Lecture 17. Lévy's continuity theorem

Paul Delatte
delatte@usc.edu
University of Southern California

Last updated: 31 December, 2022

To prove Lévy's continuity theorem, we use Prokhorov's theorem and a lemma for the application of Prokhorov's theorem, which were both proved in Lecture 16.

Theorem 1 (Lévy's Continuity Theorem). *Let $\mu, \mu_1, \mu_2, \ldots$ be Borel probability measures on $\mathbb{R}^d$. Denote the characteristic function of μ_i by $\hat{\mu}_i(\xi) = \int e^{i\langle \xi, x\rangle}\, d\mu(x)$ for all $\xi \in \mathbb{R}^d$.*

1. If $\mu_n \rightsquigarrow \mu$, then $\lim_{n\to\infty} \hat{\mu}_n(\xi) = \hat{\mu}(\xi)$ for all $\xi \in \mathbb{R}^d$.

2. If there is a function φ partially continuous at 0 (that is, continuous at 0 along each coordinate axis) such that $\lim_{n\to\infty} \hat{\mu}_n(\xi) = \varphi(\xi)$ for all $\xi \in \mathbb{R}^d$, then φ is the characteristic function of some Borel probability measure μ' on $\mathbb{R}^d$ and $\mu_n \rightsquigarrow \mu'$.

Remark. Since any characteristic function is continuous at 0 (and hence partially continuous at 0), we directly have the equivalence: $\mu_n \rightsquigarrow \mu$ if and only if $\lim_{n\to\infty} \hat{\mu}_n(\xi) = \hat{\mu}(\xi)$ for all $\xi \in \mathbb{R}^d$.

Remark. The standard proof of the indirect implication in Lévy's continuity theorem, which is the difficult part of the theorem, is a direct application of Prokhorov's theorem: (i) an integral inequality is derived to prove that the assumption in (2.) implies tightness of $(\mu_n)_{n\in\mathbb{N}}$; (ii) Prokhorov's theorem is invoked from tightness to show that every subsequence of $(\mu_n)_{n\in\mathbb{N}}$ has a convergent sub-subsequence; (iii) then the direct implication of Lévy's continuity theorem and the uniqueness theorem is used to get the uniqueness of the limit in case of convergence of the subsequence; (iv) Lemma 3 is invoked.

Remark. The proof of Lévy's continuity theorem may seem involved due to the invocation of Prokhorov's theorem (which we proved using the Riesz representation theorem), but this is an artifact of the fact that we proved Prokhorov's theorem in a much more general case than $\mathbb{R}^d$. In practice, it suffices to prove Helly's selection theorem. Simpler proofs that do not even call for this result have been worked out.

Proof. 1. Since $x \mapsto e^{i\langle \xi, x\rangle}$ is a bounded and continuous function for any $\xi \in \mathbb{R}^d$, we have by definition of weak convergence that $\lim_{n\to\infty} \hat{\mu}_n(\xi) = \hat{\mu}(\xi)$ for all $\xi \in \mathbb{R}^d$.

2. Consider the coordinate axis lines $\xi = (0, \ldots, 0, \xi_j, 0, \ldots, 0)$ for $j = 1, \ldots, d$. Since $\hat{\mu}_n$ is continuous at 0 with $\hat{\mu}_n(0) = 1$, for a given $\varepsilon > 0$, we can take $\delta < 0$ small enough so that $|\hat{\mu}_n(\xi) - 1| < \varepsilon/7d$ for $|\xi| \le \delta$ and ξ on the axes. By Lemma 4, $\mu_n(|x_j| \ge 1/\delta) \le \frac{7}{\delta}\int_0^\delta 1 - \mathrm{Re}(\hat{\mu}_{n,j}(\xi_j))\, d\xi_j$ for all $n \in \mathbb{N}$ and all j where $\mu_{n,j}$ is the distribution of the jth one-dimensional projection under μ_n. Then by (DOM) we have that $\frac{7}{\delta}\int_0^\delta 1 - \mathrm{Re}(\hat{\mu}_{n,j}(\xi_j))\, d\xi_j < \varepsilon/d$ for all $n \ge n_0$ for some large enough $n_0 \in \mathbb{N}$ and all j. Thus $\mu_n(|\xi_j| \ge 1/\delta) < \varepsilon/d$ for all $n \ge n_0$ and all j. For $n < n_0$, by

