_i}(y)d\mu_1(x) \\ &= \sum_{i \in I} \int_{\Omega_1} 1_{A_i}(x)1_{E_i}(y)d\mu_1(x) = \sum_{i \in I} \left(\int_{\Omega_1} 1_{A_i}(x)d\mu_1(x)\right) 1_{E_i}(y) \\ &= \sum_{i \in I} \mu_1(A_i)1_{E_i}(y).\end{aligned}$$

(The interchange of summation and integration is justified by the monotone convergence theorem.)

Integrating against μ_2 gives

$$\begin{aligned}\nu(A \times E) = \mu_1(A)\mu_2(E) &= \int_{\Omega_2} \mu_1(A)1_E(y)d\mu_2(y) = \int_{\Omega_2} \sum_{i \in I} \mu_1(A_i)1_{E_i}(y)d\mu_2(y) \\ &= \sum_{i \in I} \mu_1(A_i) \int_{\Omega_2} 1_{E_i}(y)d\mu_2(y) = \sum_{i \in I} \mu_1(A_i)\mu_2(E_i) = \sum_{i \in I} \nu(A_i \times E_i).\end{aligned}$$ □

* The above holds for σ-finite measure spaces as well by the same argument, but we mainly care about finite measure spaces in probability.

An induction argument easily extends Proposition 6.1 to arbitrary finite products.

Independence, Distribution, and Expectation.

We now consider the joint distribution of independent random variables.

Theorem 6.2. *If $X_1, ..., X_n$ are independent random variables with distributions $\mu_1, ..., \mu_n$, respectively, then the random vector $(X_1, ..., X_n)$ has distribution $\mu_1 \times \cdots \times \mu_n$.*

Proof. Given any sets $A_1, ..., A_n \in \mathcal{B}$, we have

$$P((X_1, ..., X_n) \in A_1 \times \cdots \times A_n) = P(X_1 \in A_1, ..., X_n \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$$
$$= \prod_{i=1}^{n} \mu_i(A_i) = (\mu_1 \times \cdots \times \mu_n)(A_1 \times \cdots \times A_n).$$

In the proof of Theorem 3.3, we showed that for any probability measures μ, ν, $\mathcal{L} = \{A : \mu(A) = \nu(A)\}$ is a λ-system. Because the collection of rectangle sets is a π-system which generates $\mathcal{B}^n$, the result follows from the π-λ Theorem. ☐

In other words random variables are independent if their joint distribution is the product of their marginal distributions.

At this point, it is appropriate to recall the theorems of Fubini and Tonelli, whose proofs can be found in any book on measure theory.

Theorem 6.3. *Suppose that $(R, \mathcal{F}, \mu)$ and $(S, \mathcal{G}, \nu)$ are σ-finite measure spaces.*

I) Tonelli: *If $f : R \times S \to [0, \infty)$ is a measurable function, then*

$$(*) \int_{R \times S} f d(\mu \times \nu) = \int_S \left(\int_R f(x, y) d\mu(x) \right) d\nu(y) = \int_R \left(\int_S f(x, y) d\nu(y) \right) d\mu(x).$$

II) Fubini: *If $f : R \times S \to \mathbb{R}$ is integrable (i.e. $\int |f| \, d(\mu \times \nu) < \infty$), then $(*)$ holds.*

In the language of probability, we have

Theorem 6.4. *Suppose that X and Y are independent with distributions μ and ν. If $f : \mathbb{R}^2 \to \mathbb{R}$ is a measurable function with $f \geq 0$ or $E|f(X, Y)| < \infty$, then*

$$E[f(X, Y)] = \int \int f(x, y) d\mu(x) d\nu(y).$$

In particular, if $g, h : \mathbb{R} \to \mathbb{R}$ are measurable functions with $g, h > 0$ or $E|g(X)|, E|h(Y)| < \infty$, then

$$E[g(X)h(Y)] = E[g(X)]E[h(Y)].$$

Proof. It follows from Theorem 6.2 and the change of variables formula (Theorem 5.8) that

$$E[f(X, Y)] = \int_{\mathbb{R}^2} f(x, y) d(\mu \times \nu)(x, y),$$

so the first statement follows from Fubini-Tonelli.

