Tonelli

Tonelli's Theorem: If $(\mathcal{X}, \mathbb{E}, \mu)$ and $(\mathcal{Y}, \mathbb{K}, \nu)$ are two σ-finite measure spaces and $f \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}, \mathbb{E} \otimes \mathbb{K})$ then

$$\int f \mathrm{d}\mu \otimes \nu = \int \left(\int f(x,y) \mathrm{d}\nu(y) \right) \mathrm{d}\mu(x) = \int \left(\int f(x,y) \mathrm{d}\mu(x) \right) \mathrm{d}\nu(y).$$

A product of Lebesgue measures is a Lebesgue measure:

$$m_k = \underbrace{m \otimes \ldots \otimes m}_{k \text{ times}}$$

and we can put parentheses as we like (associative law of $\otimes$). In particular $m_p \otimes m_q = m_{p+q}$, and we can use this to show that all hyperplanes in $\mathbb{R}^k$ are m_k-nullsets.

Fubini

Fubini's Theorem: If $(\mathcal{X}, \mathbb{E}, \mu)$ and $(\mathcal{Y}, \mathbb{K}, \nu)$ are two σ-finite measure spaces and if $f \in \mathcal{M}(\mathcal{X} \times \mathcal{Y}, \mathbb{E} \otimes \mathbb{K})$ is integrable w.r.t. $\mu \otimes \nu$ then

$$\int f \mathrm{d}\mu \otimes \nu = \int_A \left(\int f(x,y) \mathrm{d}\nu(y) \right) \mathrm{d}\mu(x) = \int_B \left(\int f(x,y) \mathrm{d}\mu(x) \right) \mathrm{d}\nu(y).$$

Here $A \in \mathbb{E}$ with $\mu(A^c) = 0$ and $B \in \mathbb{K}$ with $\nu(B^c) = 0$ are given by

$$A = \{x \in \mathcal{X} \mid \int |f(x,y)| \mathrm{d}\nu(y) < \infty\}$$

and

$$B = \{y \in \mathcal{Y} \mid \int |f(x,y)| \mathrm{d}\mu(x) < \infty\}.$$

Fubini in practice

Note that to use Fubini we need to first verify that f is integrable w.r.t. $\mu \otimes \nu$ and then compute the integral using either of the successive integration orders.

Verification of integrability follows from Tonelli's theorem.

- First verify that f is integrable by Tonelli

$$\int |f| \mathrm{d}\mu \otimes \nu = \int \left(\int |f(x,y)| \mathrm{d}\nu(y) \right) \mathrm{d}\mu(x) \leq \ldots < \infty$$

- then compute the value of the integral by Fubini

$$\int f \mathrm{d}\mu \otimes \nu = \int \left(\int f(x,y) \mathrm{d}\nu(y) \right) \mathrm{d}\mu(x).$$

This is Fubinelli's theorem.

Sums and integrals

If $f_1, f_2, \ldots : \mathcal{X} \to \mathbb{R}$ are function we can define the function $f : \mathcal{X} \times \mathbb{N} \to \mathbb{R}$ by

$$f(x,n) = f_n(x).$$

Then f is $\mathbb{E} \otimes \mathbb{P}(\mathbb{N})$-$\mathbb{B}$-measurable if and only if all the functions f_n are $\mathbb{E}$-$\mathbb{B}$-measurable (Exercise 4.8).

Recall that integration w.r.t. the counting measure τ on $\mathbb{N}$ is the same as infinite sums. Thus

$$\sum_{n=1}^{\infty} f_n(x) = \int f(x,n) \mathrm{d}\tau(n)$$

whenever the sum and the integral make sense.

Interchanging sums and integrals

Let $(\mathcal{X}, \mathbb{E})$ be a mesurable space.

Theorem: If $f_1, f_2, \ldots : \mathcal{X} \to \mathbb{R}$ are measurable functions and μ is a measure on $(\mathcal{X}, \mathbb{E})$ then if

$$\sum_{n=1}^{\infty} \int |f_n| \mathrm{d}\mu < \infty$$

each f_n is integrable, the sum $\sum_{n=1}^{\infty} f_n(x)$ is finite for all $x \in A$ where

$$A = \{x \in \mathcal{X} \mid \sum_{n=1}^{\infty} |f_n(x)| < \infty\}$$

and $\mu(A^c) = 0$. Moreover, $\sum_{n=1}^{\infty} f_n(x)$ is μ-a.e. (on A) equal to an integrable function and

$$\sum_{n=1}^{\infty} \int f_n(x) \mathrm{d}\mu(x) = \int_A \sum_{n=1}^{\infty} f_n(x) \mathrm{d}\mu(x).$$

Interchanging sums and integrals

Proof: Fubinelli with the product measure $\mu \otimes \tau$. Catch, requires a priori μ to be σ-finite!