countable additivity there is some $M_n < \infty$ such that $\mathbb{P}(|\xi_j| \geq M_n) < \varepsilon/d$ for all j. Define $M := \max(1/\delta, \max\{M_j : 0 \leq j < n_0\})$. Then $\mu_n(|\xi_j| \geq M) < \varepsilon/d$ for all $n \in \mathbb{N}$ and all $j = 1, \ldots, d$. If C is the cube $\{\xi \in \mathbb{R}^d : |\xi_j| \leq M \text{ for } j = 1, \ldots, d\}$, then $\mu_n(C) > 1 - \varepsilon$ for all $n \in \mathbb{N}$. This proves that $(\mu_n)_{n\in\mathbb{N}}$ is tight. By Prokhorov's theorem, every subsequence of $(\mu_n)_{n\in\mathbb{N}}$ has a weakly convergent sub-subsequence. If a subsequence of $(\mu_n)_{n\in\mathbb{N}}$ weakly converges to some limit μ', then (1.) in Lévy's continuity theorem and the uniqueness theorem imply that μ' is unique. By Lemma 3, we conclude that $\mu_n \rightsquigarrow \mu'$. □

Reference. T.9.8.2. in Dudley RAP p.326 or T.6.11. in Bhattacharya&Waymire BCPT p.117-118 or T.15.24. in Klenke PT p.345.

Lemma 2 (Prokhorov's Theorem)**.** *Let (X, d) be a metric space, $\mathcal{M}_1(X)$ the space of all Borel probability measures endowed with the topology of weak convergence, and $\mathcal{N}$ a subset of $\mathcal{M}_1(X)$.*

1. *If $\mathcal{N}$ is tight, then $\mathcal{N}$ is relatively sequentially compact.*
2. *If (X, d) is Polish and $\mathcal{N}$ is relatively sequentially compact, then $\mathcal{N}$ is tight.*

Proof. See Lecture 16. □

Lemma 3. *Let X be a metric space and $(x_n)_{n\in\mathbb{N}}$ a sequence in X. Suppose $(x_n)_{n\in\mathbb{N}}$ satisfies the two properties:*

1. *every sub-sequence of $(x_n)_{n\in\mathbb{N}}$ has a convergent sub-subsequence;*
2. *if $(x_{n_k})_{k\in\mathbb{N}}$ is a convergent sub-sequence of $(x_n)_{n\in\mathbb{N}}$, then its limit is x.*

Then $(x_n)_{n\in\mathbb{N}}$ is a convergent sequence and its limit is x.

Proof. See Lecture 16. □

Lemma 4. *Let μ be a Borel probability measure on $\mathbb{R}$. Then for any $a \in (0, +\infty)$,*

$$\mu\left(|x| \geq \frac{1}{a}\right) \leq \frac{7}{a} \int_0^a 1 - \mathrm{Re}(\hat{\mu}(\xi))\, d\xi.$$

Proof. We first prove that for any t with $|t| \geq 1$, we have $(\sin t)/t \leq \sin 1$. Since $t \mapsto (\sin t)/t$ is an even function, we may assume without loss of generality that $t \geq 0$. Since $\sin 1 > 0.8$, the claim holds for $t > 1.3$. For $1 \leq t \leq 1.3$, $t \mapsto (\sin t)/t$ is decreasing, hence the preliminary result. We now prove the main result. By the Tonelli–Fubini, we have

$$\frac{1}{a} \int_0^a \int_{\mathbb{R}} 1 - \cos(\xi x) d\mu(x) d\xi = \int_{\mathbb{R}} 1 - \frac{\sin(ax)}{ax} d\mu(x).$$

Then by the preliminary claim, we get

$$\int_{\mathbb{R}} 1 - \frac{\sin(ax)}{ax} d\mu(x) \geq (1 - \sin 1)\mu\left(|x| \geq \frac{1}{a}\right).$$

Since $\sin 1 < 6/7$, this yields the result. □

Reference. T.9.8.1. in Dudley RAP p.325. (Similar inequalities e.g. in T.3.3.17. in Durrett PTE p.132 or T.6.11. in Bhattacharya&Edward BCPT p.118.)

