Probability: Theory and Examples

概率论：理论与例子

更新日期：2025 年 01 月 28 日

目录

1. 测度论

1.1. 测度空间

定义 1 (σ-域、σ-代数)：设 $\mathscr{F}$ 是集合 Ω 上的集合系，称 **$\mathscr{F}$ 是 Ω 上的 σ-域或 σ-代数**，如果
1. 对 $A \in \mathscr{F}$ 有 $A^c \in \mathscr{F}$；
2. 如果 $\{A_i \in \mathscr{F}\}$ 是 Ω 的可数个子集，那么 $\bigcup_i A_i \in \mathscr{F}$。

定义 2 (测度空间)：设 Ω 是集合，$\mathscr{F}$ 是 Ω 上的 σ-域，称二元组 $(\Omega, \mathscr{F})$ 是一个**测度空间**。

定义 3 (测度、概率测度)：设 $\mathscr{F}$ 是集合 Ω 上的集合系。

称 **μ 是集合系 $\mathscr{F}$ 上的测度**，如果
1. μ 是从 $\mathscr{F}$ 到 $\overline{\mathbb{R}}$ 的函数；
2. $\mu(A) \geqslant \mu(\varnothing) = 0$ 对 $\forall A \in \mathscr{F}$ 成立；
3. 对 $\mathscr{F}$ 中的可数无交集合列 $\{A_i\}$ 有

$$\mu\left(\bigcup_i A_i\right) = \sum_i \mu(A_i).$$

称 **μ 为集合系 $\mathscr{F}$ 上的概率测度**，如果 μ 是集合系 $\mathscr{F}$ 的测度且 $\mu(\Omega) = 1$。

定义 4 (概率空间)：设 Ω 是集合，$\mathscr{F}$ 是 Ω 上的 σ-域，P 是 $\mathscr{F}$ 上的概率测度，称三元组 $(\Omega, \mathscr{F}, P)$ 是一个**概率空间**。

命题 1 (测度的性质)：设 μ 是测度空间 $(\Omega, \mathscr{F})$ 上的测度。
1. μ 有单调性，即对 $A \subset B$ 有 $\mu(A) \leqslant \mu(B)$；
2. μ 有次可数可加性，即对 $A \subset \bigcup_{i=1}^{\infty} A_i$ 有

$$\mu(A) \leqslant \sum_{i=1}^{\infty} \mu(A_i),$$

3. μ 有下连续性，即对 $A_i \uparrow A$ 有 $\mu(A_i) \uparrow \mu(A)$；
4. μ 有上连续性，即对 $A_i \downarrow A$，如果 $\mu(A_1) < \infty$ 则有 $\mu(A_i) \downarrow \mu(A)$。

命题 2 (σ-域对交的封闭性)：设 $\{\mathscr{F}_i\}_{i \in I}$ 是 Ω 上的 σ-域，那么 $\bigcap_{i \in I} \mathscr{F}$ 也是 Ω 上的 σ-域。

定义 5 (生成的 σ-域): 设 Ω 是集合，$\mathscr{A}$ 是 Ω 上的集合系，由命题 2 知道

$$\bigcap \mathscr{F}, \quad \mathscr{F} \text{ 是包含 } \mathscr{A} \text{ 的 } \sigma\text{-域}$$

也是 σ-域，称为 **$\mathscr{A}$ 生成的 σ-域**，记作 $\sigma(\mathscr{A})$。

定义 6 (Borel 集): 称 $\mathbb{R}^d$ 上所有开集生成的 σ-域为 **Borel 集**，记作 $\mathscr{R}^d$。

定义 7 (半环、半代数): 设 Ω 是集合，$\mathscr{S}$ 是 Ω 上的集合系，称 **$\mathscr{S}$ 是 Ω 上的半代数或半环**，如果
1. 对 $\forall A, B \in \mathscr{S}$ 有 $A \cap B \in \mathscr{S}$；
2. 对 $\forall A \in \mathscr{S}$ 那么 A^c 可以表示成 $\mathscr{S}$ 中有限个元素的无交并。

定义 8 (代数、域): 设 Ω 是集合，$\mathscr{F}$ 是 Ω 上的集合系，称 **$\mathscr{F}$ 是 Ω 上的代数或域**，如果
1. 对 $\forall A, B \in \mathscr{F}$ 有 $A \cup B \in \mathscr{F}$；
2. 对 $\forall A \in \mathscr{F}$ 有 $A^c \in \mathscr{F}$。

命题 3 (代数对交的封闭性): 设 $\mathscr{F}$ 是 Ω 上的代数，那么 $\forall A, B \in \mathscr{F}$ 有 $A \cap B \in \mathscr{F}$。

命题 4 (σ-代数是代数): 设 $\mathscr{F}$ 是 Ω 上的 σ-代数，那么 $\mathscr{F}$ 是 Ω 上的代数。

命题 5 (半环可以生成环): 设 $\mathscr{S}$ 是 Ω 上的半环，那么

$$\overline{\mathscr{S}} = \left\{\mathscr{S} \text{ 上有限元素的无交并}\right\}$$

是 Ω 上的环。

定义 9 (生成的环): 设 $\mathscr{S}$ 是 Ω 上的半环，那么 $\overline{\mathscr{S}}$ 称为由 $\mathscr{S}$ 生成的环。

定义 10 (σ-有限): 设 $\mathscr{A}$ 是 Ω 上的集合系，μ 是 $\mathscr{A}$ 上的测度，称 μ 是 σ-有限的，如果存在可数集合列 $\{A_i \in \mathscr{A}\}$ 使得 $\mu(A_i) < \infty$ 且 $\bigcup_i A_i = \Omega$。

命题 6 (测度的扩张): 设 $\mathscr{S}$ 是半代数，μ 是 $\mathscr{S}$ 上的函数，满足 $\mu(\varnothing) = 0$. 如果
1. 对 $\mathscr{S}$ 中的有限个无交元素 $S_i \in \mathscr{S}$ 满足 $\bigsqcup_i S_i = S \in \mathscr{S}$ 均有
$$\sum_i \mu(S_i) = \mu(S),$$
2. 对 $\mathscr{S}$ 中的可数个无交元素 $S_i \in \mathscr{S}$ 满足 $\bigsqcup_i S_i = S \in \mathscr{S}$ 均有
$$\sum_i \mu(S_i) \leqslant \mu(S),$$

那么 μ 能够唯一地扩张到 $\mathscr{S}$ 生成的代数 $\overline{\mathscr{S}}$ 上。如果 $\overline{\mathscr{S}}$ 上的测度 $\overline{\mu}$ 还是 σ-有限的，那么它还能够唯一地扩张到 $\sigma(\mathscr{S})$ 上。

练习 1.1.1: Let $\Omega = \mathbb{R}$, $\mathscr{F} =$ all subsets so that A or A^c is countable, $P(A) = 0$ in the first case and $= 1$ in the second. show that $(\Omega, \mathscr{F}, P)$ is a probability space.

证明: 只要证明 $\mathscr{F}$ 是 Ω 上的 σ-域且 P 是 $\mathscr{F}$ 上的测度。

首先有 $\varnothing$ 可数，从而 $\varnothing, \mathbb{R} \in \mathscr{F}$。考虑 $\mathscr{F}$ 中的集合列 $\{A_i \in \mathscr{F}\}$。若存在 k 使得 A_k^c 可数，那么 $|\bigcup A_i| \geqslant |A_k|$ 从而 $\bigcup A_i$ 的补集可数；若不存在上述 A_k^c，那么所有的 A_i 均为可数集合，从而它们的可数并也可数。所以 $\mathscr{F}$ 是 Ω 上的 σ-域。

设 $\{A_i\}$ 是 $\mathscr{F}$ 上的可数个不交集合列。容易看出 $\{A_i\}$ 中存在至多一个不可数集 A_k，那么若 A_k 存在则 $P(\bigcup A_i) = \sum P(A_i) = 1$；否则 $P(\bigcup A_i) = \sum P(A_i) = 0$。 □

练习 1.1.2: Rescrl the definition of $\mathcal{S}_d$ from example 1.1.5. show that $\sigma(\mathcal{S}_d) = \mathscr{R}^d$, the borel subsets of $\mathbb{R}^d$.

证明: 对于 $\mathscr{R}^d$ 的一组拓扑基
$$\{(x_1, y_1) \times (x_2, y_2) \cdots \times (x_d, y_d) : x_i, y_i \in \mathbb{R}\}$$
中的一个开集
$$(x_1, y_2) \times \cdots \times (x_d, y_d),$$
有
$$(x_1, y_2) \times \cdots \times (x_d, y_d) = \bigcup_j \left(\left(x_1, y_1 - \frac{1}{j} \right] \times \cdots \times \left(x_d, y_d - \frac{1}{j} \right] \right) \in \mathcal{S}_d,$$
从而 $\sigma(\mathcal{S}_d) \subset \sigma\left(\mathscr{R}^d\right)$。对 $\mathcal{S}_d$ 中的拓扑基
$$(x_1, y_2] \times \cdots \times (x_d, y_d],$$
有

$$(x_1, y_2] \times \cdots \times (x_d, y_d] = \bigcap_j \left(\left(x_1, y_1 + \frac{1}{j} \right) \times \cdots \times \left(x_d, y_d + \frac{1}{j} \right) \right) \in \mathscr{R}_d,$$

从而 $\sigma\left(\mathscr{R}^d\right) \subset \sigma(\mathscr{S}_d)$。 $\square$

练习 1.1.3: A σ-field $\mathscr{F}$ is said to be countably generated if there is a countable collection $\mathscr{C} \subset \mathscr{F}$ so that $\sigma(\mathscr{C}) = \mathscr{F}$. Show that $\mathscr{R}^d$ is countably generated.

证明: 只要证明 $\mathscr{R}^d$ 可以由 $\{(a, b] : a, b \in \mathbb{Q}\}$ 生成，那么只要证明 $\{(a, b] : a, b \in \mathbb{R}\}$ 可以被它生成，而对于 $\forall (a, b]$ 为实数区间，一定存在单调下降的有理数列 $b_n \downarrow b$，从而

$$(a, b] = \bigcap_n (a_n, b_n].$$

那么 $(a, b]$ 可以被生成。 $\square$

练习 1.1.4:
1. Show that if $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots$ are σ-algebras, then $\bigcap_i \mathscr{F}_i$ is an algebra.
2. Give an example to show that $\bigcup_i \mathscr{F}_i$ need not be σ-algebra.

证明:
1. 设 $\mathscr{F}_i$ 是 Ω 上的 σ-代数，那么一定有 $\varnothing, \Omega \in \mathscr{F}_i$ 对每个 i 成立。对 $\forall A \in \bigcap_i \mathscr{F}_i$，一定有 $A^c \in \bigcap_i \mathscr{F}_i$。对 $\forall \{A_n\} \in \mathscr{F}_i$，一定有 $\bigcup_n A_n \in \mathscr{F}_i$。
2. 设 $\Omega = \mathbb{N}^+, \mathscr{F}_i$ 为由 $2^{\{1,2,\cdots,i\}}$ 生成的 σ-代数。那么这时有 $\bigcup_i \mathscr{F}_i = 2^{\mathbb{N}}$。令 A_i 为 $[1, i]$ 中所有偶数的集合，那么 $A_i \in \mathscr{F}_i$ 对所有 $i \in \mathbb{N}^+$ 成立，从而 $A_i \in \bigcup_n \mathscr{F}_n$。这时 $\bigcup_i A_i$ 为 $\mathbb{N}*+$ 中所有偶数的集合，但是其并不包含于 $\bigcup_i \mathscr{F}_i$，因为它是无限集，而后者的所有元素都是有限集。

$\square$

练习 1.1.5: A set $A \subset \{1, 2, \cdots\}$ is said to have asymptotic density θ if

$$\lim_{n \to \infty} \frac{|A \cap \{1, 2, \cdots, n\}|}{n} = \theta.$$

Let $\mathscr{A}$ be the collection of sets for which the asymptotic density exists. Is $\mathscr{A}$ a σ-algebra? an algebra?

证明: $\mathscr{A}$ 不是代数。令集合 A 为 N^+ 中所有奇数的集合，那么显然有 A 具有渐进密度 $1/2$。再构造具有渐进密度为 $1/2$ 的集合 B 如下：

$$B = \{1, 4, 5, 7, 10, 12, 14, 16, 17, 19, 21, 23, \cdots\},$$

B 在所有形如 $\left(2^{2k-1}, 2^{2k}\right]$ 的区间上取所有奇数，在所有形如 $\left(2^{2k}, 2^{2k+1}\right]$ 的区间上取所有偶数。这时可以证明 $A \cup B$ 不具有渐进密度，因为在前 2^{2k-1} 个数上的密度小于 $3/4$，而在前 2^{2k} 个数上的密度大于 $3/4$。 $\square$

1.2. 分布

练习 1.2.1: Suppose X and Y are random variables on $(\Omega, \mathscr{F}, P)$ and let $A \in \mathscr{F}$. Show that if we let $Z(\omega) = X(\omega)$ for $\omega \in A$ and $Z(\omega) = Y(\omega)$ for $\omega \in A^c$, then Z is a random variable.

证明: 对任意 *Borel* 集 $B \in \mathscr{R}$，

$$\begin{aligned} Z^{-1}(B) &= \left(Z^{-1}(B) \cap A\right) \cup \left(Z^{-1}(B) \cap A^c\right) \\ &= \left(X^{-1}(B) \cap A\right) \cup \left(Y^{-1}(B) \cap A^c\right) \\ &\in \mathscr{F}. \end{aligned}$$

从而 Z 也是随机变量。 $\square$

练习 1.2.2: Let χ have the standard normal distribution. Use Theorem 1.2.6 to get upper and lower bounds on $P(\chi \geqslant 4)$.

证明:

$$P(\chi \geqslant 4) = \frac{1}{\sqrt{2\pi}} \int_4^\infty \exp\left(\frac{x^2}{2}\right) \mathrm{d}x \leqslant \frac{1}{4\sqrt{2\pi}} \exp(8),$$

$$P(\chi \geqslant 4) = \frac{1}{\sqrt{2\pi}} \int_4^\infty \exp\left(\frac{x^2}{2}\right) \mathrm{d}x \geqslant \left(\frac{1}{4} - \frac{1}{64}\right) \frac{1}{\sqrt{2\pi}} \exp(8).$$

$\square$

练习 1.2.3: Show that a distribution function has at most countably many discontinuities.

证明: 由于分布函数是单调有界函数，从数学分析相关知识知道其不连续点一定都是跳跃间断点，且不连续点至多可数。 $\square$

练习 1.2.4: Show that if $F(x) = P(X \leqslant x)$ is continuous then $Y = F(X)$ has a uniform distribution on $(0, 1)$, that is, if $y \in [0, 1]$, $P(Y \leqslant y) = y$.

