

概率论教程

任佳刚 刘继成

记号与约定

1. 具体值不重要的常数统一记为 C . 因此你会看到诸如

$$|x - y| \leq C, |y - z| \leq C \implies |x - z| \leq C$$

的推理. 不要惊诧!

- 2.

$$A := B, B =: A \text{ 将} A \text{ 定义为} B, \text{ 将} B \text{ 记为} A.$$

- 3.

$$\mathbb{N} := \{0, \pm 1, \pm 2, \dots\}, \mathbb{N}_+ := \{0, 1, 2, \dots\}, \mathbb{N}_{++} := \{1, 2, \dots\}.$$

- 4.

$$\mathbb{R} := (-\infty, \infty), \bar{\mathbb{R}} := [-\infty, \infty], \mathbb{R}_+ := [0, \infty).$$

5. \mathbb{Q} : \mathbb{R} 中的有理数全体; $\mathbb{Q}_+ = \{r \in \mathbb{Q}, r \geq 0\}$; $\mathbb{Q}_{++} = \{r \in \mathbb{Q}, r > 0\}$.

6. $\mathbb{C} := \{x + iy, x, y \in \mathbb{R}\}$ 复数的全体; $\forall z = x + iy \in \mathbb{C}$,

$$\operatorname{Re}(z) := x, \quad \operatorname{Im}(z) := y$$

分别称为 z 的实部和虚部.

- 7.

$$\delta_{ik} = \begin{cases} 1, & i = k, \\ 0, & i \neq k. \end{cases}$$

8. $\forall a \in \mathbb{R}$,

$$a^+ = a \vee 0 := \max\{a, 0\}, \quad a^- := -(a \wedge 0) := \min(a, 0),$$

分别称为 a 的正部与负部.

9. $\forall a \in \mathbb{R}_+$, $[a]$ 表示不超过 a 的最大整数.

10. $\forall a \in \mathbb{R}^n$, a' 表示向量 a 的转置向量; 若 A 为矩阵, A' 表示 A 的转置矩阵.

11. $\forall a = (a_1, \dots, a_n), b = (b_1, \dots, b_n) \in \mathbb{R}^n$,

$$a \cdot b := ab' := a_1 b_1 + \dots + a_n b_n,$$

$$|a|^2 := a_1^2 + \dots + a_n^2.$$

- 12.

$$C_0^\infty := C_0^\infty(\mathbb{R}^m) := \{f : \mathbb{R}^m \mapsto \mathbb{R} : f \text{ 无穷次可微且在某个有界集外恒等于零}\},$$

$$C_0 := C_0(\mathbb{R}^m) := \{f : \mathbb{R}^m \mapsto \mathbb{R} : f \text{ 连续且在某个有界集外恒等于零}\},$$

$$C_b := C_b(\mathbb{R}^m) := \{f : \mathbb{R}^m \mapsto \mathbb{R} : f \text{ 有界连续}\}.$$

前言

王蒙说过, 写作是对时光的一种挽留.¹

我们写作这本书时, 常常会回到我们自己初学概率论的时光. 当时我们这门课是和实变函数同时开的, 记得老师为了等待实变的一些内容先讲, 还特意停了两周的课. 但即便是这样, 概率论里头仍然有一些无法讲清的东西. 比如说, 在证明Kolmogorov不等式的时候, 设 ξ_1, ξ_2, \dots 是独立随机变量, 令

$$S_n := \sum_{i=1}^n \xi_i,$$

$$\tau := \inf\{n : |S_n| \geq \varepsilon\}.$$

那么需要用到 $S_k 1_{\{\tau=k\}}$ 和 $(\xi_{k+1}, \xi_{k+2}, \dots)$ 的独立性. 关于这点, 流行的教材中都是以“ $S_k 1_{\{\tau=k\}}$ 只与 ξ_1, \dots, ξ_k 有关”而一言以蔽之的. 但基本上这属于混过去的, 不能算数学证明. 类似的这些问题常常——曾经常常——使作为学生的我们非常困惑.

而当我们自己做了教师之后, 长期的教学实践又使我们深刻地认识到, 如果没有测度论的基础, 很多概率论的东西实际上是无法完全讲清的. 那么, 有些地方只能先混着吗? 一代代地混下去吗? 不是已经混了很多代了吗, 再混一混又何尝不可呢?

我们的理想是做到完全严格, 但这门课程的性质又决定了它不可能完全严格. 怎样在理想和现实之间找到平衡? 这是我们作为教师的一种困惑.

无须讳言, 作为学生, 概率论中最先吸引我们的, 是组合概率. 诸如一堆人的帽子混在一起, 每人随意拿一项, 都拿到自己帽子的概率是多少? n 个男生和 n 个女生随意坐在一张有 $2n$ 个座位的圆桌旁, 一个男生隔着一个女生的概率是多少?..... 这些问题是如此迷人又如此困难, 具有强烈的吸引力. 但同样无须讳言的是, 作为教师, 我们的理解是, 这些纯技巧性的问题相对于整个概率论的现代理论来说, 其实是不那么重要的. 那么, 如何处理这一部分即如何在有趣与重要性之间找到平衡呢? 这是我们的另一种困惑.

以下是我们的一些思考和实践:

1. 难以完全严格讲述的部分, 实际上就是测度的扩张, 包括以此为基础的乘积概率空间, 因为它一方面比较冗长, 一方面是其它课程如实变函数的主要内容, 如果在此花费较长时间, 无疑是有喧宾夺主了. 除此之外, 其它的似乎要严格也不是想象的那么难. 比如单调类定理, 我们觉得还是可以从容地讲清楚的, 同时它基本上也不是实变函数课程的内容. 而有了单调类定理, 很多曾经令我们困惑的东西, 比如说上面讲的独立性, 都可以比较容易地严格地讲清楚. 因此, 测度的扩张和一切赖于此的东西, 我们都直接承认, 但详细讲述了并大量使用了单调类定理.

2. 我们大量减少了古典概型的具体计算的问题, 尤其是高度技巧性的问题, 但保留了一些理论上原则上或历史上重要的问题.

在《高等数学引论》的序言里, 华罗庚说: 他的想法“绝大部分写不下来, 或者写下来就走了样”; 类似的话王蒙也说过, 大意是他想说的大部分写不下来, 写下来中的大部分读者也理解不了(很遗憾不记得原话和原话的出处了). 我们写作时也深有同感.

写作, 是对时光的一种挽留, 但时光是留不住的啊. 所以王蒙也说过: 所有的日子都去吧, 都去吧.....

¹王蒙的原话是: 文学是对时光的一种挽留. 但文学是一个学科名词, 挽留是一个动作名词或动名词, 两者不能等同. 因此我们理解他的原意应该是“写作是对时光的一种挽留”.

但同时,写作也是对未来的一种留言.我们学过、想过、教过概率论多年,我们依然还在学、在想、在教,我们希望把我们在这些过程中的心得体会记录下来,留给未来,留给未来的年轻人,希望能提高一点你们对概率论的兴趣,希望能对你们学习、掌握、研究、应用概率论提供哪怕一丁点帮助.我们希望这不是奢望.

目录

第一章 试验及事件的运算

1.1 试验及其数学描述	8
1.2 事件的关系与运算	12

第二章 离散概型

2.1 古典概型	18
2.2 有限概型	25
2.3 随机变量	29
2.4 条件概率	34
2.5 全概率公式	39
2.6 独立性	44
2.7 π -类, λ_0 -类与代数	47
2.8 重温独立性	52
2.9 可列概型	54
2.10 二项分布的Poisson分布近似: Stein-Chen方法	68

第三章 公理概型

3.1 动因	73
3.2 σ -代数, 单调类与 λ -类	74
3.3 概率的公理化	78
3.4 条件概率, 独立性与条件独立性	85

第四章 随机变量

4.1 基本概念	93
4.2 分布函数	103
4.3 分类	105
4.4 多维随机变量的分布函数	114
4.5 条件分布	121
4.6 随机变量的存在性	124
4.7 随机变量的函数	126
4.8 多维随机变量的函数	128

第五章 期望与积分

5.1 简单随机变量情形	131
5.2 一般情形	135
5.3 计算期望的例子	140

5.4	随机变量列的收敛性	148
5.5	积分(期望)号下取极限	151
5.6	Fubini定理	158
5.7	带参数的期望	160
5.8	Riemann-Stieltjes积分	162
5.9	Lebesgue-Stieltjes积分	169
5.10	方差与矩	172
5.11	几个等式与不等式	179
5.12	条件期望	184
第六章 随机变量的独立性		
6.1	基本定义及性质	187
6.2	不相关与独立的关系	198
6.3	独立随机变量之和	202
6.4	随机游动	206
6.5	条件独立性	210
第七章 大数定律		
7.1	Markov大数定律	212
7.2	强大数定律	215
7.3	Glivenko-Cantelli定理	221
第八章 特征函数		
8.1	复随机变量	223
8.2	定义及例子	224
8.3	基本性质	227
8.4	唯一性定理	234
8.5	连续性定理	240
8.6	多维情形	248
8.7	Skorokhod表现定理	254
第九章 特征函数的应用		
9.1	分布的计算	257
9.2	极限定理	261
9.3	正态分布	265
第十章 中心极限定理中的余项估计		
10.1	Lindeberg定理	268
10.2	几个分析引理	270
10.3	分布余项的积分估计	275
10.4	余项的一致估计	278
第十一章 附录		
11.1	Stirling公式	280
11.2	Bihari-LaSalle不等式及其推论	281
11.3	绝对连续函数	284

11.4 函数的磨光	284
11.5 指数函数的逼近	285
11.6 常用分布	287

1 试验及事件的运算

世界是偶然的或者说随机的. 想想看, 比如宏观方面, 关于宇宙的起源, 你如果不信上帝的话, 就得信大爆炸理论; 可是你如果信上帝的话, 上帝又是怎么来的呢? 所以你可能, 很可能, 非常可能, 还是得信大爆炸理论. 可是这么一个开天辟地的大爆炸, 怎么就炸出了个太阳? 又炸出了个地球? 地球和太阳的距离又刚好是这么远? 地球上又刚好有个大气层? 有水? 有空气? 有孕育生命的一切必要的与充分的条件? 而比如微观方面, 当时间来到某一天, 又刚好有了你一个你? “开辟鸿蒙, 谁为情种?” 又是鸿蒙, 又是谁为, 你不觉得世界来到今天这个状态, 在这个状态里有目前的这些生物, 这种可能性不是比10的负一亿次方还要小得小得多吗? “……万类霜天竞自由. 怅寥廓, 问苍茫大地, 谁主沉浮?” 谁主沉浮? 谁主沉浮? 谁主得了沉浮?

世界也是确定的. 我们从小学起就开始学习支配这个世界的确定的规律. 有数学的: $1 + 2 = 3$, 三角形的内角和为 180° , 一元二次方程的根由其系数唯一地显式地确定, $\sin(\alpha + \beta) = \sin\alpha \cos\beta + \cos\alpha \sin\beta$, ……; 有物理的: 杠杆原理, 牛顿三定律, 热力学三定律, $E = mc^2$, ……; 有化学的: 门捷列夫元素周期表, 化学反应方程式, 拉瓦锡质量守恒定律, ……; 还有, 也许我要说, 甚至还有语文的: 成语啊, 固定搭配啊, 起承转合啊, …….

但归根到底是随机的. 物理学到了微观世界是即量子物理学是随机的, 化学到了分析化学是随机的, 就更不用说生物学了. 还有文学, 有哪一篇小说是完全按照某条铁律写出来的呢? 作家笔下那些才华横溢的句子, 除了他或她的天赋之外, 难道不是, 往往是, 触发于天边的一抹晚霞, 窗外的一缕清风? 而数学呢?

数学中的概率论这门学科, 正是研究随机现象的科学, 虽然它本身并不是随机的(好像模糊数学倒有点随机的味道, 但与本课程无关). 特别有意思的是, 现代概率论整体理论的建立, 也基于一个十分偶然的因素: 概率论公理体系的奠基者、伟大的数学家Kolmogorov¹起初是作为历史系的学生进入莫斯科大学的. 仅仅因为他参加了一个讨论班, 仅仅因为他写了一篇文章得到了教授的赏识, 仅仅因为这个赏识他的教授对他说“历史学的任何一个结论都需要五个证明”而要求他增补论文, 仅仅因为他不喜欢这个要求而要找“一个结论只需要一个证明”的地方, 他才转到了数学系, 于是概率论才有了今天的样子.

另外发生在Kolmogorov身上的一件随机的事情是, 他的出生地也是随机的: 他母亲怀着他去办事, 途径一个小村子时突然临产而生了他.

真是不可思议! 这样一个本身充满了随机性的Kolmogorov成了现代概率论的开山鼻祖, 是巧合吗? 是命运也就是说是上帝吗?

世界是随机的, 随机积分和随机微分方程的创始人、著名数学家伊藤清(Kiyosi Itô)²如是说. 他的日语原话直译是, 世界是确率的. 日语里的“确率”似乎承担了汉语里的“概率”和“随机”双重意思, 并同时具有名词和形容词双重词性.

¹ Andrei Nikolaevich Kolmogorov, 1903-1987, 苏联数学家.

² 伊藤清, 1915-2008, 日本数学家.

日常生活中,当我们说概率时,大体上可分为两种不同的情况.

第一种情况是有一些很明显的客观依据的. 比如一个婴儿将要出生,我们说是男婴还是女婴的概率都是二分之一,这是因为过去数年内的数据表明,在新生儿中大体上男婴女婴各占一半,而这也预示着在未来出生的每一百个婴儿中,男女婴儿数也会都接近五十个. 当然,刚好都各为五十个可能比较罕见,也可能男婴48个,女婴52个;也可能男婴54个,女婴46个,反正是比较接近的情况比较多见,而不接近的情况,比如说男婴35个,女婴65个的情况则是很罕见的. 再比如一个士兵,说他200米的距离内射击的命中率是百分之九十,那一定是因为过去的记录中,射击的总次数中有百分之九十是命中了的,并且你如果再让他射一千次,那么你可以期望他击中的次数应该在九百次左右,虽然不太可能刚好是九百次. 而这个数字对未来的意义在于,指挥员可以期待他十次射击中命中八九次的可能性是非常大的,而基本上不可能只命中比如说四次. 此外,如果某次任务中目标非常重要,必须一枪毙命,那么指挥员必须依据每个人的命中率决定一个策略,比如至少得派几个战士同时执行任务才能稳妥地完成任

务. 这种客观概率的典型特征是,它们是从大量的试验中得出的结论,并且还可以继续做大量的试验检验这个结论的正确性.

第二种情况是并没有太明显的客观依据,或者说虽然也有客观依据,但主观判断占了很大的成分. 比如说,对明天股市的涨跌,每天晚上每个电视台都有评论员做出预测. 不可否认,有些不良评论员是昧着良心说瞎话的,但这不在我们的讨论之列. 我们假定每个评论员都是正直的. 即使这样,他们对未来的判断也不会完全一样. 比如说甲会说某股明天上涨的概率是0.8,而乙预测只有0.3. 他们的说法有客观的成分吗? 也许是有的. 但是是完全客观的吗? 这不可能! 一定是夹杂了他自己的一些主观判断. 再比如国外的总统选举,甲认为某候选人当选的可能性是百分之九十,而乙认为只有百分之六十. 这不能说完全没有客观依据,但肯定掺杂了很多主观好恶.

这种主观概率的典型特征是:它们不是从大量的试验中得出结论,并且这些结论的正确性也无法通过大量重复的试验进行检验.

在实际问题中,概率的主观性和客观性往往并不是完全绝对对立的,而是常常有统一的时候,相互补充的时候. 比如抛一枚硬币,说它是均匀的,意义是什么呢?根据是什么呢?定义又是什么呢?难道均匀性可以由每抛100次,正反面各出现50次来定义吗?当然——不行,因为如果按照这个定义,就不会有任何一枚硬币满足要求. 那么,可以由每次抛时,正反面出现的可能性都一样来定义吗?当然也不行,因为这又涉及到什么叫“可能性都一样”的问题?

那么请告诉我,什么叫均匀?答案难道不是因人而异吗?

所以,这里面就有一个主观性的问题. 也因此在此有关的一切讨论中,也就都有了主观因素掺杂其中.

还有比如说天气预报,尽管我们有各种先进的仪器,但依然要有预报员参加工作. 这时结果就很可能掺杂了预报员们的个人判断,因此也是有主观因素的.

上面的讨论似乎有点哲学的味道了,偏离了本课程的跑道. 对这个问题,我们采用邓小平同志的智慧和策略,即不争论. 作为一个现代数学分支,概率论和其它分支一样,有自己严格的公理体系,而在此基础上建立的一切结论都必须根据严格的逻辑推理获得. 上面讨论的主客观性的问题是在涉及其应用时产生的,属于统计学(包括统计物理,金融统计……)范畴. 我们此后将不纠结于这个问题,而是按公理化的方式解决这个问题——例如什么叫硬币是均匀的?正反面出现的概率都一样就叫均匀的,而概率是公理化体系中已经明确给定了的. 我们关心的,是在这个公理体系下所能得到的科学结论. 至于实际问题中一枚硬币是否均匀,那就不

是我们关心的问题，留给赌徒们自己去解决吧。

赌博当然是一种恶习。但正如毛主席所说，任何事物都是一分为二的。关于赌博的研究曾经是，也许现在仍然是，概率论发展壮大的不可缺少的动力。如果最初的研究者们因为登不了大雅之堂而停止了此项研究，那可能今天就根本没有概率论这个学科了——概率论本身的出现，也是一个随机事件啊！

1.1 试验及其数学描述

本书所指试验，皆指一般意义下的试验。比如普通物理学或化学里的试验，比如掷一枚骰子，比如猜拳，等等。

既然是试验，就会有多种不同的结果，否则就不必试验了。但有些试验人们已经基本研究清楚了，其结果受物理或化学或其它科学定律支配，结果基本上是确定的，虽然可能有一点点小小的误差（当然，这些误差有时候不重要，但有时候却非常重要。不过这不在本书的讨论范围之列）。人们通过这些试验可以验证已有的知识，也可以发现或总结新的规律。比如我们都做过测量重力加速度的试验，也都知道其值基本上会在 $9.8m/s^2$ 左右（随地点不同而略有差别）。而有些试验则不然，即其结果是完全不可预测的，漂浮不定的。比如猜拳，你能准确猜到对手下一手出什么拳？比如掷骰子，你能总结出某个规律，比如说1点过后一定出现5点？不可能的嘛。

第一类试验中的误差分析，无疑是一个很重要的研究课题，同时也是推动概率论发展的主要动力之一（例如正态分布虽然早已有之，但却因为Gauss³在研究误差分析时得到了它而又名Gauss分布）。不过这个问题基本上属于统计范畴，不在本书要讨论的问题之列。本书所指试验，是第二类试验，即其结果完全无法用现有的科学知识预测的试验，或理论上虽可以预测，但事实上很难做到的试验。

概率论的研究对象，当然不是所有这些试验，而是满足下面两个条件的试验：第一，试验是可以重复的，即在相同的环境下可反复进行同一个试验，比如说扔硬币就是如此。反之，比如说总统选举就不是可以重复的，因此不是概率论的研究对象。（但选举之前举办的民调是可以重复的，因此是概率论或者更准确地说是统计学的研究对象。）第二，大量重复这个试验时，虽然单次的结果无法预测，但会呈现一定的统计规律。比如扔硬币，如果这个硬币是均匀的，那么大量重复这个试验时，正反两面出现的次数会大致相等。为了验证这一点，历史上不少人做过这个试验，结果显示都如此。这就是统计规律。

当然，试验的结果是否可以预测，是与科学技术的发展水平相关的。就拿抛硬币为例。根据经典动力学与运动学原理，硬币自从被抛出那一刻起，它的运动轨迹就被确定下来了，包括最后的结果——哪一面朝上。因此，从理论上讲，抛硬币的试验是完全可以预测结果的。问题是影响硬币运动轨迹的参数太多，例如抛出手时的初速度、高度和角度，硬币的实际上不可避免的微小的不均匀性，空气中飘动的微风等等，而最终哪一面出现的结果又对这些参数太敏感了，它们中任何一项的任何微小的改变，包括肉眼观察不到的改变，都会导致我们肉眼观察得到的截然不同的结果。所以这个试验即使本质上是确定的，但在一般的理解中，我们仍然认为是随机的，且具有统计规律，因此我们把其归结为其结果无法用现有的科学知识预测的试验，而用概率论的方法去研究它。我们相信，随着科学的发展，会有一些目前用概率论研究的课题，将来却成为了某一科学定律。Hilbert⁴说过，我们必须知道，我们必将知道；Einstein（爱

³Johann Carl Friedrich Gauss, 1777-1855, 德国数学家。

⁴David Hilbert, 1862-1943, 德国数学家。

因斯坦)⁵说过,上帝不掷骰子.这是这些科学巨匠们的坚定信念.但在我们知道以前,在上帝告诉我们他主宰世界的方式以前,我们还是需要概率论的,哪怕只是权宜之计.所以也正是他们两人,在概率论的发展史上都做出了杰出的贡献. Hilbert第六问题是将物理学公理化,这其中包含了将概率论公理化,后者由Kolmogorov在1933年完成,为现代概率论的发展奠定了基础; Einstein关于Brown⁶运动的研究则对热力学和随机过程的研究产生了重大的推动力,而且催生了一个或多个的诺贝尔奖或同等级别的奖(例如Avogadro⁷常数的确定).

对一个试验,任何一个结果都称为一个基本事件(即不能进一步分解的事件).将其所有可能的结果即基本事件放在一起,就构成一个集合,这个集合我们一般将用 Ω 表示之.传统上,一个基本事件称为一个样本点 ω ,而 Ω 则称为样本空间.比如掷一枚骰子的试验,其样本空间便是

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

而如果将一枚骰子掷 n 次,样本空间则是

$$\Omega = \{(k_1, \dots, k_n) : k_i \in \{1, 2, 3, 4, 5, 6\}, i = 1, \dots, n\}.$$

将满足特定条件的基本事件放在一起,就构成事件.因此事件用集合论的语言来说就是 Ω 的子集.通常常用大写字母如 A, B, \dots 表示之.比如在掷一次骰子的试验中,满足“所得点数大于3”这个条件的结果便是

$$A = \{4, 5, 6\};$$

在掷 n 次骰子的试验中,满足“ n 次所得点数之和小于等于 m ”这个条件的结果便是

$$A = \{(k_1, \dots, k_n) : k_1 + \dots + k_n \leq m\}.$$

但所谓“基本”其实并不是天生不变的,它与我们所感兴趣的问题有关,也与我们对研究方法的选择有关.例如掷两枚骰子的试验,基本事件自然可选择为

$$(k_1, k_2), k_1, k_2 = 1, 2, \dots, 6.$$

此时,相应的样本空间是:

$$\Omega_1 = \{(k_1, k_2) : k_1, k_2 \in \{1, 2, 3, 4, 5, 6\}\}.$$

但是,如果我们掷两枚骰子的目的就是为了看两个骰子点数之和的大小,那么基本事件也可以取为

$$k = 2, \dots, 12.$$

此时的样本空间便为

$$\Omega_2 = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

两个相比较, Ω_1 含有更多的基本事件,且 Ω_2 中的基本事件都是由 Ω_1 的基本事件的不同组合构成的.当然,第二个问题的样本空间也可以用 Ω_1 来描述. Ω_1 和 Ω_2 的关系无非是 Ω_2 中的一个样本点对应着 Ω_1 中多个样本点而已.至于这两种选择哪个更好,则很难有一个放之四海而皆准

⁵Albert Einstein, 1879-1955, 瑞士-美国物理学家.

⁶Robert Brown, 1773-1858, 英国植物学家.

⁷Amedeo Avogadro, 1776-1856, 意大利物理学家、化学家.

的标准, 要依具体问题而定. 不过, 一般来说, 基本事件分得越细, 适用的问题就越多, 计算也相对复杂些; 基本事件分的越粗, 就越有针对性.

由此我们可以看到, 所谓样本空间也不是试验本身固有的, 而是人们为方便研究而建立的数学模型, 且这个模型的建立往往依赖于个人的品味和能力、解决问题的针对性与有效性、使用上的方便性以及数学处理上的简洁性与优雅性等等. 通常, 给出适合的样本空间并不是个简单的问题.

乍一看, 一个试验, 结果应该是有限的(谁能把一个试验做出无穷个结果呢?), 所以样本空间的元素也只有有限多个. 但其实不然. 需要想象不能实际做出的试验, 这是科学研究的基本功. Einstein不就通过电梯试验启发了广义相对论吗? 虽然他从来没有也不可能做出这个试验.

具体到我们面对的问题, 我们刚刚说过, 概率论的研究对象是可以重复的试验, 因此很容易想象将一个简单的试验重复多次, 循环往复, 以致无穷, 虽然实际上做不到无穷次, 但有必要考虑试验次数趋于无穷时发生的各种现象, 这样样本空间就有无穷多个元素.

这方面的一个典型例子是人口普查.

设一个地区共有 N 口人. 这个 N 往往是相当大的. 我们要调查这 N 口人的一些状况(比如平均身高). 当然, 最准确的方法是对每个人逐一调查, 这样最后将各人的情况平均即可. 但这样的逐一调查会耗费大量的人力物力财力, 所以有必要采取一种替代的方法, 这就是抽样调查. 所谓抽样调查就是随机地选取一些人进行调查. 但需要抽样多少次才能获得所需要的准确性? 这不是一个有现成答案的问题. 所以在建立数学模型时, 不能预先限定抽样的次数即样本的大小. 这样, 假设人群全体为 $\{\omega_i, i = 1, 2, \dots, N\}$ 这个抽样所得到的样本空间便为

$$\Omega = \{\omega = (\omega_{k_1}, \omega_{k_2}, \dots)\} = \{\omega_1, \omega_2, \dots, \omega_N\}^\infty.$$

其次, 有些试验本身就有无穷多个结果, 比如说往地上的一块区域投一根针, 就有无穷多个结果.

与无穷有关的, 我们需要区分两个概念: 可数和不可数.

可数分为两类: 有限和可列. 有限就不用解释了. 所谓可列, 是指: 1. A 有无穷多个元素; 2. 这些元素可以排成一列. 准确地说, 集合 A 称为有可列个元素, 或简称可列, 是指存在一种排列方式将 A 的元素写为 a_1, a_2, \dots . 因此

$$A = \{a_1, a_2, \dots\}.$$

所以, 全体整数的集合是可列的, 因为它们能以以下方式排列:

$$0, 1, -1, 2, -2, 3, -3, \dots$$

全体有理数集也是可列的, 因为它们能以以下方式排列:

$$0, 1, -1, \frac{1}{2}, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{3}, \frac{2}{3}, -\frac{2}{3}, \dots$$

但可以证明, 全体实数的集合不是可列的.

描述一个集合, 在该集合的元素很少时, 可以用穷举法. 例如, 若 A 表示前五个正整数, 则可直接写出:

$$A = \{1, 2, 3, 4, 5\}.$$

但只要 A 的元素一多,这种方法就不可能了.这时就用 A 的元素满足的条件来表示它.例如

$$\mathbb{N} := \{n : n \text{为整数}\}$$

表示整数全体,而

$$\mathbb{N}_+ := \{n \in \mathbb{N} : n \geq 0\}$$

则表示非负整数全体.

我们再看一些例子.

例1. 抛掷一枚硬币,以1表示出现正面,0表示出现反面.则样本空间为

$$\Omega = \{0, 1\}.$$

例2. 将一枚硬币抛 n 次,样本空间为

$$\Omega = \{0, 1\}^n := \{(a_1, \dots, a_n) : a_k = 0, 1, k = 1, \dots, n\}.$$

例3. 将一枚硬币抛无穷次,样本空间为

$$\Omega = \{0, 1\}^\infty := \{(a_1, \dots, a_n, \dots) : a_k = 0, 1, k = 1, \dots, n\}.$$

例4. 在大街上随便问一个人他的生日,样本空间为

$$\Omega = \{(m, n) : m = 1, 3, 5, 7, 8, 10, 12 \text{ 时}, 1 \leq n \leq 31; m = 2 \text{ 时}, 1 \leq n \leq 29; \\ m = 4, 6, 9, 11 \text{ 时}, 1 \leq n \leq 30\}.$$

例5. 在大街上随便问一个人他的身高,样本空间可以取为

$$\Omega = \{x \in \mathbb{R} : 0 \leq x \leq 3\}.$$

例6. 两个人玩石头剪刀布,样本空间为

$$\Omega := \{(\text{石}, \text{石}), (\text{石}, \text{剪}), (\text{石}, \text{布}), (\text{剪}, \text{石}), \\ (\text{剪}, \text{剪}), (\text{剪}, \text{布}), (\text{布}, \text{石}), (\text{布}, \text{剪}), (\text{布}, \text{布})\}$$

例7. 四个人,编号为1,2,3,4,进行射击比赛,成绩以0,1, ..., 10环记之.则样本空间为

$$\Omega := \{(i, j) : i = 1, \dots, 4, j = 0, 1, \dots, 10\}.$$

总之,描述试验及其结果的数学语言本质上是一般集合论的语言,只不过用的术语可能不同而已.前两章我们不加区别地使用集合和事件这两个名词,等第三章给出概率的公理化之后,会进一步明确事件的概念.

现在.从这些基本的对象开始我们的概率论之旅.

习题

1. 一个铁人三项运动员,每一项获得的名次都可能是1到10名.写出这个试验的样本空间.
2. 一个电码由0到9四位数字组成,但由于各种因素的干扰,发报员发出的数字都可能被接收为另外一个数字.写出发送一个电码这个试验的样本空间.

3. 张三从甲地步行到乙地, 第一次行走, 未带地图. 包括甲地, 共有 m 处分叉路口, 其中第 i 个路口有 m_i 个道路可选. 写出这个试验的样本空间.

4. 投 n 次硬币, 写出样本空间.

答案: 1. $\{-1, 1\}^n$; 2. $[0, 1]$ 上的连续函数 f 全体, $f(0) = 0$, f 在每一 $[\frac{i}{n}, \frac{i+1}{n}]$ 上为线性函数, 且斜率的绝对值为1.

5. 将扔硬币的试验无穷此重复下去, 写出这个试验的样本空间.

(答案: $[0, 1)$)

6. 一项流行病调查, 调查对象为 A, B, C, D 四种疾病, 结果为 α, β, γ 三种症状. 写出这个试验的样本空间.

1.2 事件的关系与运算

由于不可避免地要处理事件的各种组合, 所以需要用到事件的运算.

设 Ω 为样本空间, A, B, C, \dots 均为 Ω 的子集, 即事件.

1. A 与 B 的并是指这样一个事件: A 与 B 中至少一个发生. 这个事件记为 $A \cup B$, 即

$$A \cup B := \{\omega \in \Omega : \omega \in A \text{ 或 } \omega \in B\},$$

2. A 与 B 的交是指这样一个事件: A 与 B 都要发生. 这个事件记为 $A \cap B$, 即

$$A \cap B := \{\omega \in \Omega : \omega \in A \text{ 且 } \omega \in B\},$$

$A \cap B$ 也常写为 AB ;

$AB = \emptyset$ 时, 称 A 与 B 互不相容. 此时, $A \cup B$ 常记为 $A + B$. 或者反过来说, 写出 $A + B$ 时, 就自动意味着 $AB = \emptyset$.

3. A 减 B 是指这样一个事件: A 发生而 B 不发生. 这个事件记为 $A \setminus B$, 即

$$A \setminus B := \{\omega \in A \text{ 且 } \omega \notin B\}.$$

$A \subset B$ 时, $B \setminus A$ 常记为 $B - A$. 或者反过来说, 写出 $B - A$ 时, 就自动意味着 $A \subset B$.

4. A 与 B 的对称差定义为:

$$A \Delta B := (A \setminus B) \cup (B \setminus A).$$

5. A 的余集, 或称逆事件, 定义为:

$$A^c := \Omega \setminus A.$$

不含任何元素的集合称为空集或不可能事件, 用 \emptyset 表示. 显然

$$\emptyset^c = \Omega, \quad \Omega^c = \emptyset.$$

一般地有

$$(A^c)^c = A.$$

6. 两个事件 A, B , 若

$$\omega \in A \implies \omega \in B,$$

则称 A 包含于 B , 或 B 包含了 A , 记为 $A \subset B$ 或 $B \supset A$. 若 $A \subset B$ 且 $B \supset A$, 则称 A 等于 B , 记为 $A = B$.