Now suppose that $g, h : \mathbb{R} \to \mathbb{R}$ are nonnegative measurable functions. Then Tonelli's Theorem gives

$$E[g(X)h(Y)] = \int\int g(x)h(y)d\mu(x)d\nu(y) = \int h(y)\left(\int g(x)d\mu(x)\right)d\nu(y)$$
$$= \int h(y)E[g(X)]d\nu(y) = E[g(X)]\int h(y)d\nu(y) = E[g(X)]E[h(Y)].$$

If g, h are integrable, then applying the above result to $|g|, |h|$ gives $E|g(X)h(Y)| = E|g(X)|E|h(Y)| < \infty$, and we can repeat the above argument using Fubini's Theorem. □

Note that the second part of the preceding proof is typical of multiple integral arguments: One uses Tonelli's theorem to verify integrability by computing the integral of the absolute value as an iterated integral (or interchanging order of integration), and then one applies Fubini's Theorem to compute the desired integral.

Theorem 6.4 can easily be extended to handle any finite number of random variables:

Theorem 6.5. *If $X_1, ..., X_n$ are independent and have $X_i \geq 0$ for all i, or $E|X_i| < \infty$ for all i, then*

$$E\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} E[X_i].$$

Proof. Corollary 6.2 shows that X_1 and $X_2 \cdots X_n$ are independent, so Theorem 6.4 (with $g = h$ the identity function) gives

$$E\left[\prod_{i=1}^{n} X_i\right] = E[X_1]E\left[\prod_{i=2}^{n} X_i\right],$$

and the result follows by induction. □

(To make Theorem 6.5 look more like Theorem 6.4, recall that if $X_1, ..., X_n$ are independent and $f_1, ..., f_n : \mathbb{R} \to \mathbb{R}$ are measurable, then $f_1(X_1), ..., f_n(X_n)$ are independent.)

Note that it is possible that $E[XY] = E[X]E[Y]$ without X and Y being independent.

For example, let $X \sim N(0, 1)$, $Y = X^2$. Then X and Y are clearly dependent, but a little calculus shows that $E[X]$ and $E[XY] = E[X^3]$ are both 0 and $E[Y] = E[X^2] = 1$, so $E[XY] = 0 = E[X]E[Y]$.

Definition. If X and Y are random variables with $E[X^2], E[Y^2\} < \infty$ and $E[XY] = E[X]E[Y]$, then we say that X and Y are *uncorrelated*.

Often, independence is invoked solely to argue that the expectation of the product is the product of the expectations. In such cases, one can weaken the assumption from independence to uncorrelatedness.

Of course, we can obtain a partial converse to Theorem 6.4 if we require the expectation to factor over a sufficiently large class of functions.

Proposition 6.2. *X and Y are independent if $E[f(X)g(Y)] = E[f(X)]E[g(Y)]$ for all bounded continuous functions f and g.*

Proof. Given any $x, y \in \mathbb{R}$, define

$$f_n(t) = \begin{cases} 1, & t \le x \\ 1 - n(t-x), & x < t \le x + \frac{1}{n}, \\ 0, & t > x + \frac{1}{n} \end{cases} \quad g_n(t) = \begin{cases} 1, & t \le y \\ 1 - n(t-y), & y < t \le y + \frac{1}{n}. \\ 0, & t > y + \frac{1}{n} \end{cases}$$

Then bounded convergence and the assumptions give

$$\begin{aligned} P(X \le x, Y \le y) &= E\left[\lim_{n\to\infty} f_n(X) g_n(Y)\right] = \lim_{n\to\infty} E\left[f_n(X)\right] E\left[g_n(Y)\right] \\ &= E\left[\lim_{n\to\infty} f_n(X)\right]\left[\lim_{n\to\infty} g_n(Y)\right] = P(X \le x) P(Y \le y). \end{aligned}$$

□

Before moving on, we mention that the ideas in this section can be used to analyze the sum of independent random variables.

Theorem 6.6. *Suppose that X and Y are independent with distributions μ, ν and distribution functions F, G. Then $X + Y$ has distribution function*

$$P(X + Y \le z) = \int F(z - y) dG(y).$$

If X has density f, then $X + Y$ has density $h(z) = \int f(z - y) dG(y)$.
If, additionally, Y has density g, then $h(z) = \int f(z - y) g(y) dy = f * g(z)$ - that is, the density of the sum is the convolution of the densities.