证明: 对 $\forall y \in [0, 1]$，

$$P(Y \leqslant y) = P(F(X) \leqslant y) = P\left(X \leqslant F^{-1}(y)\right) = F\left(F^{-1}(y)\right) = y.$$

$\square$

练习 1.2.5: Suppose X has continuous density f, $P(\alpha \leqslant X \leqslant \beta) = 1$ and g is a function that is strictly increasing and differentiable on (α, β). Then $g(X)$ has density

$$\frac{f(g^{-1}(y))}{g'(g^{-1}(y))}$$

for $y \in (g(\alpha), g(\beta))$ and 0 otherwise. When $g(x) = ax + b$ with $a > 0$,

$$g^{-1}(y) = \frac{y-b}{a}$$

so the answer is

$$\frac{1}{a} f\left(\frac{y-b}{a}\right).$$

证明: 对 $\forall y \in [\alpha, \beta]$，有

$$P(g(X) \leqslant x) = P(X \leqslant g^{-1}(x)) = \int_{\alpha}^{g^{-1}(x)} f(t)\,\mathrm{d}t\,.$$

令 $s = g^{-1}(t)$ 换元得到

$$P(g(X) \leqslant x) = \int_{\alpha}^{g^{-1}(x)} f(t)\,\mathrm{d}t = \int_{g(\alpha)}^{x} \frac{f(g^{-1}(s))}{g'(g^{-1}(s))}\,\mathrm{d}s\,.$$

□

练习 1.2.6: Suppose X has a normal distribution. Use the previous exercise to compute the density of $\exp(X)$. (The answer is scrled the lognormal distribution.)

证明: 令 $g(X) = \exp(X), g'(X) = \exp(X), g^{-1}(y) = \log(y)$ 得到 $\exp(X)$ 的概率密度函数

$$\frac{f(g^{-1}(y))}{g'(g^{-1}(y))} = \frac{1}{\sqrt{2\pi}} \frac{\exp(-\frac{1}{2}(\log y)^2)}{\exp(\log(y))} = \frac{1}{\sqrt{2\pi}y} \exp\left(-\frac{1}{2}\log^2 y\right).$$

□

练习 1.2.7:

1. Suppose X has density function f. Compute the distribution function of X^2 and then differentiate to find its density function.
2. Work out the answer when X has a standard normal distribution to find the density of the chi-square distribution.

证明:
1. 记 X^2 的分布函数为 $G(x)$，那么

$$G(x) = P(X^2 \leqslant x) = P\big(-\sqrt{x} \leqslant X \leqslant \sqrt{x}\big) = \int_{-\sqrt{x}}^{\sqrt{x}} f(t)\,\mathrm{d}t\,.$$

即为其分布函数。记其概率密度函数为 g，则

$$g(x) = \frac{\mathrm{d}G}{\mathrm{d}x} = \frac{f\big(\sqrt{x}\big) + f\big(-\sqrt{x}\big)}{2\sqrt{x}}.$$

2. 代入正态分布概率密度函数得到

$$f_{\chi^2}(x) = \frac{1}{\sqrt{2\pi x}}\exp\Big(-\frac{x}{2}\Big).$$

□

1.3. 随机变量

练习 1.3.1: Show that if $\mathscr{A}$ generates $\mathcal{S}$, then $X^{-1}(\mathscr{A}) \equiv \{\{X \in A\} : A \in \mathscr{A}\}$ generates

$$\sigma(X) = \{\{X \in B\} : B \in \mathcal{S}\}.$$

证明: 因为 $X^{-1}(\mathscr{A}) \subset \sigma(X)$，从而 $\sigma\big(X^{-1}(\mathscr{A})\big) \subset \sigma(X)$。对于任意包含 $X^{-1}(\mathscr{A})$ 的 σ-域 $\mathscr{B}$，$X(\mathscr{B})$ 是包含 $\mathscr{A}$ 的 σ-域，从而 $\mathscr{A} \subset \mathcal{S} \subset X(\mathscr{B})$，那么 $X^{-1}(\mathscr{A}) \subset \sigma(X) \subset \mathscr{B}$，那么 $\sigma\big(X^{-1}(\mathscr{A})\big) \subset \sigma(X) \subset \mathscr{B}$，那么 $\sigma(X)$ 是包含 $X^{-1}(\mathscr{A})$ 的最小 σ-域，从而 $\sigma(X) = \sigma\big(X^{-1}(\mathscr{A})\big)$. □

练习 1.3.2: Prove Theorem 1.3.6 when $n = 2$ by checking

$$\{X_1 + X_2 < x\} \in \mathscr{F}.$$

证明: 因为 $\{(-\infty, x) : x \in \mathbb{R}\}$ 可以生成 $\mathscr{R}$，所以只要证明 $\{X_1 + X_2 < x\} \in \mathscr{F}$。容易看出 $\{x' = (x_1, x_2) \in \mathbb{R}^2 : x_1 + x_2 < x\} \in \mathscr{R}^2$，从而 $f(x_1, x_2) = x_1 + x_2$ 是可测函数，那么 $\{X_1 + X_2 < x\} \in \mathscr{F}$。 □

练习 1.3.3: Show that if f is continuous and $X_n \to X$ almost surely then $f(X_n) \to f(X)$ almost surely.

证明: 设 $X_n \to X$ 在 Ω' 上满足，其中 $P(\Omega') = 1$。那么对 $\forall \omega \in \Omega'$ 有 $X_n(\omega) \to X(\omega)$，又知道 f 连续，从而 $f(X_n(\omega)) \to f(x(\omega))$。那么 $f(X_n) \to f(X)$ 在 Ω' 上成立。 □

练习 1.3.4:

1. Show that a continuous function from $\mathbb{R}^d \to \mathbb{R}$ is a measurable map from $\left(\mathbb{R}^d, \mathscr{R}^d\right)$ to $(\mathbb{R}, \mathscr{R})$.
2. Show that $\mathscr{R}^d$ is the smallest σ-field that makes all the continuous functions measurable.

证明:
1. 设 $f:\mathbb{R}^n \to \mathbb{R}$ 是连续函数，那么因为 $\{(-\infty, x)\}$ 生成 $\mathscr{R}$，只要对 $\forall x \in \mathbb{R}$，证明 $f^{-1}((-\infty, x)) \in \mathscr{R}^d$ 即可。因为 f 连续，那么 $f^{-1}((-\infty, x))$ 为开集，从而是 *Borel* 集。
2. 只要证明

$$\mathscr{R}^d = \sigma\left(\left\{f^{-1}((-\infty, x)) : x \in \mathbb{R}, f \in C^{\infty}\left(\mathbb{R}^d\right)\right\}\right).$$

由上一问已经知道

$$\sigma\left(\left\{f^{-1}((-\infty, x)) : x \in \mathbb{R}, f \in C^{\infty}\left(\mathbb{R}^d\right)\right\}\right) \subset \mathscr{R}^d.$$

考虑任意包含 $\left\{f^{-1}((-\infty, x)) : x \in \mathbb{R}, f \in C^{\infty}\left(\mathbb{R}^d\right)\right\}$ 的 σ-域 $\mathscr{A}$。取定 $\mathscr{R}^d$ 的一组生成元 $\{(-\infty, x_1) \times \cdots \times (-\infty, x_d) : x_1, \cdots, x_d \in \mathbb{R}\}$，取其中的一个元素为 $(-\infty, x_1) \times \cdots \times (-\infty, x_d)$。那么可以构造连续函数

$$f(\boldsymbol{x}) = \operatorname{dis}(\boldsymbol{x}, (-\infty, x_1) \times \cdots \times (-\infty, x_d)),$$

其中 dis 表示距离函数。那么有 $f^{-1}((-\infty, 0)) = (-\infty, x_1) \times \cdots \times (-\infty, x_d)$，从而 $\mathscr{R}^d \subset \mathscr{A}$。那么命题得证。

□

练习 1.3.5: A function f is said to be lower semicontinuous or l.s.c. if

$$\liminf_{y \to x} f(y) \geqslant f(x)$$

and upper semicontinuous (u.s.c.) if $-f$ is l.s.c. Show that f is l.s.c. if and only if $\{x : f(x) \leqslant a\}$ is closed for each $a \in \mathbb{R}$ and conclude that semicontinuous functions are measurable.

证明: 先证必要性，假设 f 是下半连续的，只要对 $\forall a \in \mathbb{R}$，证 $\{x : f(x) > a\}$ 是开集。从中任取 x，那么有

$$\liminf_{y \to x} f(y) \geqslant f(x) > a,$$

从而存在 x 的邻域 $U(x, \delta(x))$ 使得对 $\forall y \in (x, \delta(x))$ 有

$$f(y) \geqslant \frac{f(x) + a}{2} > a,$$

从而 x 是 $\{x : f(x) > a\}$ 的内点，那么 $\{x : f(x) > a\}$ 是开集。

再证充分性。假设对 $\forall a \in \mathbb{R}$，$\{x : f(x) \leqslant a\}$ 是闭集，从而 $\{x : f(x) > a\}$ 是开集。$\forall x_0 \in \mathbb{R}$，考虑令 $y \to x_0$。对 $\forall a < f(x_0)$，知道 $f(x_0)$ 附近存在邻域 U 使得 $\forall y \in U$ 有 $f(y) > a$。令 $a \uparrow f(x_0)$ 即得到 $\liminf f(y) \geqslant f(x_0)$。

从而上半连续函数都是可测函数，类似可以证明下半连续函数都是可测函数。
□

练习 1.3.6: Let $f : \mathbb{R}^d \to \mathbb{R}$ be an arbitrary function and let

$$f^\delta(\boldsymbol{x}) = \sup\{f(\boldsymbol{y}) : |\boldsymbol{y} - \boldsymbol{x}| < \delta\}$$

and

$$f_\delta(\boldsymbol{x}) = \inf\{f(\boldsymbol{y}) : |\boldsymbol{y} - \boldsymbol{x}| < \delta\}$$

where

$$|\boldsymbol{z}| = \sqrt{\boldsymbol{z}_1^2 + \cdots + \boldsymbol{z}_d^2}.$$

Show that f^δ is l.s.c. and f_δ is u.s.c. Let $f^0 = \lim_{\delta\downarrow 0} f^\delta$, $f_0 = \lim_{\delta\downarrow 0} f_\delta$, and conclude that the set of points at which f is discontinuous $= \left\{f^0 \neq f_0\right\}$ is measurable.

follows from the fact that $f^0 - f_0$ is.

证明: 直接按定义验证得到 f^δ 是上半连续函数而 f_δ 是下半连续函数。接下来只要证明 f 的不连续点集就是 $\left\{f^0 \neq f_0\right\}$。

假设 $\boldsymbol{x} \in \mathbb{R}^d$ 是 f 的不连续点，那么存在 $\varepsilon_0 > 0$，存在收敛到 $\boldsymbol{x}$ 的点列 $\{\boldsymbol{x}_n'\}$ 和 $\{\boldsymbol{x}_n''\}$ 使得 $|\lim f(\boldsymbol{x}_n') - \lim f(\boldsymbol{x}_n'')| < \varepsilon_0$。不妨设

$$\lim f(\boldsymbol{x}_n') > \lim f(\boldsymbol{x}_n'') + \varepsilon_0.$$

对 $\forall \delta > 0$，总能找到 $\boldsymbol{x}_{n_0}', \boldsymbol{x}_{n_0}''$ 使得

$$|\boldsymbol{x}_{n_0}' - \boldsymbol{x}|, |\boldsymbol{x}_{n_0}'' - \boldsymbol{x}| < \delta$$

且

$$f\left(\boldsymbol{x}_{n_0}'\right) \geqslant f\left(\boldsymbol{x}_{n_0}''\right) + \frac{\varepsilon_0}{2},$$

从而一定有

$$f^\delta(\boldsymbol{x}) \geqslant f\left(\boldsymbol{x}_{n_0}'\right) \geqslant f\left(\boldsymbol{x}_{n_0}''\right) + \frac{\varepsilon_0}{2} \geqslant f_\delta(\boldsymbol{x}).$$

令 $\delta \downarrow 0$ 得到 $f^0(\boldsymbol{x}) \geqslant f_0(\boldsymbol{x}) + \varepsilon_0/2$。

再假设 $\boldsymbol{x}_0$ 是 f 的连续点，那么对 $\forall \varepsilon > 0$，存在 $\delta > 0$ 使得 $\forall \boldsymbol{x} \in U(\boldsymbol{x}_0, \delta)$ 有 $|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)| < \varepsilon$，从而

$$\left|f^\delta - f_\delta\right| \leqslant \left|f^\delta - f(\boldsymbol{x}_0)\right| + \left|f_\delta - f(\boldsymbol{x}_0)\right| < 2\varepsilon.$$

令 $\varepsilon \to 0$ 即可知道 $f^0 = f_0$。 $\square$

练习 1.3.7: A function $\varphi : \Omega \to \mathbb{R}$ is said to be simple if

$$\varphi(\omega) = \sum_{m=1}^{n} c_m 1_{A_m}(\omega)$$

where the c_m are real numbers and $A_m \in \mathscr{F}$. Show that the class of $\mathscr{F}$ measurable functions is the smallest class containing the simple functions and closed under pointwise limits.

证明: 设 P 是由从 Ω 到 $\mathbb{R}$ 上的部分函数构成的集合，满足所有的简单函数都属于 P 且对点态极限运算封闭。那么只要证明所有的 $\mathscr{F}$ 可测函数都在 P 中。对 $\forall f : \Omega \to \mathbb{R}$ 为 $\mathscr{F}$ 可测函数，由 $f = f^+ - f^-$ 可以不妨设 $f \geqslant 0$。那么直接构造简单函数列

$$\varphi_n(\omega) = \begin{cases} n, & f(\omega) \geqslant n \\ \dfrac{k}{2^n}, & \dfrac{k}{2^n} \leqslant f(\omega) < \dfrac{k+1}{2^n}, k = 0, 1, \cdots, n2^n - 1 \end{cases}.$$

容易看出 φ_n 点态收敛于 f，从而 $f \in P$。 $\square$

练习 1.3.8: Use the previous exercise to conclude that Y is measurable with respect to $\sigma(X)$ if and only if $Y = f(X)$ where $f : \mathbb{R} \to \mathbb{R}$ is measurable.

证明: 充分性是显然的，下证明必要性。设 Y 在 $\sigma(X)$ 上可测。那么由上一问题知道 Y 可以写成 $\sigma(X)$ 上的简单函数的点态极限，记为 $Y = \lim \varphi_n$。同样可以将 X 写为 $X = \lim \psi_n$。记

$$\varphi_n = \sum_{m=1}^{a_n} c_{n,m} 1_{A_{n,m}},$$
$$\psi_n = \sum_{m=1}^{a_n} d_{n,m} 1_{A_{n,m}},$$

可以不妨设 $\left\{d_{n,m}\right\}_m$ 两两不同。那么令

$$f_n : \mathbb{R} \to \mathbb{R}, f\left(d_{n,m}\right) = c_{n,m}, \text{否则 } f = 0.$$

那么就有 $f_n(\psi_n) = \varphi_n$。令 $n \to \infty$，则有 $f(X) = Y$，且 f 为可测函数。 $\square$

练习 1.3.9: To get a constructive proof of the last result, note that

$$\{\omega : m2^{-n} \leqslant Y < (m+1)2^{-n}\} = \left\{X \in B_{m,n}\right\}$$

for some $B_{m,n} \in \mathscr{R}$ and set $f_n(x) = m2^{-n}$ for $x \in B_{m,n}$ and show that as $n \to \infty$ $f_n(x) \to f(x)$ and $Y = f(X)$.

证明：同上题。 □

1.4. 积分

练习 1.4.1：Show that if $f \geqslant 0$ and $\int f \, d\mu = 0$ then $f = 0$ a.e.