7. 并与交均可对多个集合定义. 设 $\{A_i, i \in I\}$ 是一族集合, 定义:

$$\bigcup_{i \in I} A_i := \{\omega : \exists i \in I, \text{ 使得 } \omega \in A_i\}.$$

$$\bigcap_{i \in I} A_i := \{\omega : \forall i \in I, \omega \in A_i\}.$$

8. 设 A_1, A_2, \dots 为一列事件, 定义它们的上极限为

$$\limsup_{n \rightarrow \infty} A_n := \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k.$$

因此, $\omega \in \limsup_{n \rightarrow \infty} A_n$ 就意味着对任意 n , 都有 $k > n$ 使得 $\omega \in A_k$. 这显然等价于有无限多个 n 使得 $\omega \in A_n$. 因此, $\omega \in \limsup_{n \rightarrow \infty} A_n$ 就表示 $\{A_n\}$ 中有无限多个发生这一事件;

定义它们的下极限为

$$\liminf_{n \rightarrow \infty} A_n := \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k,$$

$\omega \in \liminf_{n \rightarrow \infty} A_n$ 就意味着存在一个 n , 使得对任意 $k > n$ 都有 $\omega \in A_k$. 这显然等价于最多只有有限多个 n 使得 $\omega \notin A_n$. 因此, $\omega \in \liminf_{n \rightarrow \infty} A_n$ 就表示 $\{A_n\}$ 中至多有限个不发生这一事件.

因此有

$$\liminf_{n \rightarrow \infty} A_n \subset \limsup_{n \rightarrow \infty} A_n.$$

这有点像数学分析里大家熟知的关系: 对任意数列 $\{a_n\}$ 有

$$\liminf_{n \rightarrow \infty} a_n \leq \limsup_{n \rightarrow \infty} a_n.$$

如果 $\limsup A_n = \liminf A_n$, 则称 $\{A_n\}$ 的极限存在, 并记为 $\lim A_n$.

若 $\{A_n\}$ 单调, 则 $\lim_{n \rightarrow \infty} A_n$ 存在: 当它们单调上升, 即 $A_1 \subset A_2 \subset A_3 \dots$ 时, 有

$$\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n;$$

当它们单调下降, 即 $A_1 \supset A_2 \supset A_3 \dots$ 时, 有

$$\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n.$$

从而我们一般地有

$$\limsup_{n \rightarrow \infty} A_n := \lim_{n \rightarrow \infty} \bigcup_{k=n}^{\infty} A_k,$$

$$\liminf_{n \rightarrow \infty} A_n := \lim_{n \rightarrow \infty} \bigcap_{k=n}^{\infty} A_k.$$

这两个式子和数学分析里下面熟知的式子又是类似的:

$$\limsup_{n \rightarrow \infty} a_n := \lim_{n \rightarrow \infty} \sup_{k \geq n} a_k,$$

$$\liminf_{n \rightarrow \infty} a_n := \lim_{n \rightarrow \infty} \inf_{k \geq n} a_k.$$

事件的运算有下面的性质:

交换律:

$$A \cup B = B \cup A, \quad AB = BA;$$

结合律:

$$A \cup B \cup C = A \cup (B \cup C) = (A \cup B) \cup C,$$

$$ABC = A(BC) = (AB)C;$$

分配律:

$$A(B \cup C) = (AB) \cup (AC),$$

这一性质类似于四则运算中的分配律, 即 $a(b+c) = ab+ac$, 这也许是人们想到用 AB 代替 $A \cap B$ 表示交的原因. 但你不能想当然地认为形式上交对应着乘积的所有运算关系. 比如我们有

$$AA = A,$$

但并没有

$$aa = a;$$

我们有

$$A \cup (BC) = (A \cup B) \cap (A \cup C),$$

但也没有

$$a + bc = (a + b)(b + c),$$

等等.

定义 1.2.1. 设 $A \subset \Omega$. 函数

$$1(A)(\omega) := 1_A(\omega) = \begin{cases} 1 & \omega \in A, \\ 0 & \omega \notin A \end{cases}$$

称为 A 的示性函数.

我们有下面的:

命题 1.2.2. (i)

$$1_A = 1 - 1_{A^c};$$

(ii)

$$A \subset B \iff 1_A \leq 1_B.$$

(iii)

$$A = B \iff 1_A = 1_B;$$

(iv)

$$A \subset B \iff 1_{B \setminus A} = 1_B - 1_A.$$

(v)

$$1_{A \cup B} = 1_A + 1_B - 1_A 1_B = 1_A \vee 1_B,$$

(vi)

$$1_{AB} = 1_A 1_B = 1_A \wedge 1_B.$$

命题 1.2.3. (i)

$$A \setminus B = A - AB;$$

(ii)

$$A \cup B = (A \setminus B) + B = (A \setminus B) + AB + (B \setminus A).$$

定理 1.2.4 (De Morgan⁸原理). 设 I 是任意指标集, 则

$$\left(\bigcup_{i \in I} A_i \right)^c = \bigcap_{i \in I} A_i^c,$$

$$\left(\bigcap_{i \in I} A_i \right)^c = \bigcup_{i \in I} A_i^c.$$

证明. 应用命题1.2.2, 有

$$\begin{aligned} 1_{(\bigcup_{i \in I} A_i)^c} &= 1 - 1_{(\bigcup_{i \in I} A_i)} \\ &= 1 - \max_{i \in I} 1_{A_i} \\ &= \min_{i \in I} (1 - 1_{A_i}) \\ &= \min_{i \in I} 1_{A_i^c} \\ &= 1_{(\bigcap_{i \in I} A_i^c)}. \end{aligned}$$

类似可证明另一式. □

上面对事件的运算完全取自集合的术语和符号. 在概率论中, 有一些专门的术语描述它们. 例如:

$AB = \emptyset$ 时, 称 A 与 B 互不相容; A^c 称为 A 的逆事件. Ω 称为必然事件, \emptyset 称为不可能事件.

习题

1. 证明:

(a)

$$A \setminus B = AB^c;$$

(b)

$$(A \cup B)C = AC \cup BC;$$

⁸Augustus De Morgan, 1806-1871, 英国数学家.

(c)

$$(A \setminus B)C = AC \setminus BC;$$

(d)

$$(A \setminus B) \cup C = (A \cup C) \setminus BC^c;$$

2. 证明:

(a)

$$1_{\liminf_{n \rightarrow \infty} A_n} = \liminf_{n \rightarrow \infty} 1_{A_n}$$

(b)

$$1_{\limsup_{n \rightarrow \infty} A_n} = \limsup_{n \rightarrow \infty} 1_{A_n}$$

(c)

$$\liminf_{n \rightarrow \infty} A_n \subset \limsup_{n \rightarrow \infty} A_n.$$

3. 定义对称差:

$$A \Delta B := (A \setminus B) \cup (B \setminus A).$$

证明:

(a)

$$A \Delta B = B \Delta A;$$

(b)

$$(A \Delta B) \Delta C = A \Delta (B \Delta C) = (C \Delta A) \Delta B;$$

所以可把它们统一地记为

$$A \Delta B \Delta C.$$

一般地, 也可定义

$$A_1 \Delta A_2 \Delta A_3 \cdots \Delta A_n.$$

(c)

$$\begin{aligned} 1_{A \Delta B} &= 1_A + 1_B - 21_{AB} \\ &= |1_A - 1_B|; \end{aligned}$$

4. 证明:

(a)

$$1_{A \cup B \cup C} = 1_A + 1_B + 1_C - 1_{A \cap B} - 1_{B \cap C} - 1_{C \cap A} + 1_{ABC};$$

(b)

$$1_{ABC} = 1 - 1_{A^c} - 1_{B^c} - 1_{C^c} + 1_{A^c B^c} + 1_{B^c C^c} + 1_{C^c A^c} - 1_{A^c B^c C^c};$$

(c)

$$1_{\bigcup_{i=1}^n A_i} = \sum_{i=1}^n 1_{A_i} - \sum_{i \neq j} 1_{A_i A_j} + \cdots + (-1)^{n-1} 1_{A_1 \cdots A_n}.$$

(d) 若 A_i 互不相交, $i = 1, 2, \dots, n, n \in \mathbb{N}_+$ 或 $n = \infty$. 则

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

(e) 设 $\{a_n(\omega)\}$ 是依赖于参数 ω 的实数列, $a(\omega)$ 是实数. 证明:

$$\{\omega : a_n(\omega) \rightarrow a(\omega)\} = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \left\{ \omega : |a_m(\omega) - a(\omega)| < \frac{1}{k} \right\}.$$

(f) 设 A_1, \dots, A_n 是事件.

- (i) 表示出事件“ A_1, \dots, A_n ”中恰有一个发生;
- (ii) 表示出事件“ A_1, \dots, A_n ”中至少有两个发生;
- (iii) 表示出事件“ A_1, \dots, A_n ”中最多有五个发生;
- (iv) 表示出事件“ A_1, \dots, A_n ”中一个都不发生;

(g) (i) 证明对称差的结合律

$$(A \Delta B) \Delta C = A \Delta (B \Delta C).$$

(ii) 证明:

$$A_1 \Delta A_2 \Delta A_3 \cdots \Delta A_n = \bigcup_{N(I)=\text{奇数}} \left(\left(\bigcup_{i \in I} A_i \right) \cap \left(\bigcup_{j \in I^c} A_j^c \right) \right).$$

(iii) 证明对称差的分配律:

$$A (A_1 \Delta A_2 \Delta A_3 \cdots \Delta A_n) = B_1 \Delta B_2 \Delta B_3 \cdots \Delta B_n,$$

其中 $B_i := A A_i$.

5. 设 Ω 是样本空间, $A \subset \Omega$. 令

$$A_n = \begin{cases} A & n = 1, 3, 5, \dots \\ A^c & n = 2, 4, 6, \dots \end{cases}$$

求 $\limsup_{n \rightarrow \infty} A_n$ 与 $\liminf_{n \rightarrow \infty} A_n$.

6. 设 $\{A_n\}$ 是一列集合, $n = 1, 2, \dots$ 或 $n = 1, 2, \dots, N$ 均可, 其中 N 为正整数. 令

$$\tau(\omega) := \inf\{n : \omega \in A_n\},$$

$$B_n := \{\omega : \tau(\omega) = n\}.$$

证明 B_1, B_2, \dots 互不相交且

$$\bigcup_n A_n = \bigcup_n B_n.$$

2 离散概型

本章讲述描述具有可数个结果的试验的概率模型, 俗称离散概型; 离散概型中, 如果样本点只是有限个, 则称为有限概型; 有限概型中, 如果各样本点出现的概率相等, 则称为古典概型, 也就是中学里讲过的概率模型. 让我们从这种最简单的概型谈起.

2.1 古典概型

定义 2.1.1. 一个试验, 如果有 n 个可能的结果, 而每个结果出现的可能性都是 $\frac{1}{n}$, 则称为古典概率模型, 简称古典概型.

我们将用 $\omega_1, \dots, \omega_n$ 表示这些结果, 用 Ω 表示它们全体:

$$\Omega := \{\omega_1, \dots, \omega_n\},$$

用 $P(\omega_i)$ 表示 ω_i 出现的概率(即可能性), 即

$$P(\omega_i) = \frac{1}{n}, \quad \forall i = 1, \dots, n.$$

古典概型最显著的特征就是每个结果出现的概率相等, 而什么时候这个概型适用, 则取决于我们的经验. 比如, 掷一枚均匀的硬币, 我们会认为出现正面或反面的概率是相等的; 在封闭的袋中摸完全一样的球, 我们会认为摸到任何一只的概率都是相等的, 等等.

我们来看几个具体例子.

例1. 设掷一枚均匀硬币, 用1表示出现正面, 0表示出现反面. 则

$$\Omega = \{0, 1\}, \quad P(i) = \frac{1}{2}, \quad i = 0, 1.$$

这个模型称为Bernoulli¹概型, 这个试验称为Bernoulli试验.

例2. 设一枚均匀硬币掷 n 次, 还是用1表示出现正面, 0表示出现反面. 则

$$\Omega = \{(i_1, i_2, \dots, i_n), i_k = 0, 1, \forall k = 1, 2, \dots, n\},$$

$$P((i_1, i_2, \dots, i_n)) = 2^{-n}, \quad \forall (i_1, i_2, \dots, i_n) \in \Omega.$$

这个模型称为 n 重Bernoulli概型, 或依旧简称为Bernoulli概型. 这个试验称为 n 重Bernoulli试验. 因为每个结果都是等可能的, 所以对事件 $A \subset \Omega$, 定义 A 发生的概率为

$$P(A) := \frac{N(A)}{N(\Omega)}, \tag{1.1}$$

其中 $N(A)$ 表示集合 A 中的元素个数. 称 $N(A)$ 为有利事件的样本点数, $N(\Omega)$ 为总样本点数.

这样定义的概率有如下性质:

¹Jacob Bernoulli, 1654-1705, 瑞士数学家

定理 2.1.2. (i) $\forall A \subset \Omega, 0 \leq P(A) \leq 1$;

(ii) $P(\Omega) = 1, P(\emptyset) = 0$;

(iii) 设 $m \in N_{++}$, 且 $\forall i = 1, \dots, m, A_i \subset \Omega$ 且两两不交. 令 $A := \sum_{i=1}^m A_i$, 则有加法公式

$$P(A) = \sum_{i=1}^m P(A_i).$$

证明是显然的.

古典概型从理论上讲是简单的, 要计算 $P(A)$, 这就归结为计算 $N(A)$ 与 $N(\Omega)$. 从而古典概型的问题其实就是计数问题. 如果选定样本空间, 这个问题看起来简单, 但在具体问题中计算却可能十分困难——你们有在中学里这些计算带来的或愉快或痛苦的记忆吗?

给你一个具体问题, 如果你能直接计算出来, 那当然是最好的; 当问题太复杂, 一时不能直接计算 $N(A)$ 时, 有一个简单而往往相当实用的方法, 即, 可以考虑把 A 分成几部分, 比如

$$A = A_1 + A_2 + \dots + A_m,$$

然后分别计算 $N(A_i)$. 由于

$$N(A) = N(A_1) + N(A_2) + \dots + N(A_m),$$

所以由此可计算出 $N(A)$. 我们将这个方法称为加法原理. 在同一个问题中, 这一原理可重复嵌套使用. 我们现在尝试用这个原理推导出已知的结果.

例1. 从 $\{1, 2, \dots, n\}$ 中取出 $m (\leq n)$ 个数, 依取出的顺序排列. 问一共有多少结果.

解. 以 Ω 表示所有结果的集合, 则

$$\Omega = \{(i_1, \dots, i_m), i_k \in \{1, 2, \dots, n\}, \forall k = 1, 2, \dots, m, i_k \neq i_l, \forall k \neq l\}.$$

以 Ω_i 表示第一个数字是 i 的结果的集合, 即

$$\Omega_i = \{(i, i_2, \dots, i_m), i_k \in \{1, 2, \dots, n\} \setminus \{i\}, \forall k = 2, \dots, m, i_k \neq i_l, \forall k \neq l\}.$$

我们有

$$\Omega = \Omega_1 + \Omega_2 + \dots + \Omega_n.$$

所以若令 $A_n^m = N(\Omega)$, 则有

$$A_n^m = N(\Omega_1) + N(\Omega_2) + \dots + N(\Omega_n).$$

但 $N(\Omega_1), \dots, N(\Omega_n)$ 都是从 $n-1$ 个元素中取 $m-1$ 个产生的结果数, 所以

$$N(\Omega_1) = N(\Omega_2) = \dots = N(\Omega_n) = A_{n-1}^{m-1}.$$

所以

$$A_n^m = nA_{n-1}^{m-1}.$$

这样递推下去, 到第 m 步终止, 最后一步是从 $n-m+1$ 个元素中取一个, 自然是有 $n-m+1$ 种取法, 即 $A_{n-m+1}^1 = n-m+1$. 从而

$$A_n^m = n(n-1) \cdots (n-m+1).$$

这就是 n 个元素中选 m 个的选排列数. 当 $m = n$ 时, 得到全排列数 $P_n = n!$.

这一模型称为无放回有顺序抽样.

例2. 如果上例中选出的这 m 个数不计顺序, 有多少种结果呢?

解. 我们可以这样考虑: 选排列可以按参与元素的不同先分类(因此是不计顺序的), 假设一共有 x 个这样的类; 在每个类中在进行全排列, 因此任何一个类中都有 $m!$ 个结果. 所以根据上面的原理有

$$n(n-1)\cdots(n-m+1) = x \cdot m!,$$

即

$$x = \frac{n!}{m!(n-m)!},$$

这就是组合数 C_n^m . 这一模型称为无放回无顺序抽样.

例3. 在 $\{1, 2, \dots, n\}$ 中取出一个, 记录下号码, 放回, 再取出一个记录下号码, 再放回. 如此重复 m 次. 问一共有多少种不同的结果.

解. 跟例1一样, 以 Ω 记所有结果全体, 则

$$\Omega = \{(i_1, \dots, i_m), i_k \in \{1, 2, \dots, n\}, \forall k = 1, 2, \dots, m\}.$$

以 Ω_i 表示第一个数字是 i 的结果的集合. 由于是有放回的, 所以

$$\Omega_i = \{(i, i_2, \dots, i_m), i_k \in \{1, 2, \dots, n\}, \forall k = 2, \dots, m\}.$$

我们有

$$\Omega = \Omega_1 + \Omega_2 + \dots + \Omega_n.$$

所以若令 $B_n^m = N(\Omega)$, 则有

$$B_n^m = N(\Omega_1) + N(\Omega_2) + \dots + N(\Omega_n).$$

由于第二次抽取时已将第一次抽出的号码放回, 所以

$$N(\Omega_i) = B_n^{m-1}, i = 1, \dots, n.$$

于是

$$B_n^m = nB_n^{m-1}.$$

所以

$$B_n^m = nB_n^{m-1} = n^2B_n^{m-2} = \dots = n^{m-1}B_n^1 = n^m.$$

这个模型称为有放回有顺序抽样.

例4. 在 $\{1, 2, \dots, n\}$ 中取出一个, 记录下号码, 放回, 再取出一个记录下号码, 再放回. 如此重复 m 次. 但最后记录下的结果不计顺序. 问一共有多少个结果?

解1. 所谓不计顺序, 是指在例如 $n = 5, m = 4$ 时, 结果 $(1, 2, 2, 5), (2, 1, 2, 5)$ 与 $(5, 2, 1, 2)$ 等视为同一个结果. 因此为确定起见, 可以把样本点写为

$$(i_1, \dots, i_m), 1 \leq i_1 \leq i_2 \leq \dots \leq i_m \leq n.$$

所以现在的样本空间为

$$\Omega = \{\omega = (i_1, \dots, i_m), 1 \leq i_1 \leq i_2 \leq \dots \leq i_m \leq n\}.$$

令

$$\Omega_k = \{\omega \in \Omega : i_1 = k\}, \quad k = 1, 2, \dots, n.$$

则

$$\Omega = \Omega_1 + \dots + \Omega_n.$$

记 $D_n^m = N(\Omega)$. 因为 Ω_1 中第一个分量已固定为1, 所以剩下的可能就是从 $\{1, \dots, n\}$ 中有放回无顺序地取 $m-1$ 次所得的结果数; 同样, 因为 Ω_2 中第一个分量已固定为2, 所以剩下的可能就是从 $\{2, \dots, n\}$ 中有放回无顺序地取 $m-1$ 次所得的结果数, 等等. 这样就有

$$D_n^m = D_n^{m-1} + D_{n-1}^{m-1} + \dots + D_1^{m-1}.$$

现在我们证明:

$$D_n^m = C_{n+m-1}^m.$$

$m = 1$ 时而 n 为任意正整数时, 显然有 n 种结果, 而 $C_n^1 = n$, 所以结论是正确的. 设对任意 n , $k = 1, 2, \dots, m$ 时, 均有

$$D_n^k = C_{n+k-1}^k.$$

则

$$\begin{aligned} D_n^{m+1} &= D_n^m + D_{n-1}^m + \dots + D_1^m \\ &= C_{n+m-1}^m + C_{n+m-2}^m + \dots + C_m^m. \end{aligned}$$

利用等式

$$C_k^{l-1} + C_k^l = C_{k+1}^l, \quad C_{m+1}^{m+1} = C_m^m,$$

上式等于

$$[C_{n+m}^{m+1} - C_{n+m-1}^{m+1}] + [C_{n+m-1}^{m+1} - C_{n+m-2}^{m+1}] + \dots + [C_{m+2}^{m+1} - C_{m+1}^{m+1}] + C_m^m = C_{n+m}^{m+1},$$

因此在 $k = m+1$ 时也成立.

解2. 换一个角度, $1, 2, \dots, n$ 可看成 n 个桶, 抽样可看成随机地往这 n 个桶里扔球. 数组 (i_1, i_2, \dots, i_m) 中 k 的个数就是 k 号桶中球的个数. 例如设 $n = 5, m = 4$, 则 $(1, 2, 2, 5)$ 表示1号桶中有1个球, 2号桶里有2个球, 5号桶里有1球, 而3、4号桶里没有球.

我们可以用下图表示这样一个结果



这就相当于两端是黑球, 然后 $4 (= 5 - 1)$ 个黑球和4个白球放入8个位置. 而一般情况是, 将 $n-1$ 个黑球和 m 个白球随机地放入 $m+n-1$ 个位置, 问一共有多少个不同的结果?

回答这个问题是容易的: 它就是组合数:

$$C_{m+n-1}^{m-1}.$$

这个模型称为有放回无顺序抽样.

在加法原理中, 如果每个 $N(A_i)$ 都一样, 例如 $N(A_i) = n$, 则有

$$N(A) = mn.$$

这一公式称为乘法原理,它是加法原理在特殊情况下的简化,一如乘法是加法在特殊情况下的简化.如同加法原理一样,这一原理在同一问题中可反复使用.

乘法原理尤其适用于下列情形:如果能将一个试验分为 n 步,第一步有 m_1 个可能的结果,再在得到的结果下进行第二步,每步又都有 m_2 个结果.以此类推,那么根据乘法原理,这个试验的总的结果数则为

$$m_1 \cdot m_2 \cdots m_n.$$

例1和例3的结果都可以用乘法原理来计算.我们再来看一个例子.本例取自[16].

例5. 有黄红黑白4种颜色的球各5颗.现随机取出10颗.设 $n_1 \geq n_2 \geq n_3 \geq n_4$, $n_1 + n_2 + n_3 + n_4 = 10$,以 $n_1 n_2 n_3 n_4$ 表示将取出的球按不同颜色的个数降序排列.求各种不同结果数及每种结果出现的概率.

解. 我们先看5500.它表示有两种颜色的球有5个,另两种颜色的没有.这就是说,现从4种颜色中选取2中颜色,可能的结果数为 C_4^2 .颜色确定后,针对每一种颜色,或者是从5个里面选5个,或者是从5个里面选0个,因此就只有一种结果.这样根据乘法原理,出现5500的结果数为

$$C_4^2 \cdot 1 = 6.$$

再看5410.这相当于现从4种颜色中确定一个有5个该颜色的,这有4种可能;再从剩下的3种颜色中确定一个有4个该颜色的,这有3种可能;再从剩下的2种颜色中确定一个有1个该颜色的,这有2种可能.最后剩下的那个颜色就是有0个该颜色的.

从5个里取5个的,有 C_5^5 种可能;从5个里取4个的,有 C_5^4 种可能;从5个里取1个的,有 C_5^1 种可能;从5个里取0个的,有 C_5^0 种可能.因此,出现5410的结果数为

$$4 \cdot 3 \cdot 2 \cdot 5 \cdot 5 = 600.$$

同理可计算其它的结果数:

$$5320: 4 \cdot 3 \cdot 2 \cdot C_5^3 \cdot C_5^2 = 2400,$$

$$5311: 4 \cdot 3 \cdot C_5^3 \cdot C_5^1 \cdot C_5^1 = 3000,$$

$$5221: 4 \cdot 3 \cdot C_5^2 \cdot C_5^2 \cdot C_5^1 = 6000,$$

$$4420: 4 \cdot 3 \cdot C_5^4 \cdot C_5^4 \cdot C_5^2 = 3000,$$

$$4411: C_4^2 \cdot C_5^4 \cdot C_5^4 \cdot C_5^1 \cdot C_5^1 = 1750,$$

$$4330: 4 \cdot 3 \cdot C_5^4 \cdot C_5^3 \cdot C_5^3 = 6000,$$

$$4321: 4 \cdot 3 \cdot 2 \cdot C_5^4 \cdot C_5^3 \cdot C_5^2 \cdot C_5^1 = 60000,$$

$$4222: 4 \cdot C_5^2 \cdot C_5^2 \cdot C_5^2 = 20000,$$

$$3322: C_4^2 \cdot C_5^3 \cdot C_5^3 \cdot C_5^2 \cdot C_5^2 = 60000,$$

$$3331: 4 \cdot C_5^3 \cdot C_5^3 \cdot C_5^3 \cdot C_5^1 = 20000.$$

如果要计算概率的话,那就还要计算总样本数,它是

$$C_{20}^{10}.$$

因此, 3322出现的概率是

$$60000/C_{20}^{10} \approx 0.324753,$$

而5500出现的概率是

$$6/C_{20}^{10} \approx 0.000032.^2$$

因此, 如果用

$$3322, 5500, 4420, \dots$$

作为基本结果的话, 那么它们出现的概率是不相等的, 故不适用于古典概型.

所以我们再强调一次, 古典概型适用的问题是只有有限个结果, 且每个结果出现的概率都相等的问题. 在建立一个问题的数学模型时, 这是两个本质的要求. 我们再看一个简单的例子.

例6. 设袋中有红球 m 只, 黑球 n 只. 从中摸一只. 假设每球被摸到的概率相同, 写出这个试验的样本空间.

解. 也许红球看起来没有差别, 黑球看起来也没有差别, 但为了研究起见, 我们给红球贴上标签: r_1, \dots, r_m , 黑球也贴上标签: b_1, \dots, b_n . 因此样本空间为

$$\{r_i, b_j, i = 1, \dots, m, j = 1, \dots, n.\}$$

由于每球被摸的概率相等, 所以这是古典概型, 即

$$P(r_i) = P(b_j) = \frac{1}{m+n}, \quad \forall i = 1, \dots, m, j = 1, \dots, n.$$

如果我们想偷点懒, 不给它们贴标签, 那么样本空间就变为

$$\Omega = \{r, b\}.$$

但此时 r, b 出现的概率显然是不一样的, 因为

$$P(r) = \frac{m}{m+n}, \quad P(b) = \frac{n}{m+n}.$$

所以这个模型不是古典概型, 因为每个样本点出现的概率不相等. 不过, 这也是一种正确的模型, 即所谓有限概型, 是我们下节要研究的对象.

在建立模型和计算时一定要小心谨慎, 不适用于古典概型的问题, 或者更明确地说, 如果你不能确认每个样本点出现的概率相同, 就绝对不能用古典概型.

除了通过直接计算事件中的样本点的个数来计算概率外, 有一个重要的途径是利用事件之间的独立性. 什么叫独立性呢? 我们从分析一个具体例子开始.

设袋中有红球 m 只, 黑球 n 只. 一个人先摸一球, 记下颜色, 将球放回, 然后再摸一次, 再记下颜色. 分别求第一次、第二次及两次皆摸出红球的概率.

先分别考察两次摸球. 第一次摸球的样本空间为

$$\Omega_1 := \{r_i, b_j, i = 1, \dots, m, j = 1, \dots, n\}.$$

摸出红球 A_1 为

$$A_1 := \{r_i, i = 1, \dots, m\}.$$

²顺便说一下, 这里的结果和王蒙先生实地观察的结果不一致, 这应该他观察的时间不够长, 或者没有认真记录结果造成的.

因此

$$P(A_1) = \frac{m}{m+n}.$$

由于第二次摸的时候, 已经把第一次摸的那个球放回去了, 所以同理有

$$\Omega_2 := \{r_i, b_j, i = 1, \dots, m, j = 1, \dots, n\}.$$

摸出红球 A_2 为

$$A_2 := \{r_i, i = 1, \dots, m\}.$$

因此

$$P(A_2) = \frac{m}{m+n}.$$

如果两次摸球同时考虑, 则

$$\Omega := \{(r_i, b_j), (b_j, r_i), (r_i, r_k), (b_j, b_l) : i, k = 1, \dots, m, j, l = 1, \dots, n\}.$$

在这个大一点的样本空间中, A_1, A_2 分别表示为

$$A_1 := \{(r_i, r_k), (r_i, b_j) : i, k = 1, \dots, m, j = 1, \dots, n\},$$

$$A_2 := \{(r_i, r_k), (b_j, r_k) : i, k = 1, \dots, m, j = 1, \dots, n\},$$

所以依然有

$$P(A_1) = P(A_2) = \frac{m}{m+n}.$$

而两次摸出红球 A 则表示为

$$A := \{(r_i, r_k), i, k = 1, \dots, m\}.$$

因此

$$P(A) = \frac{m^2}{(m+n)^2}.$$

我们注意到

$$A = A_1 A_2,$$

而

$$P(A) = P(A_1)P(A_2).$$

在这个例子中, 由于无论第一次摸到红球或者黑球, 对第二次摸球没有任何影响, 所以可以说 A_1 与 A_2 是独立的. 用数量关系来看, 就是上面最后这个等式. 因此我们来到:

定义 2.1.3. 设 A_1, A_2 是两事件. 若

$$P(A_1 A_2) = P(A_1)P(A_2),$$

则称 A_1 与 A_2 独立.

习题

1. 同时掷三枚硬币.

- (a) 写出所有可能的结果;
- (b) 所有结果出现的概率都相等吗?
- (c) 求出这些概率.

把试验换为同时掷两枚骰子, 再做上面的问题.

2. 四只黑球和四只白球排成一列. 求所有黑白球都是间隔着的概率.
3. 有3个黑球, 2个红球. 现依次挑出2个. 考虑有放回和无放回两种不同的挑法. 求下列事件的概率: 1. 第一次挑出的是红球; 2. 第二次挑出的是红球; 3. 两个都是红球; 4. 两个都是黑球; 5. 第一个是黑球第二个是红球; 6. 第一个是红球第二个是黑球.
4. 在1, 2, 3, 4, 5这5个数字中依次挑出两个, 假设20种结果都是等可能的. 求下列概率: 1. 第一次挑出的是奇数; 2. 第二次挑出的是奇数; 3. 两次都是奇数; 4. 两次都是偶数; 5. 第一次是奇数第二次是偶数; 6. 第一次是偶数第二次是奇数.
5. 将1, 3, 8三个数字随意排列, 求排出来的数字大于350的概率.
6. 将 m 个球投入 n 个盒子中($m \leq n$), 求至少有一个盒子里有不止一个球的概率. (当 $n = 365$ 时, 这就是 m 个人中至少有两个人同生日的概率.)
7. 赤橙黄绿青蓝紫, 谁持彩练当空舞. 将赤橙黄绿青蓝紫七个球随机排列, 求赤橙黄三球按从左到右的顺序紧靠在一起的概率, 不计顺序紧靠在一起的概率.
8. 将 n 个球投入 n 个桶中, 求刚好有一个桶是空桶的概率.
9. 一个人的帽子混在一堆 n 个帽子里, 他一个个地取出来. 求他在第 k 次取到自己帽子的概率($1 \leq k \leq n$).
10. 设有 n 个球, 其中一个白球, 一个红球, $n - 2$ 个黑球. 把它们随机排成一列. 求在白球和红球之间恰有 r 个黑球的概率, $r \leq n - 2$.
11. 某省有 A, B, \dots, T 共60个县. 该省的某个高校的某个班共招收50人, 全部来自本省, 且每人来自任一县的概率相等, 各人之间是独立的. 求:
 - (a) 有 k 人来自 A 县的概率, $k = 0, 1, \dots, 50$;
 - (b) 有某个县剃光头的概率;
 - (c) 有某个县不止一人的概率.

2.2 有限概型

有限概型是从古典概型向一般的概率模型过渡的第一步, 它突破了每个样本点具有相同的概率的局限. 所谓有限概型(有限概率模型), 其特点是试验只有有限个结果(这和古典概型一样), 但每个结果出现的概率不必一样. 比如掷一枚正反面不均匀的硬币; 再比如从装有 m 个黑球和 n 个红球的袋中摸出一球, 且只看摸出的球的颜色.

定义 2.2.1. 设一个试验可能的结果有 n 个:

$$\Omega = \{\omega_1, \dots, \omega_n\}.$$

设 ω_i 出现的概率为 p_i , $\sum_{i=1}^n p_i = 1$. 令

$$P(\omega_i) := p_i,$$

而对任意 $A \subset \Omega$, 令

$$P(A) := \sum_{\omega \in A} P(\omega).$$

则 (Ω, P) 称为有限概型, 称 $P(A)$ 为事件 A 的概率.

和有限概型不同, 此时在计算 $P(A)$ 时, 不能通过计算 A 中所包含的样本点的数目得到.

有限概型和古典概型都是只有有限个样本点, 它们既有区别也有联系. 区别是显然的, 因为它不要求每个样本点发生的概率相等; 除了古典概型是有限概型的特例外, 它们间还有什么联系呢?

我们知道, 同一个试验根据不同的目的, 可以选择不同的概率空间. 再看前面提到过的例子, 掷两枚骰子的试验自然的样本空间是

$$\Omega_1 = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}.$$

如果试验的目的仅仅是为了比较两个骰子点数之和的大小, 样本空间也可以取作

$$\Omega_2 = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}.$$

当然, 此时也可以用 Ω_1 来描述. 如果用 Ω_1 则是古典概型, 用 Ω_2 则是有限概型. 此时, Ω_2 中的一个样本点对应着 Ω_1 中多个样本点, 由于 Ω_2 中的一个样本点对应 Ω_1 中样本点的个数不相等, 因此 Ω_2 中样本点的概率不相等.

