



随机问题

Enlightments

作者: syqwq

时间: 2025.10.02

邮箱: 3422403944@qq.com



生如夏花之绚烂，死如秋叶之静美

Contents

1. 测度论与概率论	1
1.1. 三大重要收敛定理	1
1.2. 为什么要有概率密度函数?	5

Chapter 1

测度论与概率论

1.1 三大重要收敛定理

定理 1.1.1 单调收敛定理(Monotone Convergence Theorem - MCT)

若 $(\Omega, \mathcal{F}, \mu)$ 是一个测度空间, $\{f_i\}_{i=1}^{\infty}$ 是一族 $\Omega \rightarrow \mathbb{R}$ 的可测函数, 满足 $f_i \uparrow f$ a.e. 且 $\int f_1 d\mu > -\infty$. 则积分和极限可以交换, 且积分也单调收敛, 即

$$\int f_i d\mu \uparrow \int f d\mu$$

注解 1.1.1 热气球

- 想象 f_i 是一个热气球在时刻 i 的体积, $\int f_i d\mu$ 是它的总浮力.
- 条件 $f_i \uparrow f$ 意味着这个热气球只增不减, 不断地膨胀, 最终趋近于一个最终形态 f .
- MCT 定理说: 既然热气球一直在稳定地、单调地变大, 那么它浮力的极限, 理所当然就是它最终形态的浮力. 这个过程中没有“泄气”或者其他诡异的事情发生.

证明. 我们将证明分为两个主要部分: 首先处理更简单也更基础的非负函数情况, 然后利用它来证明一般情况.

先假设 $f_1 \geq 0$ 相当于证明等式:

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu$$

考虑一边, 即证明

$$\lim \int f_n d\mu \leq \int f d\mu$$

这是显然的. 因为 $f_n \uparrow f$ 意味着 $\forall n, f_n(\omega) \leq f(\omega)$, 从而 $\int f_n d\mu \leq \int f d\mu$. 因为 $\int f_n d\mu$ 单调有上界, 因此 $\lim \int f_n d\mu$ 存在, 且 $\lim \int f_n d\mu \leq \int f d\mu$.

接下来考虑另一侧

$$\lim \int f_n d\mu \geq \int f d\mu$$

回顾 $\int f d\mu$ 的定义是

$$\int f d\mu := \sup \left\{ \int g d\mu \mid 0 \leq g \leq f, g \text{ 为简单函数} \right\}$$

如果我们可以证明对于任意的简单函数 g , 都有 $c := \lim \int f_n d\mu \geq \int g d\mu$, 那么一定可以得到 c 比他们的上确界大, 从而得证.

为此, 任取简单函数 g s.t. $0 \leq g \leq f$, 再取一个略小于 1 的数 $a := 1 - \varepsilon$. 构造集合

$$E_n := \{\omega \mid f_n(\omega) \geq ag(\omega)\}$$

这个集合 E_n 是指那些 f_n 的值已经“长得足够高”, 至少达到了 g 的 a 倍的点 ω 的集合. 由于 f_n 单调递增, 因此这个集合也单调, 即 $E_1 \subseteq E_2 \subseteq \dots$. 因为 $f_n(\omega) \uparrow f(\omega)$ 且 $f(\omega) \geq g(\omega)$, 这个集合会不断扩大, 直至整个 g 的支撑集, 也就是

$$\bigcup_{n=1}^{\infty} E_n = \{\omega \mid g(\omega) > 0\}$$

从而, 得到

$$\int f_n d\mu \geq \int_{E_n} f_n d\mu \geq a \int_{E_n} g d\mu$$

不等式两侧取极限得到

$$c \geq a \cdot \lim_{n \rightarrow \infty} \int_{E_n} g d\mu = a \cdot \int g d\mu$$

因为 g 是任取的, 且 a 可以任意接近 1, 因此就有

$$c \geq \sup \int g d\mu = \int f d\mu$$

从而 f_1 非负部分证完了.

接下来, 证明另一部分.