Lemma 5. *The characteristic function of the Gaussian distribution $\nu = N(m, \sigma^2)$ is given for all $\xi \in \mathbb{R}$ by*

$$\hat{\nu}(\xi) = e^{i\xi m} e^{-\sigma^2 \xi^2/2}.$$

Proof. We have by definition $\hat{\nu}(\xi) = \int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/(2\sigma^2)} e^{ix\xi}\, dx$. We derive the result for $m = 0$ and $\sigma = 1$ and use the facts that $\hat{\nu}_{\sigma X+m}(\xi) = e^{i\xi m}\hat{\nu}_X(\sigma\xi)$ and that if $X \sim N(0,1)$, then $\sigma X + m \sim N(m, \sigma^2)$. By parity, the imaginary part of $\hat{\nu}$ is equal to zero. It remains to compute $f(\xi) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cos(x\xi)\, dx$. Since $|x \sin(x\xi) e^{-x^2/2}| \leq |x| e^{-x^2/2}$ which is integrable, we can differentiate under the integral sign so that $f'(\xi) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} x e^{-x^2/2} \sin(x\xi)\, dx$. By integration by parts, we find that $f'(\xi) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \xi \cos(x\xi)\, dx = -\xi f(\xi)$. The function f is thus solution to the differential equation $f'(\xi) = -\xi f(\xi)$ with initial condition $f(0) = 1$. It follows that $f(\xi) = e^{-\xi^2/2}$, which concludes the proof. □

Reference. L.8.2.3. in Le Gall's notes p.104.

Lemma 6 (Uniqueness Theorem)**.** *If μ and ν are Borel probability measures on $\mathbb{R}^d$ with same characteristic function, then $\mu = \nu$.*

Proof. Let μ be a Borel probability measure on $\mathbb{R}^d$. We equivalently show that the Fourier transform defined on Borel probability measures on $\mathbb{R}^d$ is injective, that is, $\mu \mapsto \hat{\mu}$ is injective. Suppose first that $d = 1$. For any $\sigma > 0$, denote g_σ the density of the distribution $\nu_\sigma = N(0, \sigma^2)$. Consider the convolution of measures $\mu_\sigma = \mu * \nu_\sigma$, that is, for any $A \in \mathcal{B}(\mathbb{R})$, $\mu_\sigma(A) = \int_{\mathbb{R}^2} \mathbb{1}_A(x+y)\, d\mu(x)\, d\nu_\sigma(y) = \int_A dt \int_{\mathbb{R}} g_\sigma(x-t)\, d\mu(x)$. The injectivity of the Fourier transform follows from two points: 1. if we know $\hat{\mu}$, we can compute μ_σ for all $\sigma > 0$; 2. if we know $(\mu_\sigma)_{\sigma>0}$, we can compute μ. For the first part, note from last lemma that $e^{-(x-t)^2/2\sigma^2} = \hat{\nu}_{1/\sigma}(x-t)$. Thus $\mu_\sigma(A) = \int_A dt \int_{\mathbb{R}^2} \frac{1}{\sigma\sqrt{2\pi}} e^{iy(x-t)}\, d\nu_{1/\sigma}(y)\, d\mu(x) = \int_A dt \int_{\mathbb{R}} \frac{1}{\sigma\sqrt{2\pi}} e^{-iyt)} \hat{\mu}(y)\, d\nu_{1/\sigma}(y)$, which is completely determined by $\hat{\mu}$. For the second part, we show that $\mu_\sigma \rightsquigarrow \mu$ as $\sigma \to 0$. Let h be a bounded and uniformly continuous function on $\mathbb{R}$. Let $\varepsilon > 0$. We can find $\delta > 0$ such that if $|x-y| < \delta$ then $|f(x)-f(y)| < \varepsilon$. We have $\int h\, d\mu_\sigma = \mathbb{E}(h(X+Y_\sigma))$ where $X \sim \mu$, $Y \sim \nu_\sigma$, and X and Y independent. We have that $\mathbb{P}(|Y_\sigma| \geq \eta) = 2\int_\eta^\infty \frac{1}{\sigma\sqrt{2\pi}} e^{-y^2/2\sigma^2}\, dy = \frac{2}{2\pi} \int_{\eta/\sigma}^\infty e^{-y^2/2}\, dy = \varphi(\eta/\sigma)$ is a decreasing function in η/σ with $\lim_{\sigma\to 0} \varphi(\eta/\sigma) = 0$. Moreover, we have that $|\int h\, d\mu_\sigma - \int h\, d\mu| = |\mathbb{E}(h(X+Y_\sigma) - h(X))| \leq \mathbb{E}(|f(X+Y_\sigma) - f(X)| \mid \mathbb{1}_{|Y_\sigma| \leq \eta}) + 2B\mathbb{P}(|Y_\sigma| \geq \eta) \leq \varepsilon + 2B\varphi(\eta/\sigma)$ where B is a bound for $|h|$ on $\mathbb{R}$. Then for σ sufficiently small, we have $|\int h\, d\mu_\sigma - \int h\, d\mu| \leq \varepsilon$. Thus $\lim_{\sigma\to 0} \int h\, d\mu_\sigma = \int h\, d\mu$ for any function h bounded and uniformly continuous, hence $\mu_\sigma \rightsquigarrow \mu$ as $\sigma \to 0$ by Portmanteau. This concludes the proof for $d = 1$. For $d > 1$, the proof follows similarly by considering the functions $g_\sigma^{(d)}(x_1, \ldots, x_d) = \prod_{i=1}^d g_\sigma(x_i)$ and by noting that for any $\xi \in \mathbb{R}^d$, $\int_{\mathbb{R}^d} g_\sigma^{(d)}(x) e^{i\langle \xi, x\rangle}\, dx = \prod_{i=1}^d \int g_\sigma(x_i) e^{i\xi_i x_i}\, dx_i = (2\pi\sigma^2)^{d/2} g_{1/\sigma}^{(d)}(\xi)$. □