再例如, 考虑 n 重Bernoulli试验, 我们知道可以用古典概型描述, 这时概率空间例如可取为

$$\Omega_1 := \{(i_1, \dots, i_n), i_k \in \{0, 1\}\},$$

而每个样本点都具有相同的概率 2^{-n} . 但是, 如果我们只关心这 n 重试验中成功的次数, 那么样本空间可取为

$$\Omega_2 := \{0, 1, \dots, n\}.$$

而成功 k 次的概率为

$$P(k) = C_n^k 2^{-n}, \quad k = 0, 1, \dots, n.$$

上面两个例子中的 Ω_2 虽然不是古典概型了, 但却是从古典概型脱胎出来的. 当然, 并不是所有的有限概型都能这样从古典概型脱胎而来. 比如说, 最容易想到的是, 在掷硬币的Bernoulli试验中, 硬币也有不均匀的时候.

例1. 掷一枚可能未必均匀的硬币, 仍然用1表示出现正面, 0表示出现反面. 此时仍有

$$\Omega = \{0, 1\}.$$

但

$$P(1) = p, \quad P(0) = q, \quad p, q > 0, \quad p + q = 1.$$

这个模型也称为Bernoulli概型.

例2. 将上面那枚硬币掷 n 次, 则样本空间为

$$\Omega = \{(i_1, i_2, \dots, i_n), i_k = 0, 1, \forall k = 1, 2, \dots, n\}.$$

令

$$P((i_1, i_2, \dots, i_n)) = p^{\sum_{k=1}^n i_k} q^{n - \sum_{k=1}^n i_k}, \quad \forall (i_1, i_2, \dots, i_n).$$

这个模型仍然称为 n 重Bernoulli概型, 或称为二项概型.

例3. 如果把硬币换为骰子, 那么每次可能出现的结果就不是两个而是六个. 对这样的问题, 我们可以考虑一般的多项概型. 此时可取

$$\Omega := \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \{1, 2, \dots, 6\}\},$$

$$P(\omega) := p_1^{\alpha_1(\omega)} \dots p_6^{\alpha_6(\omega)},$$

其中 $p_i > 0$, $\sum_{i=1}^6 p_i = 1$, $\alpha_i(\omega) = \sum_{k=1}^n \delta_{i, \omega_k}$, 即 $\alpha_i(\omega)$ 是 ω 的分量中等于 i 的个数.

这个模型称为多项概型.

由定义可以看出, 有限概型中的概率具有以下古典概率所具有的基本性质:

定理 2.2.2. (i) $\forall A, 0 \leq P(A) \leq 1$;

(ii) $P(\Omega) = 1, P(\emptyset) = 0$;

(iii) 单调性:

$$A \subset B \implies P(A) \leq P(B);$$

(iv) 有限可加性: $\forall k$ 及 A_1, \dots, A_k 若 $A_i A_j = \emptyset, \forall i, j$, 则有加法公式

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

证明. 显然. □

这些结论有一些直接的推论. 在最简单的 $k = 2$ 的情形, 加法公式就成为

$$P(A_1 + A_2) = P(A_1) + P(A_2).$$

特别地, 因为 $AA^c = \emptyset, A + A^c = \Omega$, 故有

$$P(A) + P(A^c) = 1.$$

与加法公式等价的是下面的减法公式: 若 $A \subset B$, 则由于 $B = (B - A) + A$, 故有

$$P(B - A) = P(B) - P(A).$$

如果 A_1 与 A_2 有交集, 加法公式就会复杂一些. 此时, 注意到

$$A_1 \cup A_2 = (A_1 - A_1 A_2) + A_2,$$

而

$$(A_1 - A_1 A_2) \cap A_2 = A_1 A_2 - A_1 A_2 A_2 = A_1 A_2 - A_1 A_2 = \emptyset,$$

故有

$$P(A_1 \cup A_2) = P(A_1 - A_1A_2) + P(A_2) = P(A_1) + P(A_2) - P(A_1A_2).$$

这个公式可以推广到多个事件的场合, 请自己写出并证明.

习题

1. 证明: 在二项概型中

$$\sum_{i_1, \dots, i_n} P((i_1, i_2, \dots, i_n)) = 1.$$

这里 \sum_{i_1, \dots, i_n} 表示对所有可能的 $i_k = 0, 1$ 求和. 对多项分布叙述平行的结果并证明之.

2. 考虑 n 重 Bernoulli 试验. 设 $1 \leq i < j \leq n$. 以 A_k 表示第 k 次掷出正面. 证明:

$$P(B_i B_j) = P(B_i)P(B_j),$$

其中 $B_k = A_k$ 或 A_k^c . 并且一般地有

$$P(B_{i_1} B_{i_2} \cdots B_{i_l}) = P(B_{i_1}) \cdots P(B_{i_l}), \quad \forall 1 \leq l \leq n, 1 \leq i_1 < i_2 < \cdots < i_l \leq n.$$

对多项概型叙述平行的结果并证明之.

3. 在二项概型中, 以 α 表示掷出正面的次数. 证明: $P(\alpha = k)$ 在 k 满足 $(n+1)p - 1 \leq k \leq (n+1)p$ 时达到最大值. 对多项分布叙述平行的结果并证明之.

4. m 个黑球 n 个白球排成一列. 求 k 个黑球前是白球的概率.

5. n 个座椅排成一排, 编号为 $1, 2, \dots, n$. 现有 n_1 个男孩和 n_2 个女孩坐上去, $n_1 + n_2 = n$. 设 $n!$ 种不同的坐法是等可能的. 以 α_1 表示前面座位是男孩的男孩人数, α_2 表示前面座位是女孩的男孩人数, α_3 表示前面座位是男孩的女孩人数, α_4 表示前面座位是女孩的女孩人数. 求 $(\alpha_1, \alpha_2, \alpha_3, \alpha_4)$ 的概率分布, 即对任意 i, j, k, h 求

$$P((\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (i, j, k, h)).$$

6. 证明: $\forall k$ 及 A_1, \dots, A_k ,

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

7. 证明:

$$P(A \Delta B) = P(A) + P(B) - 2P(AB).$$

8. 证明:

$$P\left(\bigcap_{i=1}^n A_i\right) \geq \sum_{i=1}^n P(A_i) - (n-1).$$

2.3 随机变量

Ω 上的函数, 称为随机变量.

随机变量是我们随时都会碰到的. 比如说做人口的抽样调查, 样本空间为

$$\Omega = \{\omega_1, \dots, \omega_n\}.$$

但并不是抽完样就完了的, 你总要记录下一些数据——身高啊, 年龄啊, 等等. 那么这些数据就是随机变量.

此外有的随机变量也可以没有实际意义, 但可以带来理论研究上的方便. 比如说一个集合的示性函数就是如此.

定义 2.3.1. 设 ξ 是随机变量, 定义其期望为其(加权)平均值

$$E[\xi] := \sum_{i=1}^n \xi(\omega_i)P(\omega_i).$$

在古典概型的特殊情形, 期望

$$E[\xi] = \frac{1}{n} \sum_{i=1}^n \xi(\omega_i)$$

就是普通的平均值. 一般情形下则是一个加权平均值. 之所以要加权, 是因为每个 ω_i 发生的概率不一样, 所以对不同的 i , $\xi(\omega_i)$ 出现的概率也不一样, 因而在取平均值时占的比重也应该不一样.

我们有:

定理 2.3.2. (i) 设 ξ, η 是随机变量, $a, b \in \mathbb{R}$, 则

$$E[a\xi + b\eta] = aE[\xi] + bE[\eta];$$

(ii) 设 $A \subset \Omega$, 则

$$E[1_A] = P(A).$$

(iii) 设 ξ 的值域为 $\{x_1, \dots, x_m\}$, f 为 \mathbb{R} 上的函数, 则

$$E[f(\xi)] = \sum_{k=1}^m f(x_k)P(\xi = x_k).$$

特别地, 当 $f(x) = x$ 时,

$$E[\xi] = \sum_{k=1}^m x_k P(\xi = x_k).$$

这里及以后我们使用缩写

$$\{\xi = x\} = \{\omega : \xi(\omega) = x\}.$$

证明. 前两个结果是显然的, 第三个只是简单地合并同类项, 兹说明如下. 令

$$A_k := \{i : \xi(\omega_i) = x_k\}.$$

则诸 A_k 两两不交且

$$\bigcup_{k=1}^m A_k = \{1, 2, \dots, n\}.$$

于是

$$\begin{aligned} E[f(\xi)] &= \sum_{k=1}^m \sum_{i \in A_k} f(\xi(\omega_i)) P(\omega_i) \\ &= \sum_{k=1}^m f(x_k) \sum_{i \in A_k} P(\omega_i) \\ &= \sum_{k=1}^m f(x_k) P(\xi = x_k). \end{aligned}$$

这里最后一步用到了

$$\{\xi = x_k\} = \{\omega_i, i \in A_k\}.$$

□

从第三个结果我们看到, 随机变量的期望只与其值域 $\{x_i\}$ 和概率 $p_i := P(\xi = x_i)$ 有关, 也就是说只与 $\{(x_i, p_i)\}$ 有关. 由于 $\{(x_i, p_i)\}$ 描述了 ξ 的值是以怎样的概率分布在各处的, 所以

定义 2.3.3. 设 ξ 是随机变量, 值域为 $\{x_1, \dots, x_n\}$. 令 $p_i = P(\xi = x_i)$. 则 $\{(x_i, p_i), i = 1, \dots, n\}$ 称为 ξ 的概率分布, 或概率分布列, 简称为 ξ 的分布或分布列.

注意这里 $x_i, i = 1, 2, \dots, m$ 是互异的, 并且 $\omega \mapsto \xi(\omega)$ 未必是单射, 即一个 $\{\xi = x_i\}$ 中可能含有多个 ω .

我们可以立即给出这个结果的一点应用——我们在上节末曾希望你自已得到下面的公式, 你记得吗? 得到了吗?

命题 2.3.4. 设 A_1, \dots, A_m 是事件, 则

$$\begin{aligned} P\left(\bigcup_{k=1}^m A_k\right) &= \sum_{k=1}^m P(A_k) - \sum_{1 \leq k_1 < k_2 \leq m} P(A_{k_1} A_{k_2}) + \\ &\quad + \dots + (-1)^{m-1} P(A_1 \cdots A_m). \end{aligned}$$

证明. 我们有

$$\begin{aligned}
 & 1 \left(\bigcup_{k=1}^m A_k \right) \\
 = & 1 - 1 \left(\left(\bigcup_{k=1}^m A_k \right)^c \right) = 1 - 1 \left(\bigcap_{k=1}^m A_k^c \right) \\
 = & 1 - \prod_{k=1}^m 1(A_k^c) = 1 - \prod_{k=1}^m (1 - 1(A_k)) \\
 = & \sum_{k=1}^m 1(A_k) - \sum_{1 \leq k_1 < k_2 \leq m} 1(A_{k_1})1(A_{k_2}) \\
 & + \sum_{1 \leq k_1 < k_2 < k_3 \leq m} 1(A_{k_1})1(A_{k_2})1(A_{k_3}) + \cdots + (-1)^{m-1} 1(A_1)1(A_2) \cdots 1(A_m) \\
 = & \sum_{k=1}^m 1(A_k) - \sum_{1 \leq k_1 < k_2 \leq m} 1(A_{k_1} A_{k_2}) \\
 & + \sum_{1 \leq k_1 < k_2 < k_3 \leq m} 1(A_{k_1} A_{k_2} A_{k_3}) + \cdots + (-1)^{m-1} 1(A_1 A_2 \cdots A_m).
 \end{aligned}$$

在等式两边取期望即可. □

由此可反推出一个计数公式.

推论 2.3.5. 设 A_1, \dots, A_m 是有限集. 以 $N(A)$ 表示 A 中的元素的个数, 则

$$\begin{aligned}
 N \left(\bigcup_{k=1}^m A_k \right) &= \sum_{k=1}^m N(A_k) - \sum_{1 \leq k_1 < k_2 \leq m} N(A_{k_1} A_{k_2}) + \\
 &+ \cdots + (-1)^{m-1} N(A_1 \cdots A_m).
 \end{aligned}$$

证明. 令

$$\Omega := \bigcup_{k=1}^m A_k.$$

考虑 Ω 上的古典概型. 则

$$\begin{aligned}
 P \left(\bigcup_{k=1}^m A_k \right) &= \sum_{k=1}^m P(A_k) - \sum_{1 \leq k_1 < k_2 \leq m} P(A_{k_1} A_{k_2}) + \\
 &+ \cdots + (-1)^{m-1} P(A_1 \cdots A_m).
 \end{aligned}$$

再将古典概率的定义代入即可. □

利用定理 2.3.4 有时可以计算一些难以直接计算的问题.

例 1. 某班有 n 个人. 将 n 个练习本随意分发下去, 求至少一个本子发对人的概率.

解. 以 A 表示这个事件. 令 A_i 表示第 i 个人的本子发到他或她手上. 则

$$P(A_i) = \frac{1}{n}, \quad i = 1, 2, \dots, n,$$

$$P(A_i A_j) = \frac{1}{n(n-1)}, \quad i, j = 1, \dots, n, i \neq j,$$

$$P(A_i A_j A_k) = \frac{1}{n(n-1)(n-2)}, \quad i, j = 1, \dots, n; i, j, k \text{ 互不相等},$$

$$\dots\dots\dots$$

$$P(A_1 \cdots A_n) = \frac{1}{n!}.$$

所以

$$P(A) = n \cdot \frac{1}{n} - C_n^2 \frac{1}{n(n-1)} + \cdots + (-1)^{n-1} \frac{1}{n!}$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1} \frac{1}{n!}.$$

我们顺便注意到, $n \rightarrow \infty$ 时, 这个概率趋于 e^{-1} . e 就这样和概率论联系上了, 我们以后还将看到更多的联系.

下面我们介绍几个常用的分布.

例1. Bernoulli分布. 考虑掷硬币的试验. 设出现正面的概率为 $p \in (0, 1)$. 令 $q = 1 - p$. 以1表示出现正面, 0表示出现反面, 则概率空间为

$$\Omega = \{0, 1\}, \quad P(0) = q, P(1) = p.$$

令

$$\xi(\omega) = \omega.$$

则 ξ 的分布为

$$\{(1, p), (0, q)\}$$

这个分布称为Bernoulli分布, 记为 $B(p)$. 而

$$E[\xi] = 1 \cdot p + 0 \cdot q = p.$$

从此以后, 我们将用 $\xi \sim B(p)$ 表示 ξ 服从Bernoulli分布. 其它分布也使用类似的记号.

例2. 二项分布. 将上面的硬币掷 n 次. 同样以1表示出现正面, 0表示出现反面, 则概率空间为

$$\Omega := \{\omega := (\omega_1, \dots, \omega_n), \omega_i \in \{0, 1\}\}.$$

以 ξ 表示这 n 次试验中正面出现的次数, 也称为成功的次数, 即

$$\xi(\omega) = \sum_{i=1}^n \omega_i.$$

则

$$P(\xi = k) = P\left((\omega_1, \dots, \omega_n) : \sum_{i=1}^n \omega_i = k\right) = C_n^k p^k q^{n-k}.$$

所以 ξ 的分布为

$$\{(k, C_n^k p^k q^{n-k}), k = 0, 1, \dots, n\}.$$

这个分布记为 $B(n, p)$.

下面我们求 $E[\xi]$. 当然, 你可以按照定义求, 或者用公式

$$E[\xi] = \sum_{k=0}^n k P(\xi = k).$$

但这两种方法都有点麻烦. 最简单的方法如下. 定义

$$\xi_i(\omega) = \omega_i,$$

即 ξ_i 是只依赖于第 i 次试验的随机变量: 当第 i 次正面时, $\xi_i = 1$; 反面时, $\xi_i = 0$. 而

$$P(\xi_i = 1) = \sum_{\omega_i=1} p^{\sum_{j \neq i} \omega_j} q^{n-1-\sum_{j \neq i} \omega_j} p,$$

其中 $\sum_{\omega_i=1}$ 是对所有可能的满足 $\omega_i = 1$ 的 ω 求和. 因此

$$\sum_{\omega_i=1} p^{\sum_{j \neq i} \omega_j} q^{n-1-\sum_{j \neq i} \omega_j} = (p+q)^{n-1} = 1.$$

从而

$$P(\xi_i = 1) = p.$$

类似地有

$$P(\xi_i = 0) = q.$$

所以 $\xi_i \sim B(p)$. 由于 $\xi = \sum_{i=1}^n \xi_i$. 因此

$$E[\xi] = \sum_{i=1}^n E[\xi_i] = np.$$

上面我们证明了 $P(\xi_i = 1) = p$. 这从直观上是显然的, 因为 ξ_i 只与第 i 次试验有关, 所以它自然应该服从Bernoulli分布. 我们能从定义出发直接证明这一点, 说明了我们建立的数学模型是合理的.

二项分布源于一个十分简单的试验模型, 但它刺激了不止一个重要领域的研究, 引导出了不止一个重要的分布. 事实上, 它之于概率论的重要性就像 $\{0, 1\}$ 之于计算机科学一样.

二项分布的一个自然推广是多项分布. 这里”自然的”三字既有理论上自然的含义, 也有从实际问题中自然产生的意思. 例如, 如果对产品的抽样检测只需要分为合格品与不合格品的话, 那么就是二项分布的问题; 但如果要分为一等品, 二等品, \dots , 等外品的话, 那就需要多项分布了. 其数学模型如下.

例3. 多项分布. 设样本空间为

$$\Omega = \{\omega = (\omega_1, \dots, \omega_n), \omega_i \in \{1, 2, \dots, m\}\}.$$

设 $p_i > 0$, $\sum_{i=1}^m p_i = 1$. 定义

$$p(\omega) = p_1^{\sum_{k=1}^n \delta_{1\omega_k}} p_2^{\sum_{k=1}^n \delta_{2\omega_k}} \dots p_m^{\sum_{k=1}^n \delta_{m\omega_k}}.$$

令

$$\xi_i(\omega) := \sum_{k=1}^n \delta_{i\omega_k}, \quad i = 1, 2, \dots, m.$$

则 ξ_i 表示在 n 次试验中结果 ω_i 出现的次数. 而对 $k_i \in \mathbb{N}_+$, $\sum_{i=1}^m k_i = n$, 有

$$P(\xi_1 = k_1, \dots, \xi_m = k_m) = C_n^{k_1, \dots, k_m} p_1^{k_1} \dots p_m^{k_m},$$

其中

$$C_n^{k_1, \dots, k_m} = \frac{n!}{k_1! \cdots k_m!}.$$

这个分布称为多项分布.

随机变量还可以考虑多维的. 以二维为例, 一个二维随机变量就是定义在 Ω 上取值于 \mathbb{R}^2 的函数 $\omega \mapsto (\xi(\omega), \eta(\omega))$. 对二维随机变量, 设其值域为 $\{(x_i, y_j)\}$, $p_{ij} := P((\xi, \eta) = (x_i, y_j))$, 那么其分布列就是

$$((x_i, y_j), p_{ij}).$$

习题

1. 设 $\xi \sim B(n, p)$. 计算 $E[\xi^m]$, m 为正整数.
2. 用例3的记号. 令 $\xi := \sum_{i=1}^m \alpha_i \xi_i$, 其中 α_i 是常数. 计算 $E[\xi]$.
3. 设 (ξ, η) 的分布列为 $\{(x_i, y_j), p_{ij}\}$. 求
 - (a) ξ 和 η 的分布列;
 - (b) $\xi + \eta$ 的分布列.
4. 举例说明, 有这样的二维随机变量 (ξ, η) 与 (X, Y) , 它们的概率分布不同, 但 ξ 与 X 的概率分布相同, η 与 Y 的概率分布相同.

2.4 条件概率

任何概率都是在一定条件下考虑的. 从这个意义上讲, 任何概率都是条件概率. 只不过当考虑具体问题时, 大条件就默认了, 不再重复了, 条件概率就简称概率. 这时, 如果在大条件下还有小条件, 那么, 为了区别起见, 小条件下的概率就称为条件概率. 在概率论的无论是理论还是应用中, 条件概率都是一个经常使用的必不可少的工具, 就像大景区中的小景区要另收门票一样.

下面我们来看怎么定义条件概率这个概念. 我们先从古典概型出发, 看一个具体例子.

例1. 一枚骰子连续掷两次. 求:

- (i) 两次点数之和大于7的概率;
- (ii) 若第一次已经掷出4点, 两次点数之和大于7的概率.

解. 第一个问题,

$$\Omega = \{(i, j), i, j = 1, \dots, 6\}.$$

令

$$\begin{aligned} A &= \{(i, j) : i + j > 7\} \\ &= \{(2, 6), (3, 5), (3, 6), (4, 4), (4, 5), (4, 6), (5, 3), (5, 4), \\ &\quad (5, 5), (5, 6), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}. \end{aligned}$$

则

$$N(\Omega) = 36, \quad N(A) = 15.$$

故

$$P_1 = \frac{15}{36}.$$

第二个问题, 因为第一次已经掷过, 所以只需考虑第二次掷的结果, 此时样本空间为

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

而

$$A = \{4, 5, 6\}.$$

所以

$$P_2 = \frac{3}{6} = \frac{1}{2}.$$

我们注意到 $P_2 > P_1$. 这是符合逻辑的, 因为第一次掷出了4点, 这个数字是有利于两次一共掷出大于7点这个事件的, 所以在这个条件下, 概率会加大; 反之, 如果第一次掷出的是3点, 则条件概率会减少, 大家可以自己算一下.

这个例子说明, 在有附加条件的情况下, 或者有更多的信息的条件下, 概率也会随之变化. 这就引导出了条件概率这个概念.

上面的问题可以抽象成下面的形式:

设 A 为事件. 如果我们现在确切地知道 A 已经发生, 而进一步问某个事件 B 发生的概率是多少?

怎么考虑这个问题呢? 我们先看古典概型的情况.

这时, 总样本的空间就缩小了, 但还是运用一样的原则: 用有利事件的样本点数除以总样本点数. 这时条件为 A 已发生, 同时又要 B 发生, 所以有利事件的样本点数是 $N(AB)$, 而总样本点数是 $N(A)$. 从而, 如果我们以 $P(B|A)$ 记这个概率的话, 则

$$P(B|A) = \frac{N(AB)}{N(A)}.$$

如果用原概率表示, 则有

$$P(B|A) = \frac{N(BA)/N(\Omega)}{N(A)/N(\Omega)} = \frac{P(BA)}{P(A)}.$$

下面再看有限概型中条件概率的定义. 设

$$\Omega = \{\omega_1, \dots, \omega_n\}.$$

如果 $A, B \subset \Omega$. 怎么样定义条件概率 $P(B|A)$ 呢?

此时因为不再是古典概型, 所以不能通过数样本点的个数来定义条件概率. 不过基本原则还是适用的, 即这时样本空间缩小了, 即由 Ω 变成了 A , A 变成必然事件了, 因而 A 的概率由 $P(A)$ 变成了 1. A 成为样本空间后, 仍然是有限概型. 根据比例原则——没有理由不遵循这个原则吧?—— A 中每个样本点的原概率乘以因子 $\frac{1}{P(A)}$ 就形成了 A 上的一个概率. 这样, AB 发生的(条件)概率也就变成了

$$P(B|A) = \sum_{\omega \in AB} \frac{1}{P(A)} P(\omega) = \frac{1}{P(A)} \sum_{\omega \in AB} P(\omega) = \frac{P(AB)}{P(A)}.$$

形式上它和古典概型里的定义是一致的, 但摆脱了古典概型的束缚.

总结起来, 便有了下面的定义.

定义 2.4.1. 设 A, B 为事件且 $P(A) > 0$. A 发生的条件下 B 发生的条件概率定义为

$$P(B|A) := \frac{P(AB)}{P(A)}.$$

条件概率有如下性质:

命题 2.4.2. (i)

$$P(A|A) = 1, \quad P(\emptyset|A) = 0;$$

(ii)

$$B \subset C \implies P(B|A) \leq P(C|A);$$

(iii)

$$A_i A_j = \emptyset, \forall i, j = 1, \dots, i \neq j \implies P\left(\sum_{i=1}^m A_i | A\right) = \sum_{i=1}^m P(A_i | A);$$

(iv)

$$P\left(\bigcup_{i=1}^m A_i | A\right) \leq \sum_{i=1}^m P(A_i | A);$$

这些公式都是直接从定义推出, 请自行给出证明.

因为在计算 $P(B|A)$ 时, 样本空间缩小了, 所以计算有可能方便些. 把定义条件概率的公式变一下形, 就成为

$$P(BA) = P(A)P(B|A). \quad (4.2)$$

这个公式提供了计算上的一种便利: 把一个复杂的 BA 的计算分解为计算 $P(A)$ 及 $P(B|A)$.

反复利用这个公式, 我们可得到下面的一般乘法公式:

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \cdots P(A_n|A_1 \cdots A_{n-1}).$$

当然, 我们需要假设 $P(A_1 \cdots A_n) > 0$. 不过这是废话, 如果它等于零, 还要我们算什么呢?

若 $P(A)P(B) > 0$, 由(4.2)有

$$P(A)P(B|A) = P(B)P(A|B)(= P(AB)).$$

因此

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}. \quad (4.3)$$

这就是所谓的Bayes公式³, 其中 $P(A)$ 称为先验概率, $P(A|B)$ 称为后验概率. 其用处在于: 如果把 A 理解为原因, B 理解为结果, 那么 $P(B|A)$ 就是已知原因推测结果, 而 $P(A|B)$ 就是知道结果, 推测导致该结果的原因.

例1. 一个球队, 第一场比赛获胜的概率为 $\frac{1}{2}$; 如果第一场获胜, 那么第二场获胜的概率是 $\frac{3}{4}$; 如果前两场都获胜, 那么第三场获胜的概率是 $\frac{4}{5}$. 问该球队连胜三场的概率.

解. 以 A_i 表示第 i 场获胜, 则

$$P(A_1) = \frac{1}{2}, \quad P(A_2|A_1) = \frac{3}{4}, \quad P(A_3|A_1 A_2) = \frac{4}{5}.$$

³Thomas Bayes, 1702-1761, 英国数学家.

所以

$$P(A_1A_2A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) = \frac{1}{2} \cdot \frac{3}{4} \cdot \frac{4}{5} = \frac{3}{10}.$$

使用条件概率时不能想当然. 比如, 若 $A = A_1 \cup A_2$, 且 $P(B|A_1) = P(B|A_2)$ 时, 也未必有 $P(B|A_1) = P(B|A_2) = P(B|A)$.

例2. 设一个家庭有两个孩子. 求:

- (i) 在已知有一个女孩的条件下, 两个都是女孩的概率;
- (ii) 在已知大孩子是女孩的条件下, 两个都是女孩的概率;
- (iii) 在已知小孩子是女孩的条件下, 两个都是女孩的概率.

解. 以 b 表示男孩, g 表示女孩, 则

$$\Omega = \{(b, b), (b, g), (g, b), (g, g)\},$$

$$B = \{(g, g)\}.$$

对问题(i),

$$A = \{(b, g), (g, b), (g, g)\}.$$

所以

$$P(B|A) = \frac{1}{3}.$$

对问题(ii),

$$A_1 = \{(g, b), (g, g)\}.$$

所以

$$P(B|A_1) = \frac{1}{2}.$$

对问题(iii),

$$A_2 = \{(b, g), (g, g)\}.$$

所以

$$P(B|A_2) = \frac{1}{2}.$$

注意 $A = A_1 \cup A_2$, 且 $P(B|A_1) = P(B|A_2) = \frac{1}{2}$, 但 $P(B|A) = \frac{1}{3}$.

例3. 3只口袋, 2只中有1块钱, 1只中有2块钱. 现张三先取一只, 李四打开剩下的一只, 告诉张三说是1块钱, 给张三一次和剩下那只口袋换的机会. 问张三是否要把握这次机会.

解1. 以张三取的结果建立样本空间, 则

$$\Omega = \{1, 2\}.$$

概率是

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

再令

$$A := \{1\}, \quad B := \{2\}.$$

则 A 发生等价于最后剩下2块的口袋, B 发生等价于最后剩下1块的口袋. 但

$$P(A) = \frac{2}{3}, \quad P(B) = \frac{1}{3}.$$

即最后剩下的口袋里是2块钱的概率是剩下的是1块钱的概率的2倍. 所以要换.

解2. 以试验全程的结果样本空间, 则

$$\Omega = \{\omega_1, \omega_2, \omega_3\}.$$

其中

$$\omega_1 := (1, 1, 2), \quad \omega_2 := (1, 2, 1), \quad \omega_3 := (2, 1, 1).$$

$$P(\omega_i) = \frac{1}{3}, \quad i = 1, 2, 3.$$

以A表示李四打开的是1块钱的口袋这个事件, 则

$$A = \{\omega_1, \omega_3\}.$$

于是

$$P(\omega_1|A) = P(\omega_3|A) = \frac{1}{2}.$$

所以换不换都一样.

这两种解法的结果不一样. 到底哪一种正确呢?

出现这种情况是题目本身有歧义, 即如何实现打开有1块钱的口袋?

如果试验的设计是: 李四铁定从剩下的两个里打开一个1元钱的袋子, 则第一种解答正确. 道理是: 在“李四铁定打开1块钱的袋子”这一条件下, 是换有利还是不换有利事实上已经完全由张三第一次摸的结果决定了. 所以我们建立概率模型时, 只需要考虑第一次摸这个试验.

如果试验的设计是: 李四打开的永远是第2次打开的那个袋子, 如果发现是2块钱, 则试验中止; 如果发现是1块钱, 则试验继续, 那么第二种解答正确. 道理是: 张三第一次摸的结果并不能完全决定是换有利还是不换有利, 因为还需要考虑李四再摸一次这个试验.

我们进一步具体模拟一下. 设想进行了300次这样的试验, 那么大体上 $\omega_1, \omega_2, \omega_3$ 各出现100次——让我们就假定它们刚好都是100次. 如果试验是第一种设计, 那么出现 ω_1, ω_2 时, 张三都是要换的, 所以他要换200次才对. 如果是第二种设计, 那么出现 ω_2 时, 试验就中止了, 因此只能看 ω_1 和 ω_3 出现的次数. 由于它们各是100次, 所以换不换都一样.

这个例子告诉我们, 在使用条件概率时, 是要排除一些样本点的. 具体到本例的第二种解法, 题目并没有说明李四摸到 ω_2 后怎么办, 因此在使用条件“李四打开了1块钱的口袋”时, 就直接排除了样本点 ω_2 , 而这个样本点一定是会导致“换之后拿到2元”的结果的, 所以如果是随机地打开一个袋子, 看到2块钱就终止, 看到1块钱就继续, 这实际上是一个不同的设计, 使得最后剩下2块钱的概率小于前一种设计的概率, 因此结论也就由“换”变为“不换”了.

概率论中还有一些类似的所谓“悖论”. 这些“悖论”实际上并不是逻辑上的悖论, 而是问题本身没有说清楚, 有歧义. 一旦消除歧义, 就没有什么“悖论”了.

习题

1. 设 $A = A_1 + A_2$, $P(A_1)P(A_2) > 0$. 证明:

$$P(B|A_1) \wedge P(B|A_2) \leq P(B|A) \leq P(B|A_1) \vee P(B|A_2).$$

2. 设 A, B, C 是三事件, $B \subset C$, $P(B) > 0$. 问是否一定有 $P(A|B) \geq P(A|C)$? 证明或举出反例.

3. 设 A, B 是事件, $P(A)P(A^c) > 0$. 判断下列等式是否正确, 证明或给出反例.

$$P(B|A) + P(B|A^c) = 1,$$

$$P(B|A) + P(B^c|A^c) = 1,$$

$$P(B|A) + P(B^c|A) = 1.$$

4. 设 A, B, C, D 为事件, $C \cup D = \Omega$, $P(A|C) \geq P(B|C)$, $P(A|D) \geq P(B|D)$. 问是否必有 $P(A) \geq P(B)$?

2.5 全概率公式

上面是用概率来定义条件概率, 但在理论研究和实际问题中, 方向往往是相反的, 即我们往往是在各种不同的条件下搞清了条件概率, 然后综合起来就得到无条件的概率.

如果样本空间可以分割为若干块, 例如

$$\Omega = \Omega_1 + \cdots + \Omega_n.$$

那么对任意事件 A , 都有 (见图2.1)

$$P(A) = P(A\Omega_1) + \cdots + P(A\Omega_n) \quad (5.4)$$

$$= P(\Omega_1)P(A|\Omega_1) + \cdots + P(\Omega_n)P(A|\Omega_n). \quad (5.5)$$

这个公式称为全概率公式, 它告诉我们的是, 可以先分若干情况, 在每一种情况下分别计算条件概率(让我们姑且理解为部分概率好了), 然后对这些概率加权平均, 即得到所求的概率. 相对于前面的限制于各部分的概率而言, 这是一个整体概率, 无条件的概率, 即全概率, 所以该公式就叫全概率公式了.