Proof. The change of variables formula, independence, and Tonelli's theorem give

$$\begin{aligned} P(X + Y \le z) &= \int_\Omega 1_{(-\infty,z]}(X + Y) dP = \int_{\mathbb{R}^2} 1_{(-\infty,z]}(x + y) d(\mu \times \nu)(x, y) \\ &= \int_\mathbb{R} \int_\mathbb{R} 1_{(-\infty,z]}(x + y) d\mu(x) d\nu(y) = \int_\mathbb{R} \left(\int_\mathbb{R} 1_{(-\infty,z-y]}(x) d\mu(x)\right) d\nu(y) \\ &= \int_\mathbb{R} F(z - y) d\nu(y) = \int F(z - y) dG(y). \end{aligned}$$

The final equality is just interpreting an integral against ν as a Riemann-Stieltjes integral with respect to G.
Now if X has density f, then the previous result with u-substitution and Tonelli yield

$$\begin{aligned} P(X + Y \le z) &= \int_\mathbb{R} F(z - y) d\nu(y) = \int_\mathbb{R} \int_{-\infty}^{z-y} f(x) dx d\nu(y) \\ &= \int_\mathbb{R} \int_{-\infty}^{z} f(x - y) dx d\nu(y) = \int_{-\infty}^{z} \int_\mathbb{R} f(x - y) d\nu(y) dx \\ &= \int_{-\infty}^{z} \left(\int f(x - y) dG(y)\right) dx, \end{aligned}$$

which means that the density of $X + Y$ is as claimed.
The third assertion follows from the change of variables formula for absolutely continuous random variables - which reads $dG(y) = g(y)dy$ in the present context. □

Though one can use these convolution results to derive useful facts about distributions of sums, tools such as characteristic and moment generating functions are generally much better suited for this task, so we will not pursue the issue right now.

Constructing Independent Random Variables.

To see that we have not done all of this work for nothing, we now show that independent random variables actually exist!

Given a finite collection of distribution functions $F_1, ..., F_n$, it is easy to construct independent random variables $X_1, ..., X_n$ with $P(X_i \leq x) = F_i(x)$.

Namely, let $\Omega = \mathbb{R}^n$, $\mathcal{F} = \mathcal{B}^n$, and $P = \mu_1 \times \cdots \times \mu_n$ where μ_i is the measure on $(\mathbb{R}, \mathcal{B})$ with distribution function F_i.

The product measure P is well-defined and satisfies

$$P\left((a_1, b_1] \times \cdots \times (a_n, b_n]\right) = (F_1(b_1) - F_1(a_1)) \cdots (F_n(b_n) - F_n(a_n)).$$

If we define X_i to be the projection map $X_i\left((\omega_1, ..., \omega_n)\right) = \omega_i$, then it is clear that the $X_i's$ are independent with the appropriate distributions.

In order to build an infinite sequence of independent random variables with given distribution functions, we need to perform the above construction on the infinite product space

$$\mathbb{R}^{\mathbb{N}} = \{(\omega_1, \omega_2, ...) : \omega_i \in \mathbb{R}\} = \{\text{functions } \omega : \mathbb{N} \to \mathbb{R}\}.$$

The product σ-algebra $\mathcal{B}^{\mathbb{N}}$ is generated by *cylinder sets* of the form

$$\{\omega \in \mathbb{R}^{\mathbb{N}} : \omega_i \in (a_i, b_i] \text{ for } i = 1, ..., n\},$$

and the random variables are the projections $X_i(\omega) = \omega_i$.

(In the definition of cylinders, we take $-\infty \leq a_i \leq b_i \leq \infty$ with the interpretation that $(a_i, \infty] = (a_i, \infty)$. $a_j = b_j$ for any j gives the empty set.)

Clearly, the desired measure should satisfy

$$P\left(\{\omega \in \mathbb{R}^{\mathbb{N}} : \omega_i \in (a_i, b_i] \text{ for } i = 1, ..., n\}\right) = \prod_{i=1}^{n} (F_i(b_i) - F_i(a_i))$$

on the cylinders.