Reference. T.8.2.4. in Le Gall's notes p.104. or P.2.6. in Méliot's notes p.34 or T.9.5.1. in Dudley RAP p.303 or T.6.9. in Bhattacharya&Waymire BCPT p.116.

Addendum. Some important results of measure theory and Lebesgue integration.

Lemma 7 (Convergence Theorems for Lebesgue Integration)**.** *Let $(X, \mathcal{A}, \mu)$ be a measure space.*

1. (**Monotone Convergence**) *If $(f_n)_{n\in\mathbb{N}}$ is a sequence of measurable functions from X to $[0, +\infty]$ such that $f_n \leq f_{n+1}$ for all $n \in \mathbb{N}$, then*

$$\lim_{n\to\infty} \int_X f_n\, d\mu = \int_X \lim_{n\to\infty} f_n\, d\mu.$$

2. (**Fatou**) *If $(f_n)_{n\in\mathbb{N}}$ is a sequence of measurable functions from X to $[0,+\infty]$, then*

$$\int_X \liminf_{n\to\infty} f_n \, d\mu \leq \liminf_{n\to\infty} \int_X f_n \, d\mu.$$

3. (**Dominated Convergence**) *If $(f_n)_{n\in\mathbb{N}}$ is a sequence of measurable functions from X to $[-\infty,+\infty]$ such that $\lim_{n\to\infty} f_n = f$ where $f\colon X \to [-\infty,+\infty]$ is measurable and if there exists an integrable function $g\colon X \to [0,+\infty]$ such that $|f_n| \leq g$ for all $n \in \mathbb{N}$, then f and $f_1, f_2, \ldots$ are integrable and*

$$\lim_{n\to\infty} \int_X |f_n - f| \, d\mu = 0,$$

and, in particular,

$$\lim_{n\to\infty} \int_X f_n \, d\mu = \int_X f \, d\mu.$$

Proof. 1. (MON) Write $f := \lim_{n\to\infty} f_n = \sup_n f_n$, then $f\colon X \to [0,+\infty]$ is measurable. Since the f_n are non-decreasing to f, we have from monotonicity of the integral that $\int f_n \, d\mu$ is non-decreasing and bounded above by $\int f \, d\mu$, hence $\lim_{n\to\infty} \int_X f_n \, d\mu \leq \int_X f \, d\mu$. We now show that $\int_X f \, d\mu \leq \lim_{n\to\infty} \int_X f_n \, d\mu$. By definition, it suffices to show that $\int_X g \, d\mu \leq \lim_{n\to\infty} \int_X f_n \, d\mu$ where g is a simple function such that $0 \leq g \leq f$. Without loss of generality (by vertical truncation if needed), we may assume g finite everywhere so that there exist reals $0 \leq c_i < +\infty$ and disjoint measurable sets $A_1, \ldots, A_k$ such that $g = \sum_{i=1}^k c_i \mathbb{1}_{A_i}$, and hence $\int_X g \, d\mu = \sum_{i=1}^k c_i \mu(A_i)$. Let $0 < \varepsilon < 1$. Then we have $f(x) = \sup_n f_n(x) > (1-\varepsilon)c_i$ for all $x \in A_i$. Define $A_{i,n} := \{x \in A_i : f_n(x) > (1-\varepsilon)c_i\}$. Then the $A_{i,n}$ increase to A_i and are measurable. By upwards monotonicity of measures, we get $\lim_{n\to\infty} \mu(A_{i,n}) = \mu(A_i)$. Moreover, observe that $f_n \geq \sum_{i=1}^k (1-\varepsilon) c_i \mathbb{1}_{A_{i,n}}$ for any $n \in \mathbb{N}$. By monotonicity of the integral, we obtain $\int f_n \, d\mu \geq (1-\varepsilon) \sum i = 1^k c_i \mu(A_{i,n})$. Taking limits as $n \to \infty$, we get $\lim_{n\to\infty} \int f_n \, d\mu \geq (1-\varepsilon) \sum_{i=1}^k c_i \mu(A_i)$. Taking $\varepsilon \to 0$, we conclude that $\int_X g \, d\mu \leq \lim_{n\to\infty} \int_X f_n \, d\mu$.

