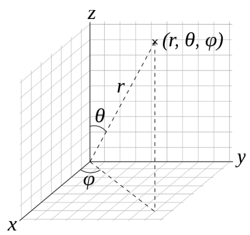
9 Radon-Nikodym theorem and conditioning

9a Borel-Kolmogorov paradox

Spherical coordinates on $\mathbb{R}^3$ may be treated as a map $\alpha : (r,\theta,\varphi) \mapsto (x,y,z)$ where[1]

(9a1)
$$\begin{gathered} x = r\sin\theta\cos\varphi, \\ y = r\sin\theta\sin\varphi, \\ z = r\cos\theta; \end{gathered}$$



this is a homeomorphism (moreover, diffeomorphism) between two open sets in $\mathbb{R}^3$:

$$(0,\infty) \times (0,\pi) \times (-\pi,\pi) \to \mathbb{R}^3 \setminus \big((-\infty,0] \times \{0\} \times \mathbb{R} \big) .$$

It does not preserve Lebesgue measure m; rather, m is the image of the measure[2]

$$\big((r,\theta,\varphi) \mapsto r^2 \sin\theta \big) \cdot m .$$

Less formally, one writes

$$\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = r^2 \sin\theta\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\varphi = (r^2\,\mathrm{d}r)(\sin\theta\,\mathrm{d}\theta)(\mathrm{d}\varphi) ,$$

a product measure. And the uniform distribution on the ball $x^2+y^2+z^2 < 1$ turns into the product of three probability measures

$$\frac{3}{4\pi}\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = (3r^2\,\mathrm{d}r)(\tfrac12 \sin\theta\,\mathrm{d}\theta)(\tfrac{1}{2\pi}\mathrm{d}\varphi)$$

[1]Picture from Wikipedia.

[2]See also Footnote 1 on page 100.

on $(0,1)\times(0,\pi)\times(-\pi,\pi)$.

According to Sect. 6d, the conditional distribution on the sphere $x^2+y^2+z^2=1$ (that is, $r=1$) is given by $(\frac12\sin\theta\,\mathrm{d}\theta)(\frac1{2\pi}\mathrm{d}\varphi)$. Further, the conditional distribution on the circle $x^2+y^2=1$, $z=0$ (that is, $r=1$, $\theta=\frac{\pi}{2}$, the equator) is given by $\frac1{2\pi}\mathrm{d}\varphi$. And the conditional distribution on the half-circle $x^2+z^2=1$, $y=0$, $x>0$ (that is, $r=1$, $\varphi=0$, a line of longitude) is given by $\frac12\sin\theta\,\mathrm{d}\theta$.

Quite strange: the result is not invariant under rotations of $\mathbb{R}^3$; why?[1]

9b Radon-Nikodym theorem

9b1 Definition. Let (X,S,μ) be a measure space. A measure ν on (X,S) is *absolutely continuous* (w.r.t. μ), in symbols $\nu\ll\mu$, if

$$\forall A\in S\;\bigl(\mu(A)=0\implies\nu(A)=0\bigr)\,.$$

If $\nu=f\cdot\mu$ for some measurable $f:X\to[0,\infty]$, then $\nu\ll\mu$ (recall Sect. 4c). If μ is σ-finite and $\nu=f\cdot\mu$ for some measurable $f:X\to[0,\infty)$, then ν is σ-finite (by 4c10(b)) and $\nu\ll\mu$. Here is the converse.

9b2 Theorem (Radon-Nikodym)**.** Let (X,S,μ) be a σ-finite measure space, and ν an absolutely continuous (w.r.t. μ) σ-finite measure on (X,S). Then $\nu=f\cdot\mu$ for some measurable $f:X\to[0,\infty)$.

9b3 Remark. If ν is not σ-finite, then still $\nu=f\cdot\mu$, but $f:X\to[0,\infty]$.

This claim fails badly without σ-finiteness of μ.

9b4 Exercise. Let (X,S) be $[0,1]$ with Borel σ-algebra, and ν the Lebesgue measure on it. Prove that ν is not of the form $f\cdot\mu$, if

(a) μ is the counting measure;

(b) $\mu=\infty\cdot\nu$.

9b5 Remark. Uniqueness of f (up to equivalence) is ensured by 7a4.

Proof *of Th. 9b2 and Remark 9b3.* WLOG, $\mu(X)<\infty$. Indeed, a σ-finite μ is equivalent to some finite measure μ_1 (by 5b8), and $\nu\ll\mu\iff\nu\ll\mu_1$ (since μ and μ_1 have the same null sets, as noted before 5b7); also, $\nu=f\cdot\mu_1\iff\nu=f\frac{\mathrm{d}\mu_1}{\mathrm{d}\mu}\cdot\mu$ (by 4b7).

• From now on, μ is finite.

