Chapter 4

Product spaces and independence

1. Product measures

<1> **Definition.** *Let $\mathcal{X}_1, \dots, \mathcal{X}_n$ be sets equipped with sigma-fields $\mathcal{A}_1, \dots, \mathcal{A}_n$. The set of all ordered n-tuples $(x_1, \dots, x_n)$, with $x_i \in \mathcal{X}_i$ for each i is denoted by $\mathcal{X}_1 \times \dots \times \mathcal{X}_n$ or $\mathsf{X}_{i \le n} \mathcal{X}_i$. It is called the **product** of the $\{\mathcal{X}_i\}$. A set of the form*

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

*with $A_i \in \mathcal{A}_i$ for each i, is called a **measurable rectangle**. The product sigma-field $\mathcal{A}_1 \otimes \dots \otimes \mathcal{A}_n$ on $\mathcal{X}_1 \times \dots \times \mathcal{X}_n$ 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 $A_1 \times A_2$.

To keep the notation simple, I will mostly consider only measures on a product of two spaces, $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{B})$. Sometimes I will abbreviate symbols like $\mathcal{M}^+(\mathcal{X} \times \mathcal{Y}, \mathcal{A} \otimes \mathcal{B})$ to $\mathcal{M}^+(\mathcal{X} \times \mathcal{Y})$, with the product sigma-field assumed, or to $\mathcal{M}^+(\mathcal{A} \otimes \mathcal{B})$, with the product space assumed. Similarly, $\mathcal{M}_{\text{bdd}}(\mathcal{A} \otimes \mathcal{B})$ will be an abbreviation for $\mathcal{M}_{\text{bdd}}(\mathcal{X} \times \mathcal{Y}, \mathcal{A} \otimes \mathcal{B})$, the vector space of all bounded, real-valued, product measurable functions on $\mathcal{X} \times \mathcal{Y}$.

Suppose μ is a finite measure on $\mathcal{A}$ and ν is a finite measure on $\mathcal{B}$. The next theorem, which is usually called Fubini's Theorem, asserts existence of a finite measure on $\mathcal{A} \otimes \mathcal{B}$ whose integrals can be calculated by iterated integrals with respect to μ and ν.

Remember the notation $\mu^x h(x, y)$ for what would be written $\int h(x, y)\, \mu(dx)$ in traditional notation, the integral of $h(\cdot, y)$ with respect to μ with y held fixed.

<2> **Theorem.** *For finite measures μ and ν, there is a uniquely determined finite measure Γ on $\mathcal{A} \otimes \mathcal{B}$ for which*

(i) $\Gamma(A \times B) = (\mu A)(\nu B)$ for each measurable rectangle.

Moreover, for each h in $\mathcal{M}_{\text{bdd}}(\mathcal{A} \otimes \mathcal{B})$, *(ii) the map $x \mapsto h(x, y)$ is $\mathcal{A}$-measurable for each fixed y and the map $y \mapsto h(x, y)$ is $\mathcal{B}$-measurable for each fixed x*

(iii) the map $x \mapsto \nu^y h(x, y)$ is $\mathcal{A}$-measurable and the map $y \mapsto \mu^x h(x, y)$ is $\mathcal{B}$-measurable

(iv) $\mu^x \left(\nu^y h(x, y)\right) = \nu^y \left(\mu^x h(x, y)\right)$

(v) the common value in (iv) is equal to Γh

REMARK. Properties (ii) and (iii) are necessary requirements for the iterated integrals in (iv) to make sense.

Proof. The method of proof is a case study in the use of the generating class argument for λ-spaces, as developed in Section 2.11. The main idea is to define the measure Γ by means of the iterated integral

Write $\mathcal{H}$ for the set of all functions h in $\mathcal{M}_{\text{bdd}}(\mathcal{A}\otimes\mathcal{B})$ for which (ii), (iii), and (iv) hold. It is very easy to check that $\mathcal{H}$ is a λ-space. For example, if $h_n \in \mathcal{H}$ and $h_n \uparrow h$ with h bounded then, by Monotone Convergence (for increasing sequences bounded from below by an integrable function), $\nu^y h(x, y) = \lim_{n\to\infty} \nu^y h_n(x, y)$, which establishes property (iii) for h. Similarly, four appeals to Monotone Convergence lead from the equality of iterated integrals for each h_n to the corresponding equality for h.