2. (Fatou) We have $\liminf f_n = \lim_{k\to\infty}(\inf_{n\geq k} f_n)$, hence by (MON) $\int \liminf f_n \, d\mu = \lim_{k\to\infty} \int \inf_{n\geq k} f_n \, d\mu$. Moreover, for all $p \geq k$, we have $\inf_{n\geq k} f_n \leq f_p$, so that by monotonicity $\int \inf_{n\geq k} f_n \, d\mu \leq \inf_{p\geq k} \int f_p \, d\mu$. By taking limits when $k \to \infty$, we get $\lim_{k\to\infty} \int \inf_{n\geq k} f_n \, d\mu \leq \lim_{k\to\infty} \inf_{p\geq k} \int f_p \, d\mu = \liminf \int f_n \, d\mu$, and this concludes the proof.

3. Since $|f_n| < g$ and $|f| < g$ and $\int g \, d\mu < \infty$, then the f_n and f are integrable. Moreover, since $|f - f_n| \leq 2g$ and $|f - f_n| \to 0$, we can use Fatou's lemma to get $\liminf \int (2g - |f - f_n|) d\mu \geq \int \liminf (2g - |f - f_n|) d\mu = 2\int g \, d\mu$. By linearity of the integral, we have $2\int g \, d\mu - \limsup \int |f - f_n| d\mu \geq 2\int g \, d\mu$. Thus $\limsup \int |f - f_n| d\mu = 0$, and so $\int |f - f_n| d\mu \to 0$. Since $|\int f \, d\mu - \int f_n \, d\mu| \leq \int |f - f_n| d\mu$, we get that $\int f_n \, d\mu \to \int f \, d\mu$. This concludes the proof. □

Reference. For (MON) see T.1.4.44. in Tao MT p.107 or T.2.1.1. in Le Gall's notes p.19. For (Fatou) see T.2.1.5. in Le Gall's notes p.22 or C.1.4.47. in Tao MT p.110. For (DOM) see T.2.2.1. in Le Gall's notes p.25 or T.1.4.49. in Tao MT p.111.

Lemma 8 (**Dynkin's $\pi-\lambda$ Theorem**). *Let X be a nonempty set and $\mathcal{C} \subseteq \mathcal{L} \subseteq 2^X$ nonempty collections of sets. Suppose that $\mathcal{C}$ is a π-system, that is, $A, B \in \mathcal{C}$ implies $A \cap B \in \mathcal{C}$. Suppose that $\mathcal{L}$ is a λ-system, that is, $X \in \mathcal{L}$, $A, B \in \mathcal{L}$ and $A \subseteq B$ imply $B \setminus A \in \mathcal{L}$, and $A_n \in \mathcal{L}$ and $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$ imply $\bigcup_{n=1}^\infty A_n \in \mathcal{L}$. Then $\sigma(\mathcal{C}) \subseteq \mathcal{L}$.*

Proof. Let $\mathcal{L}_0$ be the intersection of all λ-systems containing $\mathcal{C}$. It is a λ-system that contains $\mathcal{C}$ and that is contained in every λ-system that contains $\mathcal{C}$, hence $\mathcal{C} \subseteq \mathcal{L}_0 \subseteq \mathcal{L}$. If we can show that $\mathcal{L}_0$ is also a π-system, then it will follow that $\mathcal{L}_0$ is a σ-algebra. Indeed, a λ-system that is also a π-system is a σ-algebra: by taking complements, a λ-system is closed under finite unions, and by taking an increasing limit for sets $A_n = B_1 \cup \cdots \cup B_n$, closed under countable unions. If we prove that $\mathcal{L}_0 \supseteq \mathcal{C}$ is a σ-algebra, then from the minimality of $\sigma(\mathcal{C})$ we will have that $\mathcal{C} \subseteq \sigma(\mathcal{C}) \subseteq \mathcal{L}_0 \subseteq \mathcal{L}$, hence completing the proof. We thus have to prove that $\mathcal{L}_0$ is a π-system.

