16. PRODUCT MEASURE SPACES AND DISINTEGRATION

The goal of this section is to discuss aspects of product measure spaces. We start by reviewing Fubini-Tonelli’s theory underlying finite products, then proceed by discussing disintegration techniques and conditional measures. It is here we will see Standard Borel spaces play a useful role.

16.1 Finite product spaces and Fubini-Tonelli’s Theorem.

A classical question in the theory of multiple integrals is what condition ensures that the order of integration does not matter. This problem was first tackled by L. Euler for (Riemann) integrals of continuous functions. In 1904 H. Lebesgue proved this for (Lebesgue) integrals of bounded measurable functions. Soon thereafter G. Fubini proved it for integrable functions and L. Tonelli for non-negative measurable functions.

While we are less concerned with integrals than foundational questions of measure theory, we start by a short review of Fubini-Tonelli’s theory. Here is the first key concept:

**Definition 16.1 (Product measurable space)** Let $(X_1, F_1), \ldots, (X_n, F_n)$ be measurable spaces. The product measurable space is then the pair $(\times_{i=1}^n X_i, \otimes_{i=1}^n F_i)$ where

$$\otimes_{i=1}^n F_i := \sigma \left( \bigotimes_{i=1}^n F_i : \forall i = 1, \ldots, n \colon A_i \in F_i \right)$$

is the product $\sigma$-algebra. An alternative notation $F_1 \otimes \cdots \otimes F_n$ is used as well.

Our use of the notation with $\otimes$ is to emphasize that the resulting object involves more than just taking Cartesian products. As is easy to check, the operation of forming a product $\sigma$-algebra is associative, i.e.,

$$(F_1 \otimes F_2) \otimes F_3 = F_1 \otimes (F_2 \otimes F_3) = F_1 \otimes F_2 \otimes F_3$$

and so no parenthesizing is needed.

Fubini-Tonelli’s Theorem is based on the following idea: deduce the irrelevance of order of integration by equating the iterated integral with the integral over the full product space. As a starter, we need to ensure that the “one-dimensional” integrals of a multivariate measurable function can even be taken:

**Lemma 16.2 (Fubini-Tonelli Theorem, measurability part)** Let $\{(X_i, F_i)\}_{i=1}^n$ be measurable spaces. Then for all $A \in \otimes_{i=1}^n F_i$, all $i = 1, \ldots, n$ and all $x_i \in X_i$,

$$\left\{(\tilde{x}_1, \ldots, \tilde{x}_{i-1}, \tilde{x}_{i+1}, \ldots, \tilde{x}_n) \in \prod_{j \neq i} X_j : (\tilde{x}_1, \ldots, \tilde{x}_{i-1}, x_i, \tilde{x}_{i+1}, \ldots, \tilde{x}_n) \in A \right\} \in \otimes_{j \neq i} F_j.$$  

Given any measure space $(Z, \mathcal{H})$, for any $\otimes_{i=1}^n F_i / \mathcal{H}$-measurable function $f: \times_{i=1}^n X_i \to Z$, any $i = 1, \ldots, n$ and any $x_i \in X_i$, the function

$$(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \mapsto f(x_1, \ldots, x_n)$$

is $\otimes_{j \neq i} F_j / \mathcal{H}$-measurable.
Proof. In light of associativity (16.2), it suffices to prove the result only for product of two measurable spaces, \((X, \mathcal{F})\) and \((Y, \mathcal{G})\). Given \(x \in X\) and a set \(A \subseteq X \times Y\) we will write
\[
A^x := \{y \in Y: (x, y) \in A\}
\]
for the set in (16.3).

Consider the class of sets \(C := \{A \in \mathcal{F} \otimes \mathcal{G}: A^x \in \mathcal{G}\}\). Since \(\emptyset^x = \emptyset\) and \((E \times F)^x = F\) when \(x \in E\) and \((E \times F)^x = \emptyset\) when \(x \notin E\), we have
\[
\mathcal{F} \times \mathcal{G} \subseteq C.
\]
(16.6)

Since \((A^c)^x = (A^x)^c\) and \((\bigcup_{i \geq 1} A_i)^x = \bigcup_{i \geq 1} A_i^x\) (check these!), \(C\) is a \(\sigma\)-algebra. It follows that \(\mathcal{F} \otimes \mathcal{G} \subseteq C\), thus proving (16.3). To prove (16.4), let \(f: X \times Y \to Z\), abbreviate \(f_x(y) := f(x, y)\) and note that
\[
\forall B \subseteq Z:\quad f^{-1}_x(B) = f^{-1}(B)^x.
\]
(16.7)

The claim (16.4) thus follows from (16.3). \(\square\)

Note that (16.4) in turn implies (16.3) by taking \(f = 1_A\) while (16.4) is for \(\mathbb{R}\)-valued functions readily deduced from (16.3) with the help of the Monotone Class Lemma.