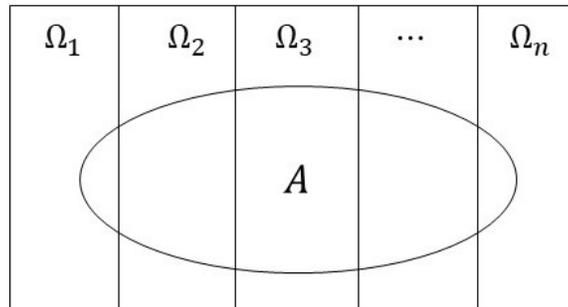


图 2.1: 分割示意图

我们应该注意到, 对给定的 A , 并不需要

$$\Omega_1 + \cdots + \Omega_n = \Omega,$$

而只需要

$$\Omega_1 + \cdots + \Omega_n \supset A,$$

公式(5.4)和(5.5)依然成立.

这个公式背后的思想非常自然：如果直接计算一个概率有困难，那么我们就分情况处理，也就是先在各种额外的条件下计算，然后将计算的结果通过加权平均综合起来，而权重正是诸 $P(\Omega_i)$.

例1. 袋中有 m 只红球 n 只白球，现无放回地连取两次. 设每个结果都是等可能的，求第二次取出的是红球的概率.

解. 令

$$A := \text{第一次红球}, \quad B := \text{第一次白球}, \quad C := \text{第二次红球}.$$

则

$$\begin{aligned} P(C) &= P(A)P(C|A) + P(B)P(C|B) \\ &= \frac{m}{m+n} \cdot \frac{m-1}{m+n-1} + \frac{n}{m+n} \cdot \frac{m}{m+n-1} \\ &= \frac{m}{m+n}. \end{aligned}$$

即第1次和第2次取得红球的概率是一样的. 所以日常生活中，为了体现公平性经常用抽签来解决问题，此时抽签次序对结果是不影响的.

例2. 一个人从装有7只红球8只白球中的袋中任取一只，每只球被取出的概率相同. 如果他取出红球，他就可以掷一枚不均匀骰子，其出现1, 2, 3的概率为 $\frac{1}{9}$ ，出现4, 5, 6的概率为 $\frac{2}{9}$ ；如果他取出白球，他就可以掷一枚均匀的骰子. 问他得到的点数大于3的概率.

解. 以 A_1 表示取得红球， A_2 表示取得白球， A 表示得到的点数大于3. 则

$$\begin{aligned} P(A) &= P(A_1)P(A|A_1) + P(A_2)P(A|A_2) \\ &= \frac{7}{15} \cdot \frac{6}{9} + \frac{8}{15} \cdot \frac{1}{2} \\ &= \frac{26}{45}. \end{aligned}$$

然而在建立实际问题的概率模型时，往往是先有条件概率，再来定义概率的. 例如，有 n 种原因 A_1, \dots, A_n ，可能引起 m 种结果 B_1, \dots, B_m . 根据科学理论及过往经验， A_i 引起 B_j 的概率为

$$P(B_j|A_i) = p_{ij}.$$

这是条件概率，即 A_i 发生的条件下引起 B_j 的条件概率. 显然有

$$\sum_{j=1}^m p_{ij} = 1, \quad \forall i.$$

再设 A_i 出现的概率为 p_i ，则

$$\sum_{i=1}^n p_i = 1.$$

为建立概率模型，令

$$\Omega := \{\omega_{ij}, i = 1, \dots, n, j = 1, \dots, m\},$$

其中

$$\omega_{ij} := A_i B_j.$$

令

$$P(\omega_{ij}) := P(A_i)P(B_j|A_i) = p_{ij}p_i.$$

则 P 是概率. 注意 P 不是随便定义的, 它的定义必须与条件概率的定义相容, 必须与全概率公式相容, 见表2.1.

表 2.1: 由条件概率建立的概率空间

		$P(B_1 A_i)$	$P(B_2 A_i)$	$P(B_3 A_i)$	\cdots	$P(B_n A_i)$	
		B_1	B_2	B_3	\cdots	B_n	
$P(A_1)$	A_1	A_1B_1	A_1B_2	A_1B_3	\cdots	A_1B_n	$P(A_iB_j)$
$P(A_2)$	A_2	A_2B_1	A_2B_2	A_2B_3	\cdots	A_2B_n	$:= P(A_i)P(B_j A_i),$
$P(A_3)$	A_3	A_3B_1	A_3B_2	A_3B_3	\cdots	A_3B_n	$i = 1, 2, \cdots, m,$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	$j = 1, 2, \cdots, n.$
$P(A_m)$	A_m	A_mB_1	A_mB_2	A_mB_3	\cdots	A_mB_n	

在实际工作中往往是要通过现象找原因. 例如医生在诊疗时, 看到的是症状(B_j), 需要的是确定原因 A_i . 因为

$$P(B_j) = \sum_{i=1}^n p_{ij}p_i,$$

所以由Bayes公式有

$$P(A_i|B_j) = \frac{P(A_iB_j)}{P(B_j)} = \frac{p_{ij}p_i}{\sum_{i=1}^n p_{ij}p_i}.$$

这是Bayes公式实际使用时常用的形式. 这一公式在理论上是简洁的, 但实际使用时就需要那些数据 p_i, p_{ij} . 这些数据的准确性直接决定了由Bayes公式得出的结论的准确性.

例如, 一个拒绝现代医学检查手段的(例如印度的)巫医在诊断时, 如果发现一个病人发烧(B), 他要判断是什么原因引起的, 即要在多个原因(比如说 A_1, \cdots, A_m)中蒙一个. 那么根据以前的经验, 他心里有一个 $P(B|A_i)$, 然后他根据当时的情况蒙一个 $P(A_i)$ (比如说, 当时出现covid-19 的概率是多少, 禽流感的概率是多少, 等等), 然后把 $P(B|A_i)$ 和 $P(A_i)$ 一乘, 看 i 等于多少时这个积最大, 就蒙是哪个原因(A_i)引起的. 不过这是我们的善意的推测, 真正的巫医应该不知道用Bayes公式. 需要用Bayes公式的, 是检查手段还不够准确时的医学科学.

例3. 有两个罐子, 一号罐里有2只红球3只白球, 二号罐里有8只红球7只白球. 某人先以各 $\frac{1}{2}$ 的概率随机取一只罐子, 然后随机取出一球, 每球被取出的概率均等. 现取出了红球. 求该球从一号罐里取出的概率.

解. 以 A_i 表示球从 i 号罐里取出, B 表示取到的是红球. 则

$$\begin{aligned} P(B) &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) \\ &= \frac{1}{2} \cdot \frac{2}{5} + \frac{1}{2} \cdot \frac{8}{15} \\ &= \frac{7}{15}, \end{aligned}$$

$$P(A_1B) = P(A_1)P(B|A_1) = \frac{1}{5},$$

所以

$$P(A_1|B) = \frac{P(A_1B)}{P(B)} = \frac{3}{7}.$$

这个数字小于 $\frac{1}{2}$. 这也符合逻辑, 因为一号罐里红球占比比较小, 所以红球是从它那里取出的可能性自然会小一些.

例4. 设在某一地区某种疾病的发病率为 p , 而用某种方法检测该疾病时, 有病的呈阳性的概率为 r_1 , 呈阴性的概率为 r_2 , 无病的呈阳性的概率为 s_1 , 无病的呈阴性的概率为 s_2 . 现随机抽查一人. 求

- (i) 他的结果呈阳性时, 真有病的概率;
- (ii) 他的结果呈阴性时, 真无病的概率;
- (iii) 他第一次呈阳性, 复检仍呈阳性时, 真有病的概率.

解. 以 B 表示检测结果为阳性, A 表示真有病. 则

$$P(A) = p, P(A^c) = 1 - p, P(B|A) = r_1, P(B|A^c) = s_1$$

因此

$$P(BA) = P(B|A)P(A) = r_1p,$$

$$P(B) = P(B|A)P(A) + P(B|A^c)P(A^c) = r_1p + s_1(1 - p).$$

故

$$P(A|B) = \frac{P(BA)}{P(B)} = \frac{r_1}{r_1 + s_1(p^{-1} - 1)}.$$

同理

$$P(A^c|B^c) = \frac{s_2(1 - p)}{r_2p + s_2(1 - p)}.$$

如果第一次检测为阳性, 那么进行第二次检测时, 依然可用同样的推理计算, 只不过由于现在已知第一次结果为阳性, 故那个初始概率 p 应换为现在的 $P(A|B)$, 其它一切数据不变. 故两次检测都为阳性时真有病的概率为

$$\frac{r_1^2}{r_1^2 + s_1^2(p^{-1} - 1)}.$$

我们来简单分析一下这个结果. 首先, $P(A|B)$ 是 p 的单调上升函数, 这就说明如果这种病的发病率越高, 那么检测为阳性时真有病的概率也就越高. 这与直观想象是一致的. 其次, 如果发病率即 p 很小, 那么 $P(A|B)$ 的表达式中的分母就会比分子大得多, 因此 $P(A|B)$ 也会比较小. 所以, 在这种情况下, 即使检测为阳性也不用太担心. 最后, 如果这是一个好的检测方法, 那么 r_1 应该比较大(接近于1)而 s_1 应该比较小(接近于0), 由此不难发现

$$\frac{r_1^2}{r_1^2 + s_1^2(p^{-1} - 1)} \gg \frac{r_1}{r_1 + s_1(p^{-1} - 1)}.$$

所以若两次检测都呈阳性, 那么得病的概率就会大幅增加.

最后我们介绍全概率公式的一种推广的形式.

此后, 我们说 $\mathcal{P} = \{\Omega_1, \dots, \Omega_n\}$ 构成 Ω 的一个分割, 是指

$$\Omega_1 + \dots + \Omega_n = \Omega.$$

全概率公式提示我们, 对一个分割 $\mathcal{P} = \{\Omega_1, \dots, \Omega_n\}$, $A \subset \Omega$, 需要同时考虑条件概率 $P(A|\Omega_1), \dots, P(A|\Omega_n)$. 对此, 引进下面的概念是方便的:

定义 2.5.1. 定义

$$P(A|\mathcal{P})(\omega) := \sum_{k=1}^n P(A|\Omega_k)1_{\Omega_k}(\omega).$$

$P(A|\mathcal{P})$ 称为给定分割 \mathcal{P} 时 A 发生的条件概率.

因此当 ω, A 均变动时, 给定 \mathcal{P} 时的条件概率是一个二元函数:

$$(\omega, A) \mapsto P(A|\mathcal{P})(\omega).$$

易见, 对固定的 $A, \omega \mapsto P(A|\mathcal{P})(\omega)$ 是取值于 $[0, 1]$ 的随机变量, 并且在每个 Ω_k 上为常数 $P(A|\Omega_k)$. 也就是说, ω 其实是起了一个指示作用, 它在哪个 Ω_k 里, $P(A|\mathcal{P})(\omega)$ 就是给定哪个 Ω_k 时的条件概率. 而对于固定的 $\omega, A \mapsto P(A|\mathcal{P})(\omega)$ 是概率, 即它满足定理2.2.2中的四条性质. 对固定的 \mathcal{P} , 以简洁一点的记号 $P(A, \omega)$ 表示这个函数, 并对任何一个随机变量 ξ , 对固定的 ω , 定义

$$E[\xi|\mathcal{P}](\omega) := E^{P(\cdot, \omega)}[\xi(\omega')] := \sum_{\omega'} \xi(\omega')P(\omega', \omega).$$

它称为关于 \mathcal{P} 的条件期望, 注意它仍是一个随机变量, 且在每个 Ω_k 上恒等于常数. 特别地, 由于

$$E[1_A|\mathcal{P}](\omega) = \sum_{\omega'} 1_A(\omega')P(\omega'|\mathcal{P})(\omega) = \sum_{k=1}^n P(A|\Omega_k)1_{\Omega_k}(\omega) = P(A|\mathcal{P})(\omega),$$

所以全概率公式可以表示为

$$P(A) = E[P(A|\mathcal{P})] = E[E[1_A|\mathcal{P}]].$$

同理, 对于一般的 ξ , 因为在 $\omega \in \Omega_k$ 上 $P(\omega'|\mathcal{P})(\omega) = P(\omega'|\Omega_k)$, 故有

$$E[\xi|\mathcal{P}](\omega) = \sum_{\omega'} \xi(\omega')P(\omega'|\Omega_k), \quad \omega \in \Omega_k.$$

因为此物在每个 Ω_k 上为常数, 故由期望公式有

$$\begin{aligned} E[E[\xi|\mathcal{P}]] &= \sum_{k=1}^n P(\Omega_k) \left(\sum_{\omega'} \xi(\omega')P(\omega'|\Omega_k) \right) \\ &= \sum_{\omega'} \xi(\omega') \left(\sum_{k=1}^n P(\omega'|\Omega_k)P(\Omega_k) \right) \\ &= \sum_{\omega'} \xi(\omega')P(\omega'). \end{aligned}$$

因此,

$$E[\xi] = E[E[\xi|\mathcal{P}]].$$

让我们照葫芦画瓢, 将这个公式称为全期望公式吧. 这个公式可以用下面的例子帮助理解: 假设一个农民预测他今年的收入, 粗略地把年成分为风调雨顺和风不调雨不顺两种情况: 风调雨顺时预计能收入10万, 风不调雨不顺时预计能收入5万. 而风调雨顺的可能性是百分之四十, 风不调雨不顺的可能性是百分之六十. 那么他今年的预期收入将是

$$10 \times 0.4 + 5 \times 0.6 = 7(\text{万元}).$$

当然,你还可以把情况分得更细致,例如风调雨不顺,风不调雨顺等等.

局部算好后加权平均,这就是全概率公式和全期望公式背后的简单思想.

习题

- 袋中有 m 只红球 n 只白球,现无放回地连取 l 次, $l \leq m+n$. 设每个结果都是等可能的,求第 l 次取出的是红球的概率.
- 在一次原子反应中,一个粒子不分裂、分裂为两个粒子、三个粒子的概率分别为 p_1, p_2, p_3 . 求经过两次反应后粒子总数的概率分布.
- 在排球比赛中,一传的到位率是 p_1 , 接住但不到位率是 p_2 , 接飞率是 p_3 ; 在一传到位的情况下,二传组织快攻的概率是 q_{11} , 组织强攻的概率是 q_{12} ; 在一传不到位的情况下,二传组织快攻的概率是 q_{21} , 组织强攻的概率是 q_{22} ; 在快攻的情况下,本队得分率是 r_1 , 在强攻的情况下,得分率是 r_2 . 现对方发球.
 - 求本队的得分率;
 - 一个观众打开电视时,看到本队得分. 求一传到位的概率.
- 在 $\{1, \dots, 10\}$ 中等可能地随机取一个数. 如果取得的是 k , 则再掷 k 次骰子.
 - 求掷出 n 次6点的概率;
 - 假设得到2次6点, 求开始取出的那个数是 m 的概率, $1 \leq m \leq 10$.
 - 求掷出的总点数的数学期望.
- 警察,小偷和证人的故事. 一次盗窃案件中,警察抓到一个嫌疑人,并研判出他作案的概率为 p . 随后找到一个证人,该证人指认嫌疑人就是作案人. 但该证人说真话的概率只有 q . 求该嫌疑人的确就是小偷的概率.
- 用数学语言精确表示如下命题: 如果在所有情况下甲事件发生的概率都大于等于乙事件发生的概率, 那么总体上甲事件发生的概率就大于等于乙事件发生的概率.
- 设 \mathcal{P} 是分割, $P(A|\mathcal{P})(\omega)$ 是给定 \mathcal{P} 时的条件概率. 证明:
 - $\forall \omega, A \mapsto P(A|\mathcal{P})(\omega)$ 都是概率;
 - $\forall A, \omega \mapsto P(A|\mathcal{P})(\omega)$ 是取值于 $[0, 1]$ 的随机变量, 并且在每个 Ω_k 上恒等于常数;
 - $\forall k = 1, \dots, n,$

$$P(\Omega_k|\mathcal{P})(\omega) = 1_{\Omega_k}(\omega).$$
- 证明本节的最后两个公式.

2.6 独立性

我们曾经对古典概型定义了独立性. 这个定义可以原封不动地移植到有限概型的情形, 即我们有:

定义 2.6.1. 设 A, B 是事件, 若

$$P(AB) = P(A)P(B),$$

则称 A, B 独立.

显然, 一个不可能事件与任何事件都是独立的, 一个必然事件与任何事件也都是独立的. 此外, 由条件概率的定义可直接推出

命题 2.6.2. 若 $P(A) > 0$, 那么 A 与 B 独立就等价于

$$P(B|A) = P(B),$$

这个命题说明, 独立就意味着无论有无 A 这个条件, B 发生的概率都不受影响——这符合独立这个词的含义, 即不依赖.

由事件的独立性概念可衍生出事件类的独立性概念. 我们先定义什么叫事件类或者集类.

定义 2.6.3. 设 Ω 为一样本空间. 以 Ω 的子集为元素的集合称为事件类或者集类.

定义 2.6.4. 设 \mathcal{A}, \mathcal{B} 为集类, 若 $\forall A \in \mathcal{A}, B \in \mathcal{B}$, A, B 都独立, 则称 \mathcal{A}, \mathcal{B} 独立.

所以两个事件类的独立, 就是任意从这两个类中各选一个, 它们都是独立的. 现在我们问:

如果 A, B 都与 C 独立, 那么 AB 、 $A \cup B$ 及 $A \setminus B$ 仍然与 C 独立吗?

首先容易证明:

命题 2.6.5. 如果 A, B , 且 $A \supset B$ 都与 C 独立, 那么 $A - B$ 也与 C 独立.

证明.

$$\begin{aligned} P((A - B)C) &= P(AC - BC) \\ &= P(AC) - P(BC) \\ &= P(A)P(C) - P(B)P(C) \\ &= [P(A) - P(B)]P(C) \\ &= P(A - B)P(C). \end{aligned}$$

□

因为 Ω 与所有事件都独立, 所以利用上述结果得到:

推论 2.6.6. 若 A, B 独立, 则 $\{A, A^c\}$ 与 $\{B, B^c\}$ 独立.

上面的命题里, 有一个条件是 $A \subset B$. 在没有这个条件时, 如果 A, B 都与 C 独立时能推出 AB 也与 C 独立, 那么就也能推出 $A \setminus B$ 也与 C 独立, 因为由命题 2.6.5

$$P((A \setminus B)C) = P((A - AB)C) = P(A - AB)P(C) = P(A \setminus B)P(C).$$

但, 不幸的是, 下面的例子表明, 上面问题的回答是否定的, 即 A, B 都与 C 独立时一般不能推出 AB 也与 C 独立.

例1. 抛2次均匀硬币的试验中, 样本空间为 $\Omega = \{(1,1), (1,0), (0,1), (0,0)\}$. 令 $A = \{(1,1), (1,0)\}$, $B = \{(1,1), (0,1)\}$, $C = \{(1,1), (0,0)\}$, 试分析事件 A 、 B 、 C 的独立性.

解. 这是古典概型, 显然 $P(A) = P(B) = P(C) = \frac{1}{2}$, $P(AB) = P(BC) = P(AC) = \frac{1}{4} = \frac{1}{2} \times \frac{1}{2}$. 因此, A, B 都与 C 独立, 但

$$P((AB)C) = P(ABC) = \frac{1}{4},$$

$$P(AB)P(C) = \frac{1}{4} \times \frac{1}{2} \neq \frac{1}{4} = P((AB)C).$$

所以 AB 与 C 不独立. 同理可证 $A \cup B$ 与 C 也不独立.

直观上可以这么理解这个反例. 事件 C 的发生与否不是没有影响事件 A 和 B , 只是这种影响没有改变其概率而已, 亦即 $P(A|C) = P(A)$, $P(B|C) = P(B)$, 这种影响造成的后果是事件 C 的发生却影响事件 AB 的概率了.

你能举出例子, 从直观上看它们不相互影响, 但却不独立吗(见3.4节的例子)?

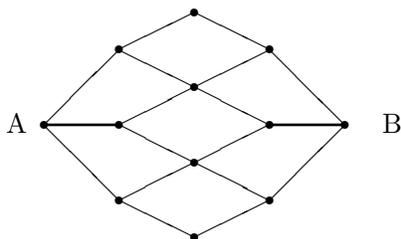
进一步的思考发现, 在此类问题中, 关键是 AB 是否与 C 独立. 一旦 AB 与 C 独立, 那么不光 $A \setminus B$, 连 $A \cup B$ 也都与 C 独立了, 因为由命题2.6.5我们有

$$\begin{aligned} P((A \cup B)C) &= P(AC \cup (B - AB)C) \\ &= P(AC) + P((B - AB)C) \\ &= P(A)P(C) + P(B - AB)P(C) \\ &= (P(A) + P(B) - P(AB))P(C) \\ &= P(A \cup B)P(C). \end{aligned}$$

这就引导我们考虑更一般的问题: 在什么情况下可由两个集类的独立性推出更大的集类的独立性? 看来交集在这里会起到关键的作用. 为准确地回答这个问题, 我们需要准备一些集类方面的知识.

习题

1. 甲乙两人分别独立地以相同的速度从A,B两地出发相向而行. 在每一个路口, 他们的每一个选择都是等可能的. 求他们在途中能会面的概率.



2. 一个人的一次旅行分为三段. 第 i 段晚点的概率为 p_i . 现朋友在终点没有如期接到此人. 求第 i 段发生晚点的概率.
3. 甲乙两队捉对厮杀, 每队各五人, 先胜三盘者获胜. 甲队一号对乙队一号的获胜概率是 p ; 在甲队前一名选手获胜的情况下, 后一名选手获胜的概率也是 p ; 在甲队前一名选手失利的情况下, 后一名选手获胜的概率是 $p/2$. 求整场比赛甲队获胜的概率.

4. 手机上有两款天气预报的APP, 独立地发出天气预报. 甲款的准确率是 p_1 , 乙款的准确率是 p_2 . 张先生学了点概率, 每天早上出门时, 根据这两款APP算出下雨的概率大于等于 p 时就带伞, 小于 p 时就不带. 问:
- p_1, p_2 满足什么关系时, 他就带伞?
 - 他白带了伞的概率是多少?
 - 天下了雨而他没带伞的概率是多少? 题目应更明确
5. 某地某种病毒性疾病的发生率为 p_1 . 做筛检时, 假阴性的概率为 p_2 , 假阳性的概率为 p_3 . 设某人检测结果为阳性, 求他的确感染了此病的概率.
6. 设有 m 个可区分的桶, 编号分别为 $1, \dots, m$. 将3个球依次投入桶中, 每球入第 i 桶的概率为 p_i ($\sum_{i=1}^m p_i = 1$). 求无任何桶中有多于1个球的概率. 又: 如果是投 n 个球呢?
7. 设 A_1, \dots, A_n 独立. 以 $P(A_i), i = 1, 2, \dots, n$ 表示出 $P(A_1 \cdots A_n)$ 及 $P(A_1 \cup \dots \cup A_n)$.
8. 甲乙两人进行射击比赛, 规定谁先击中谁获胜. 假设他们一次射击中击中的概率分别为 p_1, p_2 . 通过抽签决定谁先射, 机会均等. 问他们获胜的概率各是多少? 如果是三个人比赛呢?
9. Euler⁴素数公式的概率证明. 以 \mathcal{P} 表示素数全体. 令

$$\zeta(s) := \sum_{n=1}^{\infty} n^{-s}, \quad s > 1.$$

- 令 $\Omega := \mathbb{N}_{++}$, $P(\{n\}) := \frac{n^{-s}}{\zeta(s)}$. 证明 P 是概率;
- 令 $A_p := \{pn : n \in \mathbb{N}_{++}\}$. 计算 $P(A_p)$;
- 证明 $\{A_p, p \in \mathcal{P}\}$ 独立;
- 证明Euler公式

$$\zeta(s) = \prod_{p \in \mathcal{P}} (1 - p^{-s})^{-1}.$$

2.7 π -类, λ_0 -类与代数

我们曾经定义了事件间的各种运算. 本节我们定义对其中一些运算封闭的集类, 并讨论其基本性质. 我们将证明所谓的 $\lambda_0 - \pi$ 定理. 这个定理是后面要讲的 $\lambda - \pi$ 定理的初级形式, 且前者的证明已经包含了后者的证明中最困难的部分.

我们固定一个样本空间 Ω , 考虑在各种运算下封闭的集类.

首先是所谓 π -类.

定义 2.7.1. 一个非空集类 \mathcal{P} , 若对交封闭, 即满足

$$A, B \in \mathcal{P} \implies AB \in \mathcal{P},$$

则称为 π -类.

⁴Leonhard Euler, 1707-1783, 瑞典-俄罗斯数学家.

正如我们在上节已经看到的一样, 我们以后会更多地看到, 一个集类是否 π -类是非常重要的. 我们如果开始时能小心谨慎地选择一个 π -类作为出发点, 则以后的道路就会十分顺畅.

下面是几个例子.

例1. 设 $\Omega = \mathbb{R}^n$ (实际上可以是任何一个拓扑空间), \mathcal{P} 是所有开集(或闭集)全体. 则 \mathcal{P} 是 π -类.

例2. 设 $\Omega = \mathbb{R}$, $\mathcal{P} := \{(-\infty, a), a \in \mathbb{R}\}$, 则 \mathcal{P} 是 π -类.

例3. 设 $\Omega = [0, 1]$, $\mathcal{P} := \{[0, a], a \in [0, 1]\}$, 则 \mathcal{P} 是 π -类.

例4. 设 $\Omega = \mathbb{R}^n$, $\mathcal{P} := \{[a_1, b_1] \times \cdots \times [a_n, b_n], -\infty < a_i \leq b_i < \infty, \forall i = 1, \dots, n\}$. 则 \mathcal{P} 是 π -类.

其次是 λ_0 -类. 如果说 π -类是出发点的话, λ_0 -类就是催化剂. 一般说来 π -类里的集合是比较简单的, 但经过 λ_0 -类的催化, 就可以衍生出足够复杂足够多的集合.

定义 2.7.2. 一个集类 \mathcal{L}_0 , 若包含 Ω 且对真差封闭, 即满足

$$A, B \in \mathcal{L}_0, A \subset B \implies B - A \in \mathcal{L}_0,$$

则称为 λ_0 -类.

所以, 当然, 一个 λ_0 -类里既有 Ω 也有空集 \emptyset .

下面是几个例子.

例1. 设 $\Omega = [0, 1]$. 令

$$\mathcal{L}_0 := \left\{ \sum_{i=1}^n [a_i, b_i), 0 \leq a_1 \leq b_1 \leq a_2 \leq b_2 \leq \cdots \leq a_n \leq b_n \leq 1, n = 1, 2, \dots \right\}.$$

则 \mathcal{L}_0 为 λ_0 -类.

例2. 设 Ω 为任一空间, $A, B \subset \Omega, AB \neq \emptyset$. 令

$$\mathcal{L}_0 := \{\Omega, \emptyset, A, B, A^c, B^c\}.$$

则 \mathcal{L}_0 为 λ_0 -类. 此外容易看出, 任何一个包含了 $\{A, B\}$ 的 λ_0 -都包含了这个 \mathcal{L}_0 .

例3. 设 A 是事件, 令

$$\mathcal{L}_0 := \{B, B \text{ 与 } A \text{ 独立}\}.$$

则 \mathcal{L}_0 为 λ_0 -类.

这是因为, 当 $B, C \in \mathcal{L}_0$ 且 $B \subset C$ 时, 由命题2.6.5, 所以 $B - C \in \mathcal{L}_0$.

例4. 设 P, Q 都是定义在 Ω 的子集上的函数, $P(\Omega) = Q(\Omega)$ 且

$$AB = \emptyset \implies P(A + B) = P(A) + P(B), \quad Q(A + B) = Q(A) + Q(B).$$

令

$$\mathcal{L}_0 := \{A : P(A) = Q(A)\}.$$

则 \mathcal{L}_0 为 λ_0 -类.

读者可尝试自己写出(很简短的)证明.

显然, 若 \mathcal{L}_0 为 λ_0 -类, 则 $A \in \mathcal{L}_0$ 时有 $A^c = \Omega - A \in \mathcal{L}_0$, 即它对余封闭.

我们还可以证明:

命题 2.7.3. 一个包含了 Ω 的集类为 λ_0 类的充要条件是它对不交并及余封闭.

证明. 设 \mathcal{L}_0 为 λ_0 -类. 设 $AB = \emptyset$. 则 $B \subset A^c$. 因此

$$A^c B^c = A^c - B \in \mathcal{L}_0.$$

于是

$$A + B = (A^c B^c)^c \in \mathcal{L}_0.$$

反之, 设 \mathcal{L}_0 含有 Ω 且对不交并及余封闭. 设 $A \subset B$. 则 $AB^c = \emptyset$. 故

$$B^c + A \in \mathcal{L}_0.$$

因此

$$B - A = (B^c + A)^c \in \mathcal{L}_0.$$

□

因此, 我们可将 λ_0 类理解为含有 Ω 且对不交并与余封闭的集类. 若去掉“不交”的限制, 加强为对任意两个集合的并封闭, 则得到代数的概念.

定义 2.7.4. 一个集类 \mathcal{A} , 若满足

- (i) $\Omega \in \mathcal{A}$;
- (ii) 对并与余封闭:

$$A, B \in \mathcal{A} \implies A \cup B, A^c \in \mathcal{A},$$

则称为代数.

所以 λ_0 -类与代数的差别就在于对并封闭的要求中, 有无不交的前提.

我们看几个例子:

例5. Ω 的所有子集构成的集类为代数.

例6. 集类

$$\{\emptyset, \Omega\}$$

构成代数.

例7. 设 $N(\Omega) = +\infty$, 令

$$\mathcal{A} := \{A \subset \Omega : N(A) \wedge N(A^c) < \infty\}.$$

则 \mathcal{A} 是代数.

但读者可以自己举反例说明, 上面作为 λ_0 -类的例3与例4均不是代数.

我们立即有:

命题 2.7.5. 代数对交与差均封闭.

证明. 设 \mathcal{A} 为代数, $A, B \in \mathcal{A}$. 则

$$AB = (A^c \cup B^c)^c \in \mathcal{A},$$

$$B \setminus A = BA^c \in \mathcal{A}.$$

□

在验证一个非空集类是否为代数时, 下面的结果常常很方便.

命题 2.7.6. 一个非空集类若对余与交两种运算封闭, 则为代数.

证明. 设这集类是 \mathcal{A} , 并设 $A \in \mathcal{A}$. 因为 \mathcal{A} 对余封闭, 所以 $A^c \in \mathcal{A}$. 因为 \mathcal{A} 对交封闭, 所以 $\emptyset = AA^c \in \mathcal{A}$. 再用一次对余封闭, 得 $\Omega = (\emptyset)^c \in \mathcal{A}$.

设 $A, B \in \mathcal{A}$, 则 $A^c, B^c \in \mathcal{A}$. 因此

$$A \cup B = (A^c B^c)^c \in \mathcal{A}.$$

所以 \mathcal{A} 为代数. □

显然, 我们有

推论 2.7.7. 一个集类 \mathcal{A} 为代数的充要条件是它既是 π -类又是 λ_0 -类.

证明留作练习.

由此我们知道, 我们也可以认为, 比之于代数, λ_0 -类差的就是对交封闭. 我们下面将要证明本节最重要的结果, 即从一个 π -类出发扩张而得到的 λ_0 -类依然保留对交封闭这个重要的性质, 因此也就一定是代数. 为此先引入两个概念.

命题 2.7.8. 任给一个非空集类 \mathcal{D} , 一定存在一个代数 \mathcal{A}_0 , 使得对任意代数 $\mathcal{A} \supset \mathcal{D}$, 都有

$$\mathcal{D} \subset \mathcal{A}_0 \subset \mathcal{A}.$$

证明. 令

$$\mathcal{A}_0 = \bigcap \mathcal{A},$$

其中右边的交跑遍所有包含了 \mathcal{D} 的代数. 注意这个定义是有意义的, 因为至少有一个代数包含了 \mathcal{D} , 这就是由 Ω 的所有子集构成的代数. 作为诸代数的交, \mathcal{A}_0 自然也是代数, 且是包含了 \mathcal{D} 的最小的代数. □

定义 2.7.9. 上面命题中的 \mathcal{A}_0 称为 \mathcal{D} 生成的代数, 记为 $\alpha(\mathcal{D})$.

类似地, 我们有:

命题 2.7.10. 任给一个集类 \mathcal{D} , 存在一个 λ_0 -类 \mathcal{L}_0 , 使得对任意 λ_0 -类 $\mathcal{L} \supset \mathcal{D}$, 都有

$$\mathcal{D} \subset \mathcal{L}_0 \subset \mathcal{L}.$$

定义 2.7.11. 上面命题中的 \mathcal{L}_0 称为 \mathcal{D} 生成的 λ_0 -类, 记为 $\lambda_0(\mathcal{D})$.

