Chapter 6 —Product Measures and Fubini (Capinski–Kopp)

This chapter provides the measure-theoretic foundation for joint distributions, independence, and iterated integration. It is essential for branching processes and continuous-time stochastic models.

Roadmap. Product σ-fields; construction of product measures; Fubini and Tonelli theorems; probability applications.

Product σ-fields.

Let $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ be measure spaces. Define

$$\Omega := \Omega_1 \times \Omega_2, \qquad \mathcal{R} := \{A_1 \times A_2 : A_1 \in \mathcal{F}_1,\ A_2 \in \mathcal{F}_2\}.$$

Definition 1. *The product σ-field is*

$$\mathcal{F} := \mathcal{F}_1 \otimes \mathcal{F}_2 := \sigma(\mathcal{R}).$$

Theorem 1. *$\mathcal{F}_1 \otimes \mathcal{F}_2$ is generated by the cylinder sets*

$$\{A_1 \times \Omega_2 : A_1 \in \mathcal{F}_1\} \quad \text{and} \quad \{\Omega_1 \times A_2 : A_2 \in \mathcal{F}_2\}.$$

Equivalently, it is the smallest σ-field on Ω such that the projections $\mathrm{Pr}_1(\omega_1, \omega_2) = \omega_1$ and $\mathrm{Pr}_2(\omega_1, \omega_2) = \omega_2$ are measurable.

The key observation is that

$$A_1 \times A_2 = (A_1 \times \Omega_2) \cap (\Omega_1 \times A_2),$$

so rectangles are intersections of cylinder sets.

Construction of the product measure.

Define on rectangles

$$P(A_1 \times A_2) := P_1(A_1) P_2(A_2).$$

Theorem 2. *If P_1 and P_2 are finite measures, then this definition extends uniquely to a measure on $\mathcal{F}_1 \otimes \mathcal{F}_2$.*

The extension is obtained by verifying countable additivity on finite unions of rectangles and applying a monotone class argument.

Fubini and Tonelli theorems.

Theorem 3 (Tonelli)**.** *If $f \geq 0$ is measurable on $\Omega_1 \times \Omega_2$, then*

$$\int f \, d(P_1 \otimes P_2) = \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2) \, dP_2 \right) dP_1.$$

Theorem 4 (Fubini)**.** *If $\int |f| \, d(P_1 \otimes P_2) < \infty$, the iterated integrals exist and the order of integration may be exchanged.*

Probability viewpoint.

For random variables (X, Y), the joint distribution is

$$\mu_{X,Y}(A) := P((X, Y) \in A), \qquad A \subset \mathbb{R}^2 \text{ Borel}.$$

Independence means $\mu_{X,Y} = \mu_X \otimes \mu_Y$. Conditioning corresponds to iterated integration with respect to the product measure.

Joint distributions.

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let

$$X : \Omega \to \mathbb{R}^m, \qquad Y : \Omega \to \mathbb{R}^n$$

be random variables. The pair (X,Y) defines a random variable with values in $\mathbb{R}^{m+n}$ by

$$\omega \mapsto (X(\omega), Y(\omega)).$$

Definition 2. *The* joint distribution *of (X,Y) is the probability measure on $\mathcal{B}(\mathbb{R}^{m+n})$ defined by*

$$\mu_{X,Y}(A) := P\big((X,Y) \in A\big), \qquad A \in \mathcal{B}(\mathbb{R}^{m+n}).$$

The marginal distributions are recovered by projection:

$$\mu_X(B) = \mu_{X,Y}(B \times \mathbb{R}^n), \qquad \mu_Y(C) = \mu_{X,Y}(\mathbb{R}^m \times C).$$

Expectation via joint laws.

If $g : \mathbb{R}^{m+n} \to \mathbb{R}$ is integrable with respect to $\mu_{X,Y}$, then

$$\mathbb{E}[g(X,Y)] = \int_{\mathbb{R}^{m+n}} g(x,y)\, d\mu_{X,Y}(x,y).$$

This is a direct application of integration with respect to image measures.