With the measurable structure of the product space settled, we will now be concerned with the question what natural measures one can put on the product space. The simplest example is the content of:

**Theorem 16.3 (Product measure)** Let \(\{(X_i, \mathcal{F}_i, \mu_i)\}_{i=1}^n\) be measure spaces with \(\mu_i\) \(\sigma\)-finite for each \(i = 1, \ldots, n\). Then there exists a unique measure \(\mu_1 \otimes \cdots \otimes \mu_n\) on \((\times_{i=1}^n X_i, \otimes_{i=1}^n \mathcal{F}_i)\) such that for all \((A_1, \ldots, A_n) \in \times_{i=1}^n \mathcal{F}_i,
\[
\mu_1 \otimes \cdots \otimes \mu_n \left( A_1 \times \cdots \times A_n \right) = \prod_{i=1}^n \mu_i(A_i).
\]
(16.8)

Here we use the convention that \(0 \cdot \infty = 0\).

Proof. Consider the class of sets
\[
\mathcal{S} := \left\{ A_1 \times \cdots \times A_n: (\forall i = 1, \ldots, n: A_i \in \mathcal{F}_i) \right\}.
\]
(16.9)

Then \(\mathcal{S}\) is a semialgebra on \(\times_{i=1}^n X_i\). Define
\[
\kappa(A_1 \times \cdots \times A_n) := \prod_{i=1}^n \mu_i(A_i)
\]
(16.10)

with the convention that the right-hand side vanishes whenever \(\mu_i(A_i) = 0\) for at least one \(i\). We claim that \(\kappa\) is countably additive (and thus also finitely additive) on \(\mathcal{S}\). For this suppose that \(\{A_{1,j} \times \cdots \times A_{n,j}\}_{j \geq 1} \subseteq \mathcal{S}\) are disjoint (possibly empty) with
\[
\bigcup_{j \geq 1} (A_{1,j} \times \cdots \times A_{n,j}) = B_1 \times \cdots \times B_n
\]
(16.11)
for some \((B_1,\ldots,B_n)\in\mathcal{F}_1\times\cdots\times\mathcal{F}_n\). Then for all \((x_1,\ldots,x_n)\in\times_{i=1}^n X_i,

\[
\prod_{i=1}^n 1_{B_i}(x_i) = \sum_{j\geq 1} \prod_{i=1}^n 1_{A_{i,j}}(x_i) \tag{16.12}
\]

Integrating over \(x_i\) with respect to \(\mu_i\) for \(i\) increasing from 1 to \(n\) and using the Monotone Convergence Theorem to pass the integral around the infinite sum then gives

\[
\prod_{i=1}^n \mu_i(B_i) = \sum_{j\geq 1} \prod_{i=1}^n \mu_i(A_{i,j}), \tag{16.13}
\]

where we also use the convention that integrating any function, even one taking infinite values, over a null set yields zero integral.

It follows that \(\kappa\) is indeed countably additive on \(\mathcal{S}\). By Corollary 11.10, \(\kappa\) extends to a measure on \(\sigma(\mathcal{S}) = \bigotimes_{i=1}^n \mathcal{F}_i\). Under the assumption that \(\mu\) and \(\nu\) are both \(\sigma\)-finite, \(\kappa\) is \(\sigma\)-finite on \(\mathcal{S}\) and the extension is thus unique. \(\square\)

In light of this theorem, we put forward:

**Definition 16.4**  Let \(\mu_1,\ldots,\mu_n\) be \(\sigma\)-finite measures. We then call \(\mu_1 \otimes \cdots \otimes \mu_n\) their product measure, or a product of \(\mu_1,\ldots,\mu_n\).

Thanks to the uniqueness clause, the product is associative in the sense that

\[
(\mu_1 \otimes \mu_2) \otimes \mu_3 = \mu_1 \otimes (\mu_2 \otimes \mu_3) = \mu_1 \otimes \mu_2 \otimes \mu_3 \tag{16.14}
\]

As the proof shows, the assumption of \(\sigma\)-finiteness is needed for the uniqueness part of the claim. As shown in the next lemma, uniqueness fails in general without it.

**Lemma 16.5**  Let \(I := [0,1]\) and let \(\lambda,\text{ resp.,} \#\) be the Lebesgue, resp., counting measure on \(\mathcal{B}(I)\). Show that, for each \(t \in [0,\infty]\) there is a measure \(\kappa_t\) on \(\mathcal{B}(I) \otimes \mathcal{B}(I)\) such that

\[
\forall A,B \in \mathcal{B}(I): \kappa_t(A \times B) = \lambda(A)\#(B) \tag{16.15}
\]

and, denoting the diagonal in \(I \times I\) by \(D := \{(x,x): x \in I\},

\[
\forall C \in \mathcal{B}(I): \kappa_t((C \times C) \cap D) = t\lambda(C). \tag{16.16}
\]

In particular, there are infinitely many extensions of \(A \times B \mapsto \lambda(A)\#(B)\) to \(\mathcal{B}(I) \otimes \mathcal{B}(I)\).