证明：设 $A = \{f > 0\} = \bigcup_n \{f > 1/n\}$。若 $\mu(A) > 0$，那么由测度的连续性知道，存在足够大的 $n_0 \in \mathbb{N}^*$ 使得 $\mu(\{f > 1/n\}) > 0$。那么

$$
\begin{aligned}
\int f \, \mathrm{d}\mu &= \int_A f \, \mathrm{d}\mu + \int_{A^c} f \, \mathrm{d}\mu \\
&= \int_A f \, \mathrm{d}\mu \\
&= \int_{\{f > \frac{1}{n}\}} f \, \mathrm{d}\mu \\
&\geqslant \int_{\{f > \frac{1}{n}\}} \frac{1}{n} \, \mathrm{d}\mu \\
&= \frac{1}{n} \mu\left(\left\{f > \frac{1}{n}\right\}\right) \\
&> 0,
\end{aligned}
$$

从而矛盾。 □

练习 1.4.2：Let $f \geqslant 0$ and

$$E_{n,m} = \left\{x : \frac{m}{2^n} \leqslant f(x) < \frac{m+1}{2^n}\right\}.$$

As $n \uparrow \infty$,

$$\sum_{m=1}^{\infty} \frac{m}{2^n} \mu(E_{n,m}) \uparrow \int f \, d\mu.$$

证明：不妨设 $\lambda = \sum_m \mu\left(E_{n,m}\right)$ 有限，否则命题化为 $\infty \uparrow \infty$ 显然成立。那么有

$$
\begin{aligned}
&\int f \,\mathrm{d}\mu - \sum_{m=1}^{\infty} \frac{m}{2^n}\mu(E_{n,m}) \\
&= \sum_{m=1}^{\infty} \int_{E_{n,m}} \left(f - \frac{m}{2^n}\right) \mathrm{d}\mu \\
&\leqslant \sum_{m=1}^{\infty} \int_{E_{n,m}} \frac{1}{2^n} \,\mathrm{d}\mu \\
&\leqslant \sum_{m=1}^{\infty} \frac{\mu(E_{n,m})}{2^n} \\
&= \frac{\lambda}{2^n} \downarrow 0 \quad (n \to \infty).
\end{aligned}
$$

从而命题得证。 □

练习 1.4.3: Let g be an integrable function on $\mathbb{R}$ and $\varepsilon > 0$.

1. Use the definition of the integral to conclude there is a simple function $\varphi = \sum_k b_k 1_{A_k}$ with

$$\int |g - \varphi| \, d\mu < \varepsilon.$$

2. Use Exercise A.2.1 to approximate the A_k by finite unions of intervals to get a step function

$$q = \sum_{j=1}^{k} c_j 1_{(a_{j-1}, a_j)}$$

with $a_0 < a_1 < \cdots < a_k$, so that

$$\int |\varphi - q| < \varepsilon.$$

3. Round the corners of q to get a continuous function r so that

$$\int |q - r| \, d\mu < \varepsilon.$$

To make a continuous function replace each $c_j 1_{(a_{j-1}, a_j)}$ by a function that is 0 on $(a_{j-1}, a_j)^c$, c_j on $[a_{j-1} + \delta_j, a_j - \delta_j]$, and linear otherwise. If the δ_j are small enough and we let

$$r(x) = \sum_{j=1}^{k} r_j(x)$$

then

$$\int |q(x)-r(x)|\,d\mu=\sum_{j=1}^{k}\delta_j c_j<\varepsilon.$$

证明:

1. 由于 $g=g^+-g^-$ 不妨先证明 g^+，然后类似证明 g^-。由积分的定义知道存在可测函数 h^+，使得 $0\leqslant h^+\leqslant g$，$h^+$ 有界，且支集测度有限，且

$$\int |g-h^+|\,\mathrm{d}\mu<\frac{\varepsilon}{4}.$$

对 h^+，再由积分定义知道存在非负简单函数 $\varphi^+\leqslant h^+$，使得

$$\int |h^+-\varphi^+|\,\mathrm{d}\mu<\frac{\varepsilon}{4}.$$

那么

$$\int |g^+-\varphi^+|\,\mathrm{d}\mu<\int |g^+-h^+|\,\mathrm{d}\mu+\int |h^+-\varphi^+|\,\mathrm{d}\mu<\frac{\varepsilon}{2}.$$

类似得到 φ^- 满足

$$\int |g^--\varphi^-|\,\mathrm{d}\mu<\int |g^--h^-|\,\mathrm{d}\mu-\int |h^--\varphi^-|\,\mathrm{d}\mu<\frac{\varepsilon}{2}.$$

令 $\varphi=\varphi^+-\varphi^-$ 直接得到结果。

2. 记

$$\varphi=\sum_{i=1}^{n}b_i 1_{A_i}.$$

对每个 A_i，由测度近似定理知道存在有限个开区间 $B_{i,1},\cdots,B_{i,b_i}$ 使得对 $B_i=\bigcup_j B_{i,j}$ 有 $A_i\subset B_i$ 且

$$\mu(B_i\setminus A_i)<\frac{\varepsilon}{b_i n}.$$

那么就有

$$\varphi=\sum_{i=1}^{n}\sum_{j=1}^{b_i}b_i 1_{B_{i,j}},$$

重新排列就能够得到欲证命题。

□

练习 1.4.4: Prove the Riemann-Lebesgue lemma. If g is integrable then

$$\lim_{n\to\infty}\int g(x)\cos nx\,\mathrm{d}x=0.$$

Hint: If g is a step function, this is easy. Now use the previous exercise.

证明: 先证明对于任意开区间 (l,r) 成立

$$\lim_{n\to\infty}\int_{(l,r)}\cos nx\,\mathrm{d}x=0.$$

考虑到 $\cos nx$ 是以 $2/n$ 为周期的周期函数，且在一个周期内的积分是 0，从而知道

$$\left|\int_{(l,r)}\cos nx\,\mathrm{d}x\right|=\left|\int_{(l,l+\delta)}\cos nx\,\mathrm{d}x\right|\leqslant\int_{(l,l+\delta)}|\cos nx|\,\mathrm{d}x\leqslant\delta,$$

其中 $\delta<2/n$. 令 $n\to\infty$ 就得到积分趋近于 0. 那么对任意的 *step function* g，记

$$g=\sum_{i=1}^{k}c_i1_{(a_i,b_i)},$$

那么就有

$$\lim_{n\to\infty}\int g(x)\cos nx\,\mathrm{d}x=\sum_{i=1}^{k}c_i\lim_{n\to\infty}\int_{(a_i,b_i)}\cos nx\,\mathrm{d}x=0.$$

从而原命题在 g 为 *step function* 时成立。对于 g 为任意可积函数，$n\in\mathbb{N}^*$ 的情况，存在 *step function* 构成的函数列 $\{h_k\}$ 使得

$$\begin{aligned}&\int|g(x)\cos nx-h_k(x)\cos nx|\,\mathrm{d}x\\ \leqslant&\int|g(x)-h_k(x)|\,\mathrm{d}x\\ \to&\,0\quad(k\to\infty).\end{aligned}$$

从而有

$$\begin{aligned}&\left|\lim_{n\to\infty}\int g(x)\cos nx\,\mathrm{d}x\right|\\ \leqslant&\lim_{n\to\infty}\int|g(x)\cos nx|\,\mathrm{d}x\\ \leqslant&\lim_{n\to\infty}\lim_{k\to\infty}\int|g(x)\cos nx-h_k(x)\cos nx|+\lim_{n\to\infty}\lim_{k\to\infty}\int|h_k(x)\cos nx|\,\mathrm{d}x\\ \leqslant&\,0+0\\ =&\,0.\end{aligned}$$

□

1.5. 积分的性质

练习 1.5.1: Let

$$\|f\|_\infty = \inf\{M : \mu(\{x : |f(x)| > M\}) = 0\}.$$

Prove that

$$\int |fg| \,\mathrm{d}\mu \leqslant \|f\|_1 \, \|g\|_\infty.$$

证明:

$$\begin{aligned}
\int |fg| \,\mathrm{d}\mu &\leqslant \int |f| \, |g| \,\mathrm{d}\mu \\
&\leqslant \int |f| \, \|g\|_\infty \,\mathrm{d}\mu \\
&= \|g\|_\infty \, \mu(\Omega) \int |f| \,\mathrm{d}\mu \\
&= \|f\|_1 \, \|g\|_\infty.
\end{aligned}$$

□

练习 1.5.2: Show that if μ is a probability measure then

$$\|f\|_\infty = \lim_{p\to\infty} \|f\|_p.$$

证明: 对任意的 $\varepsilon > 0$，有

$$\begin{aligned}
\|f\|_p &= \left[\int |f|^p \,\mathrm{d}\mu\right]^{\frac{1}{p}} \\
&= \left[\int_{\{|f|^p \,>\, \|f\|_\infty \,-\varepsilon\}} |f|^p \,\mathrm{d}\mu\right]^{\frac{1}{p}} \\
&\geqslant [\mu(\{|f|^p > \|f\|_\infty - \varepsilon\})(\|f\|_\infty - \varepsilon)^p]^{\frac{1}{p}} \\
&= [\mu(\{|f|^p > \|f\|_\infty - \varepsilon\})]^{\frac{1}{p}}(\|f\|_\infty - \varepsilon).
\end{aligned}$$

令 $p \to \infty$，再令 $\varepsilon \to 0$ 得到 $\|f\|_p \geqslant \|f\|_\infty$。

另一方面有

$$
\begin{aligned}
\|f\|_p &= \left[\int_{\{|f|^p > \|f\|_\infty\}} |f|^p \,\mathrm{d}\mu + \int_{\{|f|^p \leqslant \|f\|_\infty\}} |f|^p \,\mathrm{d}\mu\right]^{\frac{1}{p}} \\
&= \left[\int_{\{|f|^p \leqslant \|f\|_\infty\}} |f|^p \,\mathrm{d}\mu\right]^{\frac{1}{p}} \\
&\leqslant \left[\int_{\{|f|^p \leqslant \|f\|_\infty\}} \|f\|^\infty \,\mathrm{d}\mu\right]^{\frac{1}{p}} \\
&= \|f\|^\infty.
\end{aligned}
$$

□

练习 1.5.3:

1. Suppose $p \in (1, \infty)$. The inequality

$$|f+g|^p \leqslant 2^p(|f|^p + |g|^p)$$

shows that if $\|f\|_p$ and $\|g\|_p$ are $< \infty$ then

$$\|f+g\|_p < \infty.$$

Apply Hölder's inequality to

$$|f||f+g|^{p-1}$$

and

$$|g||f+g|^{p-1}$$

to show

$$\|f+g\|_p \leqslant \|f\|_p + \|g\|_p.$$

2. Show that the last result remains true when $p = 1$ or $p = \infty$.

证明: 直接用 *Hölder* 不等式知道

$$
\begin{aligned}
\int |f|\ |f+g|^{p-1} \,\mathrm{d}\mu \leqslant \|f\|_p\ \|(f+g)^{p-1}\|_q, \\
\int |g|\ |f+g|^{p-1} \,\mathrm{d}\mu \leqslant \|g\|_p\ \|(f+g)^{p-1}\|_q.
\end{aligned}
$$

那么有

$$
\begin{aligned}
\int |f+g|^p \,\mathrm{d}\mu &\leqslant \int (|f||f+g|^{p-1} + |g||f+g|^{p-1}) \,\mathrm{d}\mu \\
&\leqslant (\|f\|_p + \|g\|_p)\|(f+g)^{p-1}\|_q.
\end{aligned}
$$

从而有

$$\|f\|_p + \|g\|_p \geqslant \frac{\int |f+g|^p \,\mathrm{d}\mu}{\left(\int |f+g|^{(p-1)q}\right)^{\frac{1}{q}}} = \frac{\int |f+g|^p \,\mathrm{d}\mu}{\left(\int |f+g|^p\right)^{1-\frac{1}{p}}} = \|f+g\|_p.$$

再由之前练习，*Hölder* 不等式在 $p=1$ 或 $p=\infty$ 时也成立，所以可以将 p 扩展到 $[1,\infty]$。 □

练习 1.5.4: If f is integrable and E_m are disjoint sets with union E then

$$\sum_{m=0}^{\infty} \int_{E_m} f \,\mathrm{d}\mu = \int_E f \,\mathrm{d}\mu \,.$$

So if $f \geqslant 0$, then $\nu(E) = \int_E f \,\mathrm{d}\mu$ defines a measure.

证明: 由于可以将 f 分解为 $f^+ - f^-$，故可以不妨设 $f \geqslant 0$。令

$$g = f \cdot 1_E, \quad g_n = f \cdot \sum_{i=0}^{n} 1_{E_i},$$

那么只要证明

$$\lim_{n\to\infty} \int g_n \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu \,.$$

注意到 $g_n \uparrow g$ 且 $g = f \cdot 1_E$ 可积，由单调收敛定理直接推导出以上结论。 □

练习 1.5.5: If $g_n \uparrow g$ and $\int g_1^- \,\mathrm{d}\mu < \infty$ then

$$\int g_n \,\mathrm{d}\mu \uparrow \int g \,\mathrm{d}\mu \,.$$

证明: 令 $g_n' = g_n + g_1^-$，那么有 $g_n' \geqslant g_1' = g_1^+ \geqslant 0$ 且 $g_n' \uparrow (g - g_1^-)$，由单调收敛定理知道

$$\lim \int g_n \,\mathrm{d}\mu - \int g_1^- \,\mathrm{d}\mu = \lim \int g_n' \,\mathrm{d}\mu = \int (g - g_1^-) \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu - \int g_1^- \,\mathrm{d}\mu,$$

即

$$\int g_n \,\mathrm{d}\mu \uparrow \int g \,\mathrm{d}\mu \,.$$

□

练习 1.5.6: If $g_m \geqslant 0$ then

$$\int \sum_{m=0}^{\infty} g_m \,\mathrm{d}\mu = \sum_{m=0}^{\infty} \int g_m \,\mathrm{d}\mu\,.$$

证明: 记

$$f_n = \sum_{m=1}^{n} g_m, \quad f = \lim_{n\to\infty} f_n.$$

那么有 $f_n \geqslant 0$ 且 $f_n \uparrow f$，从而由单调收敛定理知道

$$\sum_{m=0}^{\infty} \int g_m \,\mathrm{d}\mu = \lim \int f_n \,\mathrm{d}\mu \uparrow \int f \,\mathrm{d}\mu = \int \sum_{m=0}^{\infty} g_m \,\mathrm{d}\mu\,.$$

□

练习 1.5.7: Let $f \geqslant 0$.

1. Show that

$$\int f \wedge n \,\mathrm{d}\mu \uparrow \int f \,\mathrm{d}\mu$$

   as $n \to \infty$.
2. Use (1) to conclude that if g is integrable and $\varepsilon > 0$ then we can pick $\delta > 0$ so that $\mu(A) < \delta$ implies

$$\int_A |g| \,\mathrm{d}\mu < \varepsilon.$$

证明:

1. 由于 $f \wedge n \geqslant 0$ 且 $f \wedge n \uparrow f$ 直接应用单调收敛定理知道

$$\int f \wedge n \,\mathrm{d}\mu \uparrow \int f \,\mathrm{d}\mu\,.$$

2. 由上一问结论知道存在足够大的 $N \in \mathbb{N}^*$ 使得

$$\int_{|g|\,\geqslant N} |g| \,\mathrm{d}\mu = \int |g| \,\mathrm{d}\mu - \int |g| \wedge N \,\mathrm{d}\mu < \frac{\varepsilon}{2}.$$

那么令 $\delta = \varepsilon/(2N)$ 直接对 $\forall A \subset \Omega$ 且 $\mu(A) < \delta$ 有

$$
\begin{aligned}
\int_A |g| \,\mathrm{d}\mu &= \int_{A\cap\{|g| \geqslant N\}} |g| \,\mathrm{d}\mu + \int_{A\cap\{|g| < N\}} |g| \,\mathrm{d}\mu \\
&\leqslant \int_{\{|g| \geqslant N\}} |g| \,\mathrm{d}\mu + \int_A N \,\mathrm{d}\mu \\
&\leqslant \frac{\varepsilon}{2} + \frac{\varepsilon}{2N} \cdot N \\
&= \varepsilon.
\end{aligned}
$$

□

练习 1.5.8: Show that if f is integrable on $[a,b]$,

$$
g(x) = \int_{[a,x]} f(y) \,\mathrm{d}y
$$

is continuous on (a,b).