Let $A \in \mathcal{C}$. Define $\mathcal{L}_1 = \{B \in \mathcal{L}_0 : A \cap B \in \mathcal{L}_0\}$. Since $\mathcal{C}$ is closed under finite intersections, we have $\mathcal{C} \subseteq \mathcal{L}_1$. We now show that $\mathcal{L}_1$ is a λ-system: $X \in \mathcal{L}_1$ by definition; if $B, B' \in \mathcal{L}_1$ and $B \subseteq B'$, then $A \cap (B' \setminus B) = (A \cap B') \setminus (A \cap B) \in \mathcal{L}_0$, hence $B' \setminus B \in \mathcal{L}_1$; if $B_n \in \mathcal{L}_1$ for all n and $(B_n)_{n\in\mathbb{N}}$ is increasing, then $A \cap (\bigcup_n B_n) = \bigcup_n (A \cap B_n) \in \mathcal{L}_0$, hence $\bigcup_n B_n \in \mathcal{L}_1$. Since $\mathcal{L}_1$ is a λ-system that contains $\mathcal{C}$, then $\mathcal{L}_0 \subseteq \mathcal{L}_1$ by minimality of $\mathcal{L}_0$. Therefore, for all $A \in \mathcal{C}$ and all $B \in \mathcal{L}_0$, we have $A \cap B \in \mathcal{L}_0$.

The same argument can be applied again, but this time to any fixed $B \in \mathcal{L}_0$ and $\mathcal{L}_2 = \{A \in \mathcal{L}_0 : A \cap B \in \mathcal{L}_0\}$. The same arguments using the last claim in last paragraph shows that $\mathcal{L}_2$ is a λ-system that contains $\mathcal{C}$, and hence that $\mathcal{L}_0 \subseteq \mathcal{L}_2$ by minimality of $\mathcal{L}_0$. This proves that for all $A \in \mathcal{L}_0$ and all $B \in \mathcal{L}_0$, we have $A \cap B \in \mathcal{L}_0$, which concludes the proof. □

Reference. T.3.2. in Billingsley PM p.42 or T.1.4.1. in Le Gall's notes p.15.

Lemma 9. *Let μ and ν be two measures on a measurable space $(X, \mathcal{A})$. Suppose that there exists a π-system $\mathcal{C} \subseteq 2^X$ such that $\mathcal{A} = \sigma(\mathcal{C})$ and $\mu(C) = \nu(C)$ for all $C \in \mathcal{C}$.*

1. If $\mu(X) = \nu(X) < \infty$, then $\mu = \nu$.

2. If there exists a sequence of sets $(X_n)_{n\in\mathbb{N}}$ such that for all $n \in \mathbb{N}$, $X_n \in \mathcal{C}$, $X_n \subseteq X_{n+1}$, $\bigcup_{n=1}^{\infty} X_n = X$, and $\mu(X_n) = \nu(X_n) < \infty$, then $\mu = \nu$.

Proof. 1. Define $\mathcal{L} := \{A \in \mathcal{A} : \mu(A) = \nu(A)\}$. Then $\mathcal{C} \subseteq \mathcal{L}$. We show that $\mathcal{L}$ is a λ-system: $X \in \mathcal{L}$ by hypothesis; if $A, B \in \mathcal{L}$ and $A \subseteq B$, then $\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A)$, hence $B \setminus A \in \mathcal{L}$; if $A_n \in \mathcal{L}$ for all n and $(A_n)_{n\in\mathbb{N}}$ is increasing, then $\mu(\bigcup_n A_n) = \sup_n \mu(A_n) = \sup_n \nu(A_n) = \nu(\bigcup_n A_n)$, hence $\bigcup_n A_n \in \mathcal{L}$. It follows that $\mathcal{A} = \sigma(\mathcal{C}) \subseteq \mathcal{L}$ by Dynkin's $\pi - \lambda$ theorem. Thus $\mathcal{L} = \mathcal{A}$, and so $\mu = \nu$.

2. Define for all $n \in \mathbb{N}$ the restriction of μ and ν to X_n by μ_n and ν_n, that is, for all $A \in \mathcal{A}$, $\mu_n(A) := \mu(A \cap X_n)$ and $\nu_n(A) := \nu_n(A \cap X_n)$. Then apply (1.) to μ_n and ν_n, so that $\mu_n = \nu_n$. By taking an increasing limit for the sets $A \cap X_n$, we have for all $A \in \mathcal{A}$ that $\mu(A) = \sup_n \mu(A \cap X_n) = \sup_n \nu(A \cap X_n) = \nu(A)$, and so $\mu = \nu$. □

Reference. C.1.4.2. in Le Gall's notes p.16 or T.3.3. in Billingsley PM p.42.