在前面的例子中, 我们看到, 若 $A, B \subset \Omega$, $AB \neq \emptyset$, 则由 $\{A, B\}$ 生成的 λ_0 -类为

$$\{\Omega, \emptyset, A, B, A^c, B^c\}.$$

此集类并不太复杂. 但倘若在 $\{A, B\}$ 中仅仅加一个元素 AB , 则 $\{A, B, AB\}$ 生成的 λ_0 -类就为

$$\{\Omega, \emptyset, A, B, AB, A^c, B^c, (AB)^c, A \setminus B, B \setminus A, A^c \setminus B^c, B^c \setminus A^c, A \cup B, A^c B^c\}.$$

这个集类就复杂多了. 其原因是 $\{A, B\}$ 不是 π -类而 $\{A, B, AB\}$ 是 π -类. 可见一个集类是否为 π -类对其生成的 λ_0 类的结构是有重大影响的.

由于代数皆为 λ_0 -类, 所以对任意一个集类 \mathcal{C} , 皆有 $\lambda_0(\mathcal{C}) \subset \alpha(\mathcal{C})$. 然而一般说来反过来是不成立的. 不过若 \mathcal{C} 是 π -类, 反过来就成立, 即我们有

定理 2.7.12 ($\pi - \lambda_0$ 定理). 若 \mathcal{C} 为 π -类, 则 $\lambda_0(\mathcal{C}) = \alpha(\mathcal{C})$.

这是我们对前面所说的话的第一个具体例证: 选择 π -类作为出发点会使道路非常顺畅.

这个定理能够成立的理由是下面这个关键事实: 一个 π -类生成的 λ_0 -类仍然是 π -类. 下面就是详细证明.

证明. 显然 $\lambda_0(\mathcal{C}) \subset \alpha(\mathcal{C})$. 因此只需证明 $\lambda_0(\mathcal{C})$ 为代数. 由命题2.7.7, 又只需证明它是 π -类.

令

$$\mathcal{B}_1 := \{B \in \lambda_0(\mathcal{C}) : BA \in \lambda_0(\mathcal{C}), \forall A \in \mathcal{C}\}.$$

则 $\mathcal{B}_1 \supset \mathcal{C}$. 显然 $\Omega \in \mathcal{B}_1$. 再设 $B_1, B_2 \in \mathcal{B}_1, B_1 \subset B_2$. 则 $\forall A \in \mathcal{C}$ 有

$$B_1A, B_2A \in \lambda_0(\mathcal{C}), B_1A \subset B_2A.$$

从而

$$(B_2 - B_1)A = B_2A - B_1A \in \lambda_0(\mathcal{C}).$$

故 $B_2 - B_1 \in \mathcal{B}_1$. 因此 \mathcal{B}_1 是(包含了 \mathcal{C} 的) λ_0 -类. 从而 $\mathcal{B}_1 = \lambda_0(\mathcal{C})$.

再令

$$\mathcal{B}_2 := \{B \in \lambda_0(\mathcal{C}) : BA \in \lambda_0(\mathcal{C}), \forall A \in \lambda_0(\mathcal{C})\}.$$

由刚刚证明的结论, 有 $\mathcal{B}_2 \supset \mathcal{C}$. 再将刚刚用的证明方法如法泡制, 可证 \mathcal{B}_2 为 λ_0 -类. 因此 $\mathcal{B}_2 = \lambda_0(\mathcal{C})$, 而这就是说 $\lambda_0(\mathcal{C})$ 是 π -类. \square

这个证明的特点是分为了两步, 两步的步调实际上是一样的, 就如同当你过一条小溪时, 你可以在溪中央垫一块石头, 第一脚踩在这块石头上, 第二脚就过去了. 但如果你不信邪, 硬要一步过去, 那对不起, 你只能落在水里. 这也符合马克思主义关于无产阶级革命的基本原理, 即不断革命论和革命发展阶段论.

习题

1. 设 Ω 为全空间, $A \subset B \subset \Omega$, 写出含有 A, B 作为其元素的最小 λ_0 -类.

2. 在 \mathbb{R} 中, 令

$$\Pi := \{(-\infty, a) : a \in \mathbb{R}\}.$$

求 $\alpha(\Pi)$.

3. 设 \mathcal{O}, \mathcal{C} 分别表示 \mathbb{R}^d 中的开集类与闭集类. 证明它们都是 π -类, 也都不是 λ_0 -类.

4. 在 \mathbb{R} 中, 令

$$\Pi_0 := \{[a, b) : -\infty \leq a \leq b \leq \infty\},$$

其中 $[-\infty, a)$ 理解为 $(-\infty, a)$,

$$\Pi_1 := \left\{ A : A = \sum_{i=1}^n A_i, A_i \in \Pi_0, n \in \mathbb{N}_+ \right\} \left(\sum_{i=1}^0 A_i = \emptyset \right).$$

证明 Π_0 为 π -类, Π_1 为 λ_0 -类.

5. 在上题中, 将 \mathbb{R} 换成 \mathbb{R}^d , 叙述相应的结论和证明.

6. 设 S 为一有无穷多个元素的空间, 令

$$\mathcal{S} := \{A : A \subset S, N(A) \wedge N(A^c) < \infty\}.$$

证明 \mathcal{S} 为代数.

7. 设 f, g 为 \mathbb{R} 上的可积函数, 且

$$\int_{\mathbb{R}} f(x) dx = \int_{\mathbb{R}} g(x) dx.$$

以大写字母 A 代表 \mathbb{R} 上的区间. 令

$$\mathcal{A} := \left\{ A : \int_A f(x) dx = \int_A g(x) dx \right\}.$$

证明: \mathcal{A} 为 λ_0 -类, 但非 π -类.

8. 设样本空间为 Ω , $A, B \subset \Omega$.

(a) 设 $AB = \emptyset$. 写出 $\alpha(\{A, B\})$;

(b) 去掉假设 $AB = \emptyset$. 写出 $\alpha(\{A, B\})$.

9. 设样本空间为 Ω , $A_i \subset \Omega$, $i = 1, 2, \dots, n$.

(a) 设 $i \neq j$ 时 $A_i A_j = \emptyset$. 写出 $\alpha(\{A_i, i = 1, \dots, n\})$;

(b) 去掉假设 $i \neq j$ 时 $A_i A_j = \emptyset$. 写出 $\alpha(\{A_i, i = 1, \dots, n\})$.

(c) 设 $\mathcal{C} = \{A_1, \dots, A_n\}$ 且

$$\Omega = A_1 + \dots + A_n.$$

写出 $\alpha(\mathcal{C})$.

(d) 设 $A, B \subset \Omega$. 写出 $\alpha(\{A, B\})$.

(e) 两个代数之并仍然是代数吗? 证明或举出反例.

(f) 设 Ω 为一无限集合. 令

$$\mathcal{A} := \{A \subset \Omega : A \text{ 为有限集}\},$$

$$\mathcal{B} := \{A \subset \Omega : A \text{ 或 } A^c \text{ 为有限集}\}.$$

\mathcal{A}, \mathcal{B} 是代数吗? 证明你的结论.

(g) 设 $A_1, \dots, A_n \subset \Omega$, 写出 $\alpha(\{A_i, i = 1, \dots, n\})$.

2.8 重温独立性

现在可以回答前面的问题了, 即是否可以从较小的事件类间的独立性推出较大的事件类间的独立性的问题. 我们有:

定理 2.8.1. 设 \mathcal{C}_1 与 \mathcal{C}_2 是 π -类. 若 \mathcal{C}_1 与 \mathcal{C}_2 独立, 那么 $\alpha(\mathcal{C}_1)$ 与 $\alpha(\mathcal{C}_2)$ 独立.

证明. 令

$$\mathcal{B} := \{B \in \alpha(\mathcal{C}_2) : B \text{ 与 } A \text{ 独立}, \forall A \in \mathcal{C}_1\}.$$

则 $\mathcal{C}_2 \subset \mathcal{B}$. 往证 \mathcal{B} 是 λ_0 -类. 首先, 显然 $\Omega \in \mathcal{B}$. 其次, 设 $B_1, B_2 \in \mathcal{B}$, $B_1 \subset B_2$, 则 $\forall A \in \mathcal{C}_1$,

$$P((B_2 - B_1)A) = P(B_2A) - P(B_1A) = P(B_2)P(A) - P(B_1)P(A) = P(B_2 - B_1)P(A).$$

因此 $B_2 - B_1 \in \mathcal{B}$. 所以 $\mathcal{B} \supset \lambda_0(\mathcal{C}_2)$. 但由命题 2.7.12, $\lambda_0(\mathcal{C}_2) = \alpha(\mathcal{C}_2)$. 所以 $\mathcal{B} = \alpha(\mathcal{C}_2)$.

再令

$$\mathcal{A} := \{A \in \alpha(\mathcal{C}_1) : A \text{ 与 } B \text{ 独立}, \forall B \in \alpha(\mathcal{C}_2)\}.$$

用同样的方法可证 $\mathcal{A} = \alpha(\mathcal{C}_1)$.

这就证明了任取 $A \in \alpha(\mathcal{C}_1)$, $B \in \alpha(\mathcal{C}_2)$, A 与 B 都是独立的. \square

前面定义了两个事件、两个集类的独立. 自然地要问, 该如何定义三个事件、多个事件独立, 以及多个集类的独立呢? 下面就是定义.

定义 2.8.2. 设 I 为任意指标集, $\{A_i, i \in I\}$ 为一族事件. 若对任意有限的 $J \subset I$ 及 $A_i, i \in J$, 都有

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i),$$

则称 $\{A_i, i \in I\}$ 独立.

设 $\{\mathcal{A}_i, i \in I\}$ 为一族事件类. 若对任意若对任意有限的 $J \subset I$ 及 $A_i \in \mathcal{A}_i, i \in J$ 都有

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i),$$

则称 $\{\mathcal{A}_i, i \in I\}$ 独立.

类似于定理 2.8.1, 我们可以证明:

定理 2.8.3. 设 $\forall i \in I$, \mathcal{C}_i 是 π -类. 若 $\{\mathcal{C}_i, i \in I\}$ 独立, 那么 $\{\alpha(\mathcal{C}_i), i \in I\}$ 独立.

可以看出, 多个事件独立的定义不仅仅要求事件是两两独立的, 而且要求从中任意选择任意有限个都满足交的概率等于概率的乘积这个性质. 比如 A_1, \dots, A_n 独立就意味着

$$P(A_{i_1} \cdots A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}), \forall 2 \leq k \leq n, i_1 < i_2 < \cdots < i_k.$$

所以这里一共有 $2^n - n - 1$ 个等式, 且它们都是相互独立的——另一种意义上的独立, 即彼此不能互推.

在第 2.7 节的例 1 中, 三个事件两两独立, 但因为

$$P(ABC) = \frac{1}{4} \neq \frac{1}{8} = P(A) \times P(B) \times P(C),$$

因此 A, B, C 并不独立.

计算多个事件交的概率一般用乘法公式. 有了独立性, 乘法公式中的条件概率就可以直接换成无条件的概率, 因此计算事件交的概率变得非常简单.

在具体问题中,人们往往会根据对试验的直观理解,来判断事件是否独立的.后面有例子说明,这种直观是不太可靠的.用严格的概率论上的独立概念去推理,常常会出现与直观不一致的结论,因此需要特别注意.这不是因为理论同实际有矛盾,而是把理论概念和直观概念混为一谈了.

习题

1. 设 $\mathcal{A} := \{A_i, i = 1, \dots, n\}$ 与 $\mathcal{B} := \{B_j, j = 1, \dots, m\}$ 独立,且 \mathcal{A} 与 \mathcal{B} 中的元素都是两两不交的.证明 $\alpha(\mathcal{A})$ 与 $\alpha(\mathcal{B})$ 独立.
2. 设 ξ, η 是随机变量,且对任意 $a, b \in \mathbb{R}$,事件 $\{\xi < a\}$ 与 $\{\eta < b\}$ 独立.证明对任意 $-\infty < a_1 < a_2 < \dots < a_{2n} < \infty$ 及任意 $-\infty < b_1 < b_2 < \dots < b_{2m} < \infty$,事件 $\{\xi \in [a_1, a_2) \cup \dots \cup [a_{2n-1}, a_{2n})\}$ 与 $\{\eta \in [b_1, b_2) \cup \dots \cup [b_{2m-1}, b_{2m})\}$ 是独立的.
3. 构造三个事件 A_1, A_2, A_3 ,使得 $P(A_1 A_2 A_3) = P(A_1)P(A_2)P(A_3)$,但 A_1, A_2, A_3 不是独立的.

2.9 可列概型

本节我们将考虑可列概型.和有限概型相比,不同之处就是样本空间中样本点的个数是可列个的,我们主要关注它们的不同之处.

考虑 n 重Bernoulli概型 $B(n, p)$.以 ξ 表示成功的次数.我们知道

$$P(\xi = k) = C_n^k p^k q^{n-k}.$$

这个公式理论上是完美的,问题在于实际计算中,只要 n 稍微大一点,计算量就非常惊人,尤其在历史上没有电脑的时代,基本上是不可能直接计算的.即使在今天电脑普及了的时代,计算所需时间和费用也是一个需要考虑的问题.但计算问题又不能回避,因此人们就设法找一些近似计算方法.这一找不要紧,就找到了一个后来证明是非常重要的分布,即Poisson⁵分布.这颇有点像哥伦布要寻找一条通往东方的近路,结果阴差阳错,找到了一个后来被证明是非常重要的美洲一样.

定理 2.9.1. 若 $\lim_{n \rightarrow \infty} np_n = \lambda$, 则

$$\lim_{n \rightarrow \infty} C_n^k p_n^k q_n^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}, \forall k = 0, 1, 2, \dots$$

其中 $q_n = 1 - p_n$.

⁵Siméon Denis Poisson, 1781-1840, 法国数学家.

证明. 因为 $np_n \rightarrow \lambda$, 所以 $p_n \rightarrow 0$. 固定 k , 直接计算给出:

$$\begin{aligned} \lim_{n \rightarrow \infty} C_n^k p_n^k q_n^{n-k} &= \lim_{n \rightarrow \infty} \frac{n(n-1) \cdots (n-k+1)}{k!} \cdot p_n^k \cdot (1-p_n)^{n-k} \\ &= \lim_{n \rightarrow \infty} \frac{n^k}{k!} \cdot p_n^k \cdot \frac{(1-p_n)^n}{(1-p_n)^k} \\ &= \lim_{n \rightarrow \infty} \frac{(np_n)^k}{k!} \cdot (1-p_n)^n \\ &= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} e^{n \ln(1-p_n)} \\ &= \frac{\lambda^k}{k!} \lim_{n \rightarrow \infty} e^{-np_n \cdot \frac{\ln(1-p_n)}{-p_n}} \\ &= \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \end{aligned}$$

其中最后一个等式用了

$$\lim_{x \rightarrow 0} \frac{\ln(1-x)}{-x} = 1.$$

□

这说明当 p 小而 n 大时, 可用 $\frac{(np)^k}{k!} e^{-np}$ 作为 $C_n^k p_n^k q_n^{n-k}$ 的近似值. 这在实际工作中有重要意义, 比如若 $n = 800$, $p = 0.005$, $k = 3$, 则精确到小数点后4位时,

$$C_n^k p_n^k q_n^{n-k} = 0.1945,$$

而在同样的精度下也有

$$e^{-4} \frac{4^3}{3!} = 0.1945.$$

可见误差很小. 在本章最后一节我们将研究它们间的误差大小.

但更重要的是, 这个近似计算使人们发现了一个非常重要的分布, 即所谓Poisson分布. 事实上, 仔细观察一下, 我们发现

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = 1.$$

所以一个自然的问题是, 有没有一个概型, 即有没有一个样本空间 Ω 及其上的概率 P , 使得

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

这样的样本空间是有的, 且当然要包含可列个样本点. 比如说我们可取

$$\Omega := \{0, 1, \dots\},$$

并赋予概率

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

这就导出可列概型的概念.

定义 2.9.2. 设一个试验可能的结果有可列个:

$$\Omega = \{\omega_1, \omega_2, \dots\}.$$

设 ω_i 出现的概率为 p_i , $0 < p_i < 1$, $\sum_{i=1}^{\infty} p_i = 1$. 令

$$P(\omega_i) := p_i,$$

且对任意 $A \subset \Omega$, 令

$$P(A) := \sum_{\omega \in A} P(\omega).$$

则 (Ω, P) 称为可列概型, $P(A)$ 称为事件 A 的概率.

可列概型也有类似与有限概型的基本性质、有条件概率、全概率公式、独立性、随机变量、分布列、分布函数和期望之类的概念, 等等. 这些结果中主要的不同是原来的有限个样本点变成可列个样本点, 因而有限和会相应地变成级数. 但除此之外, 它们就没有什么差别了, 因此我们将有限概型和可列概型统称为离散概型.

定理 2.9.3. 概率 P 具有如下性质:

- (i) 非负性: $\forall A, P(A) \geq 0$;
- (ii) 规范性: $P(\Omega) = 1, P(\emptyset) = 0$;
- (iii) 单调性:

$$A \subset B \implies P(A) \leq P(B);$$

- (iv) 可列可加性: 若 A_1, A_2, \dots 是一列事件, 且 $A_i A_j = \emptyset, \forall i \neq j$, 则有加法公式

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

- (v) 次可列可加性: 若把上款中的两两不交条件去掉, 则

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

证明. 我们只证(v), 其它都是显然的.

令

$$\tau(\omega) := \inf\{i : \omega \in A_i\} \quad (\inf \emptyset = \infty),$$

$$B_i := \{\tau = i\} = A_1^c \cdots A_{i-1}^c A_i \quad (A_0 := \emptyset).$$

则 B_1, B_2, \dots 两两不交, $B_i \subset A_i, \forall i$, 且

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i.$$

因此

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} B_i\right) \\ &= \sum_{i=1}^{\infty} P(B_i) \quad (\text{由(iv)}) \\ &\leq \sum_{i=1}^{\infty} P(A_i) \quad (\text{由(iii)}) \end{aligned}$$

□

条件概率的定义和有限概型时是完全一样的:

定义 2.9.4. 设 A, B 是两事件, $P(B) > 0$. 定义:

$$P(A|B) := \frac{P(AB)}{P(B)},$$

称为给定 B 时(或知道 B 时, 或在 B 发生的条件下) A 的条件概率.

条件概率是将样本空间 Ω 缩小为 B 后的概率, 因此有与概率一样的性质, 兹罗列如下:

定理 2.9.5. 概率 $P(\cdot|B)$ 具有如下性质:

- (i) 非负性: $0 \leq P(A|B) \leq 1$;
- (ii) 规范性: $P(B|B) = 1, P(\emptyset|B) = 0$;
- (iii) 单调性:

$$A \subset C \implies P(A|B) \leq P(C|B);$$

- (iii) 可列可加性: 若 A_1, A_2, \dots 是一列事件, 且 $A_i A_j = \emptyset, \forall i \neq j$, 则有加法公式

$$P\left(\sum_{i=1}^{\infty} A_i|B\right) = \sum_{i=1}^{\infty} P(A_i|B).$$

- (iv) 次可列可加性: 若把上款中的两两不交条件去掉, 则

$$P\left(\bigcup_{i=1}^{\infty} A_i|B\right) \leq \sum_{i=1}^{\infty} P(A_i|B).$$

设 $\mathcal{P} = \{\Omega_1, \Omega_2, \dots\}$ 为 Ω 的一个分割, 全概率公式为

$$P(A) = \sum_{i=1}^{\infty} P(A\Omega_i) = \sum_{i=1}^{\infty} P(\Omega_i)P(A|\Omega_i).$$

Bayes公式为

$$P(\Omega_j|A) = \frac{P(A\Omega_j)}{P(A)} = \frac{P(\Omega_j)P(A|\Omega_j)}{\sum_{i=1}^{\infty} P(\Omega_i)P(A|\Omega_i)}.$$

在可列概型上也可以定义随机变量.

定义 2.9.6. Ω 上的实值函数称为随机变量.

同样地, 随机变量 ξ 的分布列定义为

$$(x_i, p_i, i = 1, 2, \dots),$$

其中

$$p_i = P(\xi = x_i) = \sum_{k: \xi(\omega_k) = x_i} P(\omega_k).$$

在必要时, 将把 p_i 写为 $p_{\xi, i}$ 以明确它是 ξ 的分布列.

现在可以写出Poisson分布的定义了.

定义 2.9.7. 设 $\lambda > 0$. 若 ξ 的分布为

$$P(\xi = n) = \frac{\lambda^n}{n!} e^{-\lambda}, \quad n = 0, 1, 2, \dots$$

则称 ξ 服从 *Poisson* 分布, 记为 $\xi \sim P(\lambda)$.

如果 $\xi \sim P(\lambda)$, 那么 $a\xi$ 服从什么分布? 这就引导出下面的定义. 这个定义本质上和 *Poisson* 分布是一样的, 但引进一个新名词有时候是方便的.

定义 2.9.8. 设 $a \in \mathbb{R}$, ξ 是取值于 $\{na, n = 0, 1, 2, \dots\}$ 的随机变量, 且

$$P(\xi = an) = \frac{\lambda^n}{n!} e^{-\lambda},$$

其中 $\lambda > 0$. 则称 ξ 服从跳为 a 参数为 λ 的 *Poisson* 分布, 记为 $\xi \sim P(\lambda, a)$.

用这个符号, 则 $a\xi \sim P(\lambda, a)$.

若 ξ 是随机变量, B 是事件, $P(B) > 0$. 定义给定 B 时 ξ 的条件分布列为

$$(x_i, p_i^B, i = 1, 2, \dots),$$

其中

$$p_i^B := P(\xi = x_i | B).$$

若 ξ, η 是两个随机变量, 定义给定 $\eta = y_j$ 时 ξ 的分布列为

$$(x_i, p_i^j, i = 1, 2, \dots),$$

其中

$$p_i^j := P(\xi = x_i | \eta = y_j).$$

定义 (ξ, η) 的联合分布列为

$$((x_i, y_j), p_{ij}, i, j = 1, 2, \dots),$$

其中

$$p_{ij} = P(\xi = x_i, \eta = y_j).$$

注意别把条件分布列和联合分布列搞混. 它们的关系是:

$$p_{ij} = p_i^j p_{\eta, j},$$

其中 $p_{\eta, j} := P(\eta = j)$.

描述概率的分布情况的另一个工具是分布函数, 其定义为

$$F(x) := \sum_{i: x_i \leq x} p_i = \sum_{\omega: \xi(\omega) \leq x} P(\omega).$$

显然, F 是定义在 \mathbb{R} 上的右连续函数, 且它和分布列是相互唯一确定的. 事实上, 由分布列唯一确定分布函数是不必说了, 由分布函数确定分布列留作习题.

容易看出, 若 ξ 是取有限个值的随机变量, 那么其分布函数是右连续的阶梯函数, 见图 2.2; 对取无限个值的随机变量, 乍一想分布函数似乎和有限概型的分布函数有差不多的形状, 无

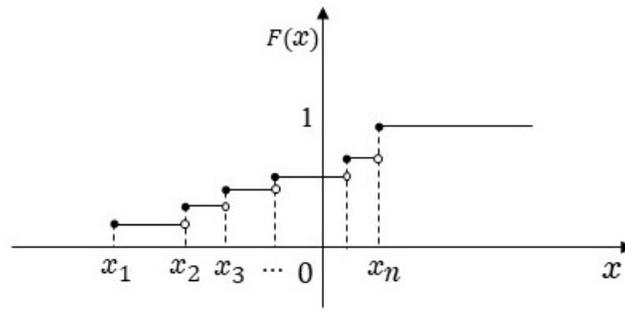


图 2.2: 分布函数的图像

非是从有限个阶梯变为可列个阶梯而已. 事实上并非如此, 因为“离散”这个词此时有点误导, 这些阶梯有可能并不是离散的, 它们可能存在聚点, 并且是很多很多的聚点.

例1. 将 $[0, 1]$ 中的二进制数按下列规则编号:

$$0, 1, \frac{1}{2}, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}, \dots, \frac{1}{2^n}, \frac{3}{2^n}, \dots, \frac{2^n - 1}{2^n}, \dots$$

设 ξ 为随机变量, 分布列为

$$P(\xi = x_i) = 2^{-i}, \quad i = 1, 2, \dots,$$

则 ξ 的分布函数为

$$F(x) = \sum_{x_i \leq x} 2^{-i}.$$

你能画出 F 的图像吗? 其图像是你想象的那样, 是一个阶梯函数?

像有限概型一样, 随机变量 ξ 的期望定义为其平均值. 但这时涉及到无穷项级数的求和, 因此有一个收敛性问题. 为此, 我们将 ξ 的正部和负部分开考虑, 即分别定义:

$$E[\xi^+] := \sum_{\omega \in \Omega} \xi^+(\omega) P(\omega),$$

$$E[\xi^-] := \sum_{\omega \in \Omega} \xi^-(\omega) P(\omega).$$

若至少其中之一有限, 则定义

$$E[\xi] := E[\xi^+] - E[\xi^-].$$

当两个都有限时, 称 ξ 可积. 显然 ξ 可积的充要条件是

$$\sum_{\omega \in \Omega} |\xi(\omega)| P(\omega) < \infty.$$

此时

$$E[\xi] = \sum_{\omega \in \Omega} \xi(\omega) P(\omega).$$

当 $E[\xi^+]$ 与 $E[\xi^-]$ 均为无限时, 差式 $E[\xi^+] - E[\xi^-]$ 没有意义, 此时称 ξ 没有期望.

当可积时, 期望有下列基本性质:

命题 2.9.9. (i) 和概率的关联性: 若 $A \subset \Omega$, 则

$$E[1_A] = P(A);$$

(ii) 单调性:

$$\xi \leq \eta \implies E[\xi] \leq E[\eta];$$

(iii) 线性性: 设 a, b 为常数, 则

$$E[a\xi + b\eta] = aE[\xi] + bE[\eta];$$

(iv) 若 ξ 的分布列为 $\{(x_i, p_i) | i = 1, 2, \dots\}$, 则 ξ 可积的充要条件是 $\sum_{i=1}^{\infty} |x_i| p_i < \infty$, 且此时有

$$E[\xi] = \sum_i x_i p_i.$$

证明. (i) 我们有

$$E[1_A] = \sum_{\omega \in A} P(\omega) = P(A).$$

(ii)

$$E[\xi] = \sum_{\omega} \xi(\omega) P(\omega) \leq \sum_{\omega} \eta(\omega) P(\omega) = E[\eta].$$

(iii)

$$\begin{aligned} E[a\xi + b\eta] &= \sum_{\omega} (a\xi(\omega) + b\eta(\omega)) P(\omega) \\ &= a \sum_{\omega} \xi(\omega) P(\omega) + b \sum_{\omega} \eta(\omega) P(\omega) \\ &= aE[\xi] + bE[\eta]. \end{aligned}$$

(iv) 由于正项级数可任意交换求和顺序, 故有

$$\sum_{\omega} |\xi(\omega)| P(\omega) = \sum_i |x_i| p_i.$$

而当它们有限时, 有

$$\begin{aligned} E[\xi] &= \sum_{\omega} \xi(\omega) P(\omega) \\ &= \sum_i \sum_{\omega: \xi(\omega)=x_i} \xi(\omega) P(\omega) \\ &= \sum_i x_i \sum_{\omega: \xi(\omega)=x_i} P(\omega) \\ &= \sum_i x_i p_i. \end{aligned}$$

注意在上述证明中, 我们隐性地用到了 $\sum_{\omega} |\xi(\omega)| P(\omega) < \infty$, 因为只有在这个条件下, 重排级数的各项才不会影响最后的和. \square

期望的两个计算公式和有限概型时的公式是相似的, 只不过是把有限和变成了级数. 但注意这里的前提条件是这个级数绝对收敛. 这是必须的, 因为如果说可列与否是客观的话, 那

么按什么样的顺序排列则是人为的, 而如果没有绝对收敛性, 那么那个定义为期望的级数是否收敛以及收敛到什么值都会依赖于顺序的选择, 也即人为的因素, 因而是不能接受的. 反之, 一旦有了绝对收敛性, 级数就可以像有限和一样运算——像重排次序啊, 合并同类项啊, 统统都没有问题.

要描述一个随机变量的总体状况, 除了期望, 就是波动了. 就像一个班的学习情况, 平均成绩, 也就是期望, 是一个指标, 而个体之间是否差异较大, 则是另一个观察点. 衡量波动大小的指标, 按说首先想到的是数量

$$E[|\xi(\omega) - E[\xi]|] = \sum_{\omega \in \Omega} |\xi(\omega) - E[\xi]| P(\omega),$$

但这个量在数学上处理起来不太方便, 比如它就不如下面的量方便:

$$D[\xi] := E[|\xi(\omega) - E[\xi]|^2] = \sum_{\omega \in \Omega} |\xi(\omega) - E[\xi]|^2 P(\omega).$$

这就像 \mathbb{R}^n 中的长度我们是用 $(x_1^2 + \cdots + x_n^2)^{\frac{1}{2}}$ 而不是用 $|x_1| + \cdots + |x_n|$ 一样.

定义 2.9.10. $D[\xi]$ 称为 ξ 的方差.

注意这里涉及的是正项级数, 所以它要么收敛, 要么发散到无穷大. 后面这种情况我们也依然认为它是方差, 等于无穷大的方差.

设 B 为事件, $P(B) > 0$. 给定 B 的条件下, 很自然地将给定 B 时 ξ 的条件期望定义为

$$E[\xi|B] := E[\xi^+|B] - E[\xi^-|B],$$

只要等式右边的两项中至少一项为有限数, 其中

$$E[\xi^+|B] := \sum_{\omega} \xi^+(\omega) P(\omega|B),$$

$$E[\xi^-|B] := \sum_{\omega} \xi^-(\omega) P(\omega|B).$$

当这两项均有限时, 易见

$$E[\xi|B] := \sum_{\omega} \xi(\omega) P(\omega|B),$$

且右边的级数绝对收敛. 显然有

$$E[\xi|B] = \sum_{i=1}^{\infty} x_i P(\xi = x_i|B) = \sum_{i=1}^{\infty} x_i p_i^B.$$

至于 $E[\xi|B]$ 何时有限, 我们有下面的简单结果:

命题 2.9.11. 若 $E[\xi]$ 有限, 则对任意 $B \subset \Omega$, $P(B) > 0$, $E[\xi|B]$ 有限, 且

$$E[\xi|B] = P(B)^{-1} E[\xi 1_B].$$

证明. 因为 $P(\omega|B) \leq P(B)^{-1} P(\omega)$, 所以 $E[|\xi|] < \infty$ 时有

$$\sum_{\omega \in \Omega} |\xi(\omega)| P(\omega|B) \leq P(B)^{-1} \sum_{\omega \in \Omega} |\xi(\omega)| P(\omega) = P(B)^{-1} E[|\xi|] < \infty.$$

所以 $E[\xi|B]$ 有限, 且

$$\begin{aligned} E[\xi|B] &= \sum_{\omega \in \Omega} \xi(\omega)P(\omega|B) \\ &= \sum_{\omega \in B} P(B)^{-1}\xi(\omega)P(\omega) \\ &= P(B)^{-1}E[\xi 1_B]. \end{aligned}$$

□

现在设 $\mathcal{P} = \{\Omega_1, \Omega_2, \dots\}$ 构成 Ω 的一个分割, 即

$$\sum_{n=1}^{\infty} \Omega_n = \Omega.$$

如同有限概型的情况一样, 引进下面的概念是方便的: 定义

$$P(A|\mathcal{P})(\omega) := \sum_{n=1}^{\infty} P(A|\Omega_n)1_{\Omega_n}(\omega).$$

$P(A|\mathcal{P})$ 称为给定分割 \mathcal{P} 时的条件概率. 它的含义是, 当 $\omega \in \Omega_n$ 时, $P(A|\mathcal{P})(\omega) = P(A|\Omega_n)$, 因此它是一种简单实用的记号, 用单独一个记号表示了所有的条件概率 $P(A|\Omega_n)$, $n = 1, 2, \dots$.

当 $A = \{\omega'\}$ 时, 我们将 $P(A|\mathcal{P})(\omega)$ 记为 $P(\omega', \omega)$. 这样, 当 $\omega \in \Omega_n$ 时, 就有

$$P(\omega', \omega) = P(\omega'|\Omega_n).$$

显然有

$$P(A|\mathcal{P})(\omega) = \sum_{\omega' \in A} P(\omega', \omega).$$

就像有限概型的情况一样, 对固定的 A , $\omega \mapsto P(A|\mathcal{P})(\omega)$ 是随机变量, 而对于固定的 ω , $A \mapsto P(A|\mathcal{P})(\omega)$ 是概率, 它满足非负性、规范性与可列可加性.