我们构造函数列 $f_n' = f_n - f_1$, 则 $f_n' \geq 0$ 且 $f_n' \uparrow f' := f - f_1$, 因此现在的 f_n' 满足上面情况的所有条件, 从而我们得到

$$\int f_n' d\mu \uparrow \int f' d\mu \Leftrightarrow \int (f_n - f_1) d\mu \uparrow \int (f - f_1) d\mu$$

由于 $\int f_1 d\mu > -\infty$ 利用积分的线性性, 得到

$$\int f_n d\mu \uparrow \int f d\mu$$

■

定理 1.1.2 Fatou 引理(Fatou's Lemma)

若 $(\Omega, \mathcal{F}, \mu)$ 是一个测度空间, $\{f_i\}_{i=1}^{\infty}$ 是一族 $\Omega \rightarrow \mathbb{R}$ 的非负可测函数. 则

$$\int \liminf_{i \rightarrow \infty} f_i d\mu \leq \liminf_{i \rightarrow \infty} \int f_i d\mu$$

注解 1.1.2 一个可能漏水的水桶

- 想象 $\int f_i d\mu$ 是一个桶在时刻 i 的总水量. f_i 是水的分布.
- f_i 不要求单调, 你可以把水在桶里晃来晃去, 甚至有些水会溅出来.
- $\liminf f_i$ 是在晃动过程中, 每个位置“最终”稳定留下的水位.
- Fatou 引理说:** 最终桶里剩下的水量 (左边), 小于等于你每次测量水量的极限 (右边) .
- 为什么是不等式? 因为在晃动的过程中 (f_i 不单调), 可能有一部分“质量”(积分值)泄露掉了或“蒸发”了(数学上叫“跑到无穷远处”). 所以你最终看到的实体 ($\liminf f_i$) 的积分, 可能会比积分的极限要小.

证明. 我们手里的工具是强大的**单调收敛定理(MCT)**, 但它的使用条件很苛刻, 要求函数序列是单调递增的. 而 Fatou 引理的 $\{f_i\}$ 序列不保证单调性. 所以, 证明的核心思想是: 从不单调的 $\{f_i\}$ 出发, 构造出一个相关的、新的、单调递增的序列 $\{g_j\}$, 然后对 $\{g_j\}$ 应用 MCT, 最后再把结果和 $\{f_i\}$ 联系起来.

我们先回顾 \liminf 的定义:

$$\liminf_i f_i := \sup_j \inf_{i \geq j} f_i$$

受到启发, 我们定义新的函数列 g_j

$$g_j := \inf\{f_i \mid i \geq j\}$$

也就是 g_j 是原序列 f_i 从第 j 项开始的“尾巴”的下确界. 由于所选取的集合越来越小, 因此下确界是递增的, 从而有 $g_1 \leq g_2 \leq \dots$, 这就意味着我们构造出了一个单调递增的序列.

而根据 \liminf 的定义, 这个单调递增序列 $\{g_j\}$ 的极限 (也就是它的上确界 $\sup g_j$) 正好就是 $\liminf f_i$. 所以我们有: $g_j \uparrow \liminf f_i$. 因为 f_i 都非负, 所以 g_j 也非负. 应用 MCT, 我们得到

$$\lim_{j \rightarrow \infty} \int g_j d\mu = \int \liminf_{i \rightarrow \infty} f_i d\mu$$

而这就是我们要证明的不等式的左边, 接下来我们考虑右边.

由于 g_j 的定义 $g_j = \inf\{f_i \mid i \geq j\}$, 于是 $\forall i \geq j, f_i \geq g_j$, 利用积分的单调性有

$$\forall i \geq j, \int f_i d\mu \geq \int g_j d\mu$$

既然 $\int g_j d\mu$ 比这里的每一项都要小, 那么一定小于等于他们的下确界, 从而有关键不等式

$$\int g_j d\mu \leq \inf_{i \geq j} \int f_i d\mu$$

我们对上面这个不等式两边同时取 $j \rightarrow \infty$ 的下极限 \liminf

$$\lim_{j \rightarrow \infty} \int g_j d\mu = \liminf_{j \rightarrow \infty} \int g_j d\mu \leq \liminf_{j \rightarrow \infty} \left(\inf_{i \geq j} \int f_i d\mu \right)$$