Lemma 10 (Existence of Product Measure). *Let μ and ν be σ-finite measures on the measurable spaces $(X, \mathcal{A})$ and $(Y, \mathcal{B})$, respectively. Define $\mathcal{A} \otimes \mathcal{B} := \sigma(\{A \times B : A \in \mathcal{A}, B \in \mathcal{B}\})$. Then there exists a unique σ-finite measure m on the product space $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ such that for all $A \in \mathcal{A}$ and all $B \in \mathcal{B}$,*

$$m(A \times B) = \mu(A) \times \nu(B).$$

The measure m is alternatively denoted $\mu \times \nu$.

Proof. (*Uniqueness.*) Define $C_n := A_n \times B_n$ where $A_n \in \mathcal{A}$ and $B_n \in \mathcal{B}$ are increasing sequences with $\mu(A_n) < \infty$, $\nu(B_n) < \infty$, $\bigcup_n A_n = X$, and $\bigcup_n B_n = Y$. Then the C_n form an increasing sequence of sets such that $\bigcup_n C_n = X \times Y$. Let m and m' be two measures satisfying the property of the product measure. Then for all $n \in \mathbb{N}$, $m(C_n) = \mu(A_n) \times \nu(B_n) = m'(C_n) < \infty$. Moreover, m and m' coincide for all measurable rectangles which form a π-system generating $\mathcal{A} \otimes \mathcal{B}$. By Dynkin's $\pi - \lambda$ theorem, we have $m = m'$.

(*Existence.*) Define tentatively for all $C \in \mathcal{A} \otimes \mathcal{B}$, $m(C) = \int_X \nu(C_x)\, d\mu(x)$ where $C_x = \{y \in Y : (x, y) \in C\}$ for any $x \in X$. It follows from the definition that $m(A \times B) = \mu(A) \times \nu(B)$ for all $A \in \mathcal{A}$ and all $B \in \mathcal{B}$.

We start by showing that the tentative definition makes sense. We first show that $\nu(C_x)$ is well-defined, that is, $C_x \in \mathcal{B}$. Let $x \in X$ and define $\mathcal{L} := \{C \in \mathcal{A} \otimes \mathcal{B} : C_x \in \mathcal{B}\}$. Then $\mathcal{L}$ contains the measure rectangles since, whenever $C = A \times B$, $C_x = B$ if $x \in A$ and $C_x = \emptyset$ if $X \notin A$. It also follows directly from the definition that $\mathcal{L}$ is a λ-system. By Dynkin's $\pi - \lambda$ theorem, we have that $\mathcal{A} \otimes \mathcal{B} \subseteq \mathcal{L}$, and hence $\mathcal{A} \otimes \mathcal{B} = \mathcal{L}$ by definition of $\mathcal{L}$. To verify that our tentative definition makes sense, we also have to prove that $x \mapsto \nu(C_x)$ is $\mathcal{A}$-measurable. Suppose first that ν is finite and define $\mathcal{L}' := \{C \in \mathcal{A} \otimes \mathcal{B} : x \mapsto \nu(C_x) \text{ is } \mathcal{A}\text{-measurable}\}$. Then $\mathcal{L}'$ contains the measurable rectangles since $\nu(C_x) = \mathbb{1}_A(x)\nu(B)$ for $C = A \times B$. Moreover, $\mathcal{L}'$ is a λ-system: $X \times Y \in \mathcal{L}'$; if $C \subseteq C'$, then $\nu((C' \setminus C)_x) = \nu(C'_x) - \nu(C_x)$ (by finiteness of ν); if the (C_n) form an increasing sequence, then $\nu((\bigcup_n C_n)_x) = \lim_n \nu((C_n)_x)$. By Dynkin's $\pi - \lambda$ theorem, we have that $\mathcal{A} \otimes \mathcal{B} \subseteq \mathcal{L}'$, and hence $\mathcal{A} \otimes \mathcal{B} = \mathcal{L}'$ by definition of $\mathcal{L}'$. This proves that $x \mapsto \nu(C_x)$ is $\mathcal{A}$-measurable. If ν is only σ-finite, we choose a sequence $(B_n)_{n \in \mathbb{N}}$ as above and consider $\nu_n(B) = \nu(B \cap B_n)$ to obtain that $x \mapsto \nu(C_x) = \lim_n \nu_x(C_x)$ is measurable for all $C \in \mathcal{A} \otimes \mathcal{B}$.