证明: $\forall x_0 \in (a,b)$. 考虑 $\forall x \in (x_0,b)$，有

$$
|g(x) - g(x_0)| = \int_{(x_0,x]} f(y) \,\mathrm{d}y \to 0 \quad (x \downarrow x_0)
$$

其中趋近的结论由上一题目得出。类似可以证明 $x \uparrow x_0$ 的情况。 □

练习 1.5.9: Show that if f has

$$
\|f\|_p = \left(\int |f|^p \,\mathrm{d}\mu\right)^{\frac{1}{p}} < \infty,
$$

then there are simple functions φ_n so that

$$
\|\varphi_n - f\|_p \to 0.
$$

证明: 由于

$$
\|\varphi_n - f\|_p \leqslant \|\varphi_n^+ - f^+\|_p + \|\varphi_n^- - f^-\|_p,
$$

故可以不妨设 $f \geqslant 0$。由简单逼近定理，存在简单函数列 φ_n 使得 $\varphi_n \leqslant f$ 并且有 $\varphi_n \uparrow f$ 几乎处处。那么

$$
\int |\varphi_n - f|^p \,\mathrm{d}\mu \leqslant \int |2f|^p \,\mathrm{d}\mu \leqslant 2^p \int |f|^p \,\mathrm{d}\mu < \infty,
$$

由 *Lebesgue* 逼近定理知道

$$
\lim \int |\varphi_n - f|^p \,\mathrm{d}\mu = \int \lim |\varphi_n - f|^p \,\mathrm{d}\mu = 0,
$$

两边同时开 p 次方就得到结论。 □

练习 1.5.10: Show that if

$$\sum_n \int |f_n| \,\mathrm{d}\mu < \infty$$

then

$$\sum_n \int f_n \,\mathrm{d}\mu = \int \sum_n f_n \,\mathrm{d}\mu \,.$$

证明: 因为可以将 f_n 分解为 $f_n^+ - f_n^-$ 并分别证明，所以可以直接不妨设 $f_n \geqslant 0$. 令

$$g_n = \sum_{m=1}^n f_m, \quad g = \sum_m f_m.$$

那么相当于已知

$$\sum_n \int |f_n| \,\mathrm{d}\mu = \lim \sum_{m=1}^n \int f_n \,\mathrm{d}\mu = \lim \int g_n \,\mathrm{d}\mu < \infty,$$

只要求证

$$\lim \int g_n \,\mathrm{d}\mu = \sum_n \int f_n \,\mathrm{d}\mu = \int \sum_n f_n \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu \,.$$

由于 $f_n \geqslant 0$，从而 $g_n \uparrow g$，从而直接应用单调收敛定理知道

$$\lim \int g_n \,\mathrm{d}\mu = \int g \,\mathrm{d}\mu < \infty.$$

□

1.6. 期望

练习 1.6.1: Suppose φ is strictly convex, i.e., $>$ holds for $\lambda \in (0,1)$. Show that, under the assumptions of Theorem 1.6.2, $\varphi(\mathrm{E}X) = \mathrm{E}\varphi(X)$ implies $X = \mathrm{E}X$ a.s. .

证明: 使用反证法。设 φ 严格凸，$\varphi(\mathrm{E}X) = \mathrm{E}\varphi(X)$ 且 $P(A = \{X \neq \mathrm{E}X\}) > 0$。那么对 $\forall \omega \in A$ 由严格凸函数性质有

$$\varphi(X(\omega)) > \varphi(\mathrm{E}X) + \varphi'(\mathrm{E}X)(X(\omega) - \mathrm{E}X),$$

也即

$$\varphi(X)1_A > (\varphi(\mathrm{E}X) + \varphi'(\mathrm{E}X)(X - \mathrm{E}X))1_A,$$

左右同时取期望得到

$$
\begin{aligned}
\text{LHS} &= \mathrm{E}(\varphi(X)1_A) \\
&= \mathrm{E}(\varphi(X)) - \mathrm{E}(\varphi(X)1_{A^c}) \\
&= \mathrm{E}(\varphi(X)) - \mathrm{E}(\varphi(\mathrm{E}X)1_{A^c}) \\
&= \mathrm{E}(\varphi(X)) - \varphi(\mathrm{E}X)P(A^c) \\
> \text{RHS} & \\
&= \mathrm{E}((\varphi(\mathrm{E}X) + \varphi'(\mathrm{E}X)(X - \mathrm{E}X))1_A) \\
&= \mathrm{E}(\varphi(\mathrm{E}X) + \varphi'(\mathrm{E}X)(X - \mathrm{E}X)) - \mathrm{E}((\varphi(\mathrm{E}X) + \varphi'(\mathrm{E}X)(X - \mathrm{E}X))1_{A^c}) \\
&= \mathrm{E}(\varphi(\mathrm{E}X) + \varphi'(\mathrm{E}X)(X - \mathrm{E}X)) - \mathrm{E}(\varphi(\mathrm{E}X)1_{A^c}) \\
&= \mathrm{E}(\varphi(\mathrm{E}X) + \varphi'(\mathrm{E}X)(X - \mathrm{E}X)) - \varphi(\mathrm{E}X)P(A^c),
\end{aligned}
$$

从而得到

$$
\begin{aligned}
\mathrm{E}(\varphi(X)) &> \mathrm{E}((\varphi(\mathrm{E}X) + \varphi'(\mathrm{E}X)(X - \mathrm{E}X))) \\
&= \varphi(\mathrm{E}X) + \varphi'(\mathrm{E}X)(\mathrm{E}X - \mathrm{E}X) \\
&= \varphi(\mathrm{E}X) \\
&= \mathrm{E}(\varphi(X)),
\end{aligned}
$$

矛盾。 □

练习 1.6.2: Suppose $\varphi : \mathbb{R}^n \to \mathbb{R}$ is convex. Imitate the proof of Theorem 1.5.1 to show

$$\mathrm{E}\varphi(X_1, \cdots, X_n) \geqslant \varphi(\mathrm{E}X_1, \cdots, \mathrm{E}X_n)$$

provided $\mathrm{E}|\varphi(X_1, \cdots, X_n)| < \infty$ and $\mathrm{E}|X_i| < \infty$ for all i.

证明: 由凸函数性质知道

$$\varphi(X_1, \cdots, X_n) \geqslant \varphi(\mathrm{E}X_1, \cdots, \mathrm{E}X_n) + \varphi'(\mathrm{E}X_1, \cdots, \mathrm{E}X_n) \cdot (X_1 - \mathrm{E}X_1, \cdots, X_n - \mathrm{E}X_n),$$

其中 $x \cdot y$ 为向量点积。直接取期望得到

$$
\begin{aligned}
\mathrm{E}\varphi(X_1, \cdots, X_n) \geqslant \varphi(\mathrm{E}X_1, \cdots, \mathrm{E}X_n) &+ \\
\varphi'(\mathrm{E}X_1, \cdots, \mathrm{E}X_n) &\cdot (\mathrm{E}X_1 - \mathrm{E}X_1, \cdots, \mathrm{E}X_n - \mathrm{E}X_n) \\
\geqslant \varphi(\mathrm{E}X_1, \cdots, \mathrm{E}X_n)&.
\end{aligned}
$$

□

练习 1.6.3: Chebyshev's inequality is and is not sharp.

1. Show that Theorem 1.6.4 is sharp by showing that if $0 < b \leqslant a$ are fixed there is an X with $\mathrm{E}X^2 = b^2$ for which

$$P(|X| \geqslant a) = \frac{b^2}{a^2}.$$

2. Show that Theorem 1.6.4 is not sharp by showing that if X has $0 < \mathrm{E}X^2 < \infty$ then

$$\lim_{a\to\infty}\frac{a^2P(|X|\geqslant a)}{\mathrm{E}X^2}=0$$

证明:

1. 令 X 为离散分布，其中 $P(X=0)=1-b^2/a^2, P(X=a)=b^2/a^2$。这时有

$$P(|X|\geqslant a)=P(X=a)=\frac{b^2}{a^2},$$

且

$$\mathrm{E}X^2=\frac{b^2}{a^2}\cdot a^2=b^2.$$

2. 只要证明 $a^2P(|X|\geqslant a)\to 0$。有

$$\begin{aligned}a^2P(|X|\geqslant a)&=a^2P(X^2\geqslant a^2)\\&=\int_{\{X^2\geqslant a^2\}}a^2\,\mathrm{d}P\\&\leqslant\int_{\{X^2\geqslant a^2\}}X^2\,\mathrm{d}P\\&\to 0\quad(a\to\infty)\qquad(\text{因为 }\mathrm{E}X^2<\infty).\end{aligned}$$

□

练习 1.6.4:

1. Let $a>b>0$, $0<p<1$ and let X have $P(X=a)=p$ and $P(X=-b)=1-p$. Apply Theorem 1.6.4 to $\varphi(x)=(x+b)^2$ and conclude that if Y is any random variable with $\mathrm{E}Y=\mathrm{E}X$ and var $(Y)=$ var (X), then $P(Y\geqslant a)\leqslant p$ and equality holds when $Y=X$.
2. Suppose $\mathrm{E}Y=0$, var $(Y)=\sigma^2$, and $a>0$. Show that

$$P(Y\geqslant a)\leqslant\frac{\sigma^2}{a^2+\sigma^2},$$

and there is a Y for which equality holds.

证明:

1. 直接将 $A=\{Y\geqslant a\},\varphi(Y)=(Y+b)^2$ 代入 *Chebychev* 不等式得到

$$\mathrm{E}((Y+b)^2;Y\geqslant a)\leqslant\mathrm{E}(Y+b)^2.$$

其中又有

$$
\begin{aligned}
\mathrm{E}(Y+b)^2 &= \mathrm{E}Y^2 + 2b\mathrm{E}Y + b^2 \\
&= \operatorname{var} Y + \mathrm{E}Y + 2b\mathrm{E}Y + b^2 \\
&= \operatorname{var} X + \mathrm{E}X + 2b\mathrm{E}X + b^2 \\
&= \mathrm{E}X^2 + 2b\mathrm{E}X + b^2 \\
&= pa^2 + (1-p)b^2 + 2b(pa + (1-p)(-b)) + b^2 \\
&= p(a+b)^2,
\end{aligned}
$$

而

$$
\mathrm{E}((Y+b)^2; Y \geqslant a) \geqslant (a+b)^2 P(Y \geqslant a).
$$

从而直接代入 *Chebychev* 不等式就得到结论。容易验证 $Y = X$ 的时候取等。

2. 代入 $p = \sigma^2/(a^2+\sigma^2), b = \sigma^2/a$ 直接得出结论。

$\square$

练习 1.6.5: Show that:

1. if $\varepsilon > 0$, $\inf\{P(|X| > \varepsilon) : \mathrm{E}X = 0, \operatorname{var}(X) = 1\} = 0$.
2. if $y \geqslant 1, \sigma^2 \in (0, \infty)$,

$$
\inf\{P(|X| > y) : \mathrm{E}X = 1, \operatorname{var}(X) = \sigma^2\} = 0.
$$

证明:

1. 令 X_n 满足 $P(X_n = n) = P(X_n = -n) = 1/(2n^2), P(X_n = 0) = 1 - 1/(2n^2)$. 这时容易验证 $\operatorname{var} X_n = 1, \mathrm{E}X_n = 0$。且在 $n \to \infty$ 时有 $P(|X_n| > \varepsilon) \to 0$ 对 $\forall \varepsilon > 0$ 成立。

2. 令 X_n 满足

$$
P(X_n = 1+n) = P(X_n = 1-n) = \frac{\sigma^2}{2n^2}, \quad P(X_n = 1) = 1 - \frac{\sigma^2}{n^2}.
$$

这时 $\mathrm{E}X_n = 1, \operatorname{var} X_n = \sigma^2$，且 $P(|X_n| > y) \to 0$ 在 $n \to \infty$ 时对 $\forall y \geqslant 1$ 成立。

$\square$

练习 1.6.6: A useful lower bound. Let $Y \geqslant 0$ with $\mathrm{E}Y^2 < \infty$. Apply the Cauchy – Schwarz inequality to $Y1_{(Y>0)}$ and conclude

$$
P(Y > 0) \geqslant \frac{(\mathrm{E}Y)^2}{\mathrm{E}Y^2}.
$$

证明: 由于 $Y \geqslant 0$ 从而 $Y1_{\{Y>0\}} = Y$。直接使用 *Cauchy-Schwarz* 不等式得到

$$
\mathrm{E}Y = \mathrm{E}Y1_{\{Y>0\}} \leqslant \sqrt{\int Y^2 \,\mathrm{d}P \int 1_{\{Y>0\}}^2 \,\mathrm{d}P} \leqslant \sqrt{\mathrm{E}Y^2 P(Y>0)},
$$

即

$$P(Y>0) \geqslant \frac{(\mathrm{E}Y)^2}{\mathrm{E}Y^2}.$$

□

练习 1.6.7: Let $\Omega = (0,1)$ equipped with the Borel sets and Lebesgue measure. Let $\alpha \in (1,2)$ and

$$X_n = n^\alpha 1_{\left(\frac{1}{n+1},\frac{1}{n}\right)} \to 0 \quad \text{a.s. .}$$

Show that Theorem 1.6.8 can be applied with $h(x) = x$ and $g(x) = |x|^{2/\alpha}$, but the X_n are not dominated by an integrable function.

证明: 首先验证三个条件。

1. $g(x) \geqslant 0$ 显然，而且当 $|x| \to \infty$ 时容易看出 $g(x) \to \infty$；

2.
$$\frac{|h(x)|}{g(x)} = |x| \cdot |x|^{-\frac{2}{\alpha}} = |x|^{1-\frac{2}{\alpha}} \to 0 \quad (x \to \infty);$$

3.
$$\begin{aligned}
\mathrm{E}g(X_n) &= n^2 \mathrm{E}\Big(1_{\left(\frac{1}{n+1},\frac{1}{n}\right)}\Big)^{\frac{2}{\alpha}} \\
&\leqslant n^2 \mathrm{E}\left(\frac{1}{n^2}\right)^{\frac{2}{\alpha}} \\
&= n^{2-\frac{4}{\alpha}} \to 0 \quad (n \to \infty),
\end{aligned}$$

从而一定有 $\mathrm{E}g(X_n) \leqslant K < \infty$.