Independence.

Definition 3. *Random variables X and Y are called* independent *if*

$$\mu_{X,Y} = \mu_X \otimes \mu_Y.$$

Equivalently, for all Borel sets $A \subset \mathbb{R}^m$ and $B \subset \mathbb{R}^n$,

$$P(X \in A,\, Y \in B) = P(X \in A)\, P(Y \in B).$$

Independence therefore means that the joint law factors as a product measure.

Conditional viewpoint (informal).

Although conditional probability is not yet defined abstractly, the product measure formalism already suggests the idea of conditioning as integrating over one coordinate while keeping the other fixed. This intuition becomes precise later via conditional expectation.

Characteristic functions.

Definition 4. *Let $X : \Omega \to \mathbb{R}^n$ be a random variable. Its* characteristic function *is*

$$\varphi_X(t) := \mathbb{E}\big[e^{i\langle t,X\rangle}\big] = \int_{\mathbb{R}^n} e^{i\langle t,x\rangle}\, d\mu_X(x), \qquad t \in \mathbb{R}^n.$$

The characteristic function is always well-defined and bounded by 1 in absolute value.

If X and Y are independent, then

$$\varphi_{(X,Y)}(s,t) = \varphi_X(s)\, \varphi_Y(t),$$

reflecting the product structure of the joint distribution.

Characteristic functions determine distributions.

Theorem 5. *If two random variables X and Y in $\mathbb{R}^n$ satisfy*

$$\varphi_X(t) = \varphi_Y(t) \quad \textit{for all } t \in \mathbb{R}^n,$$

then $\mu_X = \mu_Y$.

Thus the characteristic function uniquely determines the probability law.

Chapter 7: Radon–Nikodym Theorem (Key Notes)
(7.1) Densities and Conditioning 核心：比较两个测度，密度 =Radon–Nikodym 导数

Definition 5 (Absolute continuity / Singularity)**.** *Let λ, μ be measures on $(\Omega, \mathcal{F})$.*

- $\lambda \ll \mu$ (absolutely continuous) *if $\mu(A) = 0 \Rightarrow \lambda(A) = 0$ for all $A \in \mathcal{F}$.*
- $\lambda \perp \mu$ (singular) *if there exists $N \in \mathcal{F}$ such that $\mu(N) = 0$ and $\lambda(\Omega \setminus N) = 0$.*

Definition 6 (Radon–Nikodym derivative (density))**.** *If $\lambda \ll \mu$ and there exists an $\mathcal{F}$-measurable $h \geq 0$ such that*

$$\lambda(A) = \int_A h\, d\mu, \qquad \forall A \in \mathcal{F},$$

then h is called a Radon–Nikodym derivative *of λ w.r.t. μ, denoted by*

$$h = \frac{d\lambda}{d\mu}.$$

It is unique μ-a.s.

Remark 1 (How to read $d\lambda = h\, d\mu$ 直观：换“基准测度”)**.** *The identity $\lambda(A) = \int_A h\, d\mu$ means: the measure λ can be obtained from μ by reweighting each infinitesimal mass element by the factor h.*

Proposition 1 (Basic properties of RN derivative)**.** *Assume $\lambda \ll \mu$.*

1. *If $\lambda(A) = \int_A h\, d\mu$, then for any nonnegative (or λ-integrable) measurable f,*

$$\int_\Omega f\, d\lambda = \int_\Omega fh\, d\mu.$$

2. *(Chain rule) If $\lambda \ll \mu \ll \nu$, then $\lambda \ll \nu$ and*

$$\frac{d\lambda}{d\nu} = \frac{d\lambda}{d\mu} \cdot \frac{d\mu}{d\nu} \qquad \nu\text{-a.s.}$$

3. *If $\lambda \perp \mu$, then $\frac{d\lambda}{d\mu} = 0$ μ-a.s. (any RN derivative is 0 a.s.).*