We now show that m is a measure on $\mathcal{A} \otimes \mathcal{B}$. It follows directly from the properties of the integral and the definition of m that $m(C) \geq 0$ for all $C \in \mathcal{A} \otimes \mathcal{B}$ and $m() = 0$. For countable additivity, consider a disjoint family $(C_n)_{n \in \mathbb{N}}$ in $\mathcal{A} \otimes \mathcal{B}$. Then the $(C_n)_x$ form a disjoint family in $\mathcal{B}$ for all $x \in X$. Thus $m(\bigcup_n C_n) = \int_X \nu((\bigcup_n C_n)_x)\, d\mu(x) = \int_X \nu(\bigcup_n (C_n)_x\, d\mu(x) = \int_X \sum_n \nu((C_n)_x)\, d\mu(x) = \sum_n \int_X \nu((C_n)_x)\, d\mu(x) = \sum_n m(C_n)$. Hence m is a measure, and this concludes the proof. □

Reference. T.5.2.1. in Le Gall's notes p.58 or T.18.2. in Billingsley PM p.232-233.

Lemma 11 (**Fubini–Tonelli Theorem**). *Let μ and ν be σ-finite measures on the measurable spaces $(X, \mathcal{A})$ and $(Y, \mathcal{B})$, respectively. Let $f : X \times Y \to [0, +\infty]$ be a measurable function. Then*

(i) the functions

$$x \mapsto \int_Y f(x, y)\, d\nu(y),$$
$$y \mapsto \int_X f(x, y)\, d\mu(x)$$

are $\mathcal{A}$-measurable and $\mathcal{B}$-measurable, respectively;

(ii)

$$\int_{X\times Y} f(x,y)\,d(\mu\times\nu)(x,y) = \int_Y \left(\int_X f(x,y)\,d\mu(x)\right) d\nu(y),$$
$$= \int_X \left(\int_Y f(x,y)\,d\nu(y)\right) d\mu(x).$$

Proof. 1. Let $C \in \mathcal{A}\otimes\mathcal{B}$. If $f = \mathbb{1}_C$, we have already proved that the function $x \mapsto \int_Y f(x,y)\,d\nu(y) = \nu(C_x)$ is $\mathcal{A}$-measurable. Similarly, the function $x \mapsto \int_X f(x,y)\,d\mu(x) = \mu(C^y)$, where $C^y = \{x \in X : (x,y) \in C\}$ for any $y \in Y$, is $\mathcal{B}$-measurable. By linearity, the result is true for any positive simple function. For f positive, we can write $f = \lim_n f_n$ whee the f_n form an increasing sequence of positive simple functions. Then (MON) yields $\int_Y f(x,y)\,d\nu(y) = \lim_n \int_Y f_n(x,y)\,d\nu(y)$, from which the result follows. Similarly, we can show the measurability of $y \mapsto \int_X f(x,y)\,d\mu(x)$.

2. Let $C \in \mathcal{A}\otimes\mathcal{B}$. For $f = \mathbb{1}_C$, the result rewrites as $\mu\times\nu(C) = \int_X \nu(C_x)\,d\mu(x) = \int_Y \mu(C^y)\,d\nu(y)$. We have already proved the first equality in the proof of the existence of the product measure. The second equality can be proved in the same way and by using the uniqueness of $\mu\times\nu$. The result follows for any positive simple function by linearity, and for arbitrary positive function by taking an increasing sequence of positive simple functions and using (MON). Indeed, we note for instance that if $f = \lim_n \uparrow f_n$, then $\int_X(\int_Y f(x,y)\,d\nu(y))\,d\mu(x) = \lim_n \int_X(\int_Y f_n(x,y)\,d\nu(y))\,d\mu(x)$ by applying (MON) twice. This concludes the proof. □

Reference. T.5.3.1 in Le Gall's notes p.61.

References

Bhattacharya, R., and E. C. Waymire (2016): *A basic course in probability theory*. Springer.

Billingsley, P. (1995): *Probability and measure*. John Wiley & Sons.

Dudley, R. M. (2004): *Real analysis and probability*. Cambridge University Press.

Durrett, R. (2019): *Probability: theory and examples*, vol. 49. Cambridge university press.

Klenke, A. (2020): *Probability Theory: A Comprehensive Course*. Springer.

Le Gall, J.-F. (2006): "Integration, probabilites, et processus aleatoires," *Lecture notes*.

Meliot, P.-L. (2016): "Convergence de mesures, processus de Poisson et processus de Lévy," *Lecture notes*.

Tao, T. (2011): *An introduction to measure theory*, vol. 126. American Mathematical Soc.