由于 g_j 是单调收敛的序列，因此下极限等于极限等于 $\int \liminf_{i \rightarrow \infty} f_i d\mu$. 在这个不等式的右边，被取极限的部分正好符合 \liminf 的定义，因此就等于 $\liminf \int f_j d\mu$ ，把两侧的分析结果带入，就得到了我们要证明的 Fatou 引理

$$\int \liminf_{i \rightarrow \infty} f_i d\mu \leq \liminf_{i \rightarrow \infty} \int f_i d\mu$$

■

定理 1.1.3 控制收敛定理 (Dominated Convergence Theorem - DCT)

若 $(\Omega, \mathcal{F}, \mu)$ 是一个测度空间， $\{f_i\}_{i=1}^{\infty}$ 是一族 $\Omega \rightarrow \mathbb{R}$ 的可测函数， g 是一个可积函数。如果 $|f_i| \leq g$ a.e. 且 $\forall \omega \in \Omega, f_i(\omega) \rightarrow f(\omega)$ (逐点收敛)，则 f 也是可积的，且积分和极限可交换，即

$$\int f_i d\mu \rightarrow \int f d\mu \Leftrightarrow \int f d\mu = \int \lim_{n \rightarrow \infty} f_n d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu$$

注解 1.1.3 有盖子的桶

- 这就像 Fatou 引理里那个可能漏水的桶，但现在我们给它加了一个坚固的盖子 (g)。
- $f_i \rightarrow f$ 意味着桶里的水最终会形成一个稳定的形态 f 。
- $|f_i| \leq g$ 且 $\int g d\mu < \infty$ 这个“盖子”的作用是阻止任何水的泄露和蒸发。因为所有的水都被限制在一个总容量有限的空间里。
- DCT 定理说：既然一滴水都没有损失，那么水量的极限，自然就等于极限形态下水的总量。

证明. 证明的核心思想是构造两条“单调的夹板”来挤压 原始的序列 $\{f_i\}$ 本身不一定是单调的，所以我们不能直接用 MCT. 因此，我们从 $\{f_i\}$ 出发，构造出两个新的、单调的函数序列，一个从下方逼近 f ，一个从上方逼近 f ，像两块夹板一样把 f_i 夹在中间。然后我们分别对这两条“单调的夹板”应用 MCT.

我们定义两个辅助序列：

(1) 地板序列 f_i^{\wedge}

$$f_i^{\wedge} := \inf\{f_j \mid j \geq i\}$$

这个序列是单调递增的，并且收敛到 $\liminf f_i$. 因为 $f_i \rightarrow f$ ，所以 $f_i^{\wedge} \uparrow f$.

(2) 天花板序列 f_i^{\vee}

$$f_i^{\vee} := \sup\{f_j \mid j \geq i\}$$

这个序列是单调递减的，并且收敛到 $\limsup f_i$. 因为 $f_i \rightarrow f$ ，所以 $f_i^{\vee} \downarrow f$.

通过定义，我们得到

$$f_i^{\wedge} \leq f_i \leq f_i^{\vee} \tag{1.1.1}$$

由于 g 的存在保证了有限且可积，因此对于地板序列应用 MCT，我们得到

$$\lim_{i \rightarrow \infty} \int f_i^\wedge d\mu = \int f d\mu$$

类似的，应用 MCT（递减版），我们得到

$$\lim_{i \rightarrow \infty} \int f_i^\vee d\mu = \int f d\mu$$

对式(1.1.1)取积分，得到

$$\int f_i^\wedge d\mu \leq \int f_i d\mu \leq \int f_i^\vee d\mu$$

再通过取极限，应用夹逼定理得到

$$\lim_{i \rightarrow \infty} \int f_i d\mu = \int f d\mu$$

■

1.2 为什么要有概率密度函数？

为了接下来的讨论，我们首先要有一些概念的定义上达成共识。

定义 1.2.1 可测函数

对于两个可测空间 $(\Omega, \mathcal{F}), (S, \mathcal{S})$ ，如果存在一个函数 $X : \Omega \rightarrow S$ 满足

$$X^{-1}(B) := \{\omega \mid X(\omega) \in B\} \in \mathcal{F}, \forall B \in \mathcal{S}$$