[1] "Many quite futile arguments have raged between otherwise competent probabilists over which of these results is 'correct'." E.T. Jaynes (quote from Wikipedia).

If ν is not σ-finite, we take $A_n \in S$ such that $\nu(A_n) < \infty$ and $\mu(A_n) \to \sup_{\nu(A)<\infty} \mu(A)$; we introduce $A_\infty = \cup_n A_n$. Clearly, ν is σ-finite on A_∞; and $\nu = \infty \cdot \mu$ on $X \setminus A_\infty$ (think, why). Thus, 9b2 implies that $\nu = f \cdot \mu$ for some measurable $f : X \to [0, \infty]$.

Given a σ-finite ν, we may assume WLOG that ν is finite (similarly to μ).

• From now on, also ν is finite.

If $\nu = f \cdot (\mu + \nu)$ for some f, then $(1 - f) \cdot \nu = f \cdot \mu$, and $\nu \ll \mu$ implies $1 - f > 0$ a.e. (think, why), therefore $\nu = \frac{f}{1-f} \cdot \mu$.

• From now on, in addition, $\nu \le \mu$.

We need f such that $\nu(A) = (f \cdot \mu)(A) = \int f 1\!\!1_A \, \mathrm{d}\mu = \langle f, 1\!\!1_A \rangle_\mu$ for all $A \in S$; here the inner product is taken in $L_2(\mu)$. It is sufficient to find $f \in L_2(\mu)$ such that $\langle f, g \rangle_\mu = \int g \, \mathrm{d}\nu$ for all $g \in L_2(\mu)$ (then surely $f \ge 0$). Taking into account that $| \int g \, \mathrm{d}\nu| = |\langle g, 1\!\!1 \rangle_\nu| \le \|g\|_\nu \|1\!\!1\|_\nu = \sqrt{\nu(X) \int g^2 \, \mathrm{d}\nu} \le \sqrt{\nu(X) \int g^2 \, \mathrm{d}\mu} = \sqrt{\nu(X)} \|g\|_\mu$ we see that the linear functional $\ell : L_2(\mu) \to \mathbb{R}$ defined by $\ell(g) = \int g \, \mathrm{d}\nu$ is bounded. Thus, Th. 9b2 is reduced to the following well-known fact from the theory of Hilbert spaces. □

9b6 Lemma. For every bounded linear functional ℓ on $L_2(\mu)$ there exists $f \in L_2(\mu)$ such that

$$\forall g \in L_2(\mu) \;\; \ell(g) = \langle f, g \rangle \, .$$

Usually, $L_2(\mu)$ is separable, therefore has an orthonormal basis $(e_n)_n$, and we just take

$$f = \sum_n \ell(e_n) e_n$$

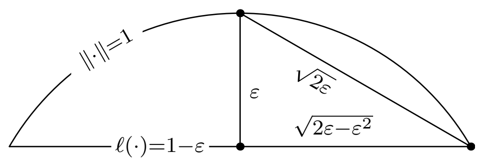
(it converges; think, why); then $\ell(g) = \langle f, g \rangle$ for $g = e_n$, therefore, for all g.

It is possible to generalize this argument to nonseparable spaces. Alternatively, a geometric proof is well-known. WLOG, the norm $\sup_{\|f\| \le 1} \ell(f)$ of ℓ is 1. For every $\varepsilon \in (0, 1)$ and f such that $\ell(f) \ge 1 - \varepsilon$ we have

(a) $|\ell(g) - \langle f, g \rangle| \le \sqrt{2\varepsilon}$ for all g of norm ≤ 1;

(b) $\|f - g\| \le 2\sqrt{2\varepsilon - \varepsilon^2}$ for all g of norm ≤ 1 such that $\ell(g) \ge 1 - \varepsilon$;

just elementary geometry on the Euclidean plane containing f and g.



Thus, every sequence $(f_n)_n$ such that $\ell(f_n) \to 1$, being Cauchy sequence, converges to some f, and $\forall g \in L_2(\mu) \;\; \ell(g) = \langle f, g \rangle$.

Theorem 9b2 is thus proved.

9b7 Remark. Let (X,S) and (Y,T) be measurable spaces, and $\varphi : X \to Y$ measurable map. If measures μ_1, ν_1 on (X,S) satisfy $\nu_1 \ll \mu_1$, then pushforward measures $\mu_2 = \varphi_*\mu_1$, $\nu_2 = \varphi_*\nu_1$ satisfy $\nu_2 \ll \mu_2$ (think, why). Therefore, every measure of the form $\varphi_*(f \cdot \mu)$ is also of the form $g \cdot \varphi_*\mu$.

9b8 Definition. Two measures μ, ν on a measure space (X,S) are *mutually singular* (in symbols, $\mu \perp \nu$) if there exists $A \in S$ such that $\mu(A) = 0$ and $\nu(X \setminus A) = 0$.