Workaround: The purpose of σ-finiteness is to assure uniqueness of the product measure as well as existence in terms of the technical measurability lemma 8.6. If we can check that by other means, Tonelli and Fubini applies! For all $A \in \mathbb{E} \otimes \mathbb{P}(\mathbb{N})$ we have the disjoint decomposition

$$A = \bigcup_{n=1}^{\infty} \mathcal{X} \times \{n\} \cap A = \bigcup_{n=1}^{\infty} A^n \times \{n\}, \quad A^n = \{x \in \mathcal{X} \mid (x, n) \in A\}$$

thus $\mu \otimes \tau(A) = \sum_{n=1}^{\infty} \mu(A^n)$, which defines this particular product measure and shows that it is uniquely specified by it's values on product sets.

Alternative proof: Use dominated convergence as in the proof of Theorem 7.10.

Interchanging the order of integration

Using Fubini on

$$f(x, y) = \sin x \, e^{-xy} \qquad \text{for } x \in (0, K), y \in (0, \infty)$$

gives

$$\int_0^K \frac{\sin x}{x} \, dx = \int_0^{\infty} \frac{1}{1 + y^2} \, dy + \ldots$$

We find from this that

$$\int_0^K \frac{\sin x}{x} \, dx \to \frac{\pi}{2} \qquad \text{for } K \to \infty$$

though

$$\int_0^{\infty} \left| \frac{\sin x}{x} \right| \, dx = \infty$$

Image measures

Let $(\mathcal{X}, \mathbb{E}, \mu)$ be a measurable space.

Let $(\mathcal{Y}, \mathbb{K})$ be a measurable, and let $t : \mathcal{X} \to \mathcal{Y}$ be $\mathbb{E}$-$\mathbb{K}$-measurable.

Definition: The image measure $t(\mu)$ is the measure on $(\mathcal{Y}, \mathbb{K})$, which is given by

$$t(\mu)(B) = \mu\left(t^{-1}(B)\right) \qquad \text{for all } B \in \mathbb{K}$$

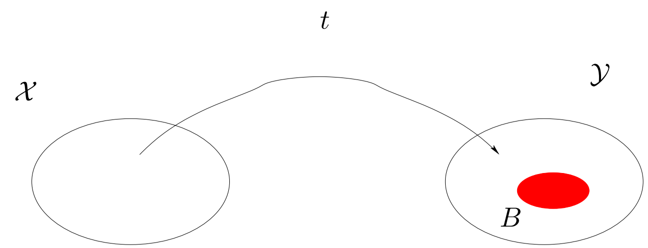


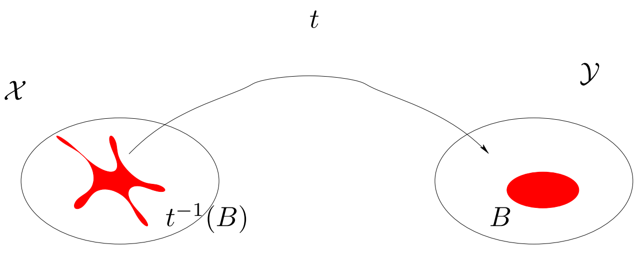
Image measures

Let $(\mathcal{X}, \mathbb{E}, \mu)$ be a measurable space.

Let $(\mathcal{Y}, \mathbb{K})$ be a measurable, and let $t : \mathcal{X} \to \mathcal{Y}$ be $\mathbb{E}$-$\mathbb{K}$-measurable.

Definition: The image measure $t(\mu)$ is the measure on $(\mathcal{Y}, \mathbb{K})$, which is given by

$$t(\mu)(B) = \mu\left(t^{-1}(B)\right) \qquad \text{for all } B \in \mathbb{K}$$



The image measure is a measure

Lemma: The image measure $t(\mu)$ is a measure on $(\mathcal{Y}, \mathbb{K})$.