那么，称 X 是一个从 (Ω, \mathcal{F}) 到 (S, \mathcal{S}) 的可测函数。

注解 1.2.1

这说的就是对于任何一个事件 $\omega \in \mathcal{F}$ ，我们固然可以得知他的结果 $X(\omega)$ ，而可测函数的定义是在说，我任意给定一个 $B \in \mathcal{S}$ ，他一定有 \mathcal{F} 中的可测事件与他对应。

定义 1.2.2 随机向量、随机变量

在定义 1.2.1 中，如果 $(S, \mathcal{S}) = (\mathbb{R}^d, \mathcal{R}^d)$ ，当 $d > 1$ 时，我们称 X 为随机向量；当 $d = 1$ 时，我们称 X 为随机变量。

我们为了研究一个事件发生的概率，例如抛出一枚硬币是否是正面朝上，一个很直观的想法就是把所有可能的情况列出来，而随机变量的作用就是给某些情况的集合一个“标签”，就如同拎葡萄，抓着柄就抓了一串，并且随机变量的定义也保证了这样拎出来的集合一定也是可测的。这样就可以让我们在讨论某些具有共性的集合的时候更加得心应手。

在抛硬币中，我们假设 X 是一个随机变量，它的取值为 0（反面）或 1（正面）。那么，我们可以很清晰地通过分布列（PMF）来研究整个样本空间中每个事件发生的概率：

	$X = 0$	$X = 1$
P	1/2	1/2

但是, 当问题来到连续型随机变量时, 这就并不奏效了.

考虑这样一个场景, 我们有一个理想的随机数生成器, 会随机地、均匀地从 $[0, 1]$ 中生成一个数字 X . 现在, 让我们尝试用分布列的逻辑来描述它. 这意味着我们要给 $[0, 1]$ 区间内的每一个点都赋予一个概率值 $P(X = x)$. 那么此时, 灾难性的问题出现了:

- 区间内有多少点? 无穷个.
- 我们应该给每个 $P(X = x)$ 赋予多大的值?
 → 假设我们给每个点的概率是一个很小的 $\varepsilon > 0$, 那么总概率

$$\sum_{x \in [0, 1]} P(X = x) = \sum_{x \in [0, 1]} \varepsilon = \infty \times \varepsilon = \infty$$

一发不可收拾了 😱

- 既然不能是正数, 那么只能是 0 了. 但是

$$\sum_{x \in [0, 1]} P(X = x) = \sum_{x \in [0, 1]} 0 = 0$$

这与 $P(\Omega) = 1$ 的要求不符, 也不对啊 😭

分布列这个工具, 它的“给每个点分配一块概率, 然后加总”的核心思想, 在面对连续变量的“不可数无穷”特性时, 彻底失效了. 无论怎么给单点分配概率, 总概率要么是无穷, 要么是 0, 永远凑不成那个必需的“1”.

既然纠结于单点的概率是死路一条, 于是我们换一个角度: 能不能不讨论单点的概率, 而是讨论一个区间的概率? 这个思想的转变, 就是引入 PDF 的直觉来源.

(1) 放弃“点的质量”, 拥抱“点的密度”

想象一根一米长的金属棒, 总质量是 1 千克. 你问: “在 0.5 米那个点上, 质量是多少?” 答案是 0. 一个没有长度的点, 自然没有质量. 但是, 你可以问: “在 0.5 米那个点附近, 质量的密度是多少?” 这就有意义了, 比如是 1.2 公斤/米.

密度这个概念, 描述的不是一个点的质量, 而是这个点周围质量的集中程度.