Proof (sketch). (1) Start with indicators $f = \mathbb{1}_A$, extend to simple functions by linearity, then to nonnegative measurable f by monotone convergence; for integrable signed f, write $f = f^+ - f^-$. (2) Plug in the identities $\lambda(A) = \int_A \frac{d\lambda}{d\mu}\, d\mu$ and $d\mu = \frac{d\mu}{d\nu}\, d\nu$, then apply (1). (3) If $\lambda \perp \mu$, then λ is supported on a μ-null set, hence any density must vanish μ-a.s. □

Theorem 6 (Lebesgue decomposition (finite measures) 分解成“绝对连续 + 奇异”)**.** *Let λ, μ be finite measures on $(\Omega, \mathcal{F})$. Then there exist unique measures λ_a, λ_s such that*

$$\lambda = \lambda_a + \lambda_s, \qquad \lambda_a \ll \mu, \quad \lambda_s \perp \mu.$$

Moreover, there exists a unique (up to μ-a.s.) measurable $h \geq 0$ such that

$$\lambda_a(A) = \int_A h\, d\mu, \qquad \forall A \in \mathcal{F}.$$

Remark 2 (你后续会用到：branching/Bessel 里“换测度、条件期望”都靠这一套)**.** *The decomposition tells you: any measure can be split into a part having a density w.r.t. μ and a part living on a μ-null set. This is the abstract version behind “density + singular part”.*

(7.2) The Radon–Nikodym Theorem 核心：$\lambda \ll \mu \Rightarrow$ 存在密度

Theorem 7 (Radon–Nikodym 最重要定理之一)**.** *Let μ be a σ-finite measure on $(\Omega, \mathcal{F})$ and let λ be a σ-finite measure such that $\lambda \ll \mu$. Then there exists an $\mathcal{F}$-measurable function $h \geq 0$ such that*

$$\lambda(A) = \int_A h \, d\mu, \qquad \forall A \in \mathcal{F}.$$

The function h is unique μ-a.s. and is denoted by $h = \frac{d\lambda}{d\mu}$.

Proof idea (finite case) 关键：在 $L^1(\mu)$ 里做"最大化/逼近". Assume first that μ, λ are finite. Consider the class

$$\mathcal{H} := \Big\{ f \geq 0 \text{ measurable} : \int_A f \, d\mu \leq \lambda(A) \ \ \forall A \in \mathcal{F} \Big\}.$$

Let $\alpha := \sup_{f \in \mathcal{H}} \int_\Omega f \, d\mu$ and choose $f_n \in \mathcal{H}$ such that $\int f_n \, d\mu \uparrow \alpha$. Define $h := \sup_n f_n$ (or $h = \lim_n f_n'$ after making it increasing). Then $h \in \mathcal{H}$ and $\int h \, d\mu = \alpha$.

Define a finite measure ν by $\nu(A) := \lambda(A) - \int_A h \, d\mu \geq 0$. We claim $\nu \equiv 0$. Otherwise there exists B with $\nu(B) > 0$. One can show (via standard approximation/normalization) that there exists $\varepsilon > 0$ and a set $C \subseteq B$ with $\mu(C) > 0$ such that adding $\varepsilon \mathbb{1}_C$ to h still keeps the inequality $\int_A (h + \varepsilon \mathbb{1}_C) \, d\mu \leq \lambda(A)$ for all A, contradicting maximality of h. Hence $\nu(A) = 0$ for all A and $\lambda(A) = \int_A h \, d\mu$.

For σ-finite measures, decompose Ω into countably many pieces of finite μ-measure and paste the densities together. □

Remark 3 (Uniqueness 为什么只到 μ-a.s.)**.** *If h_1, h_2 both satisfy $\lambda(A) = \int_A h_i \, d\mu$, then*

$$\int_A (h_1 - h_2) \, d\mu = 0, \ \forall A \ \Rightarrow \ h_1 = h_2 \ \mu\text{-a.s.}$$

(using $A = \{h_1 > h_2\}$ and $A = \{h_2 > h_1\}$).