We leave the proof of the lemma to homework. We are now ready to state:

**Theorem 16.6**  (Fubini-Tonelli)  Let \((X,\mathcal{F},\mu)\) and \((Y,\mathcal{G},\nu)\) be \(\sigma\)-finite measure spaces and let \(f: X \times Y \to \mathbb{R}\) be \(\mathcal{F} \otimes \mathcal{G}/\mathcal{B}(\mathbb{R})\)-measurable. Assume that either \(f \geq 0\) (Tonelli) or \(f \in L^1(\mu \otimes \nu)\) (Fubini). Then for \(y \mapsto f(x,y)\) is integrable with respect to \(\nu\) for all \(x \in X\) and \(x \mapsto \int f(x,y) \, \nu(\mathrm{d}y)\) is \(\mathcal{F}\)-measurable \(\tag{16.17}\)

while \(x \mapsto f(x,y)\) is integrable with respect to \(\mu\) for all \(y \in Y\) and

\[
y \mapsto \int f(x,y) \, \mu(\mathrm{d}x)\] is \(\mathcal{G}\)-measurable.  \(\tag{16.18}\)
These functions are nonnegative (Tonelli) or in \(L^1(\mu)\), resp., \(L^1(\nu)\) (Fubini). Finally, in either alternative we have
\[
\int f \, d\mu \otimes \nu = \int \left( \int f(x, y) \, \nu(dy) \right) \mu(dx) = \int \left( \int f(x, y) \, \mu(dx) \right) \nu(dy)
\]
(16.19)
and so the order of iterated integrals does not matter.

We skip a detailed proof but note that, while (16.19) for Fubini’s alternative is reduced to Tonelli’s by writing each function as the difference of its positive and negative parts, for Tonelli’s alternative (16.19) is proved using the Monotone Class Lemma and the fact that, by Lemma 16.2 and the construction of the product measure, the claim holds for indicators of measurable sets.

Every important theorem in mathematics comes with counterexamples for what happens when not all conditions are met. So, for instance, the statement of Fubini’s part fails in general when both the positive and negative parts of \(f\) are not absolutely integrable.

The standard example is:

**Example 16.7** \(X = Y := \mathbb{N}\) with \(\mathcal{F} = \mathcal{G} := 2^\mathbb{N}\), \(\mu = \nu :=\) counting measure on \(\mathbb{N}\) and
\[
f(n, m) := 1_{\{n=m\}} - 1_{\{n=m-1\}}.
\]
(16.20)
Then (writing integrals as sums)
\[
\sum_{m \geq 0} \sum_{n \geq 0} f(n, m) = 1 \quad \text{yet} \quad \sum_{n \geq 0} \sum_{m \geq 0} f(n, m) = 0.
\]
(16.21)
The catch here is that the function \(f\) fails to be absolutely integrable with respect to \(\mu \otimes \nu\).

Using the same structure except replacing \(\mathbb{N}\) by \(\mathbb{Z}\) shows that the iterated integrals can exist and be equal while the double integral does not exist.

An example also exists of a function that is separately measurable and integrable in each variable for which the iterated integrals depend on the order of integration.

**Example 16.8** Let \(X = Y\) be an uncountable totally-ordered set such that \(\{z \in X : z < x\}\) is countable for all \(x\) (such a set exists by the construction of ordinals in set theory), \(\mathcal{F} = \mathcal{G}\) is the \(\sigma\)-algebra consisting of countable and co-countable subsets of \(X\) and \(\mu = \nu\) are measures that vanish on countable sets and are equal to one on uncountable ones. Then
\[
\int \left( \int 1_{\{x<y\}} \, \mu(dx) \right) \nu(dy) = 0 \quad \text{yet} \quad \int \left( \int 1_{\{x<y\}} \, \nu(dy) \right) \mu(dx) = 1.
\]
(16.22)
The caveat here is that \(f := 1_{\{x<y\}}\) is not \(\mathcal{F} \otimes \mathcal{G}\)-measurable.

Equality of iterated integrals fails even for non-negative measurable functions when at least one of the measures is not \(\sigma\)-finite. The standard example is:

**Example 16.9** Let \(X = Y := [0, 1]\), \(\mathcal{F} := B([0, 1])\), \(\mathcal{G} := 2^Y\), \(\mu :=\) Lebesgue measure, \(\nu :=\) the counting measure and \(f(x, y) := 1_{\{x=y\}}\). Taking the integral with respect to the Lebesgue measure first yields zero while that with respect to the counting measure yields one.
16.2 Disintegration and conditional measures.

We will now move to a question that runs, in a sense, opposite to Fubini-Tonelli’s theory. Suppose we are given two measurable spaces \((X, \mathcal{F})\) and \((Y, \mathcal{G})\) and a measure \(\kappa\) on the product space \((X \times Y, \mathcal{F} \otimes \mathcal{G})\). We then ask: Can integrals with respect to \(\kappa\) can be written as iterated integrals over \(Y\) and \(X\)?