那么由定理就知道

$$\mathrm{E}X_n \to \mathrm{E}X = 0.$$

从另一方面看，X_n 不能被任何可积函数控制，这是因为如果要 $|X_n| \leqslant Y$ 对 $\forall n = 1, 2, \cdots$ 成立，那么有 Y 在 $\left\{Y \in \left(\frac{1}{n+1}, \frac{1}{n}\right)\right\}$ 上取值大于等于 $X_n = n^\alpha$. 这时

$$\begin{aligned}
\mathrm{E}Y &\geqslant \mathrm{E}\sum_n n^\alpha 1_{\left(\frac{1}{n+1},\frac{1}{n}\right)} \\
&= \sum_n n^\alpha \mathrm{E}1_{\left(\frac{1}{n+1},\frac{1}{n}\right)} \quad (\text{单调收敛定理}) \\
&\geqslant \sum_n \frac{n^\alpha}{2n^2} \\
&= \frac{1}{2}\sum_n n^{\alpha-2}
\end{aligned}$$

不收敛。

□

练习 1.6.8: Suppose that the probability measure μ has $\mu(A)=\int_A f(x)dx$ for all $A\in\mathscr{R}$. Use the proof technique of Theorem 1.6.9 to show that for any g with $g\geqslant 0$ or

$$\int |g(x)|\mu(dx)<\infty$$

we have

$$\int g(x)\mu(dx)=\int g(x)f(x)dx.$$

证明:

1. 当 g 是指示函数时，设 $g(x)=1_A$，其中 $A\in\mathscr{R}$。这时

$$\int g(x)\mu(\mathrm{d}x)=\int 1_A\mu(\mathrm{d}x)=\mu(A)=\int 1_A f(x)\,\mathrm{d}x=\int g(x)f(x)\,\mathrm{d}x;$$

2. 当 g 是简单函数时，设 $g(x)=\sum_{i=1}^n a_i 1_{A_i}$，其中 $A_i\in\mathscr{R}$。这时

$$\begin{aligned}\int g(x)\mu(\mathrm{d}x)&=\int\sum_{i=1}^n a_i 1_{A_i}\mu(\mathrm{d}x)\\&=\sum_{i=1}^n a_i\int 1_{A_i}\mu(\mathrm{d}x)\\&=\sum_{i=1}^n a_i\int 1_{A_i}f(x)\,\mathrm{d}x\\&=\int\left(\sum_{i=1}^n a_i 1_{A_i}\right)f(x)\,\mathrm{d}x\\&=\int g(x)f(x)\,\mathrm{d}x;\end{aligned}$$

3. 当 g 是非负函数时，令

$$g_n(x)=\frac{[2^n g(x)]}{2^n}\wedge n,$$

那么 $g_n(x)$ 是简单函数且 $g_n\uparrow g, g_n f\uparrow gf$，从而由单调收敛定理知道

$$\begin{aligned}\int g(x)\mu(\mathrm{d}x)&=\lim\int g_n(x)\mu(\mathrm{d}x)\\&=\lim\int g_n(x)f(x)\,\mathrm{d}x\\&=\int g(x)f(x)\,\mathrm{d}x;\end{aligned}$$

4. 当 g 为任意可积函数时，有

$$
\begin{aligned}
\int g(x)\mu(\mathrm{d}x) &= \int (g^+(x) - g^-(x))\mu(\mathrm{d}x) \\
&= \int g^+(x)\mu(\mathrm{d}x) - \int g^-(x)\mu(\mathrm{d}x) \\
&= \int g^+(x) f(x)\,\mathrm{d}x - \int g^-(x) f(x)\,\mathrm{d}x \\
&= \int (g^+(x) - g^-(x)) f(x)\,\mathrm{d}x \\
&= \int g(x) f(x)\,\mathrm{d}x\,.
\end{aligned}
$$

□

练习 1.6.9: Inclusion - exclusion formula. Let $A_1, A_2, \cdots A_n$ be events and $A = \bigcup_{i=1}^n A_i$. Prove that $1_A = 1 - \prod_{i=1}^n \left(1 - 1_{A_i}\right)$. Expand out the right hand side, then take expected value to conclude

$$
\begin{aligned}
P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=1}^n P(A_i) - \sum_{i<j} P\left(A_i \cap A_j\right) \\
&\quad + \sum_{i<j<k} P\left(A_i \cap A_j \cap A_k\right) - \cdots + (-1)^{n-1} P\left(\bigcap_{i=1}^n A_i\right).
\end{aligned}
$$

证明:

$$
\begin{aligned}
1 - \prod_{i=1}^n \left(1 - 1_{A_i}\right) &= 1 - \prod_{i=1}^n 1_{A_i^c} \\
&= 1 - \bigcap_{i=1}^n 1_{A_i^c} \\
&= 1 - 1_{\bigcap_{i=1}^n A_i^c} \\
&= 1_{\left(\bigcap_{i=1}^n A_i^c\right)^c} \\
&= 1_{\bigcup_{i=1}^n A_i} \\
&= 1_A.
\end{aligned}
$$

又知道

$$
\begin{aligned}
P\left(\bigcup_{i=1}^{n} A_i\right) &= \mathrm{E}\left(1_{\bigcup_{i=1}^{n} A_i}\right) \\
&= \mathrm{E}\left(1 - \prod_{i=1}^{n}\left(1 - 1_{A_i}\right)\right) \\
&= 1 - \prod_{i=1}^{n}\left(1 - \mathrm{E}1_{A_i}\right) \\
&= \sum_{i=1}^{n} \mathrm{E}A_i - \sum_{i<j} \mathrm{E}A_i \mathrm{E}A_j + \cdots + (-1)^{n-1}\mathrm{E}\left(\bigcap_{i=1}^{n} A_i\right).
\end{aligned}
$$

□

练习 1.6.10: Bonferroni inequalities. Let $A_1, A_2, \cdots A_n$ be events and $A = \bigcup_{i=1}^{n} A_i$. Show that $1_A \leqslant \sum_{i=1}^{n} 1_{A_i}$, etc. and then take expected values to conclude

$$
\begin{aligned}
P\left(\bigcup_{i=1}^{n} A_i\right) &\leqslant \sum_{i=1}^{n} P(A_i) \\
P\left(\bigcup_{i=1}^{n} A_i\right) &\geqslant \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P\left(A_i \cap A_j\right) \\
P\left(\bigcup_{i=1}^{n} A_i\right) &\leqslant \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P\left(A_i \cap A_j\right) + \sum_{i<j<k} P\left(A_i \cap A_j \cap A_k\right).
\end{aligned}
$$

In general, if we stop the inclusion – exclusion formula after an even (odd) number of sums, we get an lower (upper) bound.

证明: 先证明 $1_A \leqslant \sum 1_{A_i}$。这很容易证明，因为对 $\forall w \in \{1_A = 1\}$ 一定有 $\omega \in A$，从而至少存在一个 A_{i_0} 使 $\omega \in A_{i_0}$，从而

$$
\sum_{i=1}^{n} 1_{A_i}(\omega) \geqslant 1_{A_{i_0}}(\omega) = 1 = 1_A(\omega).
$$

对不等式两侧直接取期望就得到

$$
P\left(\bigcup_{i=1}^{n} A_i\right) \leqslant \sum_{i=1}^{n} P(A_i).
$$

剩余不等式可以类似证明。 □

练习 1.6.11: If $\mathrm{E}|X|^k < \infty$ then for $0 < j < k$, $\mathrm{E}|X|^j < \infty$, and furthermore

$$
\mathrm{E}|X|^j \leqslant \left(\mathrm{E}|X|^k\right)^{\frac{j}{k}}.
$$

证明: 令 $\varphi(x) = x^{\frac{j}{k}}$ 为凹函数，那么利用 *Jensen* 不等式直接得到

$$\mathrm{E}|X|^j \leqslant \left(\mathrm{E}|X|^k\right)^{\frac{j}{k}} < \infty.$$

□

练习 1.6.12: Apply Jensen's inequality with $\varphi(x) = e^x$ and $P(X = \log y_m) = p(m)$ to conclude that if

$$\sum_{m=1}^{n} p(m) = 1$$

and $p(m), y_m > 0$ then

$$\sum_{m=1}^{n} p(m) y_m \geqslant \prod_{m=1}^{n} y_m^{p(m)}.$$

When $p(m) = 1/n$, this says the arithmetic mean exceeds the geometric mean.

证明: 有

$$\varphi(\mathrm{E}X) = \prod_{m=1}^{n} y_m^{p(m)},$$
$$\mathrm{E}\varphi(X) = \sum_{m=1}^{n} y_m p(m),$$

从而直接用 *Jensen* 不等式得出结论。 □

练习 1.6.13: If $\mathrm{E}X_1^- < \infty$ and $X_n \uparrow X$ then $\mathrm{E}X_n \uparrow \mathrm{E}X$.

证明: 直接令 $X_n' = X_n + X_1^- \geqslant X_n + X_n^- = X_n^+$，还有 $X_n' \uparrow X^+$，那么就由单调收敛定理知道 $\mathrm{E}X_n' \uparrow \mathrm{E}X^+$. 对于 X_n^-，注意到 $X_n^- \leqslant X_1^-$，用 *Lebesgue* 收敛定理知道 $X_n^- \downarrow X^-$。两个式子相减就得到结论。 □

练习 1.6.14: Let $X \geqslant 0$ but do NOT assume $\mathrm{E}(1/X) < \infty$. Show

$$\lim_{y \to \infty} y\mathrm{E}\left(\frac{1}{X}; X > y\right) = 0, \quad \lim_{y \downarrow 0} y\mathrm{E}\left(\frac{1}{X}; X > y\right) = 0.$$

证明:

$$
\begin{aligned}
y\mathrm{E}\left(\frac{1}{X}; X > y\right) &\leqslant y\mathrm{E}\left(\frac{1}{y}; X > y\right) \\
&\leqslant y \cdot \frac{1}{y} \cdot P(X > y) \\
&\leqslant P(X > y) \to 0 \quad (y \to \infty).
\end{aligned}
$$

另一方面，对 $\forall \varepsilon > y$，有

$$
\begin{aligned}
y\mathrm{E}\left(\frac{1}{X}; X > y\right) &= \mathrm{E}\left(\frac{y}{X}; y < X < \varepsilon\right) + \mathrm{E}\left(\frac{y}{X}; X \geqslant \varepsilon\right) \\
&\leqslant P(y < X < \varepsilon) + \mathrm{E}\left(\frac{y}{X}; X \geqslant \varepsilon\right).
\end{aligned}
$$

令 $y \to 0$, 由有界收敛定理知道 $\mathrm{E}(y/X; X \geqslant \varepsilon) \to 0$; 再令 $\varepsilon \to 0$, 得到 $P(y < X < \varepsilon) \to 0$，从而得到不等式的右侧均收敛于 0。 □

练习 1.6.15：If $X_n \geqslant 0$ then

$$
\mathrm{E}\left(\sum_{n=0}^{\infty} X_n\right) = \sum_{n=0}^{\infty} \mathrm{E} X_n .
$$

证明：令 $Y_n = \sum_{i=0}^{n} X_n$，那么 $Y_n \uparrow Y = \sum_{i=0}^{\infty} X_n$，只要证明

$$
\mathrm{E}(Y) = \lim \mathrm{E}(Y_n),
$$

这可以显然由单调收敛定理得出。 □

练习 1.6.16：If X is integrable and A_n are disjoint sets with union A then

$$
\sum_{n=0}^{\infty} \mathrm{E}(X; A_n) = \mathrm{E}(X; A) \quad \text{i.e. ,}
$$

the sum converges absolutely and has the value on the right.

证明：对 $X^{+} 1_{A_n}$ 和 $X^{-} 1_{A_n}$ 分别使用上一题的结论，然后相减（因为 X 可积所以可以相减）直接得出。 □

1.7. 乘积测度和 Fubini 定理

练习 1.7.1：If

$$
\int_X \int_Y |f(x, y)| \mu_2(\mathrm{d}y) \mu_1(\mathrm{d}x) < \infty
$$

then

$$\int_X \int_Y f(x,y)\mu_2(\mathrm{d}y)\mu_1(\mathrm{d}x) = \int_{X\times Y} f\,\mathrm{d}(\mu_1\times\mu_2) = \int_Y\int_X f(x,y)\mu_1(\mathrm{d}x)\mu_2(\mathrm{d}y).$$

Corollary Let $X=\{1,2,\cdots\}$, $\mathscr{A}=$ all subsets of X, and $\mu_1=$ counting measure. If

$$\sum_n \int |f_n|\,\mathrm{d}\mu < \infty$$

then

$$\sum_n \int f_n\,\mathrm{d}\mu = \int \sum_n f_n\,\mathrm{d}\mu\,.$$

证明: 因为 $|f|\geqslant 0$，所以直接对 $|f|$ 用 *Fubini* 定理知道

$$\int_X\int_Y |f(x,y)|\,\mu_2(\mathrm{d}y)\mu_1(\mathrm{d}x) = \int_{X\times Y}|f(x,y)|\,\mathrm{d}(\mu_1\times\mu_2)<\infty,$$

从而 f 满足 *Fubini* 定理条件，再对 f 用 *Fubini* 定理直接得到结论。 □

练习 1.7.2: Let $g\geqslant 0$ be a measurable function on $(X,\mathscr{A},\mu)$. Use Theorem 1.7.2 to conclude that

$$\int_X g\,\mathrm{d}\mu = (\mu\times\lambda)(\{(x,y):0\leqslant y<g(x)\}) = \int_0^\infty \mu(\{x:g(x)>y\})\,\mathrm{d}y\,.$$

证明: 注意到 $f(x,y)=1_{\{g(x)>y\}}\geqslant 0$ 符合 *Fubini* 定理条件，从而有

$$\int_X\int_0^\infty 1_{\{g(x)>y\}}\,\mathrm{d}\lambda\mathrm{d}\mu = \int_{X\times[0,\infty)} 1_{\{g(x)>y\}}\,\mathrm{d}(\mu\times\lambda) = \int_0^\infty\int_X 1_{\{g(x)>y\}}\,\mathrm{d}\mu\mathrm{d}\lambda\,.$$

容易验证这就是要证的等式。 □

练习 1.7.3: Let F, G be Stieltjes measure functions and let μ,ν be the corresponding measures on $(\mathbb{R},\mathscr{R})$. Show that

1. $$\int_{(a,b]}\{F(y)-F(a)\}\,\mathrm{d}G(y) = (\mu\times\nu)(\{(x,y):a<x\leqslant y\leqslant b\});$$

2. $$\int_{(a,b]} F(y)\,\mathrm{d}G(y) + \int_{(a,b]} G(y)\,\mathrm{d}F(y)$$
$$= F(b)G(b)-F(a)G(a)+\sum_{x\in(a,b]}\mu(\{x\})\nu(\{x\});$$

3. If $F=G$ is continuous then

$$\int_{(a,b]} 2F(y)\,\mathrm{d}F(y) = F^2(b) - F^2(a).$$

To see the second term in (2) is needed, let $F(x) = G(x) = 1_{[0,\infty)}(x)$ and $a < 0 < b$.