(2) PDF 就是概率上的密度

概率密度函数 (PDF) $f(x)$ 就扮演了这个“密度”的角色. $f(x)$ 本身的值不是概率, 就像 “kg/m³” 不是 “kg” 一样. $f(x)$ 的值越大, 表示随机变量落在 x 附近的可能性就越“密集”或越“集中”.

(3) 如何从“密度”得到“概率”?

对于那根金属棒, 如何从密度得到一段区间的质量? 答案是积分! 把从 0.4 米到 0.6 米的密度函数积分, 就得到了这段的质量.

$$m = \int_L \lambda \, d\ell$$

同理, 对于连续型随机变量, 要计算它落在区间 $[a, b]$ 内的概率, 我们就需要对概率密度函数在这个区间上进行积分:

$$P(a \leq X \leq b) = \int_a^b f(x) \, dx$$

这个模型完美解决了以前的所有问题: 总概率 $\int_{-\infty}^{\infty} f(x) dx = 1$, 而单点的概率 $P(X = a) = \int_a^a f(x) dx = 0$. 所有矛盾都消失了.

因此, 我们就引出了如下的定义.

定义 1.2.3 连续型随机变量

在概率空间 (Ω, \mathcal{F}, P) 上, 如果随机变量 $X : \Omega \rightarrow \mathbb{R}$ 满足:

存在一个可测函数 $f : \mathbb{R} \rightarrow [0, \infty)$ 使得

$$P(X \leq x) = \int_{-\infty}^x f(t) dt, -\infty < x < \infty$$

那么称 X 为连续型随机变量, f 为其概率密度函数 (PDF).

但是, 这里有一个小问题: 我们之前的讨论都是对于区间 $[a, b]$ 的概率, 而定义中却是 $(-\infty, x]$ 的概率, 这是否合理呢? 答案是肯定的. 事实上, 由所有形如 $(-\infty, x]$ 的半无限区间构成的集合, 是整个实数轴上 Borel 代数 $\mathcal{B}(\mathbb{R})$ 的一个生成元, 也就是通过可数次交并补操作得到整个 $\mathcal{B}(\mathbb{R})$. 而对于具有这样性质的集合, 我们有一个很神奇的定理

定理 1.2.1 Carathéodory 扩张定理

设 \mathcal{A} 是集合 X 上的代数, μ_0 是集合 \mathcal{A} 上的一个预测度. 令 μ^* 是由 μ_0 导出的外测度. 那么所有 μ^* -可测集构成的集合 \mathcal{M} 满足:

- (1) \mathcal{M} 是一个 σ -代数
- (2) $\mathcal{A} \subseteq \mathcal{M}$
- (3) 这个测度 μ 是预测度 μ_0 的一个扩张, 即对于任何在原始代数中的集合 $A \in \mathcal{A}$, 都有 $\mu(A) = \mu_0(A)$.

如果在上述条件下, 预测度 μ_0 是 σ -有限的 (即存在一系列 $A_i \in \mathcal{A}$ s.t. $\bigcup_{i=1}^{\infty} A_i = X$ 且 $\mu_0(A_i) < \infty$), 那么这个扩张是唯一的.

既然所有我们关心的区间 (以及更复杂的集合) 都可以由 $(-\infty, x]$ 这种“基本砖块”通过集合运算搭建而成, 那么根据 定理 1.2.1 中测度的唯一性, 我们可以得出一个结论:

只要我们在“生成元”这个集合上定义好了概率, 那么所有由它生成的更复杂的集合 (所有 Borel 集) 上的概率也就被唯一地确定了.

感谢大家的补充解释, 以下是我基于个人理解的一些整理. 🎉🎉🎉

同样从金属棒入手, 想象一根一米长的密度不均匀的金属棒. 我们现在有两种测度:

- (1) 质量测度 μ : 这是一个函数, 你给它棒上的任意一段区间 A , 它会返回这段区间的质量 $\mu(A)$
- (2) 长度测度 λ : 这是一个函数, 你给它任意一段区间 A , 它会返回这段区间的长度 $\lambda(A)$. 这本质上就是勒贝格测度.