Since we are not assuming that \(\kappa\) is a product measure, the iterated integrals cannot use “independent” measures. In particular, order of iterated integrals will matter. We thus ask our question more precisely: Is there a family \(\{\nu_x : x \in X\}\) of measures on \((Y, \mathcal{G})\) and a measure \(\mu\) on \((X, \mathcal{F})\) such that, for each \(f \in L^1(\kappa)\),

\[
\int_{X \times Y} f \, d\kappa = \int_X \left( \int_Y f(x, y) \nu_x(dy) \right) \mu(dx). \tag{16.23}
\]

If this is possible, we refer to this formula as disintegration of the integral.

Let us first check the necessary conditions for (16.23) to be true. By choosing \(\mathcal{F}\)-measurable \(A \subseteq X\), \(B \subseteq Y\), \(A \times B \subseteq X \times Y\), the statement ensures that the \(A \times B\) has the form

\[
\kappa(A \times B) = \int_A v_x(B) \mu(dx), \tag{16.24}
\]

where \(\mu\) is a measure on \((X, \mathcal{F})\) and \(\{v_x : x \in X\}\) is a family of measures on \((Y, \mathcal{G})\) such that \(x \mapsto v_x(B)\) is \(\mathcal{F}\)-measurable for each \(B \in \mathcal{G}\).

The \(\mathcal{F}\)-measurability of \(x \mapsto v_x(B)\) is needed for the integral (16.24) to exist. This condition is actually sufficient to ensure existence of the integral in (16.23) as well:

**Lemma 16.11** Suppose \(\{v_x : x \in X\}\) is a family of measures on \((Y, \mathcal{G})\) such that \(x \mapsto v_x(B)\) is \(\mathcal{F}\)-measurable for each \(B \in \mathcal{G}\). Then

\[
x \mapsto \int f(x, y) v_x(dy) \quad \text{is } \mathcal{F}\text{-measurable} \tag{16.25}
\]

for each \(\mathcal{F} \otimes \mathcal{G}\)-measurable \(f : X \times Y \to [0, \infty)\).

**Proof.** The assumption ensures that the statement holds for \(f\) of the form \(f := 1_{A \times B}\), where \(A \in \mathcal{F}\) and \(B \in \mathcal{G}\). The class of sets \(C \in \mathcal{F} \otimes \mathcal{G}\) such that the statement holds for \(f := 1_C\) forms a \(\sigma\)-algebra, and since the product sets are already included, the statement holds for all such sets. The additivity of the integral along with the Monotone Class Lemma then extend the statement to all measurable \(f \geq 0\).

A problem with disintegration is that, when a solution to (16.23) exists, it is definitely not unique. Indeed, multiplying \(\nu_x\) by a strictly-positive measurable function of \(x\) and \(\mu\) by the reciprocal value of that function produces another solution. We may thus want to limit possible disintegrations further; e.g., by prescribing the measure \(\mu\) — one then talks about disintegration with respect to \(\mu\) — and/or by forcing \(\nu_x\) to be suitably normalized. Without much loss of generality, we can assume the natural normalization

\[
\forall x \in X : \quad \nu_x(Y) = 1. \tag{16.26}
\]
Then (16.24) forces $\mu$ to be the first marginal of $\kappa$,

$$\forall A \in \mathcal{F}: \quad \mu(A) := \kappa(A \times Y)$$  (16.27)

Regardless of the choice, the formula (16.24) forces that $\kappa(\cdot \times B) \ll \mu$ and so, assuming $\sigma$-finiteness, for each $B \in \mathcal{G}$ there exists $\mathcal{F}$-measurable $f_B: X \to [0, \infty]$ such that

$$\forall A \in \mathcal{F}: \quad \kappa(A \times B) = \int_A f_B \, d\mu,$$  (16.28)

and for the choice (16.27) we then even get $f_B = (1_B)_\mathcal{F}$. In any case, the Monotone Convergence Theorem shows

$$\forall \{B_i\}_{i \geq 1} \subseteq \mathcal{F} \text{ disjoint}: \quad f_{\bigcup_{i \geq 1} B_i} = \sum_{i \geq 1} f_{B_i} \mu\text{-a.e.}$$  (16.29)

so $B \mapsto f_B$ does indeed act like a measure on $(X, \mathcal{F})$. Unfortunately, $\sigma$-additivity may not hold or may be impossible to establish at any single $x \in X$ because $\mathcal{G}$ contains too many null sets. One needs to make additional assumptions on $(X, \mathcal{F})$ and $(Y, \mathcal{G})$ that prevent such scenarios in general.