It is even easier to show that $\mathcal{H} \supseteq \mathcal{G}$, where $\mathcal{G}$ denotes the set of all indicator functions $g(x, y) := \{x \in A,\ y \in B\}$ of measurable rectangles. For example, $y \mapsto g(x, y)$ is either the zero function (if $x \notin A$) or the indicator of the set B (if $x \in A$), and $\nu^y g(x, y) = \{x \in A\}(\nu B)$.

The set $\mathcal{G}$ is stable under pairwise products and it generates $\mathcal{A} \otimes \mathcal{B}$. It follows by the theorems from Section 2.11 that $\mathcal{H} = \mathcal{M}_{\text{bdd}}(\mathcal{A} \otimes \mathcal{B})$.

Now consider a function f in $\mathcal{M}^+(\mathcal{A} \otimes \mathcal{B})$. The truncated function $f_n := \min(n, f)$ belongs to $\mathcal{M}_{\text{bdd}}$ for each n in $\mathbb{N}$, which shows that $\mu^x \left(\nu^y f_n(x, y)\right) = \nu^y \left(\mu^x f_n(x, y)\right)$. Four appeals to Monotone Convergence lead to equality of the analogous iterated integrals for f. Thus

$$\Gamma(f) := \mu^x \left(\nu^y f(x, y)\right) = \nu^y \left(\mu^x f(x, y)\right)$$

is a well defined map from $\mathcal{M}^+(\mathcal{A} \otimes \mathcal{B})$ to $[0, \infty]$. It is easily checked that Γ is an increasing, linear functional with the Monotone Convergence property, that is, that Γ corresponds to an integral with respect to a measure on $\mathcal{A} \otimes \mathcal{B}$.

An appeal to the π–λ theorem, with the set of measurable rectangles as the generating class, establishes the uniquenesss. □

The measure Γ is called the ***product of the measures μ and ν***, and is denoted by $\mu \otimes \nu$.

Theorem <2> has an immediate extension to products of sigma-finite measures. (Remember that sigma-finiteness of μ means that there is a partition of $\mathcal{X}$ into countably many $\mathcal{A}$-measurable sets, $\mathcal{X} = \cup_{i\in\mathbb{N}}\mathcal{X}_i$, with $\mu\mathcal{X}_i < \infty$ for each i.) The extension is usually attributed to Tonelli.

<3> **Tonelli Theorem.** *If μ is a sigma-finite measure on $(\mathcal{X}, \mathcal{A})$, and ν is a sigma-finite measure on $(\mathcal{Y}, \mathcal{B})$, then there is a unique sigma-finite measure $\mu \otimes \nu$ on $\mathcal{A} \otimes \mathcal{B}$ with the following properties. For each f in $\mathcal{M}^+(\mathcal{X} \times \mathcal{Y}, \mathcal{A} \otimes \mathcal{B})$,*

(i) $y \mapsto f(x, y)$ is $\mathcal{B}$-measurable for each fixed x, and $x \mapsto f(x, y)$ is $\mathcal{A}$-measurable for each fixed y;

(ii) $x \mapsto \lambda^y f(x, y)$ is $\mathcal{A}$-measurable, and $y \mapsto \mu^x f(x, y)$ is $\mathcal{B}$-measurable;

(iii) $(\mu \otimes \nu) f = \mu^x (\lambda^y f(x, y)) = \lambda^y (\mu^x f(x, y))$.

Proof. Write μ_i for the finite measure obtained by restricting μ to $\mathcal{X}_i$. Define ν_j analogously. Invoke Theorem <2> to construct each $\mu_i \otimes \nu_j$ then define $(\mu \otimes \nu) f$ as $\sum_{i,j \in \mathbb{N}} \mu_i \otimes \nu_j f$. And so on. □

See Problem [LEB.COUNTING] for an example emphasizing the need for sigma-finiteness.