Proof: Obviously $t(\mu)(A) = \mu(t^{-1}(A)) \in [0, \infty]$ and we see that

$$t(\mu)(\emptyset) = \mu\left(t^{-1}(\emptyset)\right) = \mu(\emptyset) = 0$$

If $B_1, B_2, \ldots$ are disjoint $\mathbb{K}$-sets then

$$t^{-1}(B_i) \cap t^{-1}(B_j) = t^{-1}(B_i \cap B_j) = t^{-1}(\emptyset) = \emptyset$$

and therefore

$$t(\mu)\left(\bigcup_{n=1}^{\infty} B_n\right) = \mu\left(t^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right)\right) = \mu\left(\bigcup_{n=1}^{\infty} t^{-1}(B_n)\right)$$
$$= \sum_{n=1}^{\infty} \mu\left(t^{-1}(B_n)\right) = \sum_{n=1}^{\infty} t(\mu)(B_n)$$

Familiar example

Let $m|_{(0,\infty)}$ denote the Lebesgue measure restricted to the positive halfline; formally

$$m|_{(0,\infty)}(A) = m(A \cap (0, \infty))$$

for all $A \in \mathbb{B}$.

Define $t : \mathbb{R} \to \mathbb{R}$ by

$$t(x) = \begin{cases} \log(x) & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

Then

$$\begin{aligned} t(m|_{(0,\infty)})((-\infty, x]) &= m|_{(0,\infty)}(t^{-1}((-\infty, x])) \\ &= m(t^{-1}((-\infty, x]) \cap (0, \infty)) = m((0, e^x]) = e^x. \end{aligned}$$

Total mass of the image measure

Lemma: It holds that $t(\mu)(\mathcal{Y}) = \mu(\mathcal{X})$.

Proof: Observe that

$$t(\mu)(\mathcal{Y}) = \mu\left(t^{-1}(\mathcal{Y})\right) = \mu(\mathcal{X})$$

Corollary: If μ is a probability measure then $t(\mu)$ is a probability measure.

Probability theory is essentially a theory about image measures – given one measure and a map t, what is $t(\mu)$?

The abstract simplicity cheats the eye. Finding $t(\mu)$, characterizing $t(\mu)$ or just computing certain characteristics of $t(\mu)$ can by arbitrarily complicated for concrete t and μ.

Example

Let μ be the uniform distribution on

$$\mathcal{X} = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}.$$

Every point $(i,j) \in \mathcal{X}$ has probability $\frac{1}{36}$.

Interpretation: μ has something to do with throwing two dice.

The map: $t : \mathcal{X} \to \{0,1,2,3,4,5\}$ is given by

$$t(i,j) = |i-j|$$

Interpretation: t represents the difference between the two dice.

Example

Tabulating the t-values:

	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

Image measure:

$$t(\mu)(\{0\}) =$$

Example

Tabulating the t-values:

	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

Image measure:

$$t(\mu)(\{0\}) =$$

Example

Tabulating the t-values:

	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

Image measure:

$$t(\mu)(\{0\}) = \mu\big(\{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}\big) = 6 \cdot \frac{1}{36}$$

Example

Tabulating the t-values:

	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

Likewise,

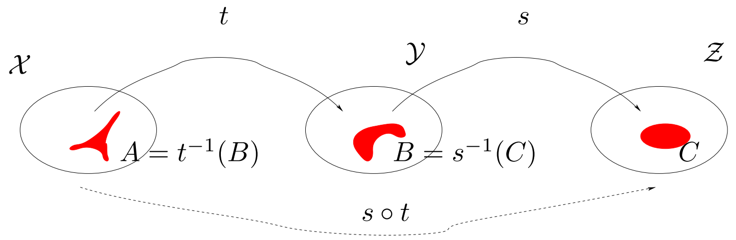
$$t(\mu)(\{1\}) = 10 \cdot \frac{1}{36}$$

Succesive transformations

Lemma: Let $t : \mathcal{X} \to \mathcal{Y}$ and $s : \mathcal{Y} \to \mathcal{Z}$ be measurable. Let μ be a measure on $\mathcal{X}$. Then

$$s\big(t(\mu)\big) = (s \circ t)(\mu)$$

Proof:



$$s\,(t(\mu))\,(C) = t(\mu)\,\left(s^{-1}(C)\right) = \mu\,\left(t^{-1}\,\left(s^{-1}(C)\right)\right)$$
$$= \mu\,\left((s \circ t)^{-1}(C)\right) = (s \circ t)(\mu)(C)$$

Example

Tabulating the t-values:

	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

And so on and so forth.

k	0	1	2	3	4	5
$t(\mu)(\{k\})$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

Example: Marginalization

If λ is a measure on $(\mathcal{X} \times \mathcal{Y}, \mathbb{E} \otimes \mathbb{K})$ the image measures

$$\hat{X}(\lambda) \quad \text{og} \quad \hat{Y}(\lambda)$$

are the marginal measures of λ's.