Before we move to precise statements, let us generalize the problem slightly. Noting that $\mathcal{G}$ can be identified with the sub-$\sigma$-algebra $\{X \times A: A \in \mathcal{G}\}$ of $\mathcal{F} \otimes \mathcal{G}$, we put forward the following concept:

**Definition 16.12 (Conditional measures)** Let $\mu$ be a measure on $(X, \mathcal{F})$ and let $\mathcal{G} \subseteq \mathcal{F}$ be a $\sigma$-algebra. A family $\{v_x: x \in X\}$ such that

1. $\forall x \in X: \quad v_x$ is a probability measure on $(X, \mathcal{F})$,  
2. $\forall A \in \mathcal{F}: \quad x \mapsto v_x(A)$ is $\mathcal{G}$-measurable and

$$\forall B \in \mathcal{G}: \quad \mu(A \cap B) = \int_B v_x(A) \mu(dx)$$  (16.30)

are the conditional measures (or conditional probabilities) for $\mu$ given $\mathcal{G}$.

To see that Definition 16.12 subsumes the notion from Definition 16.10, suppose $\kappa$ is a measure on $(X \times Y, \mathcal{F} \otimes \mathcal{G})$ and let $\{v_x: x \in X\}$ be a conditional measure of $\mu$ given the $\sigma$-algebra $\mathcal{G}' := \{X \times B: B \in \mathcal{G}\}$. For the choices $A' := A \times Y$ and $B' := X \times B$, which yield $A' \cap B' = A \times B$, (16.30) becomes

$$\kappa(A \times B) = \int_{X \times B} v_{(x,y)}(A \times Y) \kappa(dx \, dy).$$  (16.31)

The $\mathcal{G}$-measurability of $x \mapsto v_x(A)$ ensures that $v_x(A) := v_{(x,y)}(A \times Y)$ does not depend on $y$. Denoting the first marginal from (16.27) by $\mu$ we get (16.24).

The notion of conditional probability is closely related by the notion of conditional expectation introduced in Theorem 10.2. Indeed, we have:

**Lemma 16.13 (Relation to conditional expectation)** Let $\mu$ be a finite measure on $(X, \mathcal{F})$ and let $\mathcal{G} \subseteq \mathcal{F}$ be a $\sigma$-algebra. Suppose $\{v_x: x \in X\}$ is a family of conditional measures for $\mu$ given $\mathcal{G}$. Then

$$\forall f \in L^1(\mu): \quad f_{\mathcal{G}}(x) = \int f \, dv_x \quad \text{for } \mu\text{-a.e. } x \in X.$$  (16.32)
In fact, \( x \mapsto \int f \, dv_x \) is a version of \( f_G \).

**Proof.** Let \( \mathcal{M} \) be the class of non-negative measurable functions such that \( x \mapsto \int f \, dv_x \) is \( G \) measurable and

\[
\int f \, d\mu = \int \left( \int f \, dv_x \right) \mu(dx).
\]  

(16.33)

By the defining properties of the conditional measures we get \( 1_A \in \mathcal{M} \) for all \( A \in \mathcal{F} \). The additivity of the integral and the Monotone Class Lemma then show that \( \mathcal{M} \) includes all non-negative \( \mathcal{F} \)-measurable functions. Decomposing \( f = f^+ - f^- \), the defining properties of \( \mathcal{M} \) extend to all \( f \in L^1(\mu) \). Since \( f_G \) is unique \( \mu \)-a.e., we get (16.32).

The conditional measures thus realize conditional expectations as integrals. While conditional expectations given a \( \sigma \)-algebra always exist, this is not true about conditional measures. This was apparently first pointed out by J. Dieudonne in his paper “Sur le théorème de Lebesgue-Nikodym (III)” published in Annales de l’université de Grenoble, tome 23 (1947-1948), p. 25–53. Here is a clean example:

**Lemma 16.14** Abbreviate \( I := [0, 1] \) and let \( \lambda := \text{Lebesgue measure on } (X, \mathcal{F}) \) where

\[
X := I \times I \quad \text{and} \quad \mathcal{F} := \mathcal{L}(I \times I).
\]  

(16.34)

Set

\[
\mathcal{G} := \{ A \times I : A \in \mathcal{B}(I) \}.
\]  

(16.35)

Then, assuming the existence of Lebesgue non-measurable sets, there is no family of conditional measures for \( \lambda \) given \( \mathcal{G} \).

In order to prove this, we first show:

**Lemma 16.15** (Uniqueness of conditional measures) Let \( \mu \) be a finite measure on \( (X, \mathcal{F}) \) and let \( \mathcal{G} \subseteq \mathcal{F} \) be a \( \sigma \)-algebra. Assume that \( \mathcal{F} \) is countably generated, meaning that there exists \( \{ A_n : n \geq 1 \} \subseteq \mathcal{F} \) such that \( \mathcal{F} = \sigma(\{ A_n : n \geq 1 \}) \). For any families \( \{ v_x : x \in X \} \) and \( \{ \tilde{v}_x : x \in X \} \) of conditional measures for \( \mu \) given \( \mathcal{G} \) there is \( X^* \in \mathcal{G} \) such that

\[
\mu(X \setminus X^*) = 0
\]  

(16.36)

and

\[
\forall x \in X^* \quad \forall A \in \mathcal{F} : \quad v_x(A) = \tilde{v}_x(A).
\]  

(16.37)

In other words, the two families agree \( \mu \)-a.e.