这根棒的总质量 $\mu([0, 1])$ 是 1 千克. 但由于密度不均, 我们不能简单地说“质量等于长度”. 那么, 我们如何描述在每一个点 x 上, 质量相对于长度的“集中程度”呢?

对于这种问题，我们考虑“求导”的思想。在微积分中， dy/dx 描述了 y 相对于 x 的变化率。类似的，我们想定义一个“测度对于测度的导数” $d\mu/d\lambda$ 。这个导数是一个函数，我们称为密度函数 $\rho(x)$ ，他在某个点 x 的值，代表了该点单位长度上所拥有的质量。

那么怎么求这个值？我们可以想象在点 x 附近取一个极小的区间 ΔA ，然后计算这个区间的质量与长度之比： $\mu(\Delta A)/\lambda(\Delta A)$ 。当这个小区间缩向点 x 时，这个比值的极限就是点 x 的密度 $\rho(x)$ 。一旦我们求出了这个密度函数 $\rho(x)$ （也就是那个“导数”），我们就可以通过积分来反求任何一段区间的总质量：

$$\mu(A) = \int_A \rho(x) d\lambda(x)$$

这个过程，就是从导数（密度）通过积分还原原函数（质量）的过程。

那么，接下来问题回到概率论中。现在我们也有两种测度：

- (1) 概率测度 P_X ：这是由随机变量 X 导出的测度。你给它实数轴上任意一个事件（Borel 集） A ，它返回 X 落入 A 的概率 $P_{X(A)}$ 。这对应了上面例子中的“质量测度 μ ”。
- (2) 勒贝格测度 λ ：这是实数轴上标准的“长度”测度。

在进行求导之前，必须满足一个关键条件，就是测度的绝对连续性 (Absolute Continuity)。

定义 1.2.4 测度的绝对连续性

如果 μ, ν 是同一个测度空间上的两个测度，那么我们称 μ 相对于 ν 是绝对连续的，如果对于任何集合 A ，只要 $\nu(A) = 0$ ，那么 $\mu(A) = 0$ ，记为 $\mu \ll \nu$ 。形式化地说就是

$$\mu \ll \nu \Leftrightarrow (\nu(A) = 0 \Rightarrow \mu(A) = 0)$$

在这里，我们说概率测度 P_X 相对于勒贝格测度 λ 是绝对连续的，如果对于任何集合 A ，只要它的“长度”为零 $\lambda(A) = 0$ ，那么它的“概率”也必须为零 $P_X(A) = 0$ ，显然单点集就是这样的零测集。

注解 1.2.2 绝对连续性

直观含义就是“没有长度的地方，就没有概率”。这意味着概率不会“凭空”集中在一些没有长度的“点”或“尘埃”上。这完美地排除了离散概率（集中在单点上），并从根本上解释了为什么连续随机变量在单点的概率为 0。

定理 1.2.2 拉东-尼科迪姆定理 (The Radon-Nikodym Theorem)

如果测度空间 (X, Σ) 上的一个 σ -有限测度 ν 关于另一个 σ -有限测度 μ 绝对连续，那么存在一个在 X 上的非负可测函数 $f : X \rightarrow [0, \infty)$ ，也就是 μ 关于 ν 的密度，或者 Radon-Nikodym 导数 $f = d\mu/d\nu$ 。对所有的可测集合 A ，都有：

$$\mu(A) = \int_A f d\nu$$

那么在概率论的语境下，如果概率测度 P_X 相对于勒贝格测度 λ 是绝对连续的，那么根据拉东-尼科迪姆定理，必然存在一个唯一的（几乎处处唯一）非负可积函数 f_X ，使得对于任何事件 A ，都有：

$$P_X(A) = \int_A f_X(x) d\lambda(x)$$

这个必然存在的函数 f_X 就是 P_X 相对于 λ 的 Radon-Nikodym 导数, 我们给这个函数起了一个专门的名字: **概率密度函数 (PDF)**.