(7.3) Lebesgue–Stieltjes Measures 核心：分布函数 ↔ 测度

Definition 7 (Distribution function)**.** *A function $F : \mathbb{R} \to \mathbb{R}$ is called a* distribution function *(c.d.f. style) if it is nondecreasing and right-continuous, with limits $F(-\infty) = \lim_{x \to -\infty} F(x)$ and $F(+\infty) = \lim_{x \to +\infty} F(x)$. (For probability laws one has $F(-\infty) = 0$, $F(+\infty) = 1$.)*

Theorem 8 (Construction of Lebesgue–Stieltjes measure 从 F 造 μ_F)**.** *Given such F, there exists a unique measure μ_F on $\mathcal{B}(\mathbb{R})$ such that*

$$\mu_F((a, b]) = F(b) - F(a) \qquad \text{for all } a < b.$$

Proof (standard, concise). Define μ_F first on the semi-ring of half-open intervals by $\mu_F((a, b]) = F(b) - F(a)$. Finite additivity follows from telescoping. Countable additivity for disjoint unions of intervals uses monotonicity and right-continuity of F (reduce to increasing unions and use continuity from above/below). Then extend to the Borel σ-field by Carathéodory extension theorem. Uniqueness holds since half-open intervals generate $\mathcal{B}(\mathbb{R})$. □

Remark 4 (Connection to random variables 概率论链接)**.** *For a real random variable X, its law $P_X = P \circ X^{-1}$ is a Lebesgue–Stieltjes measure. Its distribution function is $F_X(x) = P(X \leq x) = P_X((-\infty, x])$.*

Definition 8 (Absolutely continuous function)**.** *A function $F : [a, b] \to \mathbb{R}$ is* absolutely continuous *if for every $\varepsilon > 0$ there exists $\delta > 0$ such that for any finite disjoint collection of intervals (x_k, y_k) with $\sum_k (y_k - x_k) < \delta$, we have $\sum_k |F(y_k) - F(x_k)| < \varepsilon$.*

Theorem 9 (FTC for absolutely continuous functions 绝对连续 ⇒ 有密度)**.** *F is absolutely continuous on $[a, b]$ iff there exists $f \in L^1([a, b])$ such that*

$$F(x) = F(a) + \int_a^x f(t) \, dt \quad \text{for all } x \in [a, b].$$

Moreover, $F'(x) = f(x)$ for a.e. x.

Proof (outline). ($\Rightarrow$) Use bounded variation + Lebesgue differentiation theorem to obtain F' a.e., then show $F(x)-F(a)=\int_a^x F'(t)\,dt$ (via absolute continuity and approximation by difference quotients). ($\Leftarrow$) If $F(x)=F(a)+\int_a^x f$, then the absolute continuity estimate follows from $\sum|F(y_k)-F(x_k)|\le\int_{\cup(x_k,y_k)}|f|$ and absolute continuity of Lebesgue integral. $\square$

Definition 9 (Bounded variation and total variation)**.** *$g:[a,b]\to\mathbb{R}$ has* bounded variation *if*

$$V_a^b(g):=\sup_{\Pi}\sum_i|g(t_i)-g(t_{i-1})|<\infty,$$

where the supremum is over all partitions $\Pi:a=t_0<\cdots<t_n=b$.