To see that we can uniquely extend this to all of $\mathcal{B}^{\mathbb{N}}$, we appeal to

Theorem 6.7 (Kolmogorov)**.** *Suppose that we are given a sequence of probability measures μ_n on $(\mathbb{R}^n, \mathcal{B}^n)$ which are* consistent *in the sense that*

$$\mu_{n+1}\left((a_1, b_1] \times \cdots \times (a_n, b_n] \times \mathbb{R}\right) = \mu_n\left((a_1, b_1] \times \cdots \times (a_n, b_n]\right).$$

Then there is a unique probability measure P on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\mathbb{N}})$ with

$$P\left(\{\omega \in \mathbb{R}^{\mathbb{N}} : \omega_i \in (a_i, b_i], i = 1, ..., n\}\right) = \mu_n\left((a_1, b_1] \times \cdots \times (a_n, b_n]\right).$$

In particular, given distribution functions $F_1, F_2, ...$, if we define the μ_n's by the condition

$$\mu_n\left((a_1, b_1] \times \cdots \times (a_n, b_n]\right) = \prod_{i=1}^{n} (F_i(b_i) - F_i(a_i)),$$

then the projections $X_n(\omega) = \omega_n$ are independent with $P(X_n \leq x) = F_n(x)$.

Proof of Theorem 6.7. Let $\{\mu_n\}_{n=1}^{\infty}$ be a consistent sequence of probability measures, let $\mathcal{S}$ be the collection of cylinder sets, and define $Q : \mathcal{S} \to [0,1]$ by

$$Q\left(\{\omega \in \mathbb{R}^{\mathbb{N}} : \omega_i \in (a_i, b_i], 1 \leq i \leq n\}\right) = \mu_n\left((a_1, b_1] \times \cdots \times (a_n, b_n]\right).$$

Let $\mathcal{A} = \{$finite disjoint unions of sets in $\mathcal{S}\}$ be the algebra generated by $\mathcal{S}$ and define $P_0 : \mathcal{A} \to [0,1]$ by $P_0\left(\bigsqcup_{k=1}^{n} S_k\right) = \sum_{k=1}^{n} Q(S_k)$ for $S_1, ..., S_n$ disjoint sets in $\mathcal{S}$.

As $\mathcal{S}$ is a semialgebra which generates $\mathcal{B}^{\mathbb{N}}$, the discussion in Section 3 shows that it suffices to prove

Claim. If $B_n \in \mathcal{A}$ with $B_n \searrow \emptyset$, then $P_0(B_n) \searrow 0$.

Proof. To further simplify our task, let $\mathcal{F}_n$ be the sub-σ-algebra of $\mathcal{B}^{\mathbb{N}}$ consisting of all sets of the form $E = E^* \times \mathbb{R} \times \mathbb{R} \times \cdots$ with $E^* \in \mathcal{B}^n$. We use this asterisk notation throughout to denote the "$\mathcal{B}^n$ component" of sets in $\mathcal{F}^n$.

We begin by showing that we may assume without loss of generality that $B_n \in \mathcal{F}_n$ for all n.

To see this, note that $B_n \in \mathcal{A}$ implies that there is a $j(n) \in \mathbb{N}$ such that $B_n \in \mathcal{F}_k$ for all $k \geq j(n)$. Let $k(1) = j(1)$ and $k(n) = k(n-1) + j(n)$ for $n \geq 2$. Then $k(1) < k(2) < \cdots$ and $B_n \in \mathcal{F}_{k(n)}$ for all n. Define $\widetilde{B}_i = \mathbb{R}^{\mathbb{N}}$ for $i < k(1)$ and $\widetilde{B}_i = B_n$ for $k(n) \leq i < k(n+1)$. Then $\widetilde{B}_n \in \mathcal{F}_n$ for all n and the collections $\{B_n\}$ and $\left\{\widetilde{B}_n\right\}$ differ only in that the latter possibly includes $\mathbb{R}^{\mathbb{N}}$ and repeats sets. The assertion follows since $\widetilde{B}_n \searrow \emptyset$ if and only if $B_n \searrow \emptyset$ and $P_0\left(\widetilde{B}_n\right) \searrow 0$ if and only if $P_0\left(B_n\right) \searrow 0$.

Now suppose that $P_0(B_n) \geq \delta > 0$ for all n. We will derive a contradiction by approximating the B_n^* from within by compact sets and then using a diagonal argument to obtain $\bigcap_n B_n \neq \emptyset$.