对任何一个随机变量 ξ , 对固定的 ω , 定义

$$E[\xi|\mathcal{P}](\omega) := E^{P(\cdot, \omega)}[\xi(\omega')] := \sum_{\omega'} \xi(\omega')P(\omega', \omega) = \sum_{n=1}^{\infty} E[\xi|\Omega_n]1_{\Omega_n}(\omega),$$

如果右边的每一个 $E[\xi|\Omega_n]$ 都存在的话. 它称为关于 \mathcal{P} 的条件期望, 仍是一个随机变量. 特别取 $\xi = 1_A$ 就得到

$$E[1_A|\mathcal{P}](\omega) = P(A|\mathcal{P})(\omega).$$

曾记否, 我们曾经对无条件的期望注意过、使用过这个公式? 而无条件的期望是条件期望的一种特殊情况, 即 $\mathcal{P} = \{\Omega\}$ 时的条件期望, 不是吗?

关于分割的条件期望依然有一个是否有限的问题. 我们有:

命题 2.9.12. 若 $E[|\xi|] < \infty$, 则对任意分割 \mathcal{P} , $E[\xi|\mathcal{P}]$ 对每一个 ω 都有限.

事实上, 当命题的条件满足时, 对任意 n , $E[\xi|\Omega_n]$ 有限, 因此 $E[\xi|\mathcal{P}]$ 有限. 我们还有:

定理 2.9.13. 当 $E[\xi]$ 有限时有

$$E[E[\xi|\mathcal{P}]] = E[\xi].$$

特别地, 当 $\xi := 1_A$ 时, 我们得到全概率公式另一种简洁表达式:

$$P(A) = E[E[1_A|\mathcal{P}]] = E[P(A|\mathcal{P})],$$

证明. 直接计算可知:

$$\begin{aligned} E[E[\xi|\mathcal{P}]] &= \sum_{n=1}^{\infty} E[\xi|\Omega_n]P(\omega_n) \\ &= \sum_{n=1}^{\infty} E[\xi 1_{\Omega_n}] \\ &= E[\xi]. \end{aligned}$$

□

这个公式有什么用? 我们看一个例子.

假设一次人口普查中, $\xi(\omega)$ 表示 ω 的身高. 现在需要计算全国成年男子的平均身高. 那么

$$\Omega = \{\omega_1, \dots, \omega_n\}.$$

则平均身高为

$$E[\xi] = \frac{1}{n} \sum_{i=1}^n \xi(\omega_i).$$

不过想想看, 如果 n 有十几亿之多, 这个计算该有多费力!

现设一共 m 个省, 每个省有 n_m 人, $n_1 + \dots + n_m = n$. 那么上面的公式告诉你, 国家可以要求每个省上报平均身高, 比如说是 η_1, \dots, η_m , 然后对这 m 个数据按每个省的人数比例加权平均, 即得到所要的数据, 即

$$E[\xi] = \sum_{k=1}^m \eta_k \frac{n_k}{n}.$$

这里右边的 m 个数据 $\eta_k, k = 1, \dots, m$ 可以同步得到, 这样工作效率就可以大大提高. 并且, 为得到诸 η_k , 每个省又可以将任务分解到各个县, 然后是各个乡……, 这样最后的工作效率不知要提高多少!

整体平均等于局部平均后再平均, 这就是上面这个公式背后的简单思想.

现在假设 η 是另一随机变量, 值域为 $\{y_j, j = 1, 2, \dots\}$. 令

$$\Omega_j := \{\eta = y_j\}, \quad p_i^{\eta=y_j} := P(\xi = i | \eta = y_j), \quad \mathcal{P} := \{\Omega_n, n = 1, 2, \dots\}.$$

利用上述结果, 当 $E[|\xi|] < \infty$ 时, 可定义条件期望 $E[\xi|\eta = y_j]$ 且有

$$E[\xi|\eta = y_j] = \sum_{i=1}^{\infty} x_i p_i^{\eta=y_j}.$$

而不固定 j , ξ 关于 η 的条件期望则定义为随机变量.

$$E[\xi|\eta] := E[\xi|\mathcal{P}]$$

显然

$$E[\xi|\eta] = \sum_{j=1}^{\infty} E[\xi|\eta = y_j] \mathbf{1}_{\eta=y_j},$$

$$E[\xi] = E[E[\xi|\eta]].$$

最后我们要谈的是独立性. 这个概念是有限概型时同一概念的自然延申.

定义 2.9.14. (i) 设 A, B 是事件. 若

$$P(AB) = P(A)P(B),$$

则称 A 与 B 独立;

设 $\{A_i, i \in I\}$ 是事件族, 其中 I 是任意指标集. 若对任意有限 $J \subset I$, 都有

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in I} P(A_i),$$

则称 $\{A_i, i \in I\}$ 独立.

(ii) 设 ξ, η 是随机变量. 如果

$$P(\xi = x_i, \eta = y_j) = P(\xi = x_i)P(\eta = y_j), \quad \forall i, j,$$

则称 ξ 与 η 独立;

设 $\{\xi_i, i \in I\}$ 是随机变量族, 其中 I 是任意指标集. 若对任意有限 $J \subset I$ 都有

$$P\left(\bigcap_{i \in J} \{\xi_i = x_i\}\right) = \prod_{i \in J} P(\xi_i = x_i)$$

其中 x_i 跑遍 ξ_i 的值域, 则称为独立的.

我们来看几个例子.

例1. 同时掷红白两枚骰子, 样本空间为

$$\Omega = \{(i, j) : i, j = 1, \dots, 6\}.$$

令

$$\xi((i, j)) = i, \quad \eta((i, j)) = j.$$

则对任意 $1 \leq i, j \leq 6$,

$$P(\xi = i) = P(\eta = j) = \frac{1}{6},$$

$$P(\xi = i, \eta = j) = \frac{1}{36}.$$

故

$$P(\xi = i, \eta = j) = P(\xi = i)P(\eta = j), \quad \forall i, j.$$

因此 ξ 与 η 独立.

例2. 设 (Ω, P) 与 (Ω', P') 为离散概率空间. 令

$$\Omega \times \Omega' := \{(\omega, \omega') : \omega \in \Omega, \omega' \in \Omega'\},$$

$$P \times P'((\omega, \omega')) := P(\omega)P(\omega').$$

易证 $(\Omega \times \Omega', P \times P')$ 也为离散概率空间. 设 ξ, η 分别是定义在 Ω, Ω' 上的随机变量. 在 $\Omega \times \Omega'$ 上定义

$$\tilde{\xi}((\omega, \omega')) = \xi(\omega), \quad \tilde{\eta}((\omega, \omega')) := \eta(\omega').$$

则对任意 x, y ,

$$\begin{aligned} P \times P'(\tilde{\xi} = x, \tilde{\eta} = y) &= \sum_{(\omega, \omega'):\xi(\omega)=x, \eta(\omega')=y} P \times P'((\omega, \omega')) \\ &= \sum_{(\omega, \omega'):\xi(\omega)=x, \eta(\omega')=y} P(\omega)P(\omega') \\ &= \sum_{\omega:\xi(\omega)=x} \sum_{\omega':\eta(\omega')=y} P(\omega)P(\omega') \\ &= \sum_{\omega:\xi(\omega)=x} P(\omega) \cdot \sum_{\omega':\eta(\omega')=y} P(\omega') \\ &= P(\xi = x)P(\eta = y). \end{aligned}$$

类似可得

$$P(\xi = x) = P \times P'(\tilde{\xi} = x), \quad P(\eta = y) = P \times P'(\tilde{\eta} = y).$$

因此

$$P \times P'(\tilde{\xi} = x, \tilde{\eta} = y) = P(\tilde{\xi} = x)P(\tilde{\eta} = y).$$

故 $\tilde{\xi}$ 与 $\tilde{\eta}$ 独立.

例3. 设 (Ω_i, P_i) 是离散概率空间, $i = 1, 2, \dots, n$. 令

$$\Omega := \Omega_1 \times \dots \times \Omega_n := \{(\omega_1, \dots, \omega_n) : \omega_i \in \Omega_i, i = 1, 2, \dots, n\},$$

$$P((\omega_1, \dots, \omega_n)) := P_1(\omega_1) \cdots P_n(\omega_n).$$

设 ξ_i 是 Ω_i 上的函数, 定义 Ω 上的函数 $\tilde{\xi}_i$, 使得

$$\tilde{\xi}_i((\omega_1, \dots, \omega_n)) = \xi_i(\omega_i), \quad \forall (\omega_1, \dots, \omega_n).$$

则与上例类似, 可证 $\tilde{\xi}_1, \dots, \tilde{\xi}_n$ 独立.

独立性是概率论中最重要的概念之一. 我们有下列重要结果:

定理 2.9.15. 设 ξ, η 独立, 且 $E[\xi], E[\eta]$ 存在, 则 $E[\xi\eta]$ 也存在且

$$E[\xi\eta] = E[\xi]E[\eta].$$

证明. 设 ξ 的分布列为 $\{(x_i, p_i)\}$, η 的分布列为 $\{(y_j, q_j)\}$. 若 $\sum_{\omega} |\xi(\omega)\eta(\omega)|P(\omega) < \infty$, 则

$$\begin{aligned} E[\xi\eta] &= \sum_{\omega} \xi(\omega)\eta(\omega)P(\omega) \\ &= \sum_{i,j} \sum_{\omega: \xi(\omega)=x_i, \eta(\omega)=y_j} \xi(\omega)\eta(\omega)P(\omega) \\ &= \sum_{i,j} x_i y_j \sum_{\omega: \xi(\omega)=x_i, \eta(\omega)=y_j} P(\omega) \\ &= \sum_{i,j} x_i y_j P(\xi(\omega)=x_i, \eta(\omega)=y_j) \\ &= \sum_{i,j} x_i y_j P(\xi(\omega)=x_i)P(\eta(\omega)=y_j) \\ &= \sum_i x_i p_i \cdot \sum_j y_j q_j \\ &= E[\xi]E[\eta]. \end{aligned}$$

至于 $\sum_{\omega} |\xi(\omega)\eta(\omega)|P(\omega)$ 的收敛性, 在上述推理中将 ξ 与 η 分别换为 $|\xi|$ 与 $|\eta|$ 就可以得到. \square

习题

1. 证明: $\xi \sim P(\lambda)$ 的充要条件是: 对 \mathbb{N}_+ 上的任意有界函数 f , 有

$$E[\xi f(\xi)] = \lambda E[f(\xi + 1)].$$

2. 设 $\xi \sim P(\lambda)$. 计算 $E[\xi]$ 与 $E[(\xi - E[\xi])^2]$.
3. 设 ξ 为随机变量, 分布列为 $\{(x_i, p_i), i = 1, 2, \dots\}$; $f: \mathbb{R} \mapsto \mathbb{R}$. 证明:

$$\sum_{\omega \in \Omega} |f(\xi(\omega))|P(\omega) < \infty \iff \sum_{i=1}^{\infty} |f(x_i)|p_i < \infty,$$

且在此情况下有

$$E[f(\xi)] = \sum_{i=1}^{\infty} f(x_i)p_i.$$

对二维随机变量 (ξ, η) 和函数 $f: \mathbb{R}^2 \mapsto \mathbb{R}$ 写出并证明类似的结果.

4. 设 f 为 \mathbb{R}^2 上的实值函数. 证明在条件 $\xi = x_i$ 下, $f(\xi, \eta)$ 与 $f(x_i, \eta)$ 具有相同的条件分布.
5. 设 ξ 的值域为 $\{x_i, i = 1, 2, \dots\}$, 分布列为 $\{(x_i, p_i)\}$. 证明

$$E[\eta] = \sum_{i=1}^{\infty} E[\eta | \xi = x_i]p_i.$$

6. 设 \mathcal{P}_1 与 \mathcal{P}_2 是两个分割, 且 \mathcal{P}_2 是 \mathcal{P}_1 的加细, 即对任意 $A \in \mathcal{P}_2$, 都有 $B \in \mathcal{P}_1$, 使得 $A \subset B$. 证明:

$$E[\xi | \mathcal{P}_1] = E[E[\xi | \mathcal{P}_2] | \mathcal{P}_1].$$

7. 设 ξ, η 独立, 分别有值域 R_1 与 R_2 . 证明:

$$P(\xi \in A, \eta \in B) = P(\xi \in A)P(\eta \in B), \quad \forall A \subset R_1, B \subset R_2.$$

对多个随机变量叙述并证明平行的结果.

8. 设 ξ, η, ζ 独立. 证明: 对任意函数 $f: \mathbb{R}^2 \mapsto \mathbb{R}^1$, $f(\xi, \eta)$ 与 ζ 独立.

叙述并证明这一结果到多个随机变量的推广.

9. 设 ξ, η 是随机变量, $f: \mathbb{R}^2 \mapsto \mathbb{R}$, 且 $E[f(\xi, \eta)]$ 存在. 证明

(a)

$$\begin{aligned} E[f(\xi, \eta)|\eta = y_j] &= E[f(\xi, y_j)|\eta = y_j], \\ E[f(\xi, \eta)] &= \sum_j E[f(\xi, y_j)|\eta = y_j]P(\eta = y_j). \end{aligned}$$

(b)

$$E[\xi] = E[E[\xi|\eta]].$$

(c) 设 $f: \mathbb{R} \mapsto \mathbb{R}$, 则

$$\begin{aligned} E[\xi f(\eta)|\eta = y_j] &= f(y_j)E[\xi|\eta = y_j], \\ E[\xi f(\eta)|\eta] &= f(\eta)E[\xi|\eta], \\ E[\xi f(\eta)] &= E[f(\eta)E[\xi|\eta]]. \end{aligned}$$

(d) 若再假设 ξ 与 η 独立, 则

$$E[f(\xi, \eta)] = E[E[f(\xi, y)|y=\eta]];$$

特别地,

$$E[\xi\eta] = E[\xi]E[\eta].$$

10. 设 ξ_1, \dots, ξ_n 独立, 且 $E[\xi_i^2]$ 存在, $E[\xi_i] = 0, \forall i = 1, 2, \dots, n$. 证明

$$E \left[\left| \sum_{i=1}^n \xi_i \right|^2 \right] = \sum_{i=1}^n E[\xi_i^2].$$

11. 证明条件期望的线性性, 即

$$E[a\xi_1 + b\xi_2|\eta] = aE[\xi_1|\eta] + bE[\xi_2|\eta],$$

其中 ξ_1, ξ_2, η 是随机变量, $a, b \in \mathbb{R}$, 且涉及到的期望均有限.

12. 设 $\{(x_i, p_i), i = 1, \dots, m\}$ 和 $\{(y_i, q_i), i = 1, \dots, n\}$ 为分布列, 且有同样的分布函数. 证明: $m = n$ 且存在 $\{1, \dots, n\}$ 到自身的双射 φ 使得 $(x_i, p_i) = (y_{\varphi(i)}, q_{\varphi(i)}), \forall i = 1, \dots, n$.

13. 设 ξ 为取非负整数值的随机变量, 定义其母函数为

$$g(s) := \sum_{n=0}^{\infty} P(\xi = n)s^n.$$

证明:

- (a) $g(s)$ 至少对 $|s| \leq 1$ 是有定义的;
 (b) 若 $g(s)$ 在 $[1, 1 + \delta)$ 上有定义,则 ξ 的任意阶矩存在,且

$$E\xi = g'(1), \quad E\xi^2 - E\xi = g''(1).$$

你还可以写出更多的矩和高阶导数之间的关系.

- (c) 求Poisson分布的母函数.

14. 设 ξ 是随机变量, F 是其分布函数. 证明:

$$P(\xi = x) = F(x) - F(x-), \quad \forall x \in \mathbb{R}.$$

15. 设 ξ, η 独立, $f, g: \mathbb{R} \mapsto \mathbb{R}$ 为任意函数. 证明 $f(\xi), g(\eta)$ 独立.

2.10 二项分布的Poisson分布近似: Stein-Chen方法

我们曾经证明了二项分布可以用Poisson分布近似, 即

$$C_n^k p^k (1-p)^{n-k} \sim \frac{\lambda^k}{k!} e^{-\lambda}.$$

不过我们没有对误差进行估计, 而这就是本节要研究的问题. 我们所用的方法是Stein⁶先对正态分布做的, 然后Chen⁷将其扩展到Poisson分布.

我们从一个一般的概念开始. 设 ξ, η 都是取非负整数值的随机变量. 定义

$$d(\xi, \eta) = \sup_{A \subset \mathbb{N}_+} |P(\xi \in A) - P(\eta \in A)|.$$

注意这个量给出的是 ξ 和 η 的值的分布之间的距离, 而不是其值之间的距离. 完全有可能 ξ 与 η 是定义在不同的空间上的, 因此谈不上其值之间的距离问题; 而即使它们定义在同一个空间上, 也有可能他们取值相等的点完全不同, 但其分布一样, 即 $d(\xi, \eta) = 0$. 例如, 下面就是这种情况:

$$\Omega := \{0, 1\}, \quad P(0) = P(1) = \frac{1}{2}, \quad \xi(0) = \eta(1) = 1, \quad \xi(1) = \eta(0) = -1.$$

我们的目的是估计二项分布和Poisson分布之间的距离.

我们从一个更一般性的问题入手. 由上节的习题1我们知道, $\xi \sim P(\lambda)$ 的充要条件是: 对 \mathbb{N}_+ 上的有界函数 f , 有

$$E[\xi f(\xi)] = \lambda E[f(\xi + 1)].$$

所以对任意取值于 \mathbb{N}_+ 的随机变量 η 及适当选择的 \mathbb{N}_+ 上的函数 f , 量

$$|E[\eta f(\eta)] - \lambda E[f(\eta + 1)]|$$

的大小也许就代表了 η 的分布和Poisson分布间的距离?

为将此量与上面的距离 $d(\xi, \eta)$ 联系起来, 找到合适的 f 并研究其性质就成了问题的关键. 为此, 对任意 $A \subset \mathbb{N}_+$, $f_A(0) = 0, \forall k \geq 0$, 归纳定义

$$f_A(k+1) := \lambda^{-1} \left(1_A(k) - P(\xi \in A) + k f_A(k) \right).$$

⁶Charles M. Stein, 1920-2016, 美国数学家.

⁷Louis Chen, 陈晓云, 1940年生, 新加坡数学家.

则 f_A 为 \mathbb{N}_+ 上的函数且满足

$$\lambda f_A(k+1) - k f_A(k) = 1_A(k) - P(\xi \in A), \quad \forall k \in \mathbb{N}_+.$$

从而

$$\lambda f_A(\eta+1) - \eta f_A(\eta) = 1_A(\eta) - P(\xi \in A).$$

注意 $f_A(\eta+1) - \eta f_A(\eta)$ 是有界函数. 取期望然后取绝对值即得

$$|P(\eta \in A) - P(\xi \in A)| = |E[\lambda f_A(\eta+1) - \eta f_A(\eta)]|.$$

所以的确可以通过计算右边来估计 $d(\xi, \eta)$.

下面我们具体实施这一方案. 为叙述顺畅起见, 我们先准备下面的引理, 其证明是直接的.

引理 2.10.1. 设 $\mathbf{b} = \{b_k\}$ 是数列, $\xi \sim P(\lambda)$. 定义新数列

$$\begin{aligned} f_{\mathbf{b}}(0) &= 0, \\ f_{\mathbf{b}}(k+1) &= \lambda^{-1}(b_k + k f_{\mathbf{b}}(k)). \end{aligned}$$

则 (i)

$$f_{\mathbf{b}}(k+1) = \frac{1}{\lambda P(\xi = k)} \sum_{i=0}^k P(\xi = i) b_i.$$

(ii) 设 \mathbf{b} 与 \mathbf{c} 为两数列, 定义 $\mathbf{b} + \mathbf{c} := \{b_k + c_k\}$. 则

$$f_{\mathbf{b}+\mathbf{c}} = f_{\mathbf{b}} + f_{\mathbf{c}}.$$

下面我们证明:

引理 2.10.2. $\forall A \subset \mathbb{N}_+$, 有

$$|f_A(k+1) - f_A(k)| \leq \min\left(1, \frac{1}{\lambda}\right), \quad \forall k \in \mathbb{N}_+.$$

证明. 记 $f_n(k) := f_{\{n\}}(k)$. 由上一引理, 则

$$\begin{aligned} f_n(k+1) &= \frac{1}{\lambda P(\xi = k)} \sum_{i=0}^k P(\xi = i)(1_{i=n} - P(\xi = n)) \\ &= \frac{P(\xi = n)}{\lambda P(\xi = k)} [1_{n \leq k} - P(\xi \leq k)]. \end{aligned}$$

当 $k \geq n$ 时,

$$f_n(k+1) = P(\xi = n) \frac{P(\xi > k)}{\lambda P(\xi = k)}.$$

简单计算可知

$$\frac{P(\xi > k)}{\lambda P(\xi = k)} = \sum_{i=1}^{\infty} \frac{k!}{(i+k)!} \lambda^i.$$

易见它是 \mathbb{N}_+ 上的单调下降函数. 所以 $k \geq n$ 时, $f_n(k)$ 为 k 的单调下降函数.

再来看 $k < n$ 的情况. 此时

$$f_n(k+1) = -P(\xi = n) \frac{P(\xi \leq k)}{\lambda P(\xi = k)}.$$

简单计算可知

$$\frac{P(\xi \leq k)}{P(\xi = k)} = \sum_{m=0}^k \frac{\lambda^{-m} k!}{(k-m)!}$$

是单调上升的, 因此 $f_n(k)$ 也是 k 的单调下降函数.

因此, 对每个 n , 只有当 $k = n$ 时, $f_n(k+1) - f_n(k)$ 才有可能非负. 此时有

$$\begin{aligned} & f_n(n+1) - f_n(n) \\ &= \frac{P(\xi = n)}{\lambda P(\xi = n)} (1 - P(\xi \leq n)) - \frac{P(\xi = n)}{\lambda P(\xi = n-1)} (-P(\xi \leq n-1)) \\ &= \frac{P(\xi > n)}{\lambda} + \frac{P(\xi \leq n-1)P(\xi = n)}{\lambda P(\xi = n-1)} \\ &= \frac{P(\xi > n)}{\lambda} + \frac{P(\xi \leq n-1)}{n} \\ &= \frac{P(\xi > n)}{\lambda} + \sum_{i=0}^{n-1} \frac{P(\xi = i)}{n} \\ &= \frac{P(\xi > n)}{\lambda} + \sum_{i=0}^{n-1} \frac{P(\xi = i+1)}{\lambda n} (i+1) \\ &\leq \frac{P(\xi > n)}{\lambda} + \frac{P(0 < \xi \leq n)}{\lambda} \\ &= \frac{P(\xi \neq 0)}{\lambda} \\ &= \frac{1 - P(\xi = 0)}{\lambda} \\ &= \frac{1 - e^{-\lambda}}{\lambda} \\ &\leq \min\left(1, \frac{1}{\lambda}\right). \end{aligned}$$

所以对任意 $k \in \mathbb{N}_+$ 有

$$f_A(k+1) - f_A(k) = \sum_{n \in A} (f_n(k+1) - f_n(k)) \leq f_k(k+1) - f_k(k) \leq \min\left(1, \frac{1}{\lambda}\right).$$

又显然有 $f_A + f_{A^c} = f_{\mathbb{N}_+} \equiv 0$, 所以

$$f_A(k) - f_A(k+1) = f_{A^c}(k+1) - f_{A^c}(k) \leq \min\left(1, \frac{1}{\lambda}\right).$$

所以

$$|f_A(k+1) - f_A(k)| \leq \min\left(1, \frac{1}{\lambda}\right).$$

□

推论 2.10.3. $\forall A \subset \mathbb{N}_+$, 有

$$|f_A(m) - f_A(n)| \leq \min\left(1, \frac{1}{\lambda}\right) |m - n|, \quad \forall m, n \in \mathbb{N}_+.$$

现在我们可以叙述本节的主要结果了.

定理 2.10.4. 设 $\eta = \sum_{i=1}^n \eta_i$, 其中 $\eta_i \sim B(p_i)$ 且相互独立. 令 $\lambda = \sum_{i=1}^n p_i$. 再设 $\xi \sim P(\lambda)$. 则

$$d(\xi, \eta) \leq \min\left(1, \frac{1}{\lambda}\right) \sum_{i=1}^n p_i^2. \quad (10.6)$$

证明. 首先注意 η 只取有限个值, 因此所涉及的期望存在且有限.

令 $\zeta_i := \sum_{j \neq i} \eta_j$. 注意对任意有界函数 g 有

$$\begin{aligned} E[\eta_i g(\eta)] &= E[E[\eta_i g(\eta) | \eta_i]] \text{ (上节习题9(ii))} \\ &= E[0 | \eta_i = 0] q_i + E[g(\zeta_i + 1) | \eta_i = 1] p_i \\ &= E[g(\zeta_i + 1)] p_i, \end{aligned}$$

$q_i = 1 - p_i$, 这里最后一步用到了 ζ_i 与 η_i 的独立性. 于是, 对任意 $A \subset \mathbb{N}_+$, 我们有

$$\begin{aligned} |\lambda E[f_A(\eta + 1)] - E[\eta f_A(\eta)]| &= \left| \sum_{i=1}^n p_i E[f_A(\eta + 1)] - \sum_{i=1}^n E[\eta_i f_A(\eta)] \right| \\ &= \left| \sum_{i=1}^n p_i E[f_A(\eta + 1)] - \sum_{i=1}^n E[f_A(\zeta_i + 1)] p_i \right| \\ &\leq \sum_{i=1}^n p_i |E[f_A(\eta + 1)] - E[f_A(\zeta_i + 1)]| \\ &\leq \min(1, 1/\lambda) \sum_{i=1}^n p_i E[|\eta - \zeta_i|] \\ &= \min(1, 1/\lambda) \sum_{i=1}^n p_i E[|\eta_i|] \\ &= \min(1, 1/\lambda) \sum_{i=1}^n p_i^2. \end{aligned}$$

所以

$$\begin{aligned} d(\xi, \eta) &\leq \sup_A |\lambda E[f_A(\eta + 1)] - E[\eta f_A(\eta)]| \\ &\leq \min(1, 1/\lambda) \sum_{i=1}^n p_i^2. \end{aligned}$$

□

特别地, 取 $\xi_i \sim B(p)$, 则得到

推论 2.10.5. 设 $p \in (0, 1)$. 令 $\lambda := np$, 则

$$\sup_{0 \leq k \leq n} \left| C_n^k p^k (1-p)^{n-k} - \frac{\lambda^k}{k!} e^{-\lambda} \right| \leq \frac{\min(\lambda, \lambda^2)}{n}.$$

易见, 这个不等式右边是 $\min(p, np^2)$. 所以一般来说, 它只是对比较小的 p 有用.

这个结果是纯分析的, 但证明却是概率的. 用概率方法证明分析的结果, 是一个很有意思的方法, 因为它使用了分析学里没有的概念, 因而使人有别开生视角之感.

习题

1. 证明定理2.10.4的下述推广: 设 $\eta_i \sim B(p_i)$, $\xi \sim P(\lambda)$, 其中 $\lambda = \sum_{i=1}^n p_i$. 令 $\eta := \sum_{i=1}^n \eta_i$. 设 μ_i 和 $(\eta - 1 | \eta_i = 1)$ 同分布(即 μ_i 的分布与 $\eta - 1$ 在 $\eta_i = 1$ 下的条件分布相同), 则

$$d(\xi, \eta) \leq \min\left(1, \frac{1}{\lambda}\right) \sum_{i=1}^n p_i E[|\eta - \mu_i|].$$

2. 在上题的记号下, 若进一步假设 $\eta \geq \mu_i, \forall i$, 则有

$$d(\xi, \eta) \leq 1 - \frac{D[\eta]}{E[\eta]}.$$

3. 设Covid-19的某种变种的死亡率为千分之二. 求10000个病人中死亡人数不超过8个的概率.
4. 设在某个爪哇国的某个城市, 每条街道发生暴恐袭击的可能性都是百分之一, 各条街道是否发生相互独立. 问下面两种防暴警察配备法中, 哪种更有利于及时处置暴恐袭击:
- 每20条街道配备1队防暴警;
 - 每90条街道配备3队防暴警.

3 公理概型

3.1 动因

离散概型是不是已经够用了呢? 答案依然是否定的. 我们看两个简单的例子.

例1 (可列重Bernoulli试验). 反复投一枚硬币, 各次投掷相互独立, 以 τ 记第一次得到正面时投掷的次数. 求 $P(\tau = n)$, $n = 1, 2, \dots$.

我们已经知道 τ 服从几何分布, 即 $P(\tau = n) = q^{n-1}p$. 当 n 固定时, 样本空间可取为

$$\Omega_n := \{(i_1, \dots, i_n) : i_k = 0, 1, k = 1, \dots, n\}.$$

所以 Ω 是与 n 有关的.

现在我们问: 能建立一个与 n 无关的, 能描述可列重Bernoulli试验的样本空间——概率空间吗?

这个问题的关键是我们预先不知道到底投多少次才能得到正面, 因此有限重Bernoulli试验的概型已经不能满足要求, 因为样本空间必须代之以

$$\Omega := \{(i_1, i_2, \dots), i_k = 0, 1, \forall k\}.$$

这个 Ω 有多少个元素? 若 (i_1, i_2, \dots) 中含有无穷多个0, 则它可对应于 $[0, 1)$ 中的数 $\sum_{k=1}^{+\infty} i_k 2^{-k}$; 而其中只含有有限多个0的元素个数是可数个. 因此, Ω 与 $[0, 1)$ 等势, 即具有连续统的势, 故其中元素已不止可列个, 因此不再是离散概型. 该模型称为可列重Bernoulli试验的概率模型.

这个概型和可列概型的本质区别在于已不可能通过赋予每个 ω 一个 $P(\omega)$ 来定义概率. 因为若设单次投掷中正面出现的概率为 p , $q := 1 - p$, 那么由独立性, 每个 ω 出现的概率都为

$$P(\omega) = p^{\sum_{k=1}^{\infty} i_k} q^{\sum_{k=1}^{\infty} (1-i_k)} = 0,$$

因为 $\sum_{k=1}^{\infty} i_k$ 与 $\sum_{k=1}^{\infty} (1 - i_k)$ 中, 至少有一个是无穷大, 也可能两个都是无穷大.

这个例子告诉我们, 应该在更一般的框架下定义样本空间和概率. 当然, 驱动人们这样做的远远不止这个例子. 事实上, 上世纪二三十年代统计物理和量子力学的发展从更宏大更重要的问题出发也提出了这个要求.

例2. 在 $[0, 1]$ 上任取一点, 求这点落在某集合 A 中的概率.

考虑这个问题时, 我们碰到的第一个难题是, 跟上一个例子一样, 我们依然不能通过对每个 ω 指定一个 $P(\omega)$ 来确定 ω , 因为每个 ω 都是等可能出现的, 所以 $P(\omega)$ 必须是0, 但这个对确定落在 A 中的概率 $P(A)$ 没有任何帮助.

那好, 我们避开这个问题, 直接对 A 定义 $P(A)$. 直观告诉我们, $P(A)$ 应该等于 A 的长度. 但对区间我们可以谈长度, 对有限个区间的并集可以谈长度, 甚至对一般的Borel¹集也可以谈长度的推广——Lebesgue²测度, 但终究不可测量长度的集合是存在的, 我们不可能对这些集合谈长度亦即Lebesgue测度. 这就给我们一个启示: 不是对 Ω 的任何子集都可以谈概率. 这是与可列概型最大的不同, 在那里我们对 Ω 的所有子集 A 都定义了概率 $P(A)$.

此时, 既然不能对所有 Ω 的子集定义概率, 那应该对哪部分子集定义概率呢? 一方面, 我们应该对感兴趣的子集(想要研究的事件)定义概率, 另一方面, 为了计算的便利, 我们也不可避免地会对这些感兴趣的子集进行运算, 因此也需要确保对运算后的子集也是有概率可言的. 基于这样的考虑, 我们面临的任务往往是, 从 Ω 的某个包含感兴趣子集的集类出发, 构造一个包含该集类且关于某些运算封闭的更大的集类, 然后在这个大集类上定义概率.

完成这个任务的可能性来自于Lebesgue测度论的启示: Lebesgue测度正是把长度的概念从一些特殊的集合即区间的长度推广到一般Borel可测集的测度. 因此从某种意义上说, 概率的公理化是建立在Lebesgue的测度论的思想之上的. Kolmogorov之所以看得比别人远一点, 是因为他站在了Lebesgue的肩膀上——这是牛顿原理的合理推论. 当然, 这只是因素之一, Kolmogorov的天才是另一个, 应该说是更主要的, 因素. 换了阿猫阿狗, 能爬上Lebesgue的肩膀吗? 而即使他们运气好, 有人把他们扶上去了, 估计也就是个扶不起的阿斗.

因此现在我们需要做的, 是研究由一个相对简单的集类出发, 如何获得包含该集类的关于某些运算封闭的更大的集类. 这里可以分两步走. 第一步先研究对有限次运算封闭的集类, 这就产生了 π -类, λ_0 -类和代数的概念. 这一步在上一章已完成了.