<4> **Example.** Let μ be a sigma-finite measure on $\mathcal{A}$. For f in $\mathcal{M}^+(\mathcal{X}, \mathcal{A})$ and each constant $p \geq 1$, we can express $\mu(f^p)$ as an iterated integral,

$$\mu\left(f^p\right) = \mu^x\left(\mathfrak{m}^y\left(py^{p-1}\{y : f(x) > y > 0\}\right)\right),$$

where $\mathfrak{m}$ denotes Lebesgue measure on $\mathcal{B}(\mathbb{R})$. It is not hard—although a little messy, as you will see from Problem [SPECIAL.CASE]—to show that the function $g(x, y) := py^{p-1}\{f(x) > y > 0\}$, on $\mathcal{X} \times \mathbb{R}$, is product measurable. Tonelli lets us reverse the order of integration. Abbreviating $\mu^x\{y : f(x) > y > 0\}$ to $\mu\{f > y\}$ and writing the Lebesgue integral in traditional notational, we then conclude that

$$\mu\left(f^p\right) = p\int_0^\infty y^{p-1}\mu\{f > y\}\,dy.$$

In particular, if $\mu(f^p) < \infty$ then $\mu\{f > y\}$ must decrease to zero faster than y^{-p} as $y \to \infty$. □

The definition of product measures, and the Tonelli Theorem, can be extended to collections of more than two sigma-finite measures.

<5> **Example.** Apparently every mathematician is supposed to know the value of the constant $C := \int_{-\infty}^{\infty} \exp(-x^2)\,dx$. With the help of Tonelli, you too will discover that $C = \sqrt{\pi}$. Let $\mathfrak{m}$ denote Lebesgue measure on $\mathcal{B}(\mathbb{R})$ and $\mathfrak{m}_2 = \mathfrak{m} \otimes \mathfrak{m}$ denote Lebesgue measure on $\mathcal{B}(\mathbb{R}^2)$. Then

$$C^2 = \mathfrak{m}^x \mathfrak{m}^y \exp(-x^2 - y^2) = \mathfrak{m}_2^{x,y}\left(\mathfrak{m}^z\{x^2 + y^2 \leq z\}e^{-z}\right).$$

The $\mathfrak{m}_2$ measure of the ball $\{x^2 + y^2 \leq z\}$, for fixed positive z, equals πz. A change in the order of integration leaves $\mathfrak{m}^z\left(\pi\{0 \leq z\}ze^{-z}\right) = \pi$ as the value for C^2. □

The Tonelli Theorem is often invoked to establish integrability of a product measurable (extended-) real-valued function f, by showing that at least one of the iterated integrals $\mu^x(\nu^y|f(x, y)|)$ or $\nu^y(\mu^x|f(x, y)|)$ is finite. In that case, the Theorem also asserts equality for pairs of iterated integrals for the positive and negative parts of the function:

$$\mu^x \nu^y f^+(x, y) = \nu^y \mu^x f^+(x, y) < \infty,$$

with a similar assertion for f^-. As a consequence, the $\mathcal{A}$-measurable set

$$N_\mu := \{x : \nu^y f^+(x,y) = \infty \text{ or } \nu^y f^-(x,y) = \infty\}$$

has zero μ-measure, and the analogously defined $\mathcal{B}$-measurable set N_ν has zero ν measure. For $x \notin N_\mu$, the integral $\nu^y f(x,y) := \nu^y f^+(x,y) - \nu^y f^-(x,y)$ is well defined and finite. If we replace f by the product measurable function $\widetilde{f}(x,y) := f(x,y)\{x \notin N_\mu,\ y \notin N_\nu\}$, the negligible sets of bad behavior disappear, leaving an assertion similar to the Tonelli Theorem but for integrable functions taking both positive and negative values. Less formally, we can rely on the convention that a function can be left undefined on a negligible set without affecting its integrability properties.