See 3d5 for a nonatomic measure on $[0,1]$ that is singular to Lebesgue measure.

9b9 Exercise. Two σ-finite measures μ, ν on (X,S) are mutually singular if and only if $\frac{d\mu}{d(\mu+\nu)} \in \{0,1\}$ a.e.

Prove it.

9b10 Theorem (Lebesgue's decomposition theorem)**.** Let (X,S,μ) be a σ-finite measure space, and ν a σ-finite measure on (X,S). Then ν can be expressed uniquely as a sum of two measures, $\nu = \nu_a + \nu_s$, where $\nu_a \ll \mu$ and $\nu_s \perp \mu$.

9b11 Exercise. Prove Theorem 9b10.[1]

9c Conditioning

9c1 Definition. Given a probability space $(\Omega, \mathcal{F}, P)$, a measurable space (E,S) and a measurable map $\varphi : \Omega \to E$ from $(\Omega, \mathcal{F})$ to (E,S), we define the *conditional expectation* $\mathbb{E}\big(X \,\big|\, \varphi\big)$ of an integrable $X : \Omega \to \mathbb{R}$

(a) for $X : \Omega \to [0,\infty)$, as a measurable $g : E \to [0,\infty)$ such that $\varphi_*(X \cdot P) = g \cdot \varphi_* P$;

(b) in general, by $\mathbb{E}\big(X \,\big|\, \varphi\big) = \mathbb{E}\big(X_+ \,\big|\, \varphi\big) - \mathbb{E}\big(X_- \,\big|\, \varphi\big)$.

9c2 Remark. Existence of $\mathbb{E}\big(X \,\big|\, \varphi\big)$ is ensured by 9b7, uniqueness (up to equivalence) by 9b5. The equivalence class of $\mathbb{E}\big(X \,\big|\, \varphi\big)$ is uniquely determined by the equivalence class of X.

9c3 Exercise. The conditional expectation is a linear operator from $L_1(P)$ to $L_1(\varphi_* P)$, and $\|\mathbb{E}\big(X \,\big|\, \varphi\big)\| \le \|X\|$, and $\mathbb{E}_1\big(\mathbb{E}\big(X \,\big|\, \varphi\big)\big) = \mathbb{E}\, X$ (where $\mathbb{E}_1$ is the integral w.r.t. $\varphi_* P$).

Prove it.[2]

[1] Hint: consider $\frac{d\mu}{d(\mu+\nu)}$.

[2] Recall the proof of 4d2.

Some convenient notation:

(9c4) $\mathbb{P}\big(A\,\big|\,\varphi\big) = \mathbb{E}\big(1\!\!1_A\,\big|\,\varphi\big)$ for $A \in \mathcal{F}$ ("conditional probability");

(9c5) $$\mathbb{E}\big(X\,\big|\,\varphi = b\big) = \mathbb{E}\big(X\,\big|\,\varphi\big)(b) \quad \text{for } b \in E\,.$$

By 4c21, $\varphi_*\big((f\circ\varphi)\cdot P\big) = f\cdot\varphi_*P$; applying this to f_+, f_- we get for a φ_*P-integrable f

(9c6) $$\mathbb{E}\big(f\circ\varphi\big|\varphi\big) = f\,,$$

that is,

(9c7) $$\mathbb{E}\big(f(\varphi)\big|\varphi = b\big) = f(b)\,.$$

Moreover, assuming integrability of X, $f\circ\varphi$ and $(f\circ\varphi)X$,

(9c8) $$\mathbb{E}\big((f\circ\varphi)X\,\big|\,\varphi\big) = f\,\mathbb{E}\big(X\,\big|\,\varphi\big)\,,$$

since for $X \ge 0$, $f \ge 0$ (otherwise, take f_+, f_-, X_+, X_-)

$$\begin{aligned}\varphi_*\big((f\circ\varphi)X\cdot P\big) = \varphi_*\big((f\circ\varphi)\cdot(X\cdot P)\big) = f\cdot\varphi_*(X\cdot P) = \\ = f\cdot\big(\mathbb{E}\big(X\,\big|\,\varphi\big)\cdot\varphi_*P\big) = \big(f\,\mathbb{E}\big(X\,\big|\,\varphi\big)\big)\cdot\varphi_*P\,.\end{aligned}$$

That is,

(9c9) $$\mathbb{E}\big(f(\varphi)X\,\big|\,\varphi = b\big) = f(b)\,\mathbb{E}\big(X\,\big|\,\varphi = b\big)$$

("taking out what is known", or "pulling out known factors").