**Proof.** Let \( \mathcal{A} \) be the set of finite unions of sets in \( \{ A_n : n \geq 1 \} \). Then \( \mathcal{A} \) is still countable and \( \sigma(\mathcal{A}) = \mathcal{F} \). For each \( A \in \mathcal{A} \) define

\[
X_A := \{ x \in X : v_x(A) = \tilde{v}_x(A) \}.
\]  

(16.38)

Then \( X_A \in \mathcal{G} \) and, by (16.32) (which implies \( v_x(A) = (1_A)_G(x) \) for \( \mu \)-a.e. \( x \)) and the uniqueness of the conditional expectation, \( \mu(X \setminus X_A) = 0 \). Set

\[
X^* := \bigcap_{A \in \mathcal{A}} X_A.
\]  

(16.39)
The intersection is over countable index and so we still have $X^* \in \mathcal{G}$ and $\mu(X \setminus X^*) = 0$, thus proving (16.36). As $\mu$ is finite,

$$\mathcal{L} := \{ A \in \mathcal{F} : (\forall x \in X^* : \nu_x(A) = \nu_x(A)) \}$$

is readily checked to be a $\lambda$-system. As $A \subseteq \mathcal{L} \subseteq \mathcal{F}$ and $\mathcal{F} = \sigma(A)$, Dynkin’s $\pi/\lambda$-Theorem gives $\mathcal{L} = \mathcal{F}$, and so (16.37) holds as well.

With uniqueness under control, we can give:

**Proof of Lemma 16.14.** Write $\lambda_1$ for the one-dimensional Lebesgue measure on $(I, \mathcal{B}(I))$ and $\lambda_2$ for the two-dimensional Lebesgue measure on $(I \times I, \mathcal{L}(I \times I))$ or its subspace $(I \times I, \mathcal{B}(I \times I))$. Let $\mathcal{G}$ be as in (16.35) and assume, for the sake of contradiction, that $\{\nu_{(x,y)} : (x,y) \in I \times I\}$ is a family of conditional measures for $\lambda_2$ given $\mathcal{G}$. Since

$$\mathcal{G} \subseteq \mathcal{B}(I \times I) \subseteq \mathcal{L}(I \times I),$$

$\{\nu_{(x,y)} : (x,y) \in I \times I\}$ are also the conditional measures for $\lambda_2$ on $(I \times I, \mathcal{B}(I \times I))$ given $\mathcal{G}$. But $\mathcal{B}(I \times I) = \mathcal{B}(I) \otimes \mathcal{B}(I)$ is a product space and $\lambda_2 = \lambda_1 \otimes \lambda_1$ is a product measure on $(I \times I, \mathcal{B}(I \times I))$, so

$$\nu_{(x,y)}(A) := \lambda_1(A^x),$$

where $A^x := \{ y \in I : (x,y) \in A \}$, is another family of conditional measures for $\lambda_2$ given $\mathcal{G}$. As $\mathcal{B}(I \times I)$ is countably generated, Lemma 16.15 shows there is a set $J^* \in \mathcal{G}$ of full $\lambda_2$-measure such that

$$\forall (x,y) \in J^* \forall A \in \mathcal{B}(I \times I) : \nu_{(x,y)}(A) = \lambda(A^x).$$

(16.43) Pick any $(x^*, y^*) \in J^*$ and noting that $\{x^*\} \times A \in \mathcal{B}(I \times I)$ for all $A \in \mathcal{B}(I)$, we have

$$\forall A \in \mathcal{B}(I) : \nu_{(x^*,y^*)}(\{x^*\} \times A) = \lambda_1(A).$$

(16.44) But the completeness of $\lambda_2$ ensures

$$\forall E \subseteq I : \{x^*\} \times E \in \mathcal{L}(I \times I)$$

(16.45) and so $E \mapsto \nu_{(x^*,y^*)}(\{x^*\} \times E)$ is well defined and $\sigma$-additive on $2^I$. In light of (16.44), this means that $\lambda_1$ on $(I, \mathcal{B}(I))$ can be extended to all subsets of $I$, in contradiction with our assumptions. \qed

Recall that Lebesgue non-measurable sets exist assuming the Axiom of Choice thanks to Vitali’s Theorem (see Theorem 12.6). Still, nothing particularly specific to the Lebesgue measure plays a role in the above example; the key point is having a measure that is the completion of a product of non-atomic measures one of which cannot be extended to all sets. The main culprit here is the completion, of course. Indeed, (as we will show momentarily) the product measure admits conditional measures as soon as the underlying space is nice, while the abundance of null sets makes this considerably harder for completed measures. The moral is that, although sometimes convenient, automatic completion of measure spaces is not always the right thing to do.