证明:
1. 注意到 $f(x,y) = 1_{\{x \leqslant y\}} \geqslant 0$ 符合 *Fubini* 定理条件，从而有

$$\int_{(a,b]} \int_{(a,b]} 1_{\{x \leqslant y\}}\,\mathrm{d}F\,\mathrm{d}G = \int_{(a,b]^2} 1_{\{x \leqslant y\}}\,\mathrm{d}(\mu \times \nu),$$

也即欲证的等式。
2. 由于 F, G 是 *Stieltjes* 测度函数，所以 F, G 均只在至多可数个点处不连续，所以

$$\sum_{x \in (a,b]} \mu(\{x\})\nu(\{x\})$$

为至多可数个点相加，从而有意义。设 F, G 分别对应定义在 $(\Omega, \mathscr{F}, \gamma)$ 上的随机变量 X, Y。设 F, G 在 $(a,b]$ 上共同的不连续点集为 A，那么在 $(a,b] \setminus A$ 上有

$$\int_{(a,b]\setminus A} F\,\mathrm{d}G + \int_{(a,b]\setminus A} G\,\mathrm{d}F = \int_{(a,b]\setminus A} \mathrm{d}(FG)\,.$$

再考虑在 A 上的积分情况。有

$$\begin{aligned}\int_A F\,\mathrm{d}G &= \int_{G^{-1}(A)} F \circ G\,\mathrm{d}\gamma \\ &= \sum_{\omega \in G^{-1}(A)} \gamma(\{\omega\}) \cdot F \circ G(\omega) \\ &= \sum_{x \in A} \nu(\{x\}) \cdot \mu(\{x\}).\end{aligned}$$

同理对 $\int_A G\,\mathrm{d}F$ 得到相同的结果。那么相加就得到所求结论。
3. 因为 $F = G$ 连续所以它们 *Riemann* 可积，那么直接应用 *Riemann* 积分结论直接得出。

□

练习 1.7.4: Let μ be a finite measure on $\mathbb{R}$ and $F(x) = \mu((-\infty, x])$. Show that

$$\int (F(x+c) - F(x))\,\mathrm{d}x = c\mu(\mathbb{R}).$$

证明: 注意到 $1_{\{x < y \leqslant x+c\}} \geqslant 0$ 满足 *Fubini* 定理的条件，从而直接应用 *Fubini* 定理得到

$$\int_{\mathbb{R}}\int_{\mathbb{R}} 1_{\{x<y\leqslant x+c\}}\mu(\mathrm{d}y)\lambda(\mathrm{d}x)=\int_{\mathbb{R}}\int_{\mathbb{R}} 1_{\{x<y\leqslant x+c\}}\lambda(\mathrm{d}x)\mu(\mathrm{d}y),$$

其中 λ 为 $(R,\mathscr{R})$ 上的 *Lebesgue* 测度。上式中

$$\begin{aligned}
\mathrm{LHS}&=\int_{\mathbb{R}}\int_{(x,x+c]}\mathrm{d}F\,\lambda(\mathrm{d}x)\\
&=\int_{\mathbb{R}}(F(x+c)-F(x))\,\mathrm{d}x,\\
\mathrm{RHS}&=\int_{\mathbb{R}}\int_{[y-c,y)}\mathrm{d}x\,\mu(\mathrm{d}y)\\
&=c\int_{\mathbb{R}}\mu(\mathrm{d}y)\\
&=c\int_{\mathbb{R}}\mathrm{d}F\\
&=c\mu(\mathbb{R}).
\end{aligned}$$

□

练习 1.7.5: Show that $e^{-xy}\sin x$ is integrable in the strip $0<x<a,0<y$. Perform the double - integral in the two orders to get:

$$\int_0^a\frac{\sin x}{x}\,\mathrm{d}x=\arctan(a)-(\cos a)\int_0^\infty\frac{e^{-ay}}{1+y^2}\,\mathrm{d}y-(\sin a)\int_0^\infty\frac{ye^{-ay}}{1+y^2}\,\mathrm{d}y$$

and replace $1+y^2$ by 1 to conclude

$$\left|\int_0^a\frac{\sin x}{x}\,\mathrm{d}x-\arctan(a)\right|\leqslant\frac{2}{a}$$

for $a\geqslant 1$.

证明: 有

$$\begin{aligned}
\int_0^a\int_0^\infty e^{-xy}\sin x\,\mathrm{d}y\,\mathrm{d}x&=-\int_0^a\frac{\sin x}{x}\int_0^\infty e^{-xy}\,\mathrm{d}(-xy)\,\mathrm{d}x\\
&=\int_0^a\frac{\sin x}{x}\,\mathrm{d}x\,.
\end{aligned}$$

由于

$$\int_0^a\sin x\,\mathrm{d}x\text{ 有界},\quad\frac{1}{x}\downarrow 0,$$

由 *Dirichlet* 判别法知道

$$\int_0^a \frac{\sin(x)}{x} \, \mathrm{d}x$$

Riemann 可积，从而 *Lebesgue* 可积。那么可以对 $e^{-xy} \sin x$ 使用 *Fubini* 定理得到

$$\int_0^a \int_0^\infty e^{-xy} \sin x \, \mathrm{d}y \mathrm{d}x = \int_0^\infty \int_0^a e^{-xy} \sin x \, \mathrm{d}x \mathrm{d}y \, .$$

上式中

$$\begin{aligned}
\mathrm{LHS} &= \int_0^a \frac{\sin x}{x} \, \mathrm{d}x, \\
\mathrm{RHS} &= \int_0^\infty \frac{-e^{-xy}(\cos x + y \sin x)}{1+y^2} \bigg|_0^a \, \mathrm{d}y \\
&= \int_0^\infty \left(\frac{1}{1+y^2} - \frac{e^{-ay} \cos a}{1+y^2} - \frac{e^{-ay} y \sin a}{1+y^2} \right) \mathrm{d}y
\end{aligned}$$

即为欲证之结论。 □

2. 大数定律

2.1. 独立性

定义 11 (独立):
1. 称**事件 $\boldsymbol{A}, \boldsymbol{B}$ 独立**，如果 $P(A \cap B) = P(A)P(B)$。
2. 称**随机变量 $\boldsymbol{X}, \boldsymbol{Y}$ 独立**，如果对 $\forall C, D \in \mathscr{R}$ 有

$$P(X \in C, Y \in D) = P(X \in C)P(Y \in D).$$

3. 称 **$\boldsymbol{\sigma}$-域 $\boldsymbol{\mathscr{F}}, \boldsymbol{\mathscr{G}}$ 独立**，如果对 $\forall A \in \mathscr{F}$ 和 $B \in \mathscr{G}$ 事件 A, B 独立。

练习 2.1.1: Suppose $(X_1, \cdots, X_n)$ has density $f(x_1, x_2, \cdots, x_n)$, that is

$$P((X_1, X_2, \cdots, X_n) \in A) = \int_A f(x) \, \mathrm{d}x \quad \text{for } A \in \mathscr{R}^n.$$

If $f(x)$ can be written as $g_1(x_1) \cdots g_n(x_n)$ where the $g_m \geqslant 0$ are measurable, then $X_1, X_2, \cdots, X_n$ are independent. Note that the g_m are not assumed to be probability densities.

证明: 对 $\Omega = \Omega_1 \times \cdots \times \Omega_n$ 中的一个长方体 $A = A_1 \times \cdots \times A_n$，有

$$
\begin{aligned}
P(X_1 \in A_1, \cdots, X_n \in A_n) &= \int_A f(x_1, \cdots, x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n \\
&= \int_A \prod_{i=1}^n g_i(x_i)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n \\
&= \prod_{i=1}^n \int_{A_i} g_i(x_i)\,\mathrm{d}x_i \quad \text{(Fubini 定理)}.
\end{aligned}
$$

对 $\forall k \in \{1, 2, \cdots, n\}$ 有

$$
\begin{aligned}
P(X_k \in A_k) = P(X_1 \in \Omega_1, \cdots, X_k \in A_k, \cdots, X_n \in \Omega_n) \\
= \left(\int_{A_k} g_k(x_k)\,\mathrm{d}x_k\right) \cdot \prod_{i \neq k} \int_{\Omega_i} g_i(x_i)\,\mathrm{d}x_i\,.
\end{aligned}
$$

又知道

$$
1 = \int_\Omega f\,\mathrm{d}x_1 \cdots \mathrm{d}x_n = \prod_{i=1}^n \int_{\Omega_i} g_i(x_i)\,\mathrm{d}x_i,
$$

那么

$$
\begin{aligned}
\prod_{i=1}^n P(X_i \in A_i) &= \left(\prod_{i=1}^n \int_{A_i} g_i(x_i)\,\mathrm{d}x_i\right) \cdot \left(\prod_{i=1}^n \int_{\Omega_i} g_i(x_i)\,\mathrm{d}x_i\right)^{n-1} \\
&= \left(\prod_{i=1}^n \int_{A_i} g_i(x_i)\,\mathrm{d}x_i\right) \cdot 1^{n-1} \\
&= \prod_{i=1}^n \int_{A_i} g_i(x_i)\,\mathrm{d}x_i \\
&= P(X_1 \in A_1, \cdots, X_n \in A_n).
\end{aligned}
$$

□

练习 2.1.2: Suppose $X_1, \cdots, X_n$ are random variables that take values in countable sets $S_1, \cdots, S_n$. Then in order for $X_1, \cdots, X_n$ to be independent, it is sufficient that whenever $x_i \in S_i$,

$$
P(X_1 = x_1, \cdots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i).
$$

证明: 令 $\mathscr{A}_i$ 为 S_i 中所有单元素集和 $\varnothing$ 构成的集合。那么 $\mathscr{A}_i$ 是 π-系，只要证明 $\mathscr{A}_1, \mathscr{A}_2, \cdots, \mathscr{A}_n$ 独立，那么 $\sigma(\mathscr{A}_1), \cdots, \sigma(\mathscr{A}_n)$ 自然互相独立，从而得证。而从题目条件自然得出 $\mathscr{A}_i$ 是独立的。 □

练习 2.1.3: Let $\rho(x,y)$ be a metric.
1. Suppose h is differentiable with $h(0)=0$, $h'(x)>0$ for $x>0$ and $h'(x)$ decreasing on $[0,\infty)$. Then $h(\rho(x,y))$ is a metric.
2. $h(x)=x/(x+1)$ satisfies the hypotheses in (1).

证明:
1. 条件说明 h 为单调增的凹函数。$\forall x,y,z$，只要证明

$$h(\rho(x,y))+h(\rho(y,z))\geqslant h(\rho(x,z)).$$

这是因为

$$\begin{aligned}h(\rho(x,y))+h(\rho(y,z))&\geqslant h(\rho(x,y)+\rho(y,z))\\&\geqslant h(\rho(x,z)).\end{aligned}$$

2. 直接验证即可。

□

练习 2.1.4: Let $\Omega=(0,1)$, $\mathscr{F}=$ Borel sets, $P=$ Lebesgue measure. Then

$$X_n(\omega)=\sin(2\pi n\omega),\quad n=1,2,\cdots$$

are uncorrelated but not independent.

证明: 对 $\forall n,m\in\mathbb{N}^+$ 且 $n\neq m$ 有

$$\begin{aligned}\mathrm{E}X_n&=\int_0^1\sin(2\pi nx)\,\mathrm{d}x=0,\\\mathrm{E}X_m&=0,\\\mathrm{E}X_nX_m&=\int_0^1\sin(2\pi nx)\sin(2\pi mx)\,\mathrm{d}x\\&=\int_0^1\frac{\cos(2\pi(n-m)x)+\cos(2\pi(n+m)x)}{2}\,\mathrm{d}x=0.\end{aligned}$$

所以有 $\mathrm{E}X_nX_m=\mathrm{E}X_n\mathrm{E}X_m$，从而它们无关。但是又考虑 X_2,X_3 在 $A=(0,1/3)$ 上的期望有

$$\begin{aligned}\mathrm{E}(X_2;A)&=\frac{3}{8\pi},\\\mathrm{E}(X_3;A)&=0,\\\mathrm{E}(X_2X_3;A)&=\frac{3\sqrt{3}}{20\pi},\end{aligned}$$

从而它们无关。

□

练习 2.1.5:
1. Show that if X and Y are independent with distributions μ and ν then

$$P(X+Y=0)=\sum_y \mu(\{-y\})\nu(\{y\}).$$

2. Conclude that if X has continuous distribution $P(X=Y)=0$.

证明:
1.

$$\begin{aligned}P(X+Y=0) &= \iint 1_{\{X+Y=0\}}\mu(\mathrm{d}x)\nu(\mathrm{d}y)\\ &= \int \mu(\{-y\})\nu(\mathrm{d}y)\\ &= \sum_y \mu(\{-y\})\nu(\{y\}).\end{aligned}$$

2. 与上一问同理可以证明

$$P(X=Y)=\sum_y \mu(\{y\})\nu(\{y\}),$$

等式右侧只有在 μ,ν 都不为 0,即 X,Y 的分布都不连续的地方才不等于 0,而这永不成立,因为 X 的分布是连续的。

□

练习 2.1.6: Prove directly from the definition that if X and Y are independent and f and g are measurable functions then $f(X)$ and $g(Y)$ are independent.

证明: 对 $\forall A,B\in\mathscr{R}$,有

$$\begin{aligned}P(f(X)\in A)P(g(Y)\in B) &= P(X\in f^{-1}(A))P(Y\in g^{-1}(B))\\ &= P(X\in f^{-1}(A),Y\in g^{-1}(B))\\ &= P(f(X)\in A,g(Y)\in B).\end{aligned}$$

□

练习 2.1.7: Let $K\geqslant 3$ be a prime and let X and Y be independent random variables that are uniformly distributed on $\{0,1,\cdots,K-1\}$. For $0\leqslant n<K$, let

$$Z_n = X+nY \bmod K.$$

Show that $Z_0,Z_1,\cdots,Z_{K-1}$ are pairwise independent, i.e., each pair is independent. They are not independent because if we know the values of two of the variables then we know the values of all the variables.

证明: 由数论相关知识知道对于 $\forall 0 \leqslant n < K$ 有 $Z_n = X + nY = p$ 当且仅当 $Y = q(X)$ 存在 $q(X)$，即此时的 q 是唯一的，所以

$$
\begin{aligned}
P(Z_n = p) &= \sum_x P(X = x, Y = q(x)) \\
&= \sum_x P(X = x)P(Y = q(x)) \\
&= \sum_x \frac{1}{K^2} = \frac{1}{K}
\end{aligned}
$$

对 $\forall p \in \{0, 1, \cdots, K-1\}$ 成立。那么对 $\forall a, b \in \{0, 1, \cdots, K-1\}$ 且 $a \neq b$，考虑 Z_a 和 Z_b 之间的独立性。有

$$
P(Z_a = p) = P(Z_b = q) = \frac{1}{K}, \quad P(Z_a = p, Z_b = q) = \frac{1}{K^2}
$$

从而 Z_a, Z_b 独立。 □

练习 2.1.8: Find four random variables taking values in $\{-1, 1\}$ so that any three are independent but all four are not. Hint: Consider products of independent random variables.

证明: 记 X, Y, Z, W 为满足题意的随机变量，且 $A = \{X = 1\}, B = \{Y = 1\}, C = \{Z = 1\}, D = \{W = 1\}, P(A) = a, P(B) = b, P(C) = c, P(D) = d$。以下推导所有符合题意的随机变量。

设 $P(A \cap B \cap C \cap D) = x$，那么有

1. $$P(A^c \cap B \cap C \cap D) = bcd - x;$$
2. $$P(A \cap B^c \cap C \cap D) = acd - x,$$
$$P(A^c \cap B^c \cap C \cap D) = (1-a)cd - bcd + x;$$
3. $$
\begin{aligned}
P(A \cap B \cap C^c \cap D) &= abd - x, \\
P(A \cap B^c \cap C^c \cap D) &= a(1-b)d - acd + x, \\
P(A^c \cap B \cap C \cap D) &= (1-a)bd - bcd + x, \\
P(A^c \cap B^c \cap C \cap D) &= (1-a)(1-b)d - (1-a)cd + bcd - x;
\end{aligned}
$$
4. 类似推导所有带有 D^c 的情况。这样只要令 $x \neq abcd$ 的值即可保证
$$P(A \cap B \cap C \cap D) = x \neq abcd = P(A)P(B)P(C)P(D),$$
但是任意三个集合是相互独立的。

□

练习 2.1.9: Let $\Omega = \{1, 2, 3, 4\}$, $\mathscr{F} =$ all subsets of Ω, and $P(\{i\}) = 1/4$. Give an example of two collections of sets $\mathscr{A}_1$ and $\mathscr{A}_2$ that are independent but whose generated σ-fields are not.