Theorem 10 (Jordan decomposition for BV functions BV= 两个单调函数之差)**.** *g has bounded variation on $[a,b]$ iff there exist increasing functions g^+,g^- such that $g=g^+-g^-$. Moreover one can take*

$$g^+(x)=\frac{V_a^x(g)+g(x)-g(a)}{2},\qquad g^-(x)=\frac{V_a^x(g)-g(x)+g(a)}{2}.$$

Proof (concise). Define $v(x):=V_a^x(g)$; then v is increasing and $|g(x)-g(y)|\le v(x)-v(y)$. Set $g^\pm$ as above; both are increasing because they are sums/differences of increasing functions with the Lipschitz-type control from v. Direct algebra gives $g=g^+-g^-$. Conversely, difference of two increasing functions has bounded variation with $V_a^b(g)\le(g^+(b)-g^+(a))+(g^-(b)-g^-(a))$. $\square$

Definition 10 (Signed measure, total variation)**.** *A* signed measure *ν is a countably additive set function $\nu:\mathcal{F}\to\mathbb{R}$ with $\nu(\emptyset)=0$. Its* total variation *is the measure*

$$|\nu|(A):=\sup\Big\{\sum_i|\nu(A_i)|:\ A=\bigsqcup_i A_i,\ A_i\in\mathcal{F}\Big\}.$$

Theorem 11 (Hahn decomposition and Jordan decomposition 签名测度的正负部分)**.** *For a signed measure ν there exist disjoint measurable sets P,N with $P\cup N=\Omega$ such that $\nu(A)\ge0$ for all $A\subseteq P$ and $\nu(A)\le0$ for all $A\subseteq N$ (Hahn decomposition). Define*

$$\nu^+(A):=\nu(A\cap P),\qquad \nu^-(A):=-\nu(A\cap N).$$

Then ν^+,ν^- are measures, $\nu=\nu^+-\nu^-$, and this decomposition is unique up to null sets. Moreover $|\nu|=\nu^++\nu^-$.

Proof (sketch). Use standard Hahn decomposition theorem (construct maximal positive set via Zorn/approximation). Then define $\nu^\pm$ as above; countable additivity follows from that of ν. Uniqueness follows since any two Hahn decompositions differ only on a ν-null set, giving the same $\nu^\pm$. For total variation, refine partitions by intersecting with P and N to see $\sup\sum|\nu(A_i)|=\nu^+(A)+\nu^-(A)$. $\square$

(7.4) Probability 核心：条件期望 =Radon–Nikodym 导数；martingale 是离散时间 SDE 的骨架

Definition 11 (Conditional expectation w.r.t. a σ-field 定义 =RN 导数)**.** *Let $(\Omega,\mathcal{F},P)$ be a probability space, $\subseteq\mathcal{F}$ a sub-σ-field, and let $X\in L^1(P)$. The* conditional expectation $E[X\ |]$ *is any -measurable r.v. Y such that*

$$\int_A Y\,dP=\int_A X\,dP,\qquad \forall A\in.$$

It is unique P-a.s.

Existence/uniqueness (RN argument) 核心：把 $A\mapsto\int_A XdP$ 看成测度. Define a signed measure on $(\Omega,)$ by $\nu(A)=\int_A X\,dP$ for $A\in$. Then $\nu\ll P|$ since $P(A)=0\Rightarrow\nu(A)=0$. By Radon–Nikodym, there exists -measurable Y such that $\nu(A)=\int_A Y\,dP$ for all $A\in$, i.e. $Y=E[X\ |]$. Uniqueness is RN uniqueness: if Y_1,Y_2 both work then $\int_A(Y_1-Y_2)dP=0$ for all $A\in$ implies $Y_1=Y_2$ a.s. $\square$

Proposition 2 (Core properties of conditional expectation)**.** *Let $X,X_n\in L^1$ and $\subseteq\mathcal{F}$.*