Let $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2$, and consider $\hat{X} : \mathcal{X} \times \mathcal{X}_3 \to \mathcal{X}$, $\hat{X}_1 : \mathcal{X}_1 \times (\mathcal{X}_2 \times \mathcal{X}_3) \to \mathcal{X}_1$ and $\hat{X}_{01} : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{X}_1$. Then

$$\hat{X}_1 = \hat{X}_{01} \circ \hat{X}$$

and by the theorem on successive transformation

$$\hat{X}_1(\lambda) = \hat{X}_{01}(\hat{X}(\lambda)).$$

Moral: For marginalization on multiple product spaces it does not matter if we do several succesive marginalizations or one combined marginalization.

Example: Marginalization

Example: Consider $m_2 = m \otimes m$ on $(\mathbb{R}^2, \mathbb{B}_2)$. Then

$$\hat{X}(m_2)(A) = m_2(A \times \mathbb{R}) = m(A)m(\mathbb{R}) = \begin{cases} \infty & \text{if } m(A) > 0 \\ 0 & \text{if } m(A) = 0 \end{cases}$$

Example: If λ is a probability measure on $(\mathcal{X} \times \mathcal{Y}, \mathbb{E} \otimes \mathbb{K})$ then the marginals are probability measures and λ is a product measure if and only if it is a product of it's marginals;

$$\lambda = \hat{X}(\lambda) \otimes \hat{Y}(\lambda).$$

Integral transformation for $\mathcal{M}^+$

Theorem: Let $t : \mathcal{X} \to \mathcal{Y}$ be measurable. Let μ be a measure on $\mathcal{X}$. Then

$$\int g \, d\, t(\mu) = \int g \circ t \, d\mu$$

for all $g \in \mathcal{M}^+(\mathcal{Y}, \mathbb{K})$.

Proof: Strategy: Show the formula for

1) indicator functions

2) simple functions

3) $\mathcal{M}^+$-functions

Point 3) is shown from 2) via monotone convergence.

Integral transformation for $\mathcal{M}^+$

Theorem: Let $t : \mathcal{X} \to \mathcal{Y}$ be measurable. Let μ be a measure on $\mathcal{X}$. Then

$$\int g \, d\, t(\mu) = \int g \circ t \, d\mu$$

for all $g \in \mathcal{M}^+(\mathcal{Y}, \mathbb{K})$.

Indicator functions: Let $B \in \mathbb{K}$. Then

$$\int 1_B \, d\, t(\mu) = t(\mu)(B) = \mu\left(t^{-1}(B)\right) = \int 1_{t^{-1}(B)} \, d\mu = \int 1_B \circ t \, d\mu$$

So the formula is correct in this case.

Integral transformation for $\mathcal{M}^+$

Theorem: Let $t : \mathcal{X} \to \mathcal{Y}$ be measurable. Let μ be a measure on $\mathcal{X}$. Then

$$\int g \, d\, t(\mu) = \int g \circ t \, d\mu$$

for all $g \in \mathcal{M}^+(\mathcal{Y}, \mathbb{K})$.

Simple functions: Consider

$$g = \sum_{i=1}^{n} c_i \, 1_{B_i}$$

Then

$$\int g \, d\, t(\mu) = \sum_{i=1}^{n} c_i \int 1_{B_i} \, d\, t(\mu) = \sum_{i=1}^{n} c_i \int 1_{B_i} \circ t \, d\mu = \int \sum_{i=1}^{n} c_i \, 1_{B_i} \circ t \, d\mu$$

So the formula is correct in this case.

Integral transformation for $\mathcal{M}^+$

Theorem: Let $t : \mathcal{X} \to \mathcal{Y}$ be measurable. Let μ be a measure on $\mathcal{X}$. Then

$$\int g \, d\, t(\mu) = \int g \circ t \, d\mu$$

for all $g \in \mathcal{M}^+(\mathcal{Y}, \mathbb{K})$.

All functions: Let $g \in \mathcal{M}^+$ and choose $\mathcal{S}^+$-functions s_n such that $s_n \nearrow g$. Then $s_n \circ t \nearrow g \circ t$ and by the theorem on monotone convergence

$$\int g \, d\, t(\mu) = \lim_{n\to\infty} \int s_n \, d\, t(\mu) = \lim_{n\to\infty} \int s_n \circ t \, d\mu = \int g \circ t \, d\mu$$

In conclusion, the formula holds for all $g \in \mathcal{M}^+$.