Since B_n is nonempty and belongs to $\mathcal{A} \cap \mathcal{F}_n$, we can write

$$B_n = \bigcup_{k=1}^{K(n)} \{\omega : \omega_i \in (a_{i,k}, b_{i,k}], i = 1, ..., n\} \text{ where } -\infty \leq a_{i,k} < b_{i,k} \leq \infty.$$

By a continuity from below argument, we can find a set $E_n \subseteq B_n$ of the form

$$E_n = \bigcup_{k=1}^{K(n)} \left\{\omega : \omega_i \in [\widetilde{a}_{i,k}, \widetilde{b}_{i,k}], i = 1, ..., n\right\}, \quad -\infty < \widetilde{a}_{i,k} < \widetilde{b}_{i,k} < \infty,$$

with $\mu_n\left(B_n^* \setminus E_n^*\right) \leq \frac{\delta}{2^{n+1}}$.

Let $F_n = \bigcap_{m=1}^{n} E_m$. Since $B_n \subseteq B_m$ for any $m \leq n$, we have

$$B_n \setminus F_n = B_n \cap \left(\bigcup_{m=1}^{n} E_m^C\right) = \bigcup_{m=1}^{n} \left(B_n \cap E_m^C\right) \subseteq \bigcup_{m=1}^{n} \left(B_m \cap E_m^C\right),$$

hence

$$\mu_n(B_n^* \setminus F_n^*) \leq \sum_{m=1}^{n} \mu_m(B_m^* \setminus E_m^*) \leq \frac{\delta}{2}.$$

Since $\mu_n(B_n^*) = P_0(B_n) \geq \delta$, this means that $\mu_n(F_n^*) \geq \frac{\delta}{2}$, hence F_n^* is nonempty.

Moreover, E_n^* is a finite union of closed and bounded rectangles, so

$$F_n^* = E_n^* \cap (E_{n-1}^* \times \mathbb{R}) \cap \cdots \cap (E_1 \times \mathbb{R}^{n-1})$$

is compact.

For each $m \in \mathbb{N}$, choose some $\omega^m \in F_m$. As $F_m \subseteq F_1$, ω_1^m (the first coordinate of ω^m) is in F_1^*.

By compactness, we can find a subsequence $m(1,j) \geq j$ such that $\omega_1^{m(1,j)}$ converges to a limit $\theta_1 \in F_1^*$.

For $m \geq 2$, $F_m \subseteq F_2$, so $(\omega_1^m, \omega_2^m) \in F_2^*$. Because F_2^* is compact, we can find a subsequence of $\{m(1,j)\}$, which we denote by $m(2,j)$, such that $\omega_2^{m(2,j)}$ converges to a limit θ_2 with $(\theta_1, \theta_2) \in F_2^*$.

In general, we can find a subsequence $m(n,j)$ of $m(n-1,j)$ such that $\omega_n^{m(n,j)}$ converges to θ_n with $(\theta_1, ..., \theta_n) \in F_n^*$.

Finally, define the sequence $\omega(i) = \omega^{m(i,i)}$. Then $\omega(i)$ is a subsequence of each $\omega^{m(i,j)}$, so $\lim_{i\to\infty} \omega(i)_k = \theta_k$ for all k. Since $(\theta_1, ..., \theta_n) \in F_n^*$ for all n, $\theta = (\theta_1, \theta_2, ...) \in F_n$ for all n, hence

$$\theta \in \bigcap_{n=1}^{\infty} F_n \subseteq \bigcap_{n=1}^{\infty} B_n,$$

a contradiction!

□

Note that the proof of Theorem 6.7 used certain topological properties of $\mathbb{R}^n$.

As one might expect, the theorem does not hold for infinite products of arbitrary measurable spaces $(S, \mathcal{G})$.

However, one can show that it does hold for *nice spaces* where $(S, \mathcal{G})$ is said to be nice if there exists an injection $\varphi : S \to \mathbb{R}$ such that φ and φ^{-1} are measurable.

The collection of nice spaces is rich enough for our purposes. For example, if S is (homeomorphic to) a complete and separable metric space and $\mathcal{G}$ is the collection of Borel subsets of S, then $(S, \mathcal{G})$ is nice.