We are now ready to state the main result of this section:
Theorem 16.16 (Existence of conditional measures) Let \((X, \mathcal{B}(X))\) be a standard Borel space, \(\mu\) a finite measure on \((X, \mathcal{B}(X))\) and \(\mathcal{G} \subseteq \mathcal{B}(X)\) a \(\sigma\)-algebra. Then there exists a family \(\{\nu_x : x \in X\}\) of conditional measures for \(\mu\) given \(\mathcal{G}\) and this family is unique \(\mu\)-a.e. Moreover, if \(\mathcal{G}\) is countably generated then there is a version of \(\{\nu_x : x \in X\}\) and \(X' \in \mathcal{G}\) such that

\[
\forall x \in X' \forall A \in \mathcal{G} : \quad \nu_x(A) = 1_A(x) \tag{16.46}
\]

and \(\mu(X \setminus X') = 0\). (We call such a version proper.)

Proof. Let \((X, \mathcal{B}(X))\) be a standard Borel space. By Kuratowski’s Theorem there is a bimeasurable injection \(f : X \to \mathbb{R}\) whose range is a subset of \(\mathbb{N} \subseteq \mathbb{R}\) if \(X\) is finite or countably infinite or \([0, 1]\) if \(X\) is uncountable. As finite or countable sets are Borel, we can treat all three cases under the same footing. We will use that \(f \geq 0\), though.

For each \(q \in \mathbb{Q}\), consider a version of \(1_{\{f \leq q\}}\) which, as \(\mu\) is finite, we may assume to be everywhere finite. We will assume that this version vanishes identically for \(q < 0\). Assemble these to define a function \(F : \mathbb{Q} \times X \to [0, \infty)\) by

\[
F(q, x) := (1_{\{f \leq q\}})_\mathcal{G}(x) \tag{16.47}
\]

We will now show that, on a set of full \(\mu\)-measure, \(q \mapsto F(q, x)\) is non-decreasing and right continuous. For this set, for each \(p, q \in \mathbb{Q}\) with \(p < q\),

\[
X_{p,q} := \{x \in X : F(p, x) \leq F(q, x)\}. \tag{16.48}
\]

Then \(X_{p,q} \in \mathcal{G}\) while (since \(p < q\))

\[
\int_{\{F(p, \cdot) > F(q, \cdot)\}} [F(q, \cdot) - F(p, \cdot)] \, d\mu = \int_{\{F(p, \cdot) > F(q, \cdot)\}} (1_{\{f \leq q\}} - 1_{\{f \leq p\}}) \, d\mu \geq 0 \tag{16.49}
\]

shows \(\mu(X \setminus X_{p,q}) = 0\). Setting

\[
X^* := \bigcap_{\substack{p, \mu \in \mathbb{Q} \\ p < q}} X_{p,q} \tag{16.50}
\]

we get \(\mu(X \setminus X^*) = 0\). Note that \(q \mapsto F(q, x)\) is non-decreasing for all \(x \in X^*\).

Next we need to ensure right continuity. For this we use Dominated Convergence and the fact that \(F(q + 1/n, \cdot) \leq F(q + 1, \cdot)\) \(\mu\)-a.e. for all \(n \geq 1\) with \(F(q + 1, \cdot) \in L^1(\mu)\) (the integral equals \(\mu(f \leq q + 1)\) which is finite) to get

\[
\mu(f \leq q) = \lim_{n \to \infty} \mu(f \leq q + 1/n) = \lim_{n \to \infty} \int F(q + 1/n, \cdot) \, d\mu = \int F(q, \cdot) \, d\mu = \mu(f \leq q), \tag{16.51}
\]

where the first limit exists by Downward Monotone Convergence for (finite) measures while the limit under integral exists \(\mu\)-a.e. by the \(\mu\)-a.e. monotonicity we proved above. It follows that

\[
\lim_{n \to \infty} F(q + 1/n, \cdot) = F(q, \cdot), \quad \mu\text{-a.e.} \tag{16.52}
\]

Defining, for each \(q \in \mathbb{Q}\),

\[
X^*_q := \{x \in X^* : \limsup_{n \to \infty} F(q + 1/n, x) = F(q, x)\} \tag{16.53}
\]
we conclude that $X_q^* \in \mathcal{G}$ and $\mu(X \setminus X_q^*) = 0$ for all $q \in \mathcal{Q}$. Setting

$$X^{**} := \bigcap_{q \in \mathcal{Q}} X_q^*$$

(16.54)

we then have $X^{**} \in \mathcal{G}$ with $\mu(X \setminus X^{**}) = 0$. Note that $q \mapsto F(q, x)$ non-decreasing and right continuous on $\mathcal{Q}$ for each $x \in X^{**}$.