The equality $\varphi_*(X\cdot P) = g\cdot\varphi_*P$ may be rewritten as

(9c10) $$\int_{\varphi^{-1}(B)} X\,\mathrm{d}P = \int_B g\,\mathrm{d}\varphi_*P \qquad \text{for all } B \in S$$

or, using (4c22), as

(9c11) $$\int_{\varphi^{-1}(B)} X\,\mathrm{d}P = \int_{\varphi^{-1}(B)} g\circ\varphi\,\mathrm{d}P \qquad \text{for all } B \in S\,.$$

Introducing the σ-algebra $\mathcal{F}_\varphi$ ("generated by φ") by

$$\mathcal{F}_\varphi = \{\varphi^{-1}(B) : B \in S\}\,,$$

we rewrite (9c11) as $\int_A X\,\mathrm{d}P = \int_A g\circ\varphi\,\mathrm{d}P$ for all $A \in \mathcal{F}_\varphi$, that is, $(X\cdot P)|_{\mathcal{F}_\varphi} = \big((g\circ\varphi)\cdot P\big)|_{\mathcal{F}_\varphi}$; also, $g\circ\varphi$ is measurable on $(\Omega, \mathcal{F}_\varphi)$.

Thus, we may forget φ, consider instead a sub-σ-algebra $\mathcal{F}_1 \subset \mathcal{F}$, and define $\mathbb{E}\big(X \,\big|\, \mathcal{F}_1\big)$ as an integrable function on $(\Omega, \mathcal{F}_1, P_{\mathcal{F}_1})$ such that[1]

$$(X \cdot P)|_{\mathcal{F}_1} = \mathbb{E}\big(X \,\big|\, \mathcal{F}_1\big) \cdot P|_{\mathcal{F}_1} \quad \text{for } X \ge 0\,,$$

and in general,

$$\int_A X \,\mathrm{d}P = \int_A \mathbb{E}\big(X \,\big|\, \mathcal{F}_1\big) \,\mathrm{d}P \quad \text{for all } A \in \mathcal{F}_1\,,$$

that is,

$$\mathbb{E}\left(X 1\!\!1_A\right) = \mathbb{E}\left(\mathbb{E}\big(X \,\big|\, \mathcal{F}_1\big) 1\!\!1_A\right) \quad \text{for all } A \in \mathcal{F}_1\,.$$

This approach may seem to be more general, but in fact, it is not. Given $\mathcal{F}_1 \subset \mathcal{F}$, we may take $(E, S) = (\Omega, \mathcal{F}_1)$ and $\varphi = \mathrm{id}$. Thus, all formulas written in terms of $\mathbb{E}\big(\cdot \,\big|\, \varphi\big)$ may be rewritten (and still hold!) in terms of $\mathbb{E}\big(\cdot \,\big|\, \mathcal{F}_1\big)$. In particular, (9c6)–(9c9) turn into

(9c12) $$\mathbb{E}\big(f \,\big|\, \mathcal{F}_1\big) = f \quad \text{for } \mathcal{F}_1\text{-measurable, integrable } f\,;$$

(9c13) $$\mathbb{E}\big(fX \,\big|\, \mathcal{F}_1\big) = f\,\mathbb{E}\big(X \,\big|\, \mathcal{F}_1\big) \text{ for } \mathcal{F}_1\text{-measurable } f$$

(integrability of f, X and fX is assumed, integrability of $f\,\mathbb{E}\big(X \,\big|\, \mathcal{F}_1\big)$ follows).

Also, by 9c3, the conditional expectation is a linear operator $L_1(\Omega, \mathcal{F}, P) \to L_1(\Omega, \mathcal{F}_1, P|_{\mathcal{F}_1}) \subset L_1(\Omega, \mathcal{F}, P)$, and

(9c14) $$\|\mathbb{E}\big(X \,\big|\, \mathcal{F}_1\big)\|_1 \le \|X\|_1\,,$$

(9c15) $$\mathbb{E}\left(\mathbb{E}\big(X \,\big|\, \mathcal{F}_1\big)\right) = \mathbb{E}\,X$$

("law of total[2] expectation").

By 5f4, $L_2(P) \subset L_1(P)$. Let us consider $Y = \mathbb{E}\big(X \,\big|\, \mathcal{F}_1\big)$ for $X \in L_2(P)$. For every $\mathcal{F}_1$-measurable $Z \in L_2(P)$ we know that XZ is integrable, and (9c13) gives $\mathbb{E}\big(ZX \,\big|\, \mathcal{F}_1\big) = Z\mathbb{E}\big(X \,\big|\, \mathcal{F}_1\big) = ZY$. Using (9c14), $\|ZY\|_1 \le \|ZX\|_1 \le \|Z\|_2 \|X\|_2$, which implies $\|Y\|_2 \le \|X\|_2$ (take $Z_n \to Y$, $|Z_n| \le |Y|$), thus, $Y \in L_2$. Using (9c15), $\mathbb{E}\,(ZX) = \mathbb{E}\,(ZY)$, that is, $\langle Z, X\rangle = \langle Z, Y\rangle$. We see that $X - Y$ is orthogonal to the subspace $L_2(\Omega, \mathcal{F}_1, P|_{\mathcal{F}_1})$ of $L_2(\Omega, \mathcal{F}, P)$, and Y belongs to this subspace, which shows that