第二步, 是研究对无穷次运算封闭的集类. 这里的无穷是任意无穷吗? 不是! 这里的无穷只能是可列无穷, 因为若指标集 I 是不可列的, 那么 $\sum_{i \in I} P(A_i)$ 是无法定义的. 就像每个点的长度都是0, 而你无法把 $[0, 1)$ 中所有点的长度加起来而得到整个区间的长度.

所以, 我们只需考虑对可列次运算封闭的集类.

3.2 σ -代数, 单调类与 λ -类

我们固定一个空间 Ω , 下面所有的集合均是其子集.

如果一个集类中有无穷多个元素, 则有必要也有可能考虑它们的可列多次运算, 因此自然就有相应的对可列多次运算封闭的集类的概念.

定义 3.2.1. 一个代数, 若对可列并封闭, 则称为 σ -代数.

更明确地说, 一个代数 \mathcal{F} , 若满足

$$A_n \in \mathcal{F}, n = 1, 2, \dots \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{F},$$

则称为 σ -代数. 由定义结合De Morgan法则直接推出 σ -代数对可列交也是封闭的.

定义 3.2.2. 一个集类 \mathcal{M} , 若对单调极限封闭, 即若

$$A_n \in \mathcal{M}, n = 1, 2, \dots, \{A_n\} \text{单调} \implies \lim_{n \rightarrow \infty} A_n \in \mathcal{M},$$

则称为单调类.

¹Émile Félix-Edouard-Justin Borel, 1871-1956, 法国数学家.

²Henri Léon Lebesgue, 1875-1941, 法国数学家.

所谓单调, 无非是单调上升与单调下降两种情况. 因此上述定义又可复述为:

$$A_n \in \mathcal{M}, n = 1, 2, \dots, A_n \uparrow \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{M},$$

$$A_n \in \mathcal{M}, n = 1, 2, \dots, A_n \downarrow \implies \bigcap_{n=1}^{\infty} A_n \in \mathcal{M}.$$

我们有下面简单的结果.

命题 3.2.3. 一个代数为 σ -代数的充要条件是它是单调类.

证明. 必要性是显然的, 往证充分性. 设代数 \mathcal{F} 是单调类, $A_n \in \mathcal{F}$. 令 $B_n = \cup_{i=1}^n A_i$. 则 $B_n \in \mathcal{F}$ 且 $\{B_n\}$ 单调上升. 于是

$$\bigcup_{n=1}^{\infty} A_n = \lim_{n \rightarrow \infty} B_n \in \mathcal{F}.$$

所以 \mathcal{F} 是 σ -代数. □

因此代数性质和单调类性质是 σ -代数的两要素, 即我们有

$$\sigma\text{-代数} \equiv \text{代数} + \text{单调类}.$$

定义 3.2.4. 非空集类 \mathcal{L} 如果满足以下条件则称为 λ -类:

- (i) $\Omega \in \mathcal{L}$;
- (ii) $E, F \in \mathcal{L}, E \supset F \implies E \setminus F \in \mathcal{L}$;
- (iii) $E_n \in \mathcal{L}, E_n \uparrow \implies \lim_n E_n \in \mathcal{L}$.

此定义表面上只说了 λ -类对单调上升的极限是封闭的, 但很容易推出它对单调下降的极限也是封闭的. 事实上, 设 $E_n \in \mathcal{L}$ 且单调下降, 则 $E_n^c \in \mathcal{L}$ 且单调上升. 于是,

$$\bigcap_{n=1}^{\infty} E_n = \left(\bigcup_{n=1}^{\infty} E_n^c \right)^c \in \mathcal{L}.$$

所以 λ -类一定也是单调类.

命题 3.2.5. (i) 对任意集类 \mathcal{C} , 都存在唯一的 σ -代数 \mathcal{F} , 使得对任意 σ -代数 $\mathcal{F}' \supset \mathcal{C}$, 均有 $\mathcal{C} \subset \mathcal{F} \subset \mathcal{F}'$.

(ii) 对任意集类 \mathcal{C} , 都存在唯一的单调类 \mathcal{M} , 使得对任意单调类 $\mathcal{M}' \supset \mathcal{C}$, 均有 $\mathcal{C} \subset \mathcal{M} \subset \mathcal{M}'$.

(iii) 对任意集类 \mathcal{C} , 都存在唯一的 λ -类 \mathcal{L} , 使得对任意 λ -类 $\mathcal{L}' \supset \mathcal{C}$, 均有 $\mathcal{C} \subset \mathcal{L} \subset \mathcal{L}'$.

证明. 以(i)为例. 取

$$\mathcal{F} := \bigcap \mathcal{G},$$

其中的交是对所有包含了 \mathcal{C} 的 σ -代数而取. 易证这样的 \mathcal{F} 即符合要求.

(ii)和(iii)的证明类似. □

定义 3.2.6. 上面的 \mathcal{F} , \mathcal{M} 与 \mathcal{L} 分别称为 \mathcal{C} 生成的 σ -代数、单调类与 λ -类, 记为 $\sigma(\mathcal{C})$, $\mathcal{M}(\mathcal{C})$ 与 $\lambda(\mathcal{C})$.

由于 σ -代数一定是 λ -类, 而 λ -类一定是单调类, 所以下面的命题是显然的.

命题 3.2.7. 对任何集类 \mathcal{C} 有

$$\mathcal{M}(\mathcal{C}) \subset \lambda(\mathcal{C}) \subset \sigma(\mathcal{C}).$$

从任意一个集类生成 σ -代数的过程过于复杂, 用处不大——如果不是完全没有用处的话. 所以一般都是由 π -类或代数出发生成 σ -代数. 这时我们有两个非常有用的结果——它们都称为单调类定理.

首先我们回头看看命题3.2.3: 一个 σ -代数的两条命脉分别是代数和单调类. 这样下面的定理就呼之欲出了.

定理 3.2.8 (单调类定理1). 代数生成的单调类即为它生成的 σ -代数.

证明. 设这集类为 \mathcal{A} , 它生成的单调类和 σ -代数分别记为 $\mathcal{M}(\mathcal{A})$ 和 $\sigma(\mathcal{A})$. 显然, $\mathcal{M}(\mathcal{A}) \subset \sigma(\mathcal{A})$. 因此只需证 $\mathcal{M}(\mathcal{A})$ 是代数. 由命题2.7.6, 我们只需要证明它对余和交封闭. 先证它对余封闭. 为此, 我们利用证明命题2.7.12时用的方法, 即令

$$\mathcal{B}_1 := \{B \in \mathcal{M}(\mathcal{A}) : B^c \in \mathcal{M}(\mathcal{A})\}.$$

则 $\mathcal{M}(\mathcal{A}) \supset \mathcal{B}_1 \supset \mathcal{A}$. 往证 \mathcal{B}_1 是单调类. 设 $\{B_n\} \subset \mathcal{B}_1$, $B_n \uparrow$ (或 \downarrow) B , 要证 $B \in \mathcal{B}_1$. 则 $\{B_n^c\} \subset \mathcal{M}(\mathcal{A})$, 且 $B_n^c \downarrow$ (或 \uparrow) B^c . 于是, 用 $\mathcal{M}(\mathcal{A})$ 是单调类, 就有 $B^c \in \mathcal{M}(\mathcal{A})$. 所以 $B \in \mathcal{B}_1$. 这就证明了 \mathcal{B}_1 是单调类. 因此 $\mathcal{B}_1 = \mathcal{M}(\mathcal{A})$, 即 $\mathcal{M}(\mathcal{A})$ 对余封闭.

现在证它对交也封闭. 我们用类似的方法, 即令

$$\mathcal{B}_2 := \{B \in \mathcal{M}(\mathcal{A}) : AB \in \mathcal{M}(\mathcal{A}), \forall A \in \mathcal{A}\}.$$

显然 $\mathcal{B}_2 \supset \mathcal{A}$. 同上面类似, 可证 \mathcal{B}_2 是单调类. 因此 $\mathcal{B}_2 = \mathcal{M}(\mathcal{A})$. 这就是说, 对任意 $A \in \mathcal{A}$, $B \in \mathcal{M}(\mathcal{A})$, 有 $AB \in \mathcal{M}(\mathcal{A})$.

再令

$$\mathcal{B}_3 := \{B \in \mathcal{M}(\mathcal{A}) : AB \in \mathcal{M}(\mathcal{A}), A \in \mathcal{M}(\mathcal{A})\}.$$

则由刚刚证明了的, 有 $\mathcal{B}_3 \supset \mathcal{A}$, 且同样地可以证明 \mathcal{B}_3 为单调类. 因此 $\mathcal{B}_3 = \mathcal{M}(\mathcal{A})$. 这就说明 $\mathcal{M}(\mathcal{A})$ 对交是封闭的.

所以 $\mathcal{M}(\mathcal{A})$ 是代数. □

在应用中有时单调类定理之下的形式更方便一些, 它是由Dynkin³通过改造原来的定理得到的.

定理 3.2.9 (单调类定理2, 或称(Dynkin) π - λ 定理). π -类生成的 λ -类即为它生成的 σ -代数.

证明. 记此 π -类为 \mathcal{P} . 则 $\lambda(\mathcal{P}) \subset \sigma(\mathcal{P})$. 但又有

$$\begin{aligned} \lambda(\mathcal{P}) &= \lambda(\lambda(\mathcal{P})) \\ &\supset \mathcal{M}(\lambda(\mathcal{P})) \\ &\supset \mathcal{M}(\lambda_0(\mathcal{P})) \\ &= \mathcal{M}(\alpha(\mathcal{P})) \text{ (由命题2.7.12)} \\ &= \sigma(\alpha(\mathcal{P})) \text{ (由上一定理)} \\ &\supset \sigma(\mathcal{P}). \end{aligned}$$

³Eugene Borisovich Dynkin, 1924-2014, 俄罗斯-美国数学家.

□

习题

1. 设 Ω 有可数个元素, \mathcal{F} 为其子集 σ -代数且 $\{\omega\} \in \mathcal{F}, \forall \omega \in \Omega$. 证明 Ω 的所有子集均属于 \mathcal{F} .
2. 设 E 是一个有无穷个元素的空间. 令

$$\mathcal{E} := \{A \subset E : A \text{ 与 } A^c \text{ 中有一个可数}\}.$$

证明 \mathcal{E} 为 σ -代数.

3. 证明单调类对单调下降的序列的极限也封闭.
4. 考虑 \mathbb{R} 上的集类

$$\mathcal{S}_1 := \{(-\infty, b) : -\infty < b < \infty\},$$

$$\mathcal{S}_2 := \{[a, b) : -\infty < a \leq b < \infty\},$$

$$\mathcal{O} := \{O : O \text{ 是开集}\},$$

$$\mathcal{C} := \{C : C \text{ 是闭集}\}.$$

证明: $\sigma(\mathcal{S}_1) = \sigma(\mathcal{S}_2) = \sigma(\mathcal{O}) = \sigma(\mathcal{C})$. 这个集类称为 \mathbb{R} 上的Borel σ -代数, 记为 \mathcal{B} , 而这个集类中的元素称为Borel集. 叙述并证明 \mathbb{R}^2 及 \mathbb{R}^n 上的平行结果.

5. 设 $O \subset \mathbb{R}^n$ 为开集.

- (a) 证明 $\forall x \in O$, 存在 $q \in \mathbb{Q}^n$ 及 $k \in \mathbb{N}_{++}$, 使得

$$x \in B(q, k^{-1}) \subset O,$$

其中 $B(q, r) := \{y \in \mathbb{R}^n : \max_{1 \leq i \leq n} |y_i - q_i| < r\}$ 或者 $B(q, r) := \{y \in \mathbb{R}^n : \sum_{i=1}^n |y_i - q_i|^2 < r^2\}$.

- (b) 证明存在 $q_i \in \mathbb{Q}^n$ 及 $k_i \in \mathbb{N}_{++}, i = 1, 2, \dots$, 使得

$$O = \bigcup_{i=1}^{\infty} B(q_i, k_i^{-1}).$$

- (c) 证明 $\mathcal{B}^n = \sigma(B(x, r) : x \in \mathbb{R}^n, r > 0)$, 其中 \mathcal{B}^n 为 \mathbb{R}^n 上的Borel σ -代数.

- (d) 证明 $\mathcal{B}^n = \sigma(\mathcal{A}_i), i = 1, 2, 3, 4$, 其中

$$\mathcal{A}_1 = \{(-\infty, b_1) \times \cdots \times (-\infty, b_n), b_k \in \mathbb{R}, \forall k\},$$

$$\mathcal{A}_2 = \{(-\infty, b_1] \times \cdots \times (-\infty, b_n], b_k \in \mathbb{R}, \forall k\},$$

$$\mathcal{A}_3 = \{(a_1, b_1) \times \cdots \times (a_n, b_n), a_k, b_k \in \mathbb{R}, a_k \leq b_k, \forall k\},$$

$$\mathcal{A}_4 = \{(a_1, b_1] \times \cdots \times (a_n, b_n], b_k \in \mathbb{R}, \forall k\}.$$

(e) 令

$$\Pi := \{A \times B : A, B \in \mathcal{B}\}.$$

证明: $\Pi \subset \mathcal{B}^2$ 且 $\mathcal{B}^2 = \sigma(\Pi)$.

一般地, 对任意正整数 m, n , 令

$$\Pi := \{A \times B : A \in \mathcal{B}^m, B \in \mathcal{B}^n\}.$$

证明: $\Pi \subset \mathcal{B}^{m+n}$ 且 $\sigma(\Pi) = \mathcal{B}^{m+n}$.

6. 设 P, Q 都是概率, 令 $\mathcal{A} := \{A : P(A) = Q(A)\}$. 证明 \mathcal{A} 是 λ -类.

7. 设 \mathcal{C} 为 π -类, P_1 与 P_2 都是定义在 $\sigma(\mathcal{C})$ 上的实值非负函数, 且对任意互不相交的 $A_n \in \sigma(\mathcal{C})$, 有

$$P_i \left(\sum_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} P_i(A_n), \quad i = 1, 2.$$

证明: 若 $P_1|_{\mathcal{C}} = P_2|_{\mathcal{C}}$, 则 $P_1 \equiv P_2$.

3.3 概率的公理化

随着各种具体概率模型的丰富, 其它学科(例如量子物理)的需要, 促使Kolmogorov在1930年代考虑了建立概率论的严格的数学基础的问题. 这就需要摆脱具体模型的束缚, 把各种模型的共性提炼出来, 形成公理化体系.

此时数学自身的发展也恰好为Kolmogorov准备了合适的基础, 这就是Lebesgue的测度理论. 这个理论无疑为Kolmogorov提供了重要的启示, 其中最重要的莫过于Lebesgue测度理论将长度的概念推广到了Borel集, 并保留了下面的可列可加性:

$$l \left(\sum_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} l(A_n), \quad \forall A_n \in \mathcal{B}.$$

这里 l 即Lebesgue测度, \mathcal{B} 为 \mathbb{R} 上的Borel σ -代数. 而当 $A = [a, b)$ 时, $l(A) = b - a$. 作为具体例证, 我们取 $A_1 = [0, \frac{1}{2})$, $A_2 = [\frac{1}{2}, \frac{3}{4})$, $A_3 = [\frac{3}{4}, \frac{7}{8})$, \dots , 上式即成为

$$1 = 2^{-1} + 2^{-2} + \dots$$

Kolmogorov在这个基于可列可加性而搭建的舞台上, 发展了概率论最核心的一些概念. 这个舞台和这些概念是此后在这个舞台上上演的一幕幕大戏的必不可少的前提条件.

不过, 事情也有另一面. 据说Kolmogorov本人曾经(在1963年)抱怨过(见[13])概率论的公理化在纯数学上太成功了, 以致于许多人失去了搞清楚怎样在实际问题中应用概率论的兴趣. 这是数典忘祖吗? 是忘了初心不管使命吗? 是异化吗? 该怎样回答Kolmogorov的抱怨呢?

也许你能做的, 我们能做的, 是在遇到每一个概念、每一个结果时, 花个几分钟想想, 它们的实际意义是什么? 适用于什么样的实际问题?

让我们永远不忘初心.

同时我们还要指出, Kolmogorov的公理体系并不是万能的, 正如欧几里得的公理体系不是万能的一样. 比如说它就不适用于理论物理中的一些问题, 因为这些问题涉及无界测度及算子代数. 为解决这些问题, 人们尝试了不同的观点, 发展了不同的工具. 不过这些知识已超出本课程的范围, 就不在此介绍了.

下面我们就介绍Kolmogorov的这个公理化体系.

定义 3.3.1. 设 Ω 为样本空间, \mathcal{F} 为 Ω 的部分子集构成的 σ -代数. 如果存在定义在 \mathcal{F} 上而取值于 $[0, 1]$ 的函数 P , 满足

(i) 规范性: $P(\Omega) = 1$;

(ii) 对 $A_i \in \mathcal{F}, i = 1, 2, \dots$, 且 $A_i A_j = \emptyset, i \neq j$, 有加法公式

$$P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i),$$

则称 \mathcal{F} 为事件域, P 为概率, $P(A)$ 为事件 $A \in \mathcal{F}$ 的概率, 三元组 (Ω, \mathcal{F}, P) 为概率空间.

若 $A \in \mathcal{F}$ 且 $P(A) = 0$, 则称 A 为可略集, 简称 A 可略.

若 $A \in \mathcal{F}$ 且 $P(A) = 1$, 则称 A 几乎必然发生, 记为 A a.s..

显然, 由下面的命题的(iii)知, A a.s.等价于 A^c 可略.

从定义可立即推出 P 有下述性质.

命题 3.3.2. (i) $P(\emptyset) = 0$;

(ii) 对 $A_i \in \mathcal{F}, i = 1, 2, \dots, n$, 且 $A_i A_j = \emptyset, i \neq j$, 有

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

(iii) 设 $A \in \mathcal{F}, P(A^c) = 1 - P(A)$;

(iv) 设 $A, B \in \mathcal{F}, P(A \setminus B) = P(A) - P(AB)$;

特别地, 若 $A \subset B$, 则 $P(A - B) = P(A) - P(B)$ 且 $P(A) \geq P(B)$;

(v) 设 $A, B \in \mathcal{F}, P(A \cup B) = P(A) + P(B) - P(AB)$; 特别地, $P(A) + P(B) - P(AB) \in [0, 1]$;

证明. (i) 因为

$$\emptyset = \emptyset + \emptyset + \dots,$$

所以

$$P(\emptyset) = P(\emptyset) + P(\emptyset) + \dots.$$

于是

$$P(\emptyset) \geq 2P(\emptyset) \geq P(\emptyset).$$

因此 $P(\emptyset) = 0$.

(ii) 补充 $A_k = \emptyset, k = n + 1, n + 2, \dots$. 则由可列可加性有

$$P\left(\sum_{k=1}^n A_k\right) = P\left(\sum_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} P(A_k) = \sum_{k=1}^n P(A_k).$$

(iii) 因为 $A + A^c = \Omega$, 所以

$$P(A) + P(A^c) = P(\Omega) = 1.$$

(iv) 因为 $AB + A \setminus B = A$, 所以

$$P(AB) + P(A \setminus B) = P(A).$$

(iv) 因为

$$A \cup B = A + (B \setminus A),$$

所以

$$P(A \cup B) = P(A) + P(B \setminus A) = P(A) + P(B) - P(AB).$$

□

此命题之(ii)说明, 从 P 的可列可加性可以推出有限可加性. 那么, 有限可加性比可列可加性究竟差了多少呢? 回答是: 差一个单调连续性. 我们先说明什么是单调连续性.

定义 3.3.3. 设 $A_n, A \in \mathcal{F}$. 若

$$A_n \uparrow A \implies P(A_n) \uparrow P(A),$$

则称 P 在 A 处下连续; 若

$$A_n \downarrow A \implies P(A_n) \downarrow P(A),$$

则称 P 在 A 处上连续; 若 P 在任意 $A \in \mathcal{F}$ 处上(下)连续, 则称 P 上(下)连续.

下面我们证明:

命题 3.3.4. 设 P 是 \mathcal{F} 上的取值于 $[0, 1]$ 的函数, 满足规范性和有限可加性. 则 P 的可列可加性和下列条件中的任意一个等价:

- (i) 下连续性;
- (ii) 在 Ω 处的下连续性;
- (iii) 在 \emptyset 处的上连续性;
- (iv) 上连续性.

证明. 我们用(0)表示可列可加性.

(0) \implies (i); 设 $A_n \uparrow A$. 令

$$B_1 := A_1, \quad B_n := A_n - A_{n-1}, \quad n \geq 2.$$

则 $\sum_{i=1}^n B_i = A_n, \quad \sum_{n=1}^{\infty} B_n = A$. 因此

$$P(A_n) = \sum_{i=1}^n P(B_i) \uparrow \sum_{i=1}^{\infty} P(B_i) = P\left(\sum_{i=1}^{\infty} B_i\right) = P(A).$$

(i) \implies (ii): 显然.

(ii) \implies (iii): 设 $A_n \downarrow \emptyset$, 则 $A_n^c \uparrow \Omega$. 所以 $P(A_n^c) \uparrow 1$. 由有限可加性, 有 $P(A_n) = 1 - P(A_n^c) \downarrow 0$.

(iii) \implies (iv): 设 $A_n \downarrow A$, 则 $(A_n - A) \downarrow \emptyset$, 故 $(P(A_n) - P(A)) \downarrow 0$, 即 $P(A_n) \downarrow P(A)$.

(iv) \implies (0): 设 $\sum_{n=1}^{\infty} A_n = A$. 令 $B_n := A - \sum_{i=1}^n A_i$. 则 $B_n \downarrow \emptyset$. 因此 $P(A) - \sum_{i=1}^n P(A_i) = P(B_n) \downarrow 0$. 故

$$P(A) - \sum_{i=1}^{\infty} P(A_i) = P(A) - \lim_{n \rightarrow \infty} \sum_{i=1}^n P(A_i) = 0.$$

□

我们来看几个被定义3.3.1所覆盖的概型.

例1. 离散概型. 设 $\Omega = \{\omega_1, \omega_2, \dots\}$, Ω 有可数个元素. $p_i > 0$, $\sum_i p_i = 1$. 以 \mathcal{F} 表示 Ω 的所有子集构成的 σ -代数. 对 $A \in \mathcal{F}$, 令

$$P(A) := \sum_{\omega_i \in A} p_i.$$

易证 P 具有规范性与可列可加性, 因此 (Ω, \mathcal{F}, P) 为概率空间.

例2. 可列重Bernoulli试验的概率模型. 设 $p \in (0, 1)$, $q = 1 - p$. 令

$$\Omega_n := \{(\omega_1, \omega_2, \dots, \omega_n), \omega_i \in \{0, 1\}, \forall i = 1, \dots, n\},$$

$$\mathcal{F}_n := \{\Omega_n \text{ 的所有子集}\},$$

对 $\omega^n = (\omega_1, \dots, \omega_n) \in \Omega_n$, 令

$$P_n(\omega^n) := p^{\sum_{i=1}^n \omega_i} q^{1 - \sum_{i=1}^n \omega_i},$$

而对 $A_n \in \mathcal{F}_n$, 令

$$P_n(A_n) := \sum_{\omega^n \in A_n} P_n(\omega^n).$$

则由上例知 $(\Omega_n, \mathcal{F}_n, P_n)$ 为概率空间, $\forall n$. 再令

$$\Omega := \{(\omega_1, \omega_2, \dots, \omega_n, \dots), \omega_i \in \{0, 1\}, \forall i \in \mathbb{N}_{++}\},$$

$$\mathcal{F} := \sigma(\mathcal{A}),$$

其中

$$\mathcal{A} = \{A : \exists n, A_n \in \mathcal{F}_n, A = \{\omega : (\omega_1, \dots, \omega_n) \in A_n\}\},$$

则易证(请证!) \mathcal{A} 为代数. 可以证明有唯一一个定义在 \mathcal{F} 上的概率测度 P , 使得对写为上面形式的 A 有

$$P(A) = P_n(A_n).$$

P 的存在性的证明的超出了本课程的范围, 因此我们就直接承认好了. 唯一性的证明如下: 设 P, Q 为 \mathcal{F} 上的两个概率且在 \mathcal{A} 上相等. 令

$$\mathcal{G} := \{A \in \mathcal{F} : P(A) = Q(A)\}.$$

由概率的上下连续性知 \mathcal{G} 为单调类. 但又有 $\mathcal{G} \supset \mathcal{A}$, 所以由单调类定理1有 $\mathcal{G} \supset \sigma(\mathcal{A})$. 从而 $\mathcal{G} = \mathcal{F}$.

不过可列重Bernoulli试验自古, 至少是自Bernoulli就开始用了, 那时应该还不能严格建立其数学模型, 不知古人是怎么蒙混过关的. 当然这在科学发展史上是通例, 没什么好大惊小怪的: 你也可以说关于微积分, 不知牛顿和莱布尼茨是怎么蒙混过关的. 反正就是这样混过来了, 后人再补充了其所缺少的严谨性. 生活之树常青, 而理论永远是个跟屁虫.

公理化概型也包含了一些我们尚未接触过的概型, 例如下面的几何概型.

例3. 取 $\Omega = [0, 1]$, \mathcal{F} 是 $[0, 1]$ 上的Borel σ -代数, P 是Lebesgue测度. 所谓Lebesgue测度就是定义在 \mathcal{F} 上的, 取值于 $[0, 1]$ 的, 满足可列可加性, 且当 $0 \leq a < b \leq 1$ 时, $P([a, b]) = b - a$ 的函数.

这样的 P , 也即Lebesgue测度, 是唯一存在的. 与前一个例子一样, 其存在性的证明要用到比较复杂的测度论知识, 且比较冗长, 在一般实变函数论的书中均可找到, 我们在此略过了. 唯一性的证明比较简单. 事实上, 设有另一个函数 Q 具有同样的性质. 令

$$\mathcal{G} := \{B \in \mathcal{F} : P(B) = Q(B)\}.$$

则易证 \mathcal{G} 为 λ -类. 但 $\mathcal{G} \supset \{[a, b] : 0 \leq a \leq b \leq 1\}$, 而后者是个生成了 \mathcal{F} 的 π -类, 由单调类定理2有 $\mathcal{G} = \mathcal{F}$.

这种概型可一般化:

例4. 取 Ω 为 \mathbb{R}^n 的一个体积有限的区域(或更一般地, 一个Lebesgue测度有限的Borel集), \mathcal{F} 为 Ω 中的Borel集全体, l 为Lebesgue测度. 这里的Lebesgue测度是定义在 \mathcal{F} 上, 取值于 $[0, 1]$, 满足可列可加性, 且对任意区域 A , $l(A)$ 都等于 A 的体积. 定义

$$P(A) := \frac{l(A)}{l(\Omega)}.$$

则 (Ω, \mathcal{F}, P) 为概率空间.

同上例一样, 这个 P 也是唯一存在的.

例5. 乘积概率空间.

设 $(\Omega_i, \mathcal{F}_i, P_i)$, $i = 1, 2$, 为概率空间. 定义

$$\Omega := \Omega_1 \times \Omega_2 := \{\omega = (\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\},$$

$$\mathcal{F} := \mathcal{F}_1 \times \mathcal{F}_2 := \sigma(A_1 \times A_2, A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2),$$

其中

$$A_1 \times A_2 = \{\omega = (\omega_1, \omega_2), \omega_1 \in A_1, \omega_2 \in A_2\}.$$

可以证明(但该证明超出了本课程的范围, 故不在此证明)在 (Ω, \mathcal{F}) 上存在唯一的概率 P , 使得

$$P(A_1 \times A_2) = P_1(A_1) \times P_2(A_2), \quad \forall A_i \in \mathcal{F}_i.$$

(Ω, \mathcal{F}, P) 称为 $(\Omega_1, \mathcal{F}_1, P_1)$ 与 $(\Omega_2, \mathcal{F}_2, P_2)$ 的乘积(概率)空间, 记为

$$(\Omega, \mathcal{F}, P) = (\Omega_1, \mathcal{F}_1, P_1) \times (\Omega_2, \mathcal{F}_2, P_2).$$

类似地, 设 $(\Omega_i, \mathcal{F}_i, P_i)$, $i = 1, 2, \dots, n$ 为概率空间. 令

$$\Omega := \prod_{i=1}^n \Omega_i := \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \Omega_i\},$$

$$\mathcal{F} := \prod_{i=1}^n \mathcal{F}_i = \sigma \left\{ \prod_{i=1}^n A_i : A_i \in \mathcal{F}_i \right\}.$$

则在 \mathcal{F} 上存在唯一一个概率 P , 使得

$$P \left(\prod_{i=1}^n A_i \right) = \prod_{i=1}^n P_i(A_i), \quad \forall A_i \in \mathcal{F}_i, i = 1, \dots, n.$$

更一般地, 设 $(\Omega_i, \mathcal{F}_i, P_i), i = 1, 2, \dots$, 为概率空间. 令

$$\Omega := \prod_{i=1}^{\infty} \Omega_i := \{\omega = (\omega_1, \dots, \omega_n, \dots) : \omega_i \in \Omega_i\},$$

$$\mathcal{F} := \prod_{i=1}^{\infty} \mathcal{F}_i = \sigma \left\{ \prod_{i=1}^n A_i \times \Omega_{n+1} \times \Omega_{n+2} \times \dots : A_i \in \mathcal{F}_i, i = 1, \dots, n, n \in \mathbb{N}_{++} \right\}.$$

则在 \mathcal{F} 上存在唯一一个概率 P , 使得

$$P \left(\prod_{i=1}^n A_i \times \Omega_{n+1} \times \Omega_{n+2} \times \dots \right) = \prod_{i=1}^n P_i(A_i), \quad \forall A_i \in \mathcal{F}_i, i = 1, \dots, n, n \in \mathbb{N}_{++}.$$

习题

1. 以 Ω 表示 $[0, 1]$ 上的所有有理数的集合, \mathcal{F} 表示其全体子集. 证明: 不存在 Ω 上的概率 P , 使得对任意 $0 \leq a < b \leq 1$ 有

$$P(\Omega \cap [a, b]) = b - a.$$

2. 设 \mathcal{C} 为 π -类, P_1 与 P_2 都是定义在 $\sigma(\mathcal{C})$ 上的概率. 证明: 若 P_1 与 P_2 限制在 \mathcal{C} 上相等, 则它们恒等.
3. 设 P_n 是定义在同一样本空间上的概率, 且 $\forall A \in \mathcal{F}, \lim_{n \rightarrow \infty} P_n(A) =: P(A)$ 存在. 证明 P 也是这个空间上的概率.
4. 设 P_n 是定义在同一样本空间上的概率, $a_n \geq 0, \sum_{n=1}^{\infty} a_n = 1$. 证明 $\sum_{n=1}^{\infty} a_n P_n$ 也是定义在这个空间上的概率.

5. 安得倚天抽宝剑, 把汝裁为三截. 现在把 $[0, 1]$ 裁为三截, 裁法是在 $[0, 1]$ 上随机地独立取两点, 每点均服从 $[0, 1]$ 上的均匀分布.

(a) 求这三截长度的分布;

(b) 这三截能构成一个三角形的概率是多少?

(c) 在最小的一截小于 x 的条件下, 它们能构成三角形的概率又是多少?

6. (Buffon的针) 设两条平行线间的距离为 l . 现将一条长度为 l 的针投到这两条线中间, 求该针与其中一条相交的概率.
7. (约会问题) 两人约好7点到8点之间在某地会面, 并约定先到的那个等待十分钟, 过时不候. 如果两人均是随机地在7点至8点之间到达, 问他们能见面的概率是多少?

8. 设 P 是概率. 证明

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

9. 设 P 是概率. 证明

$$P \left(\bigcup_{i=1}^n A_i \right) = \sum_{i=1}^n (-1)^{i-1} \sum_{1 \leq k_1 < \dots < k_i \leq n} P(A_{k_1} \cdots A_{k_i}).$$

10. 设 P 是概率, $\lim_{n \rightarrow \infty} A_n = A$ 存在. 证明

$$P(A) = \lim_{n \rightarrow \infty} P(A_n).$$

11. 设 $\{A_n, n \geq 1\}$ 是概率空间 (Ω, \mathcal{F}, P) 上的一列事件. 证明:

$$\sum_{n=1}^{\infty} P(A_n) < \infty \implies P\left(\limsup_{n \rightarrow \infty} A_n\right) = 0.$$

这个结果称为Borel-Cantelli第一引理.