For all $x \in X^{**}$ define

$$\widetilde{F}(t, x) := \inf_{q \in \mathcal{Q} \cap (t, \infty]} F(q, x)$$

(16.55)

while for $x \notin X^{**}$ set $\widetilde{F}(t, x) := 1_{[0, \infty)}(t)$. The definition ensures that $t \mapsto \widetilde{F}(t, x)$ is non-decreasing and right-continuous for all $x \in X$. As $\widetilde{F}(t, \cdot) = 0$ for $t < 0$, for each $x \in X$ there exists a Radon measure $\nu_x$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that

$$\forall x \in X \forall t \in \mathbb{R}: \quad \nu_x((-\infty, t]) = \widetilde{F}(t, x).$$

(16.56)

Since $\widetilde{F}(q, x) = F(q, x)$ for all $q \in \mathcal{Q}$ and $x \in X^{**}$, we get that $x \mapsto \nu_x((-\infty, t])$ is a version of $(1_{\{f \leq q\}})_G$ for every $q \in \mathcal{Q}$.

We now invoke a standard (by now) extension argument: The properties of conditional expectation ensure that

$$\mathcal{L} := \left\{ A \in \mathcal{B}(X) : x \mapsto \nu_x(A) \text{ is a version of } (1_A)_G \right\}.$$  

(16.57)

is a $\lambda$-system. Since $\mathcal{L}$ contains $\mathcal{P} := \{ \{ f \leq q \} : q \in \mathcal{Q} \}$ and $\mathcal{P}$ is a $\pi$-system with $\sigma(\mathcal{P}) = \mathcal{B}(X)$, Dynkin’s $\pi/\lambda$-Theorem gives $\mathcal{B}(X) \subseteq \mathcal{L}$. This means that

$$\forall A \in \mathcal{B}(X) \forall B \in \mathcal{G}: \quad \mu(A \cap B) = \int_B \nu_x(A) \, d\mu$$

(16.58)

Since being a version of $(1_A)_G$ implies that $x \mapsto \nu_x(A)$ is $\mathcal{G}$-measurable, $\{\nu_x : x \in X\}$ is a family of conditional probabilities given $\mathcal{G}$.

Since $\mathcal{B}(X)$ is countably generated (due to the fact that the topology on $X$ is countably generated), Lemma 16.15 ensures that the conditional probabilities are unique $\mu$-a.e. It remains to prove that there is a proper version when $\mathcal{G}$ is countably generated. Let $\{A_n : n \geq 1\} \subseteq \mathcal{G}$ be such that $\mathcal{G} = \sigma(\{A_n : n \geq 1\})$ and let $\mathcal{A}$ be the algebra generated by $\{A_n : n \geq 1\}$. Let $\{\nu_x : x \in X\}$ be as constructed above. As

$$\forall A \in \mathcal{G}: \quad (1_A)_G = 1_A \text{ $\mu$-a.e.,}$$

(16.59)

the fact that $\mathcal{A}$ is countable shows that

$$X' := \bigcap_{A \in \mathcal{A}} \{ x \in X : \nu_x(A) = 1_A(x) \}$$

(16.60)

obeys $X' \in \mathcal{G}$ and $\mu(X \setminus X') = 0$. Since

$$\mathcal{L}' := \left\{ A \in \mathcal{B}(X) : (\forall x \in X' : \nu_x(A) = 1_A(x)) \right\}$$

(16.61)

is a $\lambda$-system, the fact that $\mathcal{A}$ is a $\pi$-system with $\mathcal{A} \subseteq \mathcal{L}'$ implies $\mathcal{G} = \sigma(\mathcal{A}) \subseteq \mathcal{L}'$ by Dynkin’s $\pi/\lambda$-Theorem. This shows that (16.46) holds indeed. 

\[ \square \]
We note that the exceptional set \( X \setminus X' \) in (16.46) cannot be removed in general; the paper by D. Blackwell and C. Ryll-Nadzewski (“Non-Existence of Everywhere Proper Conditional Distributions,” Ann. Math. Statist. 34 (1963), no. 1, 223–225) gives necessary and sufficient conditions for this to hold.

In probability, the push-forward of a measure by a measurable function \( f \) is referred to as the distribution of \( f \). The above argument is usually presented by saying that any measurable \( f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \) has a well-defined conditional distribution with respect to any \( \sigma \)-algebra \( \mathcal{G} \subseteq \mathcal{F} \). Thanks to Kuratowski’s Theorem, the same is true about measurable functions taking values in an arbitrary standard Borel space.

D. Blackwell’s influential article “On a class of probability spaces” (Proc. Third Berkeley Symp. on Math. Statist. and Prob., Vol. 2 (Univ. of Calif. Press, 1956), 1-6) discusses ways to go beyond the setting of standard Borel spaces. For instance, he shows the existence of conditional distributions when the underlying measure space is the so-called Lusin space, which is a concept based on the notion of analytic sets.