1. *(Linearity) $E[aX+bZ\ |]=aE[X\ |]+bE[Z\ |]$.*

2. *(Monotonicity) If $X \le Z$ a.s., then $E[X \mid] \le E[Z \mid]$ a.s.*

3. *(Taking out what is known) If H is -measurable and bounded, then*

$$E[HX \mid] = H\,E[X \mid].$$

4. *(Tower property) If $\subseteq \ \subseteq \mathcal{F}$, then*

$$E\big[E[X \mid] \mid \ \text{"030}B] = E[X \mid].$$

5. *(Normalization) If X is -measurable then $E[X \mid] = X$ a.s.*

Proof (all by testing against $A \in$). Each identity follows from the defining property $\int_A E[\cdot \mid]\,dP = \int_A \cdot\,dP$ for all $A \in$, plus measurability and basic integral properties. For (3) use $\int_A HE[X \mid] = \int_A HX$ for all A. For (4) test against $A \in$ (hence $A \in$) and apply the defining property twice. □

Theorem 12 (Conditional Jensen 很常用)**.** *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be convex and $X \in L^1$ with $\varphi(X) \in L^1$. Then*

$$\varphi\big(E[X \mid]\big) \le E[\varphi(X) \mid] \qquad a.s.$$

Proof (supporting lines). For convex φ, we can write $\varphi(x) = \sup_{(a,b)\in\mathcal{S}}(ax+b)$ as the supremum of affine minorants. Then

$$\varphi(E[X \mid]) = \sup_{(a,b)}\big(aE[X \mid] + b\big) = \sup_{(a,b)} E[aX + b \mid] \le E\Big[\sup_{(a,b)}(aX+b)\Big|\Big] = E[\varphi(X) \mid].$$

□

Definition 12 (Filtration and adapted process)**.** *A* filtration *is an increasing family $(\mathcal{F}_n)_{n\ge 0}$ with $\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \mathcal{F}$. A process (X_n) is* adapted *if X_n is $\mathcal{F}_n$-measurable for each n.*

Definition 13 (Martingale / submartingale / supermartingale)**.** *Let (X_n) be adapted and integrable: $E|X_n| < \infty$.*

- *(X_n) is a* martingale *if $E[X_{n+1} \mid \mathcal{F}_n] = X_n$ a.s.*
- *It is a* submartingale *if $E[X_{n+1} \mid \mathcal{F}_n] \ge X_n$ a.s.*
- *It is a* supermartingale *if $E[X_{n+1} \mid \mathcal{F}_n] \le X_n$ a.s.*

Proposition 3 (Equivalent increment form 记忆版本)**.** *Let (X_n) be integrable and adapted. Then (X_n) is a martingale iff*

$$E[X_{n+1} - X_n \mid \mathcal{F}_n] = 0 \quad a.s.\ \text{for all } n.$$

证明. Immediate by linearity of conditional expectation. □

Theorem 13 (Optional stopping (bounded stopping time) 足够应付很多基础题)**.** *Let (X_n) be a martingale and let τ be a stopping time w.r.t. $(\mathcal{F}_n)$ such that $\tau \le N$ a.s. Then*

$$E[X_\tau] = E[X_0].$$

Proof (complete for bounded case). Define the stopped process $Y_n := X_{\tau\wedge n}$. Then Y_n is adapted and integrable. For each n,

$$E[Y_{n+1} \mid \mathcal{F}_n] = E[X_{\tau\wedge(n+1)} \mid \mathcal{F}_n] = \mathbb{1}_{\{\tau\le n\}}X_\tau + \mathbb{1}_{\{\tau>n\}}E[X_{n+1} \mid \mathcal{F}_n] = \mathbb{1}_{\{\tau\le n\}}X_\tau + \mathbb{1}_{\{\tau>n\}}X_n = Y_n,$$

so (Y_n) is a martingale. Taking expectations gives $E[Y_N] = E[Y_0]$. Since $\tau \le N$, $Y_N = X_\tau$ and $Y_0 = X_0$, hence $E[X_\tau] = E[X_0]$. □

Remark 5 (和 Bessel/branching 的关系)**.**

- ***Bessel SDE**: 常用到“换测度/密度/条件期望/鞅” (Girsanov、local martingale 等) 的语言；*
- ***Branching**: 生成函数/分布经常通过“测度与密度”表述；很多关键对象天然是 (次) 鞅。*