<6> **Corollary (Fubini Theorem).** *For sigma-finite measures μ and ν, and a product measurable function f with $(\mu \otimes \nu)\,|f| < \infty$,*

(i) $y \mapsto f(x,y)$ is $\mathcal{B}$-measurable for each fixed x; and $x \mapsto f(x,y)$ is $\mathcal{A}$-measurable for each fixed y;

(ii) the integral $\nu^y f(x,y)$ is well defined and finite μ almost everywhere, and $x \mapsto \nu^y f(x,y)$ is μ-integrable; the integral $\mu^x f(x,y)$ is well defined and finite ν almost everywhere, and $y \mapsto \mu^x f(x,y)$ is ν-integrable;

(iii) $(\mu \otimes \nu) f = \mu^x\left(\nu^y f(x,y)\right) = \nu^y\left(\mu^x f(x,y)\right)$.

REMARKS. If we add similar almost sure qualifiers to assertion (i), then the Fubini Theorem also works for functions that are measurable with respect to $\mathcal{F}$, the $\mu \otimes \nu$ completion of the product sigma-field. The result is easy to deduce from the Theorem as stated, because each $\mathcal{F}$-measurable function f can be sandwiched between two product measurable functions, $f_0 \le f \le f_1$, with $f_0 = f_1$, a.e. $[\mu \otimes \nu]$. Many authors work with the slightly more general version, stated for the completion, but then the Tonelli Theorem also needs almost sure qualifiers.

Without integrability of the function f, the Fubini Theorem can fail, as shown by Problem [PLUS.MINUS]. Strictly speaking, the sigma-finiteness of the measures is not essential, but little is gained by eliminating it from the assumptions of the Theorem. As explained in Chapter 4, under the traditional definition of products for general measures, integrable functions must almost concentrate on a countable union of measurable rectangles each with finite product measure.

<7> **Example.** Recall from Section 2.2 the definition of the distribution function F_X and its corresponding quantile function for a random variable X:

$$F_X(x) = \mathbb{P}\{X \le x\} \qquad \text{for } x \in \mathbb{R},$$
$$q_X(u) = \inf\{t : F_X(t) \ge u\} \qquad \text{for } 0 < u < 1.$$

The quantile function is almost an inverse to the distribution function, in the sense that $F_X(x) \ge u$ if and only if $q_X(u) \le x$. As a random variable on $(0,1)$ equipped with its Borel sigma-field and Lebesgue measure $\widetilde{\mathbb{P}}$, the function $\widetilde{X} := q_X(u)$ has the same distribution as X. Similarly, if Y has distribution function F_Y and quantile function q_Y, the random variable $\widetilde{Y} := q_Y(u)$ has the same distribution as Y.

Notice that $\widetilde{X}$ and $\widetilde{Y}$ are both defined on the same $(0,1)$, even though the original variables need not be defined on the same space. If X and Y do happen to be defined on the same Ω their joint distribution need not be the same as the joint distribution for $\widetilde{X}$ and $\widetilde{Y}$. In fact, the new variables are closer to each other, in various senses. For example, several applications of Tonelli will show that $\mathbb{P}|X-Y|^p \ge \widetilde{\mathbb{P}}|\widetilde{X}-\widetilde{Y}|^p$ for each $p \ge 1$.

As a first step, calculate an inequality for tail probabilities.

$$\begin{aligned}\mathbb{P}\{X > x,\ Y > y\} &\le \min\left(\mathbb{P}\{X > x\},\ \mathbb{P}\{Y > y\}\right)\\ &= \min\left(1 - F_X(x), 1 - F_Y(y)\right)\\ &= 1 - F_X(x) \vee F_Y(y)\\ &= \int_0^1 \{u > F_X(x) \vee F_Y(y)\}\, du\\ &= \int_0^1 \{x < q_X(u),\ y < q_Y(u)\}\, du\\ &= \widetilde{\mathbb{P}}\{\widetilde{X} > x,\ \widetilde{Y} > y\}\end{aligned} \qquad \text{<8>}$$

We also have $\mathbb{P}\{X > x\} = \widetilde{\mathbb{P}}\{\widetilde{X} > x\}$ and $\mathbb{P}\{Y > y\} = \widetilde{\mathbb{P}}\{\widetilde{Y} > y\}$, from equality of the marginal distributions. By subtraction,

$$\begin{aligned}&\mathbb{P}\big(\{X > x\} + \{Y > y\} - 2\{X > x,\ Y > y\}\big)\\ &\qquad \ge \widetilde{\mathbb{P}}\big(\{\widetilde{X} > x\} + \{\widetilde{Y} > y\} - 2\{\widetilde{X} > x,\ \widetilde{Y} > y\}\big) \qquad \text{for all } x \text{ and } y.\end{aligned} \qquad \text{<9>}$$