证明: 令 $\mathscr{A}_1 = \{\{1,2\},\{1,3\}\}, \mathscr{A}_2 = \{\{2,3\}\}$，那么

$$P(\{1,2\}\cap\{2,3\}) = \frac{1}{2} = P(\{1,2\})P(\{2,3\}),$$
$$P(\{1,3\}\cap\{2,3\}) = \frac{1}{2} = P(\{1,3\})P(\{2,3\}),$$

从而 $\mathscr{A}_1$ 和 $\mathscr{A}_2$ 独立。但是有 $\{2,3\}\in\sigma(\mathscr{A}_1)$，这时

$$P(\{2,3\}\cap\{2,3\}) = P(\{2,3\}) = \frac{1}{2} \neq P(\{2,3\})^2,$$

从而 $\sigma(\mathscr{A}_1)$ 和 $\sigma(\mathscr{A}_2)$ 不独立。 □

练习 2.1.10: Show that if X and Y are independent, integer-valued random variables, then

$$P(X+Y=n) = \sum_m P(X=m)P(Y=n-m).$$

证明: 因为 $1_{\{X+Y=n\}} \geqslant 0$，所以可以直接应用 *Fubini* 定理，从而有

$$\begin{aligned}
P(X+Y=n) &= \iint 1_{\{x+y=n\}}\,\mathrm{d}\mu\mathrm{d}\nu \\
&= \int\left(\int 1_{\{x+y=n\}}\,\mathrm{d}\mu\right)\mathrm{d}\nu \\
&= \int \mu(n-y)\,\mathrm{d}\nu \\
&= \sum_m \mu(n-m)\nu(m) \\
&= \sum_m P(X=m)P(Y=n-m).
\end{aligned}$$

□

练习 2.1.11: In Example 1.6.13, we introduced the Poisson distribution with parameter λ, which is given by

$$P(Z=k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

for $k=0,1,2,\cdots$. Use the previous exercise to show that if $X=\text{Poisson}(\lambda)$ and $Y=\text{Poisson}(\mu)$ are independent then $X+Y=\text{Poisson}(\lambda+\mu)$.

证明: $\forall n\in\mathbb{N}$，则有

$$
\begin{aligned}
P(X+Y=n) &= \sum_m P(X=m)P(Y=n-m) \\
&= \sum_{m=0}^{n} \frac{e^{-\lambda}\lambda^m}{m!}\frac{e^{-\mu}\mu^{n-m}}{(n-m)!} \\
&= e^{-(\lambda+\mu)}\sum_{m=0}^{n}\frac{\lambda^m\mu^{n-m}}{m!(n-m)!} \\
&= e^{-(\lambda+\mu)}\frac{(\lambda+\mu)^n}{n!}.
\end{aligned}
$$

最后一步利用了二项式定理。 □

练习 2.1.12: X is said to have a Binomial(n,p) distribution if

$$
P(X=m) = \binom{n}{m}p^m(1-p)^{n-m}.
$$

1. Show that if $X = \text{Binomial}(n,p)$ and $Y = \text{Binomial}(m,p)$ are independent then $X+Y = \text{Binomial}(n+m,p)$.
2. Look at Example 1.6.12 and use induction to conclude that the sum of n independent Bernoulli(p) random variables is Binomial(n,p).

证明:

1.

$$
\begin{aligned}
P(X+Y=x) &= \sum_{y=0}^{x} P(X=y)P(Y=x-y) \\
&= \sum_{y=0}^{x}\binom{n}{y}p^y(1-p)^{n-y}\binom{m}{x-y}p^{x-y}(1-p)^{m-x+y} \\
&= p^x(1-p)^{n+m-x}\sum_{y=0}^{x}\binom{n}{y}\binom{m}{x-y} \\
&= \binom{n+m}{x}p^x(1-p)^{n+m-x}.
\end{aligned}
$$

2. $n=1$ 的情况是显然的。对 $n>1$ 的情况，记 $X = X_1+\cdots+X_{n-1}$ 为 $(n-1)$ 个 *Bernoulli*(p) 分布的随机变量的和，归纳假设 X 服从 Binomial$(n-1,p)$. 设 X_n 也服从 Bernoulli(p) 分布，只要证明 $X+X_n$ 服从 Binomial(n,p) 分布。有

$$
\begin{aligned}
P(X_n+X=x) &= P(X=x)P(X_n=0)+P(X=x-1)P(X_n=1) \\
&= \binom{n-1}{x}p^x(1-p)^{n-x-1}(1-p)+\binom{n-1}{x-1}p^{x-1}(1-p)^{n-x}p \\
&= \left(\binom{n-1}{x}+\binom{n-1}{x-1}\right)p^x(1-p)^{n-x} \\
&= \binom{n}{x}p^x(1-p)^{n-x}.
\end{aligned}
$$

$\square$

练习 2.1.13: It should not be surprising that the distribution of $X+Y$ can be $F*G$ without the random variables being independent. Suppose $X, Y \in \{0,1,2\}$ and take each value with probability $1/3$.

1. Find the distribution of $X+Y$ assuming X and Y are independent.
2. Find all the joint distributions (X, Y) so that the distribution of $X+Y$ is the same as the answer to (1).

证明:

1.

$$
\begin{aligned}
P(X+Y=0) &= P(X=0)P(Y=0) = \frac{1}{9},\\
P(X+Y=1) &= P(X=0)P(Y=1) + P(X=1)P(Y=0) = \frac{2}{9},\\
P(X+Y=2) &= \frac{3}{9},\\
P(X+Y=3) &= \frac{2}{9},\\
P(X+Y=4) &= \frac{1}{9}.
\end{aligned}
$$

分布函数略。

2. 直接设 $P(X=1, Y=0) = x \leqslant \frac{2}{9}$，然后根据题意和上一问列方程求解得到

$X \setminus Y$	**0**	**1**	**2**
0	$\frac{1}{9}$	$\frac{2}{9}-x$	x
1	x	$\frac{1}{9}$	$\frac{2}{9}-x$
2	$\frac{2}{9}-x$	x	$\frac{1}{9}$

即为所求的所有联合分布。

$\square$

练习 2.1.14: Let $X, Y \geqslant 0$ be independent with distribution functions F and G. Find the distribution function of XY.

证明: 有

$$
\begin{aligned}
P(XY \leqslant p) &= \iint 1_{\{xy\leqslant p\}}\,\mathrm{d}F\,\mathrm{d}G \\
&= \int_0^{\infty}\int_0^{p/y}\mathrm{d}F\,\mathrm{d}G \\
&= \int_0^{\infty} F\left(\frac{p}{y}\right)\mathrm{d}G\,.
\end{aligned}
$$

□

练习 2.1.15: If we want an infinite sequence of coin tossings, we do not have to use Kolmogorov's theorem. Let Ω be the unit interval $(0,1)$ equipped with the Borel sets $\mathscr{F}$ and Lebesgue measure P. Let $Y_n(\omega)=1$ if $[2^n\omega]$ is odd and 0 if $[2^n\omega]$ is even. Show that $Y_1, Y_2, \cdots$ are independent with $P(Y_k=0)=P(Y_k=1)=1/2$.

证明: $P(Y_k=0)=1/2$ 容易证明。下证明 $Y_1, Y_2, \cdots$ 独立。对 $\forall A\in\{1,2,\cdots\}$，其中 A 为有限集，只要证明 $\{Y_a : a\in A\}$ 独立。记 $A=\{a_1,\cdots,a_m\}$，其中 $a_1<\cdots<a_m$，那么只要证明

$$
P\left(Y_{a_1}=1, Y_{a_2}=1,\cdots,Y_{a_m}=1\right)=P\left(Y_{a_1}=1\right)P\left(Y_{a_2}=1,\cdots,Y_{a_m}=1\right),
$$

然后归纳即可。而

$$
\begin{aligned}
P\left(Y_{a_1}=1, Y_{a_2}=1,\cdots,Y_{a_m}=1\right) &= \frac{1}{2}P\left(Y_{a_1}=1, Y_{a_2}=1,\cdots,Y_{a_{m-1}}=1\right) \\
&= \frac{1}{2^2}P\left(Y_{a_1}=1, Y_{a_2}=1,\cdots,Y_{a_{m-2}}=1\right) \\
&= \cdots \\
&= \frac{1}{2^m}, \\
P\left(Y_{a_2}=1, Y_{a_2}=1,\cdots,Y_{a_m}=1\right) &= \frac{1}{2^{m-1}},
\end{aligned}
$$

那么自然有

$$
\begin{aligned}
P\left(Y_{a_1}=1, Y_{a_2}=1,\cdots,Y_{a_m}=1\right) &= \frac{1}{2}P\left(Y_{a_2}=1,\cdots,Y_{a_m}=1\right) \\
&= P\left(Y_{a_1}=1\right)P\left(Y_{a_2}=1,\cdots,Y_{a_m}=1\right).
\end{aligned}
$$

□

2.2. 弱大数定律

练习 2.2.1: Let $X_1, X_2, \cdots$ be uncorrelated with $\mathrm{E}X_i=\mu_i$ and $\operatorname{var}(X_i)/i\to 0$ as $i\to\infty$. Let $S_n=X_1+\cdots+X_n$ and $\nu_n=\mathrm{E}S_n/n$ then as $n\to\infty$,

$$\frac{S_n}{n} - \nu_n \to 0$$

in L^2 and in probability.

证明:

$$\begin{aligned}
\mathrm{E}\left(\frac{S_n}{n} - \mathrm{E}\frac{S_n}{n}\right)^2 &= \operatorname{var}\,\frac{S_n}{n} \\
&= \frac{\operatorname{var}\,S_n}{n^2} \\
&= \frac{\sum_i \operatorname{var}\,X_i}{n^2} \\
&\leqslant \frac{\sum_i \frac{\operatorname{var}\,X_i}{i}}{n}.
\end{aligned}$$

令 $n \to \infty$ 并用 *Stolz* 定理知道

$$\mathrm{E}\left(\frac{S_n}{n} - \mathrm{E}\frac{S_n}{n}\right)^2 \to 0,$$

从而题目中随机变量 L^2 收敛到 0，进而依测度收敛。 □

练习 2.2.2: The L^2 weak law generalizes immediately to certain dependent sequences. Suppose $\mathrm{E}X_n = 0$ and $\mathrm{E}X_nX_m \leqslant r(n-m)$ for $m \leqslant n$ (no absolute value on the left - hand side!) with $r(k) \to 0$ as $k \to \infty$. Show that

$$\frac{X_1 + \cdots + X_n}{n} \to 0$$

in probability.

证明:

$$\begin{aligned}
\mathrm{E}\left(\frac{(X_1 + \cdots + X_n)^2}{n}\right) &= \frac{\sum_{i \leqslant j} \mathrm{E}(X_iX_j)}{n^2} \\
&\leqslant \frac{\sum_{i \leqslant j} r(j-i)}{n^2} \\
&\leqslant \frac{nr(0) + (n-1)r(1) + \cdots + r(n-1)}{n^2} \\
&\to \frac{r(0) + r(1) + \cdots + r(n)}{2n+1} \\
&\to \frac{r(n)}{2} \to 0 \quad (n \to \infty).
\end{aligned}$$

其中两次使用了 *Stolz* 定理。那么它 L^2 收敛，从而依测度收敛。 □

练习 2.2.3: Monte Carlo integration.

1. Let f be a measurable function on $[0,1]$ with
$$\int_0^1 |f(x)|\,\mathrm{d}x < \infty.$$
Let $U_1, U_2, \cdots$ be independent and uniformly distributed on $[0,1]$, and let
$$I_n = \frac{f(U_1) + \cdots + f(U_n)}{n}.$$
Show that
$$I_n \to I \equiv \int_0^1 f\,\mathrm{d}x$$
in probability.
2. Suppose
$$\int_0^1 |f(x)|^2\,\mathrm{d}x < \infty.$$
Use Chebyshev's inequality to estimate
$$P\left(|I_n - I| > \frac{a}{n^{\frac{1}{2}}}\right).$$

证明:

1. 因为 $U_1, U_2, \cdots$ 是独立同分布序列，所以 $f(U_1), f(U_2), \cdots$ 独立同分布，又因为 $\mathrm{E}(f(U_n)) < \infty$，所以可以直接应用弱大数定律得到结论。

2.
$$\begin{aligned}
P\left(|I_n - I| > \frac{a}{\sqrt{n}}\right) &= P\left((I_n - I)^2 > \frac{a^2}{n}\right) \\
&\leqslant \frac{n\ \mathrm{var}(I_n)}{a^2} \\
&= \frac{n \cdot \frac{n\sigma_{f(U_i)}}{n^2}}{a^2} \\
&= \frac{\sigma^2_{f(U_i)}}{a^2}.
\end{aligned}$$

□

练习 2.2.4: Let $X_1, X_2, \cdots$ be i.i.d. with

$$P\big(X_i = (-1)^k k\big) = \frac{C}{k^2 \log k}$$

for $k \geqslant 2$ where C is chosen to make the sum of the probabilities $= 1$. Show that $\mathrm{E}|X_i| = \infty$, but there is a finite constant μ so that $S_n/n \to \mu$ in probability.

证明: 容易看出

$$\mathrm{E}|X_i| = \sum_{k=0}^{\infty} \frac{Ck}{k^2 \log k} = \sum_{k=0}^{\infty} \frac{c}{k \log k} = \infty.$$

再证明下一结论，考虑使用弱大数定律。对足够大的 n，有

$$\begin{aligned} nP(|X_i| > n) &= n \sum_{k=n}^{\infty} \frac{1}{k^2 \log k} \\ &\leqslant \frac{n}{\log n} \sum_{k=n}^{\infty} \frac{1}{k^2} \\ &\leqslant \frac{n}{\log n} \int_{n-1}^{\infty} \frac{1}{x^2} \,\mathrm{d}x \\ &= \frac{n}{(n-1)\log n} \to 0 \quad (n \to \infty), \end{aligned}$$

从而满足弱大数定律条件，从而有

$$\left| \frac{S_n}{n} - \mu_n \right| \xrightarrow{P} 0,$$

其中

$$\begin{aligned} \mu_n &= \mathrm{E} X_i 1_{\{|X_i| \leqslant n\}} \\ &= \sum_{k=0}^{n} \frac{(-1)^k k}{k^2 \log k} \\ &= \sum_{k=0}^{n} \frac{(-1)^k}{k \log k} \end{aligned}$$

为交错级数，从而收敛。设 $\mu_n \to \mu$，则有

$$\left| \frac{S_n}{n} - \mu \right| \leqslant \left| \frac{S_n}{n} - \mu_n \right| + |\mu_n - \mu| \xrightarrow{P} 0.$$

□

练习 2.2.5: Let $X_1, X_2, \cdots$ be i.i.d. with

$$P(X_i > x) = \frac{e}{x \log x}$$

for $x \geqslant e$. Show that $\mathrm{E}|X_i| = \infty$, but there is a sequence of constants $\mu_n \to \infty$ so that

$$\frac{S_n}{n} - \mu_n \to 0$$

in probability.