Example

Consider the measure μ with

$$\mu((-\infty, x]) = e^x$$

for all $x \in \mathbb{R}$. We know that $\mu = t(m|_{(0,\infty)})$

The function $x \mapsto e^{-\alpha x}$ is $\mathcal{M}^+$ and we find that

$$\begin{aligned} \int e^{-\alpha x} \mathrm{d}\mu(x) &= \int_0^\infty e^{-\alpha t(x)} \mathrm{d}m(x) \\ &= \int_0^\infty e^{-\alpha \log(x)} \mathrm{d}m(x) \\ &= \int_0^\infty x^{-\alpha} \mathrm{d}m(x) = \infty \end{aligned}$$

for all $\alpha \in \mathbb{R}$

Example

The function $x \mapsto 1_{(0,\infty)}(x) e^{-\alpha x}$ is $\mathcal{M}^+$ and we find that

$$\begin{aligned} \int_0^\infty e^{-\alpha x} \mathrm{d}\mu(x) &= \int_0^\infty 1_{(0,\infty)}(t(x)) e^{-\alpha t(x)} \mathrm{d}m(x) \\ &= \int_1^\infty e^{-\alpha \log(x)} \mathrm{d}m(x) \\ &= \int_1^\infty x^{-\alpha} \mathrm{d}m(x) = \begin{cases} \frac{1}{\alpha - 1} & \text{for } \alpha > 1 \\ \infty & \text{for } \alpha \le 1 \end{cases} \end{aligned}$$

Integral transformation for $\mathcal{L}$

Theorem: Let $t : \mathcal{X} \to \mathcal{Y}$ be measurable. Let μ be a measure on $\mathcal{X}$. A function $g \in \mathcal{M}(\mathcal{Y}, \mathbb{K})$ is integrable w.r.t. $t(\mu)$ if and only if

$$\int |g \circ t| \, d\, \mu < \infty,$$

in which case

$$\int g \, d\, t(\mu) = \int g \circ t \, d\mu.$$

Proof: We know that g is integrable if and only if $\int |g| \, dt(\mu) < \infty$. But this integral is computed from the previous theorem. In case we have integrability, the actual integral is computed as:

$$\begin{aligned} \int g \, d\, t(\mu) &= \int g^+ \, d\, t(\mu) - \int g^- \, d\, t(\mu) = \int g^+ \circ t \, d\mu - \int g^- \circ t \, d\mu \\ &= \int (g \circ t)^+ \, d\mu - \int (g \circ t)^- \, d\mu = \int g \circ t \, d\mu \end{aligned}$$

Translations in $\mathbb{R}^k$

Definition: The translation with $w \in \mathbb{R}^k$ is the map $\tau_w : \mathbb{R}^k \to \mathbb{R}^k$ defined by

$$\tau_w(x) = x + w \qquad \text{for all } x \in \mathbb{R}^k \,,$$

Definition: A measure μ on $(\mathbb{R}^k, \mathbb{B}_k)$ is translation invariant if

$$\tau_w(\mu) = \mu \qquad \text{for all } \ w \in \mathbb{R}^k$$

The Lebesgue measure

Theorem: A measure μ on $(\mathbb{R}^k, \mathbb{B}_k)$, which is finite on bounded sets is translation invariant if and only if

$$\mu = c m_k$$

for some constant $c \geq 0$.

Translation and integrals

Example If $f \in \mathcal{M}^+(\mathbb{R}^k)$ or if $f \in \mathcal{L}(\mathbb{R}^k, m_k)$ then in general

$$\int f(x+w)\mathrm{d}\mu(x) = \int f \circ \tau_w(x)\mathrm{d}\mu(x) = \int f(x)\mathrm{d}\tau_w(\mu)(x)$$

and for the Lebesgue measure

$$\int f(x+w)\,dx = \int f(x)\,dx \qquad \text{for all } w \in \mathbb{R}^k.$$

Linear transformations on $\mathbb{R}^k$

A linear map $s : \mathbb{R}^k \mapsto \mathbb{R}^k$ is given in terms of a $k \times k$ matrix A,

$$s(x) = Ax.$$

The map is an isomorphism if A is invertible, that is, if $\det A \neq 0$, in which case the inverse map is given by A^{-1}.

Theorem: If s is a linear transformation given by an invertible matrix A then

$$s(m_k) = |\det A^{-1}| m_k.$$

Remark: Note that for an orthonormal matrix Q ($QQ^T = Q^TQ = I$) we have $\det Q = \pm 1$, which shows that m_k is invariant under orthonormal transformations.