(9c16) $$\mathbb{E}\big(X \,\big|\, \mathcal{F}_1\big) \text{ is the orthogonal projection of } X \text{ to } L_2(\Omega, \mathcal{F}_1, P|_{\mathcal{F}_1})$$

(in other words, the best approximation...), whenever $X \in L_2(P)$. Taking into account that $L_2(P)$ is dense in $L_1(P)$ we may say that the conditional expectation is the orthogonal projection extended by continuity to $L_1(P)$.[3]

[1]For a $\mathcal{F}_1$-measurable f we have $\int f \,\mathrm{d}P = \int f \,\mathrm{d}(P|_{\mathcal{F}_1})$, as was noted before 4c24.

[2]Or "iterated".

[3]The continuity in L_1 metric does not follow just from continuity in L_2 metric; specific properties of this operator are used.

9c17 Exercise. (a) Let $b \in E$ be an atom of $\varphi_* P$, that is, $\{b\} \in S$ and $P\big(\varphi^{-1}(b)\big) > 0$. Then

$$\mathbb{P}\big(A \big| \varphi = b\big) = \frac{P(A \cap \varphi^{-1}(b))}{P(\varphi^{-1}(b))} \,.$$

(b) Let B be an atom of $P|_{\mathcal{F}_1}$, that is, $B \in \mathcal{F}_1$, $P(B) > 0$, and

$$\forall C \in \mathcal{F}_1 \quad \big(C \subset B \quad \Longrightarrow \quad P(C) \in \{0, P(B)\} \,\big) \,.$$

Then

$$\mathbb{P}\big(A \big| \mathcal{F}_1\big) = \frac{P(A \cap B)}{P(B)} \;\text{ on } B \,.$$

Prove it.

We see that an atom leads to a conditional measure,

$$P_b : A \mapsto \frac{P(A \cap \varphi^{-1}(b))}{P(\varphi^{-1}(b))} \,, \qquad \text{or} \qquad P_B : A \mapsto \frac{P(A \cap B)}{P(B)} \,,$$

a probability measure concentrated on $\varphi^{-1}(b)$, or B; and in this case, the conditional expectation is the integral w.r.t. the conditional measure,

$$\mathbb{E}\big(X \big| \varphi = b\big) = \int X \, \mathrm{d}P_b \,, \qquad \text{or} \qquad \mathbb{E}\big(X \big| \mathcal{F}_1\big) = \int X \, \mathrm{d}P_B \;\text{ on } B$$

(check it). Also, an atom is "self-sufficient": in order to know its conditional measure we need to know only B (or $\varphi^{-1}(b)$) rather than the whole $\mathcal{F}_1$ (or φ).

In the general theory, existence of conditional measures is problematic.[1] But in specific (non-pathological) examples it usually exists and may be calculated (more or less) explicitly.

9c18 Example. The special case treated in Sect. 6d: $(\Omega, \mathcal{F}, P) = (\Omega_1, \mathcal{F}_1, P_1) \times (\Omega_2, \mathcal{F}_2, P_2)$ and $\varphi(\omega_1, \omega_2) = \omega_1$. The conditional measure P_{ω_1} is the image of P_2 under the embedding $\omega_2 \mapsto (\omega_1, \omega_2)$.

9c19 Example. Let Ω be the unit disk $\{(x, y) : x^2 + y^2 < 1\}$ on $\mathbb{R}^2$, with the Lebesgue σ-algebra $\mathcal{F}$ and the uniform distribution P (with the constant density $1/\pi$); and let $\varphi(x, y)$ be the polar angle,

$$\begin{aligned} x &= r \cos\theta \\ y &= r \sin\theta \end{aligned} \qquad \text{where} \qquad \begin{aligned} r &= \sqrt{x^2 + y^2} \\ \theta &= \varphi(x, y) \,. \end{aligned}$$

[1] It holds for standard probability spaces, and may fails otherwise.

(We neglect the origin.)