证明:

$$\begin{aligned}
\mathrm{E}|X_i| &= \mathrm{E}X_i \\
&= \int_0^\infty P(X_i > y)\,\mathrm{d}y \\
&= \int_0^\infty \frac{e}{y \log y}\,\mathrm{d}y \\
&= \infty.
\end{aligned}$$

因为

$$xP(X_i > x) = \frac{e}{x} \to 0 \quad (x \to \infty),$$

所以 $X_1, X_2, \cdots$ 满足弱大数定律条件，那么取

$$\begin{aligned}
\mu_n &= \mathrm{E}X_1 1_{\{|X_1| \leqslant n\}} \\
&\to \mathrm{E}X_1 = \infty \quad (\text{单调收敛定理})
\end{aligned}$$

即可，由弱大数定律直接得到

$$\frac{S_n}{n} - \mu_n \to 0.$$

$\square$

练习 2.2.6:

1. Show that if $X \geqslant 0$ is integer - valued

$$\mathrm{E}X = \sum_{n \geqslant 1} P(X \geqslant n).$$

2. Find a similar expression for $\mathrm{E}X^2$.

证明:

1.
$$\begin{aligned}
\mathrm{E}X &= \int_0^\infty P(X > y)\,\mathrm{d}y \\
&= \sum_{n=0}^\infty \int_n^{n+1} P(X \geqslant n+1)\,\mathrm{d}y \\
&= \sum_{n=0}^\infty P(X \geqslant n+1) \\
&= \sum_{n=1}^\infty P(X \geqslant n).
\end{aligned}$$

2.
$$\begin{aligned}
\mathrm{E}X^2 &= \int_0^\infty yP(X > y)\,\mathrm{d}y \\
&= \sum_{n=0}^\infty P(X \geqslant n+1) \int_n^{n+1} y\,\mathrm{d}y \\
&= \sum_{n=0}^\infty \left(n + \frac{1}{2}\right) P(X \geqslant n+1) \\
&= \sum_{n=1}^\infty \left(n - \frac{1}{2}\right) P(X \geqslant n).
\end{aligned}$$

□

练习 2.2.7: Generalize Lemma 2.2.13 to conclude that if

$$H(x) = \int_{(-\infty,x]} h(y)\,\mathrm{d}y$$

with $h(y) \geqslant 0$, then

$$\mathrm{E}H(X) = \int_{-\infty}^\infty h(y)P(X \geqslant y)\,\mathrm{d}y\,.$$

An important special case is $H(x) = \exp(\theta x)$ with $\theta > 0$.

证明:

$$\begin{aligned}
\mathrm{E}H(X) &= \int_\Omega H(X(\omega))\,\mathrm{d}P \\
&= \int_\Omega \int_{-\infty}^\infty h(t) 1_{\{t \leqslant X(\omega)\}}\,\mathrm{d}t\,\mathrm{d}P \\
&= \int_{-\infty}^\infty \int_\Omega h(t) 1_{\{t \leqslant X(\omega)\}}\,\mathrm{d}P\,\mathrm{d}t \quad (\text{Fubini 定理}) \\
&= \int_{-\infty}^\infty h(t)P(X \geqslant t)\,\mathrm{d}t\,.
\end{aligned}$$

□

练习 2.2.8: An unfair "fair game". Let

$$p_k = \frac{1}{2^k k(k+1)},$$

$k = 1, 2, \cdots$ and $p_0 = 1 - \sum_{k \geqslant 1} p_k$.

$$\sum_{k=1}^{\infty} 2^k p_k = \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \cdots = 1,$$

so if we let $X_1, X_2, \cdots$ be i.i.d. with $P(X_n = -1) = p_0$ and

$$P\big(X_n = 2^k - 1\big) = p_k \quad \text{for} \quad k \geqslant 1$$

then $\mathrm{E}X_n = 0$. Let $S_n = X_1 + \cdots + X_n$. Use Theorem 2.2.11 with $b_n = 2^{m(n)}$ where $m(n) = \min\big\{m : 2^{-m} m^{-3/2} \leqslant n^{-1}\big\}$ to conclude that

$$\frac{S_n}{n/\log_2 n} \to -1 \quad \text{in probability.}$$

证明: 要使用三角随机变量列的弱大数定律，需要先验证 $\sum_{k=1}^n P(X_k > b_n) \to 0$。这是因为

$$\begin{aligned}
\sum_{i=1}^n P(X_k > b_n) = nP(X_1 > 2^m) \\
&= n \sum_{k=m+1}^{\infty} \frac{1}{2^k k(k+1)} \\
&\leqslant 2^m m^{3/2} \sum_{k=m+1}^{\infty} \frac{1}{2^k k^2} \\
&\leqslant \frac{1}{\sqrt{m}} \sum_{k=m+1}^{\infty} \frac{1}{2^{k-m}} \\
&= \frac{2}{\sqrt{m}} \\
&\to 0 \quad (n \to \infty).
\end{aligned}$$

上式最后一步中 $n \to \infty$ 时 $m \to \infty$。另外还需要验证 $b_n^{-2} \sum_{k=1}^n \mathrm{E}\overline{X}_k^2 \to 0$，这是因为

$$
\begin{aligned}
\frac{\sum_{k=1}^{n}\mathrm{E}\overline{X}_k^2}{b_n^2} &= \frac{n\Big(P(X_1=-1)+\sum_{k=1}^{m}(2^{2k}-1)p_k\Big)}{2^{2m}} \\
&\leqslant \frac{n\Big(1+\sum_{k=1}^{m}2^k/k^2\Big)}{2^{2m}} \\
&= Cn\frac{2^{m+1}}{m^2 2^{2m}} \\
&\leqslant 2^m C m^{3/2}\frac{2^{m+1}}{m^2 2^{2m}} \\
&= \frac{2}{\sqrt{m}} \to 0 \quad (n\to\infty).
\end{aligned}
$$

其中 C 为一常数。 □

练习 2.2.9: Weak law for positive variables. Suppose $X_1, X_2, \cdots$ are i.i.d., $P(0\leqslant X_i<\infty)=1$ and $P(X_i>x)>0$ for all x. Let

$$
\mu(s)=\int_0^s x\,\mathrm{d}F(x)
$$

and

$$
\nu(s)=\frac{\mu(s)}{s(1-F(s))}.
$$

It is known that there exist constants a_n so that $S_n/a_n\to 1$ in probability, if and only if $\nu(s)\to\infty$ as $s\to\infty$. Pick $b_n\geqslant 1$ so that $n\mu(b_n)=b_n$ (this works for large n), and use Theorem 2.2.11 to prove that the condition is sufficient.

证明: 先解释如何取到的 $n\mu(b_n)=b_n$。三角形式的弱大数定律的结论为

$$
\frac{S_n-\mathrm{E}\overline{S}_n}{a_n}\to 0,
$$

与题目中所要证明的形式对照得到

$$
a_n=\mathrm{E}\overline{S}_n=n\mathrm{E}\overline{X}_n=n\mu(a_n).
$$

所以只要取 $n\mu(b_n)=b_n=a_n$ 并使用弱大数定律，即可导出欲求结论。再证明 $n\to\infty$ 时 $b_n\to\infty$。这是容易的，因为

$$
\nu(b_n)=\frac{\mu(b_n)}{b_n(1-F(b_n))}=\frac{1}{n(1-F(b_n))}\to\infty,
$$

那么 $F(b_n)\to 1$，结合 $P(X_i>x)>0$ 知道 $b_n\to\infty$。

再证明弱大数定律的下一条件。

$$\sum_{k=1}^{n} P(X_k > b_n) = nP(X_1 > b_n) = n(1 - F(b_n)) = \frac{1}{\nu(b_n)} \to 0.$$

再证明弱大数定律的另一个条件。

$$\frac{\sum_{k=1}^{n} \mathrm{E}\overline{X}_k^2}{b_n^2} = \frac{n\mathrm{E}\overline{X}_1^2}{b_n^2}.$$

下考虑对 b_n^2/n 进行放缩（**重要**），有

$$\int_0^{b_n} \mu(x)\,\mathrm{d}x \leqslant b_n\mu(b_n) = \frac{b_n^2}{n},$$

那么有

$$\begin{aligned}\frac{\sum_{k=1}^{n} \mathrm{E}\overline{X}_k^2}{b_n^2} &= \frac{n\mathrm{E}\overline{X}_1^2}{b_n^2} \\ &\leqslant \frac{\displaystyle\int_0^{b_n} 2x(1-F(x))\,\mathrm{d}x}{\displaystyle\int_0^{b_n} \mu(x)\,\mathrm{d}x} \\ &= \frac{2b_n(1-F(b_n))}{\mu(b_n)} \\ &= 2n(1-F(b_n)) \\ &\to 0 \quad (n \to \infty).\end{aligned}$$

上式先后使用了洛必达法则、等式 $n\mu(b_n) = b_n$ 和 $\nu(b_n) \to \infty$。

综上，弱大数定律成立，从而明所欲证。 □

2.3. Borel-Cantelli 引理

练习 2.3.1: Prove that

$$P(\limsup A_n) \geqslant \limsup P(A_n)$$

and

$$P(\liminf A_n) \leqslant \liminf P(A_n).$$

证明: 有 $\liminf 1_{A_n} = 1_{\liminf A_n}$，然后直接用 *Fatou* 引理即可。另一个结论同理。
□

练习 2.3.2: Prove the first result in Theorem 2.3.4 directly from the definition.

练习 2.3.3: Let ℓ_n be the length of the head run at time. See Example 8.3.1 for the precise definition. Show that

$$\limsup_{n\to\infty} \frac{\ell_n}{\log_2 n} = 1, \quad \liminf_{n\to\infty} \ell_n = 0 \text{ a.s.} .$$

练习 2.3.4: Suppose $X_m \geqslant 0$ and $X_n \to X$ in probability. Show that

$$\liminf_{n\to\infty} \mathrm{E}X_n \geqslant \mathrm{E}X.$$

练习 2.3.5: Suppose $X_n \to X$ in probability, and:
1. $|X_n| \leqslant Y$ with $\mathrm{E}Y < \infty$, or
2. There is a continuous function g with $g(x) > 0$ for large x with

$$\frac{|x|}{g(x)} \to 0 \quad (x \to \infty)$$

   so that $\mathrm{E}g(x) \leqslant C < \infty$ for all n. Show that $\mathrm{E}X_n \to \mathrm{E}X$.

练习 2.3.6: Show:
1. $$d(X, Y) = \mathrm{E}\frac{|X - Y|}{1 + |X - Y|}$$

   defines a metric on the set of random variables;
2. $d(X_n, X) \to 0$ as $n \to \infty$ if and only if $X_n \xrightarrow{P} X$.

练习 2.3.7: Show that random variables are a complete space under the metric defined in the previous exercise, i.e. , if $d(X_m, X_n) \to 0$ whenever $m, n \to \infty$ then there is a r.v. X_∞ so that $X_n \to X_\infty$ in probability.

练习 2.3.8: Let A_n be a sequence of independent events with $P(A_n) < 1$ for all n. Show that $P(\cup A_n) = 1$ implies $\sum_n P(A_n) = \infty$ and hence $P(A_n \text{ i.o.}) = 1$.

练习 2.3.9:
1. If $P(A_n) \to 0$ and

$$\sum_{n=1}^{\infty} P(A_n^c \cap A_{n+1}) < \infty$$

then $P(A_n \text{ i.o.}) = 0$.

2. Find an example of a sequence A_n to which the result in (1) can be applied but the Borel - Cantelli lemma cannot.

练习 2.3.10: Kochen - Stone lemma. Suppose $\sum P(A_k) = \infty$. Use Exercises 1.6.6 and 2.3.1 to show that if

$$\limsup_{n\to\infty} \frac{\left(\sum_{k=1}^{n} P(A_k)\right)^2}{\sum_{1\leqslant j,k\leqslant n} P(A_j \cap A_k)} = \alpha > 0$$

then $P(A_n \text{ i.o.}) \geqslant \alpha$. The case $\alpha = 1$ contains Theorem 2.3.7.

练习 2.3.11: Let $X_1, X_2, \cdots$ be independent with $P(X_n = 1) = p_n$ and $P(X_n = 0) = 1 - p_n$. Show that:

1. $X_n \to 0$ in probability if and only if $p_n \to 0$, and
2. $X_n \to 0$ a.s. if and only if $\sum p_n < \infty$.

练习 2.3.12: Let $X_1, X_2, \cdots$ be a sequence of r.v.'s on $(\Omega, \mathscr{F}, P)$ where Ω is a countable set and $\mathscr{F}$ consists of all subsets of Ω. Show that $X_n \to X$ in probability implies $X_n \to X$ a.s.

练习 2.3.13: If X_n is any sequence of random variables, there are constants $c_n \to \infty$ so that $X_n / c_n \to 0$ a.s.

练习 2.3.14: Let $X_1, X_2, \cdots$ be independent. Show that $\sup X_n < \infty$ a.s. if and only if

$$\sum_n P(X_n > A) < \infty$$

for some A.

练习 2.3.15: Let $X_1, X_2, \cdots$ be i.i.d. with $P(X_i > x) = e^{-x}$, let

$$M_n = \max_{1 \leqslant m \leqslant n} X_m.$$

Show that

1.
$$\limsup_{n \to \infty} \frac{X_n}{\log n} = 1 \text{ a.s.}$$

and

2.
$$\frac{M_n}{\log n} \to 1 \text{ a.s. }.$$

练习 2.3.16: Let $X_1, X_2, \cdots$ be i.i.d. with distribution F, let $\lambda_n \uparrow \infty$, and let

$$A_n = \left\{ \max_{1 \leqslant m \leqslant n} X_m > \lambda_n \right\}.$$

Show that $P(A_n \text{ i.o.}) = 0$ or 1 according as $\sum_{n \geqslant 1} (1 - F(\lambda_n)) < \infty$ or $= \infty$.

练习 2.3.17: Let $Y_1, Y_2, \cdots$ be i.i.d. Find necessary and sufficient conditions for:

1. $Y_n / n \to 0$ almost surely;
2.
$$\frac{\max_{m \leqslant n} Y_m}{n} \to 0$$

almost surely;

3.
$$\frac{\max_{m \leqslant n} Y_m}{n} \to 0$$

in probability;

4.
$$\frac{Y_n}{n} \to 0$$

in probability.

练习 2.3.18: Let $0 \leqslant X_1 \leqslant X_2 \cdots$ be random variables with $EX_n \sim an^\alpha$ with $a, \alpha > 0$, and var $(X_n) \leqslant Bn^\beta$ with $\beta < 2\alpha$. Show that

$$\frac{X_n}{n^\alpha} \to a \text{ a.s. }.$$

练习 2.3.19: Let X_n be independent Poisson r.v.'s with $EX_n = \lambda_n$, and let $S_n = X_1 + \cdots + X_n$. Show that if $\sum \lambda_n = \infty$ then

$$\frac{S_n}{ES_n} \to 1 \text{ a.s. } .$$

练习 2.3.20: Show that if X_n is the outcome of the nth play of the St. Petersburg game (Example 2.2.16) then

$$\limsup_{n\to\infty} \frac{X_n}{n \log_2 n} = \infty \text{ a.s.}$$

and hence the same result holds for S_n. This shows that the convergence

$$\frac{S_n}{n \log_2 n} \to 1$$

in probability proved in Section 2.2 does not occur a.s. .