The left-hand side can be rewritten as

$$\mathbb{P}^\omega\left(\{X(\omega) > x,\ y \ge Y(\omega)\} + \{X(\omega) \le x, y < Y(\omega)\}\right),$$

a nonnegative function just begging for an application of Tonelli. For each real constant s, put $y = x + s$ then integrate over x with respect to Lebesgue measure $\mathfrak{m}$ on $\mathcal{B}(\mathbb{R})$. Tonelli lets us interchange the order of integration, leaving

$$\begin{aligned}&\mathbb{P}^\omega\left(\mathfrak{m}^x\{X(\omega) > x \ge Y(\omega) - s\} + \mathfrak{m}^x\{X(\omega) \le x < Y(\omega) - s\}\right)\\ &\quad = \mathbb{P}^\omega\left((X(\omega) - Y(\omega) + s)^+ + (Y(\omega) - s - X(\omega))^+\right)\\ &\quad = \mathbb{P}|X - Y + s|.\end{aligned}$$

Argue similarly for the right-hand side of <9>, to deduce that

$$\mathbb{P}|X - Y + s| \ge \widetilde{\mathbb{P}}|\widetilde{X} - \widetilde{Y} + s| \qquad \text{for all real } s.$$

For each nonnegative t, invoke the inequality for $s = t$ then $s = -t$, then add.

$$\mathbb{P}\left(|X - Y + t| + |X - Y - t|\right) \ge \widetilde{\mathbb{P}}\left(|\widetilde{X} - \widetilde{Y} + t| + |\widetilde{X} - \widetilde{Y} - t|\right) \qquad \text{for all } t \ge 0.$$

An appeal to the identity $|z + t| + |z - t| = 2t + 2\left(|z| - t\right)^+$, for $z \in \mathbb{R}$ and $t \ge 0$, followed by a cancellation of common terms, then leaves us with a useful relationship, which neatly captures the idea that $\widetilde{X}$ and $\widetilde{Y}$ are more tightly coupled than X and Y.

$$\mathbb{P}\left(|X - Y| - t\right)^+ \ge \widetilde{\mathbb{P}}\left(|\widetilde{X} - \widetilde{Y}| - t\right)^+ \qquad \text{for all } t \ge 0. \qquad \text{<10>}$$

Various interesting inequalities follow from <10>. Putting t equal to zero we get $\mathbb{P}|X - Y| \ge \widetilde{\mathbb{P}}|\widetilde{X} - \widetilde{Y}|$. For $p > 1$, note the identity

$$D^p = p(p-1)\int_0^D (D - t)t^{p-2}\, dt = p(p-1)\mathfrak{m}_0^t\left(t^{p-2}\left(D - t\right)^+\right) \qquad \text{for } D \ge 0,$$

where $\mathfrak{m}_0$ denotes Lebesgue measure on $\mathcal{B}\left(\mathbb{R}^+\right)$. Temporarily write Δ for $|X-Y|$ and $\widetilde{\Delta}$ for $|\widetilde{X}-\widetilde{Y}|$. Two more appeals to Tonelli then give

$$\begin{aligned}\mathbb{P}|X-Y|^p &= p(p-1)\mathfrak{m}_0^t\left(t^{p-2}\mathbb{P}^\omega\left(\Delta(\omega)-t\right)^+\right)\\ &\geq p(p-1)\mathfrak{m}_0^t\left(t^{p-2}\widetilde{\mathbb{P}}^u(\widetilde{\Delta}(u)-t)^+\right)=\widetilde{\mathbb{P}}|\widetilde{X}-\widetilde{Y}|^p.\end{aligned}$$

See Problem [QUANTILE.ORLICZ] for the analogous inequality, $\mathbb{P}\Psi(|X-Y|)\geq\widetilde{\mathbb{P}}\Psi(|\widetilde{X}-\widetilde{Y}|)$, for every convex, increasing function Ψ on $\mathbb{R}^+$. □