We have a homeomorphism (moreover, diffeomorphism) between two open sets in $\mathbb{R}^2$:

$$\alpha : (0,1) \times (-\pi, \pi) \to \Omega \setminus \big((-1,0] \times \{0\} \big) . \quad \alpha(r,\theta) = (x,y) .$$

Using elementary geometry,

$$P\big(\alpha\big((0,r) \times (\theta_1, \theta_2)\big)\big) = \frac{1}{\pi} \frac{\theta_2 - \theta_1}{2} r^2 = \bigg(\int_{\theta_1}^{\theta_2} \frac{\mathrm{d}\theta}{2\pi} \bigg) \bigg(\int_0^r 2\rho \, \mathrm{d}\rho \bigg)$$

for $-\pi \le \theta_1 \le \theta_2 \le \pi$ and $0 \le r \le 1$, which means that P is the image of the product measure $\frac{\mathrm{d}\theta}{2\pi} 2r \, \mathrm{d}r$ on $(0,1) \times (-\pi, \pi)$. (Indeed, the latter measure coincides with $(\alpha^{-1})_* P$ on the algebra generated by boxes.)[1]

Neglecting the null set $(-1,0] \times \{0\} \subset \Omega$ we see that conditioning on the map $\varphi : \Omega \to (-\pi, \pi)$, $\varphi(x,y) = \theta$, is equivalent[2] to conditioning on the projection $(0,1) \times (-\pi,\pi) \to (-\pi,\pi)$, $(r,\theta) \mapsto \theta$. Treated as random variables, r and θ are independent, and the distribution of r has the density $2r$; the same is the conditional distribution of r given θ. Thus,

$$\mathbb{E}\big(X \,\big|\, \varphi = \theta \big) = \int_0^1 X(r\cos\theta, r\sin\theta) \, 2r \, \mathrm{d}r \, ;$$

$$\mathbb{E}\big(X \,\big|\, \mathcal{F}_\varphi \big)(x,y) = \int_0^1 X\bigg(\frac{rx}{\sqrt{x^2+y^2}}, \frac{ry}{\sqrt{x^2+y^2}} \bigg) \, 2r \, \mathrm{d}r \, .$$

9c20 Example. Still, the same Ω (the disk), $\mathcal{F}$ and P, but now let φ be the projection $(x,y) \mapsto x$ from Ω to $(-1,1)$.

Treating P as a measure on $\mathbb{R}^2$ we see that it is not a product measure (think, why), but it has a density $\frac{1}{\pi} \mathbb{1}_\Omega$ w.r.t. the product measure $m_2 = m_1 \times m_1$. Thus,

$$\int X \, \mathrm{d}P = \int \mathrm{d}x \int \mathrm{d}y \, X(x,y) \frac{1}{\pi} \mathbb{1}_\Omega(x,y) \, ;$$

for $X \ge 0$ we see that $\varphi_*(X \cdot P)$ has the density $x \mapsto \int X(x,y) \frac{1}{\pi} \mathbb{1}_\Omega(x,y) \, \mathrm{d}y$ w.r.t. m_1. In particular, taking $X = 1$ we see that $\varphi_*(P)$ has the density

---

[1] By the way, this is a special case of a well-known change of variable theorem from Analysis-3: if $U, V \subset \mathbb{R}^d$ are open sets and $\varphi : U \to V$ a diffeomorphism, then $\int_U (f \circ \varphi) |\det D\varphi| \, \mathrm{d}m = \int_V f \, \mathrm{d}m$ for every compactly supported continuous function f on V. A limiting procedure gives $\int_B |\det D\varphi| \, \mathrm{d}m = m(\varphi(B))$ for every box B such that $\overline{B} \subset U$. It follows that $(\varphi^{-1})_* m = |\det D\varphi| \cdot m$ on every B, and therefore, on the whole U.

[2] See also 9c22.

$x \mapsto \int \frac{1}{\pi} 1\!\!1_\Omega(x,y)\,\mathrm{d}y = \frac{2}{\pi}\sqrt{1-x^2}$ (and 0 if $x^2 > 1$) w.r.t. m_1. Thus, $\varphi_*(X \cdot P)$ has the density[1]

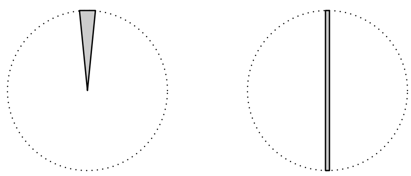
$$\frac{\pi}{2\sqrt{1-x^2}} \int X(x,y) \frac{1}{\pi} 1\!\!1_\Omega(x,y)\,\mathrm{d}y = \frac{1}{2\sqrt{1-x^2}} \int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} X(x,y)\,\mathrm{d}y$$

w.r.t. $\varphi_*(P)$. It means that

$$\mathbb{E}\big(X \,\big|\, \varphi = x\big) = \frac{1}{2\sqrt{1-x^2}} \int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}} X(x,y)\,\mathrm{d}y \quad \text{for } -1 < x < 1$$

(just the mean value on the section) for $X \ge 0$, and therefore for arbitrary X.

We observe another manifestation of the Borel-Kolmogorov paradox: by 9c19, the conditional density of y given $\theta = \pi/2$ is proportional to y, while by 9c20, the conditional density of y given $x = 0$ is constant.



As noted after 9c17, a condition of positive probability is self-sufficient. Now we see that a condition of zero probability is not. Being unable to divide by zero, we need a limiting procedure, involving a neighborhood of the given condition.

9c21 Exercise. Let $(\Omega, \mathcal{F}, Q) = (\Omega_1, \mathcal{F}_1, Q_1) \times (\Omega_2, \mathcal{F}_2, Q_2)$ (probability spaces), $P \ll Q$ another probability measure on $(\Omega, \mathcal{F})$, and $\varphi : \Omega \to \Omega_1$ the projection $\varphi(\omega_1, \omega_2) = \omega_1$. Then, on $(\Omega, \mathcal{F}, P)$, the conditioning is

$$\mathbb{E}\big(X \,\big|\, \varphi = \omega_1\big) = \int_{\Omega_2} \frac{f(\omega_1, \cdot)}{f_1(\omega_1)} X(\omega_1, \cdot)\,\mathrm{d}Q_2$$

where $f = \frac{\mathrm{d}P}{\mathrm{d}Q}$ and $f_1(\omega_1) = \int_{\Omega_2} f(\omega_1, \cdot)\,\mathrm{d}Q_2$.

Formulate it accurately, and prove.[2]

In this case we have conditional measures, and moreover, conditional densities (w.r.t. Q_2, not w.r.t. $Q_1 \times Q_2$).

[1] Indeed, if $\nu = f \cdot \mu$, $0 < f < \infty$, and $\xi = g \cdot \mu$, then $\mu = \frac{1}{f} \cdot \nu$ and so $\xi = g \cdot \big(\frac{1}{f} \cdot \nu\big) = \frac{g}{f} \cdot \nu$.

[2] Hint: similar to 9c20; what about $f_1(\omega_1) = 0$?

9c22 Exercise. Let $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ be probability spaces, $\alpha : \Omega_1 \to \Omega_2$ a measure preserving map, (E, S) a measurable space, and $\varphi : \Omega_2 \to E$ a measurable map from $(\Omega_2, \mathcal{F}_2)$ to (E, S). Then

$$\mathbb{E}\big(X \circ \alpha \,\big|\, \varphi \circ \alpha \big) = \mathbb{E}\big(X \,\big|\, \varphi \big)$$

for all $X \in L_1(P_2)$.

Prove it.

9c23 Exercise. Let the joint distribution $P_{X,Y}$ of two random variables X, Y be absolutely continuous (w.r.t. the two-dimensional Lebesgue measure m_2). Then

$$\mathbb{E}\big(Y \,\big|\, X = x \big) = \int y \, p_{Y|X=x}(y) \, \mathrm{d}y$$

where

$$p_{Y|X=x}(y) = \frac{p_{X,Y}(x,y)}{p_X(x)} \, , \quad p_X(x) = \int p_{X,Y}(x,y) \, \mathrm{d}y \, , \quad p_{X,Y} = \frac{\mathrm{d}P_{X,Y}}{\mathrm{d}m_2} \, .$$

Formulate it accurately, and prove.[1]

Back to the "great circle puzzle" of Sect. 9a. Suppose that a random point is distributed uniformly on the sphere. What is the conditional distribution on a given great circle?

This question cannot be answered without asking first, how is this great circle obtained from the random point.[2]

One case: there is a special (nonrandom) point ("the North Pole"), and we are given the great circle through the North Pole and the random point. Then the conditional density is $\frac{1}{2} \sin \theta$, where θ is the angle to the North Pole.

Another case: the given great circle is chosen at random among all great circles containing the random point. Equivalently: the "North Pole" is chosen at random, uniformly, independently of the random point. Then the conditional density is constant, $\frac{1}{2\pi}$.[3]

Having conditional measures, it is tempting to define conditional expectation of X as the integral w.r.t. the conditional measure, requiring just integrability of X w.r.t. almost all conditional measures (which is necessary and

[1] Hint: 9c21, 9c22.

[2] "... the term 'great circle' is ambiguous until we specify what limiting operation is to produce it. The intuitive symmetry argument presupposes the equatorial limit; yet one eating slices of an orange might presuppose the other." E.T. Jaynes (quote from Wikipedia).

[3] The proof involves the invariant measure on the group of rotations ("Haar measure").

not sufficient for unconditional integrability, since the conditional expectation of $|X|$ need not be integrable). Then, however, strange things happen. For example, it may be that $\mathbb{E}\big(X \,\big|\, \mathcal{F}_1\big) > 0$ a.s., but $\mathbb{E}\big(X \,\big|\, \mathcal{F}_2\big) < 0$ a.s. An example (sketch): $\mathbb{P}\big(X = n, Y = n+1\big) = \mathbb{P}\big(X = n+1, Y = n\big) = 0.5p^n(1-p)$ for $n = 0, 1, 2, \dots$; then $\mathbb{E}\big(a^Y \,\big|\, X = x\big) = \frac{pa+a^{-1}}{1+p} a^x$ for $x = 1, 2, \dots$; we take $ap > 1$ and get $\mathbb{E}\big(a^Y \,\big|\, X\big) > a^X$ a.s., but also $\mathbb{E}\big(a^X \,\big|\, Y\big) > a^Y$ a.s.[1] Would you prefer to gain a^X or a^Y in a game?

9d More on absolute continuity

9d1 Proposition. Let (X, S, μ) be a measure space, and ν a finite measure on (X, S). Then

$$\nu \ll \mu \quad \Longleftrightarrow \quad \forall \varepsilon > 0 \, \exists \delta > 0 \, \forall A \in S \, \big(\mu(A) < \delta \quad \Longrightarrow \quad \nu(A) < \varepsilon \big) .$$

Proof. "$\Longleftarrow$" is easy: $\mu(A) = 0$ implies $\forall \varepsilon \; \nu(A) < \varepsilon$.

"$\Longrightarrow$": Otherwise we have ε and $A_n \in S$ such that $\mu(A_n) \to 0$ but $\nu(A_n) \ge \varepsilon$. WLOG, $\sum_n \mu(A_n) < \infty$. Taking $B_n = A_n \cup A_{n+1} \cup \dots$ we have $\mu(B_n) \to 0$, $\nu(B_n) \ge \varepsilon$, and $B_n \downarrow B$ for some B. Thus, $\mu(B) = 0$, but $\nu(B) \ge \varepsilon$ (due to finiteness of ν), in contradiction to $\nu \ll \mu$. □

9d2 Proposition. Let (X, S, μ) be a measure space, $\mathcal{E} \subset S$ a generating algebra of sets, μ be $\mathcal{E}$-σ-finite,[2] and ν a finite measure on (X, S). Then

$$\nu \ll \mu \quad \Longleftrightarrow \quad \forall \varepsilon > 0 \, \exists \delta > 0 \, \forall E \in \mathcal{E} \, \big(\mu(E) < \delta \quad \Longrightarrow \quad \nu(E) < \varepsilon \big) .$$

Proof. "$\Longrightarrow$" follows easily from 9d1 (since $\mathcal{E} \subset S$).

"$\Longleftarrow$": By 9d1 it is sufficient to prove that $\mu(A) < \frac{1}{2}\delta \implies \nu(A) < 2\varepsilon$. Given $A \in S$ such that $\mu(A) < \frac{1}{2}\delta$, 7b4 applies to $\mu + \nu$ (think, why) giving $E \in \mathcal{E}$ such that $(\mu + \nu)(E \triangle A) < \min\big(\frac{1}{2}\delta, \varepsilon\big)$. Then $\mu(E) \le \mu(A) + \mu(E \triangle A) < \frac{1}{2}\delta + \frac{1}{2}\delta = \delta$, whence $\nu(E) < \varepsilon$ and $\nu(A) \le \nu(E) + \nu(E \triangle A) < \varepsilon + \varepsilon = 2\varepsilon$. □

In particular, we may take (X, S, μ) to be $\mathbb{R}$ (or $\mathbb{R}^d$) with Lebesgue measure (or arbitrary locally finite measure), and $\mathcal{E}$ the algebra generated by intervals (or boxes).

9d3 Definition. A continuous function $F : [a, b] \to \mathbb{R}$ is *absolutely continuous,* if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for every n and disjoint intervals $(a_1, b_1), \dots, (a_n, b_n) \subset [a, b]$,

$$\sum_{k=1}^{n} (b_k - a_k) < \delta \quad \Longrightarrow \quad \sum_{k=1}^{n} |F(b_k) - F(a_k)| < \varepsilon \, .$$

[1] Recall 1b1: $-\frac{1}{2} = \frac{1}{2} - 1 + 1 - 1 + \dots = +\frac{1}{2}$.

[2] As defined before 7b4.

9d4 Proposition. A finite nonatomic measure μ on $\mathbb{R}$ is absolutely continuous (w.r.t. Lebesgue measure) if and only if the function

$$F_\mu : x \mapsto \mu\big((-\infty, x]\big)$$

is absolutely continuous on every $[a, b]$.

9d5 Exercise. Prove Prop. 9d4.

9d6 Corollary. An increasing continuous function F on $[a, b]$ is absolutely continuous if and only if there exists $f \in L_1[a, b]$ such that $F(x) = \int_a^x f \, \mathrm{d}m$ for all $x \in [a, b]$.

Taking $F = F_\mu$ for μ of 3d5 we get a continuous but not absolutely continuous increasing function on $[0, 1]$.[1]

Index

[1] Known as "Cantor function", "Cantor ternary function", "Lebesgue's singular function", "the Cantor-Vitali function", "the Cantor staircase function" and even "the Devil's staircase", see Wikipedia.