12. 证明下列不等式:

(a) Boole不等式:

$$P\left(\bigcap_{i=1}^n A_i\right) \geq 1 - \sum_{i=1}^n P(A_i^c) \geq \sum_{i=1}^n P(A_i) - (n-1);$$

(b) Kounias不等式:

$$P\left(\bigcup_{i=1}^n A_i\right) \leq \min_j \left\{ \sum_{i=1}^n P(A_i) - \sum_{i \neq j} P(A_i A_j) \right\}.$$

(c) Chung-Erdős不等式:

$$P\left(\bigcup_{i=1}^n A_i\right) \geq \frac{(\sum_{i=1}^n P(A_i))^2}{\sum_{i,j=1}^n P(A_i A_j)}.$$

13. 证明对任意事件列 $\{A_n\}$ 有:

$$\lim_{n \rightarrow \infty} P\left(A_n \setminus \limsup_{n \rightarrow \infty} A_n\right) = \lim_{n \rightarrow \infty} P\left(\liminf_{n \rightarrow \infty} A_n \setminus A_n\right) = 0.$$

14. 设 $\{A_n\}$ 为事件列. 证明:

(a)

$$1\left(\liminf_{n \rightarrow \infty} A_n\right) = \liminf_{n \rightarrow \infty} 1(A_n), \quad 1\left(\limsup_{n \rightarrow \infty} A_n\right) = \limsup_{n \rightarrow \infty} 1(A_n);$$

(b)

$$1\left(\bigcup_{n=1}^{\infty} A_n\right) = \max_n 1(A_n), \quad 1\left(\bigcap_{n=1}^{\infty} A_n\right) = \min_n 1(A_n);$$

(c)

$$P\left(\liminf_{n \rightarrow \infty} A_n\right) \leq \liminf_{n \rightarrow \infty} P(A_n) \leq \limsup_{n \rightarrow \infty} P(A_n) \leq P\left(\limsup_{n \rightarrow \infty} A_n\right).$$

3.4 条件概率, 独立性与条件独立性

设 (Ω, \mathcal{F}, P) 为概率空间, $A, B \in \mathcal{F}$, $P(A) > 0$. 依据前面的经验, 定义 A 发生的条件下 B 发生的条件概率为

$$P(B|A) = \frac{P(AB)}{P(A)}.$$

现设 $P(A)P(B) > 0$. 由于

$$P(B|A)P(A) = P(A|B)P(B)(= P(AB)),$$

所以

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$

这个公式称为Bayes公式. 它是离散概型时的Bayes公式的自然延伸.

设 $\mathcal{S} := \{\Omega_1, \dots, \Omega_n, \dots\}$ 构成 Ω 的一个分割, 即

$$\Omega = \sum_{n=1}^{\infty} \Omega_n, \quad \Omega_n \in \mathcal{F}, \quad P(\Omega_n) > 0, \quad \forall n.$$

则全概率公式依然成立. 即 $\forall A \in \mathcal{F}$, 有

$$P(A) = \sum_{n=1}^{\infty} P(\Omega_n)P(A|\Omega_n).$$

在对实际问题建立概率模型时, 往往是先有条件概率, 再通过乘法公式决定概率. 我们看一个具体例子.

例1. 考虑下面的试验: 一个人首先在装有 m 个红球, n 个白球的袋中摸一个; 如果摸到红球, 则转到装有 p 个蓝球, q 个黑球的袋中再摸一次; 如果摸到白球, 则转到装有 r 个黄球, s 个紫球的袋中再摸一次. 建立这个试验的数学模型.

解. 以 A_1 表第一次摸红球, A_2 表第一次摸白球, 以 B_1, B_2, B_3, B_4 分别表第二次摸蓝黑黄紫球. 则样本空间为

$$\Omega = \{A_i B_j, i = 1, 2, j = 1, 2, 3, 4\}.$$

我们需要赋予每个 $A_i B_j$ 以概率. 注意这个概率不是随便赋予的, 它必须满足乘法公式.

由题意有

$$\begin{aligned} P(A_1) &= \frac{m}{m+n}, \quad P(A_2) = \frac{n}{m+n}, \\ P(B_1|A_1) &= \frac{p}{p+q}, \quad P(B_2|A_1) = \frac{q}{p+q}, \quad P(B_3|A_1) = P(B_4|A_1) = 0, \\ P(B_1|A_2) &= P(B_2|A_2) = 0, \quad P(B_3|A_2) = \frac{r}{r+s}, \quad P(B_4|A_2) = \frac{s}{r+s}. \end{aligned}$$

据此可用乘法公式计算出 $P(A_i B_j)$, 即

$$P(A_i B_j) = P(B_j|A_i)P(A_i).$$

然后可算出

$$P(B_j|A_i) = \frac{P(A_i B_j)}{P(A_i)}.$$

最后再根据全概率公式计算每个 B_j 的概率:

$$P(B_j) = P(B_j|A_1)P(A_1) + P(B_j|A_2)P(A_2).$$

以 $\sigma(\mathcal{P})$ 记 \mathcal{P} 生成的 σ -代数. 定义

$$h(A, \omega) := \sum_{n=1}^{\infty} P(A|\Omega_n)1_{\Omega_n}(\omega), \quad A \in \mathcal{F}, \omega \in \Omega.$$

h 称为给定 $\sigma(\mathcal{P})$ 时 A 的条件概率. 注意 h 是二元函数: 当固定 ω 时, 它作为定义在 \mathcal{F} 上的函数是一个概率, 且对 $\omega \in \Omega_n$ 就是给定 Ω_n 时的条件概率; 当固定 A 时, 作为 ω 的函数是一个在每个 Ω_n 上取常值 $P(A|\Omega_n)$ 的函数. 当然, h 还依赖于 \mathcal{P} . 为完整地表示出这种依赖关系, 我们将用 $P(A|\mathcal{P})(\omega)$ 表示这个函数.

独立性的概念和离散时也是一样的.

定义 3.4.1. 设 A, B 是事件. 若

$$P(AB) = P(A)P(B),$$

则称 A, B 独立, 或相互独立.

设 A_1, \dots, A_n 是事件. 若对任意 $2 \leq k \leq n$ 及 $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$, 有

$$P(A_{i_1} \cdots A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}),$$

则称 A_1, \dots, A_n 独立.

设 A_1, \dots, A_n, \dots 是事件. 若 $\forall n, A_1, \dots, A_n$ 独立, 则称 A_1, \dots, A_n, \dots 独立.

定义 3.4.2. 设 \mathcal{C}, \mathcal{D} 是集类. 若对任意 $A \in \mathcal{C}, B \in \mathcal{D}, A, B$ 都独立, 则称 \mathcal{C}, \mathcal{D} 独立.

设 $\mathcal{C}_1, \mathcal{C}_2, \dots$ 是集类. 若对任意 n , 有

$$P(A_1 \cdots A_n) = P(A_1) \cdots P(A_n),$$

这里 $A_i \in \mathcal{C}_i \cup \{\Omega\}, 1 \leq i \leq n$, 则称 $\mathcal{C}_1, \mathcal{C}_2, \dots$ 独立.

注意, 因为这里允许 $A_i = \Omega$, 而不管是否 $\Omega \in \mathcal{C}_i$, 所以本定义自然意味着对任意 n , 任意 $i_1 < i_2 < \dots < i_n$, 任意 $A_{i_k} \in \mathcal{C}_{i_k}, A_{i_1}, \dots, A_{i_n}$ 都是独立的, 因为对缺失的指标, 取相应的集合为 Ω , 再用定义即可.

下面的结论非常有用:

命题 3.4.3. 若 $\mathcal{C}_1, \mathcal{C}_2, \dots$ 是 π -类且相互独立, 则 $\sigma(\mathcal{C}_1), \sigma(\mathcal{C}_2), \dots$ 相互独立.

证明. 我们证两个的情况, 一般情况类似.

因此设 \mathcal{C} 与 \mathcal{D} 是两个相互独立的 π -类. 令

$$\mathcal{F}_1 := \{A \in \sigma(\mathcal{C}) : A \text{ 与 } \mathcal{D} \text{ 独立}\}.$$

则 $\mathcal{F}_1 \supset \mathcal{C}$ 且 \mathcal{F}_1 为 λ -类. 因此由定理3.2.9, $\mathcal{F}_1 = \sigma(\mathcal{C})$.

再令

$$\mathcal{F}_2 := \{B \in \sigma(\mathcal{D}) : B \text{ 与 } \sigma(\mathcal{C}) \text{ 独立}\}.$$

则 $\mathcal{F}_2 \supset \mathcal{D}$, 且 \mathcal{F}_2 是 λ -类. 因此由定理3.2.9, $\mathcal{F}_2 = \sigma(\mathcal{D})$. □

如果在独立性的定义中以条件概率代替原概率, 则得到条件独立性的概念.

定义 3.4.4. 设 A, B, C 是事件, $P(A) > 0$. 若

$$P(BC|A) = P(B|A)P(C|A),$$

则称 B, C 在给定 A 时条件独立.

利用条件概率的定义, B, C 在给定 A 时条件独立即为

$$P(ABC)P(A) = P(AB)P(AC),$$

而 B, C 独立则为

$$P(BC) = P(B)P(C).$$

因此由独立不能得到条件独立, 反正亦然. 关于这一事实的解释是: 条件独立性是限制在某种条件下的独立性, 和整体独立性没有直接的关系. 当然, 这并不是说它们完全没有关系. 有的时候它们也有很密切的关系. 比如, 容易看出, 若 A, B, C 独立, 那么在条件 A 下, B, C 也是条件独立的.

我们来看一些例子.

设古典概型的样本空间为

$$\Omega = \{(0, 0), (0, 1), (1, 0), (1, 1)\},$$

三个事件为

$$A = \{(0, 1), (1, 0)\}, \quad B = \{(0, 0), (0, 1)\}, \quad C = \{(0, 0), (1, 0)\},$$

则

$$P(B) = \frac{1}{2}, \quad P(C) = \frac{1}{2}, \quad P(BC) = \frac{1}{4}.$$

因此 $P(BC) = P(B)P(C)$, 知 B, C 独立. 但是由于

$$P(B|A) = \frac{1}{2}, \quad P(C|A) = \frac{1}{2}, \quad P(BC|A) = 0,$$

因此 $P(BC|A) \neq P(B|A)P(C|A)$, 知 B, C 在给定 A 时不条件独立.

另一方面, 设古典概型的样本空间为

$$\Omega = \{(0, 0, 0), (0, 1, 0), (1, 0, 0), (1, 1, 0), (0, 0, 1), (0, 1, 1), (1, 0, 1), (1, 1, 1)\},$$

三个事件为

$$A = \{(1, 1, 0), (1, 1, 1)\}, \quad B = \{(1, 0, 0), (1, 1, 0), (1, 0, 1), (1, 1, 1)\}, \quad C = \{(1, 1, 1)\},$$

则

$$P(B) = \frac{1}{2}, \quad P(C) = \frac{1}{8}, \quad P(BC) = \frac{1}{8}.$$

因此 $P(BC) \neq P(B)P(C)$, 知 B, C 不独立. 但是由于

$$P(B|A) = 1, \quad P(C|A) = \frac{1}{2}, \quad P(BC|A) = \frac{1}{2},$$

因此 $P(BC|A) = P(B|A)P(C|A)$, 知 B, C 在给定 A 时条件独立.

公理化概型是漂亮的数学模型. 但在应用到实际问题时, 必须十分小心. 尤其是独立性与条件独立性这两个概念的区别虽然在数学定义上是清楚的, 但在应用到实际问题中往往不是那么清楚, 需要仔细斟酌, 马虎不得. 我们来看两个具体例子.

例1. 设某个地区某种流行病的流行率是 p . 现有两家公司该病的检测. 甲公司在真阳性的情况下, 检测出阳性的概率为 p_1 ; 在真阴性的情况下, 检测出阳性的概率为 p_2 . 乙公司在真阳性的情况下, 检测出阳性的概率为 q_1 ; 在真阴性的情况下, 检测出阳性的概率为 q_2 . 两公司独立运作. 现设某个人

(i) 在甲公司检出了阳性, 问他是真阳性的概率是多少?

(ii) 在两家公司均检出了阳性. 问他是真阳性的概率是多少?

解. 以 A_1 表示真阳性, A_2 表示真阴性; B_1 表示甲公司检出阳性, B_2 表示甲公司检出阴性; C_1 表示乙公司检出阳性, C_2 表示乙公司检出阴性. 由题意有

$$P(A_1) = p, P(A_2) = 1 - p;$$

$$P(B_1|A_1) = p_1, P(B_2|A_1) = 1 - p_1, P(B_1|A_2) = p_2, P(B_2|A_2) = 1 - p_2;$$

$$P(C_1|A_1) = q_1, P(C_2|A_1) = 1 - q_1, P(C_1|A_2) = q_2, P(C_2|A_2) = 1 - q_2.$$

则

$$P(B_1) = P(B_1|A_1)P(A_1) + P(B_1|A_2)P(A_2) = p_1p + p_2(1 - p),$$

$$P(A_1B_1) = P(B_1|A_1)P(A_1) = p_1p.$$

所以

$$P(A_1|B_1) = \frac{p_1p}{p_1p + p_2(1 - p)}.$$

然后, 问题来了: B_i 与 C_j 到底是独立还是条件独立?

独立操作意味着什么呢? 它意味着当检测一个人时, 无论在此人是感染了还是没有感染的情况下, 两家得到的结果都是独立的. 因此, 这里的独立性表现为条件独立性, 即

$$P(B_iC_j|A_k) = P(B_i|A_k)P(C_j|A_k) \quad \forall i, j, k.$$

因此有

$$\begin{aligned} P(B_1C_1) &= P(B_1C_1|A_1)P(A_1) + P(B_1C_1|A_2)P(A_2) \\ &= P(B_1|A_1)P(C_1|A_1)P(A_1) + P(B_1|A_2)P(C_1|A_2)P(A_2) \\ &= p_1q_1p + p_2q_2(1 - p), \end{aligned}$$

$$P(A_1B_1C_1) = P(B_1C_1|A_1)P(A_1) = p_1q_1p.$$

所以

$$P(A_1|B_1C_1) = \frac{p_1q_1p}{p_1q_1p + p_2q_2(1 - p)}.$$

在正常情况下, 即 $p_2/p_1 \ll 1$ 与 $q_2/q_1 \ll 1$ 的情况下, 第二个概率要大于第一个. 这是符合逻辑的.

读者可以自己算算, 看 B_i 与 C_j 是否独立.

另外, 注意到

$$P(A_1|B_1) = \frac{P(B_1|A_1)P(A_1)}{P(B_1|A_1)P(A_1) + P(B_1|A_2)P(A_2)},$$

而

$$\begin{aligned} P(A_1|B_1C_1) &= \frac{P(B_1|A_1)P(C_1|A_1)P(A_1)}{P(B_1|A_1)P(C_1|A_1)P(A_1) + P(B_1|A_2)P(C_1|A_2)P(A_2)} \\ &= \frac{P(C_1|A_1)P(A_1|B_1)P(B_1)}{P(C_1|A_1)P(A_1|B_1)P(B_1) + P(C_1|A_2)P(A_2|B_1)P(B_1)} \\ &= \frac{P(C_1|A_1)P(A_1|B_1)}{P(C_1|A_1)P(A_1|B_1) + P(C_1|A_2)P(A_2|B_1)}, \end{aligned}$$

因此计算 $P(A_1|B_1C_1)$ 时, 就是在 $P(A_1|B_1)$ 的计算公式中将 $P(B_1|A_1)$, $P(B_1|A_2)$, $P(A_1)$, $P(A_2)$, 分别直接换为 $P(C_1|A_1)$, $P(C_1|A_2)$, $P(A_1|B_1)$, $P(A_2|B_1)$ 来计算即可. 回顾2.5节例4, 那时我们已经利用直观这么做过. 现在看来, 当时也是自觉地假设了那种检测方法在已知他有病(或无病)条件下, 先后的两次检测结果是条件独立的.

例2. 设一个家庭有两个小孩, 现在有甲乙两人分别去看了一下, 都看到了一个女孩. 问两个小孩都是女孩的概率是多少?

我们先看看一种解法.

以 B_1 表示两个都是男孩, B_2 表示两个都是女孩, B_3 表示一个男孩一个女孩. 则

$$P(B_1) = P(B_2) = \frac{1}{4}, \quad P(B_3) = \frac{1}{2}.$$

以 A_1 表甲看到的是女孩, A_2 表乙看到的是女孩. 则

$$P(A_i|B_1) = 0, \quad P(A_i|B_2) = 1, \quad P(A_i|B_3) = \frac{1}{2}, \quad i = 1, 2.$$

由全概率公式, 所以

$$\begin{aligned} P(A_i) &= P(A_i|B_1)P(B_1) + P(A_i|B_2)P(B_2) + P(A_i|B_3)P(B_3) \\ &= \frac{1}{4} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}, \quad i = 1, 2. \end{aligned}$$

另一方面

$$P(B_2) = \frac{1}{4}.$$

直观上, A_1 是甲看到女孩, A_2 是乙看到女孩, 两者相互不受影响, 因此 A_1 与 A_2 应该是独立的. 又由于 $P(A_i|B_2) = 1$, 则 $B_2 \subset A_i$, $i = 1, 2$. 所以

$$P(B_2|A_1A_2) = \frac{P(B_2A_1A_2)}{P(A_1A_2)} = \frac{P(B_2)}{P(A_1A_2)} = \frac{1/4}{1/4} = 1.$$

但这个答案显然是错误的. 然而问题出在哪里呢? A_1 与 A_2 独立吗?

令

$$\omega_1 = (b, b), \omega_2 = (g, g), \omega_3 = (b, g), \omega_4 = (g, b).$$

虽然没有明说,但在上面的解法中,我们心目中的样本空间是

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}.$$

然而在这个模型下, A_1, A_2 都不是事件(你把它表示为事件给我瞧瞧?), 所以此后的一切计算都是错误的.

那么什么是正确的样本空间呢?

不忘初心. 我们应该回忆一下什么是样本空间. 所谓样本空间, 是所有可能的结果的集合. 但上面这个 Ω , 只展现了结果的一部分, 即该家庭的孩子的情况, 而没有展现两个观察者看到的情况, 因此信息是不完整的, 从而无法支撑起正确的数学模型.

那么什么是正确的模型? 正确的模型必须包含完整的信息. 具体到本例, 它必须完整地呈现该家庭孩子的情况和观察者看到的情况. 因此正确的模型如下.

以 α_1 表示甲看到的是男孩, α_2 表示甲看到的是女孩, β_1 表示乙看到的是男孩, β_2 表示乙看到的是女孩. 令

$$\Omega := \{(\omega_i, \alpha_j, \beta_k), i = 1, 2, 3, 4, k, j = 1, 2\},$$

其中 ω_i 还是上面的 ω_i . 由题意, 合理的概率应定义为

$$\begin{aligned} P(\omega_1, \alpha_1, \beta_1) &= \frac{1}{4}, \\ P(\omega_1, \alpha_k, \beta_j) &= 0, \quad k \vee j = 2, \\ P(\omega_2, \alpha_2, \beta_2) &= \frac{1}{4}, \\ P(\omega_2, \alpha_k, \beta_j) &= 0, \quad k \wedge j = 1, \\ P(\omega_i, \alpha_k, \beta_j) &= \frac{1}{16}, \quad i = 3, 4, k, j = 1, 2. \end{aligned}$$

因为

$$A_1 A_2 = \{(\omega_2, \alpha_2, \beta_2), (\omega_3, \alpha_2, \beta_2), (\omega_4, \alpha_2, \beta_2)\},$$

所以

$$P(A_1 A_2) = \frac{1}{4} + 2 \times \frac{1}{16} = \frac{3}{8}.$$

又

$$B_2 = \{(\omega_2, \alpha_2, \beta_2)\},$$

因此

$$P(B_2 | A_1 A_2) = \frac{1/4}{3/8} = \frac{2}{3}.$$

这才是正确答案.

我们顺便指出, 在这个例子中, 对每个 $i = 1, 2, 3, 4$, 在给定 ω_i 的条件下, A_1 与 A_2 都是独立的, 但 A_1 与 A_2 整体上不是独立的. 这是因为

$$\begin{aligned} P(A_1 A_2) &= \frac{3}{8} \neq \frac{1}{2} \times \frac{1}{2} = P(A_1)P(A_2), \\ P(A_1 A_2 | \omega_1) &= 0 = P(A_1 | \omega_1)P(A_2 | \omega_1), \\ P(A_1 A_2 | \omega_2) &= 1 = P(A_1 | \omega_2)P(A_2 | \omega_2), \\ P(A_1 A_2 | \omega_i) &= \frac{1}{4} = \frac{1}{2} \times \frac{1}{2} = P(A_1 | \omega_i)P(A_2 | \omega_i), \quad i = 3, 4. \end{aligned}$$

也可以在缩小的样本空间里直接计算条件概率后, 用全概率公式来计算. 显然

$$P(A_1A_2|B_1) = 0, \quad P(A_1A_2|B_3) = 1.$$

现在计算 $P(A_1A_2|B_2)$. 对 B_2 观察2次, 此时样本空间应该是

$$\Omega_2 = \{(0, 0), (0, 1), (1, 0), (1, 1)\},$$

因此 $P(A_1A_2|B_2) = 1/4$. 由全概率公式,

$$\begin{aligned} P(A_1A_2) &= P(A_1A_2|B_1)P(B_1) + P(A_1A_2|B_2)P(B_2) + P(A_1A_2|B_3)P(B_3) \\ &= \frac{1}{4} \times \frac{1}{2} + \frac{1}{4} = \frac{3}{8}. \end{aligned}$$

因此

$$P(B|A_1A_2) = \frac{P(BA_1A_2)}{P(A_1A_2)} = \frac{P(B)}{P(A_1A_2)} = \frac{\frac{1}{4}}{\frac{3}{8}} = \frac{2}{3}.$$

习题

1. 设一个家庭有两个小孩. 现有 n 个人分别去看, 其中有 k 个人看到的是男孩 ($0 \leq k \leq n$). 求该家庭的两个孩子都是男孩, 女孩以及一个男孩一个女孩的概率.
2. 设 $\{A_1, A_2, \dots\}$ 为 Ω 的一个分割, B 与 C 为事件. 设 $\forall i$, 在给定 A_i 时, B 与 C 条件独立. 再设 $\forall i$, C 与 A_i 独立. 证明 B 与 C 独立. 举例说明 C 与 A_i 独立的假设不可去掉.
3. 设 A_1, \dots, A_5 是独立事件. 证明:
 - (a) $(A_1 \cup A_2)A_3$ 与 $A_4^c \cup A_5^c$ 独立;
 - (b) $A_1 \cup A_2, A_3A_4$ 与 A_5^c 独立.
4. 设 $\mathcal{A} := \{A_i, i = 1, \dots\}$ 与 $\mathcal{B} := \{B_j, j = 1, \dots\}$ 独立, 且 \mathcal{A} 与 \mathcal{B} 中的元素都是两两不交的. 证明 $\sigma(\mathcal{A})$ 与 $\sigma(\mathcal{B})$ 独立.
5. 设从甲地到乙地有三条公路可走, 选择它们时能按时到达的概率分别为 p_1, p_2, p_3 ; 现分别以 q_1, q_2, q_3 的概率选择它们. 建立这个试验的概率模型并求按时到达的概率.
6. 证明: 若 $\mathcal{C}_1, \dots, \mathcal{C}_n, \dots$ 均是 π -类且相互独立, 则 $\sigma(\mathcal{C}_1), \dots, \sigma(\mathcal{C}_n), \dots$ 也相互独立.
7. 设 A_1, \dots, A_n, \dots 相互独立且 $P(A_i) = p_i$. 用诸 p_i 表示出 $P(B)$, 其中 $B = \cup_{i=1}^{\infty} A_i$.
8. 设 A, B 独立, $P(A) = p, P(B) = q$. 计算 A, B 中至少 k 个发生, 刚好 k 个发生, 至多 k 个发生的概率, $k = 0, 1, 2$.
9. 将上题推广到一般情况: 设 A_1, \dots, A_n 独立, $P(A_i) = p_i$. 计算诸 A_i 中至少 k 个发生, 刚好 k 个发生, 至多 k 个发生的概率, $k = 0, 1, \dots, l, l \leq n$.
10. 设 $A \subset B$ 且 A 与 B 独立. 证明: 或者 $P(A) = 0$, 或者 $P(B) = 1$. 特别地, 若 A 与它自己独立, 则 $P(A) = 0$ 或 1 .
11. 设 A, B, C 是三个事件, $P(C)P(C^c) > 0$. 判断下面陈述是否正确, 证明之或举反例否定之.

(a)

$$P(A) > P(B) \implies P(A|C) > P(B|C), P(A|C^c) > P(B|C^c);$$

(b)

$$P(A|C) > P(B|C), P(A|C^c) > P(B|C^c) \implies P(A) > P(B).$$

12. 证明对条件概率的全概率公式: 设 A, B 是事件, C_1, C_2, \dots 是 Ω 的分割, 则

$$P(A|B) = \sum_{i=1}^{\infty} P(A|BC_i)P(C_i|B),$$

只要 $P(BC_i) > 0, \forall i = 1, 2, \dots$.

13. 设一家有两个孩子, 老大是男孩, 而有人又随机地看到这家的一个男孩. 问这家的两个都是男孩的概率是多少?

14. 设 \mathcal{F}, \mathcal{G} 分别为 Ω_1 与 Ω_2 上的 σ -代数, 且分别为 π -类 \mathcal{C}, \mathcal{D} 所生成, 而 $\Omega_1 \in \mathcal{C}, \Omega_2 \in \mathcal{D}$. 在乘积空间:

$$\Omega_1 \times \Omega_2 := \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}$$

上定义乘积 σ -代数:

$$\mathcal{F} \times \mathcal{G} := \sigma(A \times B : A \in \mathcal{F}, B \in \mathcal{G}).$$

证明:

$$\mathcal{F} \times \mathcal{G} = \lambda(\{A \times B : A \in \mathcal{C}, B \in \mathcal{D}\}).$$

叙述并证明多个 σ -代数的情形.

15. 设 $\{A_n, n \geq 1\}$ 独立. 证明:

$$\sum_{n=1}^{\infty} P(A_n) = \infty \implies P(\limsup A_n) = 1.$$

这个结果称为 Borel-Cantelli 第二引理.

16. 设 $\mathcal{F}_1, \mathcal{F}_2, \dots$ 独立. 定义其尾 σ -代数:

$$\mathcal{G} := \bigcap_{n=1}^{\infty} \sigma(\mathcal{F}_{n+1}, \mathcal{F}_{n+2}, \dots).$$

证明:

(a) $\forall n < m, \sigma(\mathcal{F}_1, \dots, \mathcal{F}_n)$ 与 $\sigma(\mathcal{F}_{n+1}, \dots, \mathcal{F}_{n+m})$ 独立;

(b) $\forall n, \sigma(\mathcal{F}_1, \dots, \mathcal{F}_n)$ 与 $\sigma(\mathcal{F}_{n+1}, \mathcal{F}_{n+2}, \dots)$ 独立;

(c) $\forall n, \sigma(\mathcal{F}_1, \dots, \mathcal{F}_n)$ 与 \mathcal{G} 独立;

(d) $\sigma(\mathcal{F}_1, \mathcal{F}_2, \dots)$ 与 \mathcal{G} 独立;

(e) \mathcal{G} 与 \mathcal{G} 独立;

(f) $\forall A \in \mathcal{G}, P(A) = 0$ 或 1 .

最后这个结论叫 Kolmogorov 0-1 律. 恭喜你, 你也会证了!

4 随机变量

我们已经有了离散概率空间上的随机变量的概念. 本章我们要在一般的概率空间上定义随机变量并研究它们的各种性质.

如果想做某种联想, 把概率空间类比于欧氏空间的话, 那么随机变量就可类比于欧氏空间上的函数: 在数学分析里是连续函数或分段连续函数, 在实变函数里是可测函数. 不过类比于后者更合适一些, 因为定义连续函数首先要有两点是否靠近的概念(f 连续是指当 x 和 y 充分靠近时, $f(x)$ 和 $f(y)$ 也充分靠近), 而在概率空间上并没有这个概念; 然而在实变函数中, 主要研究对象是可测函数, 这个不需要有“靠近”的概念, 而只需要“可测”的概念. 概率空间足够支撑起这个概念.

4.1 基本概念

我们先从一些一般概念讲起. 设 S_1, S_2 是两个集合, $f: S_1 \mapsto S_2$. 对 $A \subset S_2$, 令

$$f^{-1}(A) := \{x \in S_1 : f(x) \in A\},$$

称为 A 在 f 下的原像. 注意原像不是逆映射, 虽然它们常常用同一个记号 f^{-1} 表示. 逆映射是 S_2 到 S_1 的映射. 只有当 f 是单射时, 逆映射才存在, 且定义域为

$$f(S_1) := \{y \in S_2 : \exists x \in S_1 \text{ 使得 } y = f(x)\}.$$

(所以只有当 f 为满射时, 逆映射的定义域才是整个 S_2). 而原像永远都是存在的, 它把 S_2 的任意子集映射成 S_1 的子集, 大不了是个空集, 而空集也是 S_1 的子集. 本书中大部分时间, f^{-1} 都表示原像; 在表示逆映射的很少部分时间, 我们会加以说明.

原像的一个简单有用的性质是它保持所有的运算关系不变, 见习题1.

设 \mathcal{S}_2 是 S_2 中的集类. 记

$$f^{-1}(\mathcal{S}_2) := \{f^{-1}(A) : A \in \mathcal{S}_2\}.$$

则 $f^{-1}(\mathcal{S}_2)$ 是 S_1 中的集类. 由习题1, 当 \mathcal{S}_2 本身为 σ -代数时, $f^{-1}(\mathcal{S}_2)$ 也为 σ -代数. 我们还有

命题 4.1.1. 设 \mathcal{S}_2 是 S_2 中的集类, 则

$$\sigma(f^{-1}(\mathcal{S}_2)) = f^{-1}(\sigma(\mathcal{S}_2)).$$

证明. 由习题1, $f^{-1}(\sigma(\mathcal{S}_2))$ 为 σ -代数, 且 $f^{-1}(\sigma(\mathcal{S}_2)) \supset f^{-1}(\mathcal{S}_2)$. 所以

$$f^{-1}(\sigma(\mathcal{S}_2)) \supset \sigma(f^{-1}(\mathcal{S}_2)).$$

为证反包含, 令

$$\mathcal{G} := \{A \in \sigma(\mathcal{S}_2) : f^{-1}(A) \in \sigma(f^{-1}(\mathcal{S}_2))\}.$$

易见 $\mathcal{S}_2 \subset \mathcal{G}$, 且由习题1易证 \mathcal{G} 为 σ -代数. 因此 $\mathcal{G} = \sigma(\mathcal{S}_2)$. 所以 $f^{-1}(\sigma(\mathcal{S}_2)) \subset \sigma(f^{-1}(\mathcal{S}_2))$. \square

在欧氏空间上有一个经典的 σ -代数, 称为 Borel σ -代数. 我们现在来介绍它. 让我们从最简单的一维情况 \mathbb{R} 开始.

对 $x \in \mathbb{R}$, $\rho > 0$, 令

$$B(x, \rho) := \{y : |y - x| < \rho\},$$

称为以 x 为中心 ρ 为半径的开球.

设 $O \subset \mathbb{R}$. 如果

$$x \in O \implies \exists \rho > 0, \text{ 使得 } B(x, \rho) \subset O,$$

则 O 称为开集.

与开集对立的观念是闭集. 一个集合 C , $x \in \mathbb{R}$, 如果 $\forall \rho > 0$, 都有 $C \cap B(x, \rho) \neq \emptyset$, 称 x 为 C 的聚点. 如果 C 包含它所有的聚点, 则称为闭集. 我们有:

命题 4.1.2. C 为闭集 $\iff C^c$ 为开集.

证明. 设 C 为闭集. 设 $x \in C^c$, 则 $x \notin C$ 且 x 不是 C 的聚点. 所以存在 $\rho > 0$ 使得 $C \cap B(x, \rho) = \emptyset$. 所以 $B(x, \rho) \subset C^c$. 因此 C^c 是开集.

反之, 设 C^c 是开集. 则 $\forall x \in C^c$, 有 $\rho > 0$ 使得 $B(x, \rho) \subset C^c$. 因此 x 不是 C 的聚点. 因此 C 包含了它所有的聚点. \square

以 \mathcal{O} 表示 \mathbb{R} 上的开集全体, \mathcal{B} 表示 \mathcal{O} 生成的 σ -代数:

$$\mathcal{B} := \sigma(\mathcal{O}),$$

称为(一维)Borel σ -代数.

以 \mathcal{O}^c 表示闭集全体. 由上一命题, 我们有

命题 4.1.3.

$$\mathcal{B} = \sigma(\mathcal{O}^c).$$

我们注意到 $\forall x \in \mathbb{R}$, x 是闭集, 所以任意单点集 $\{x\} \in \mathcal{B}$.

不管开集或者闭集, 其结构都相当复杂. 我们当然希望找到一些结构简单的集合来生成 \mathcal{B} . 这就是我们下面要做的事情.

$\forall -\infty \leq a \leq b \leq \infty$, 以 (a, b) 表示开区间:

$$(a, b) := \{x \in \mathbb{R} : a < x < b\},$$

Π_1 表示所有这样的开区间构成的集类:

$$\Pi_1 := \{(a, b) : -\infty \leq a \leq b \leq \infty\}.$$

因为 $(a, b) \cap (c, d) = (a \vee c, b \wedge d)$ (若 $a \vee c \geq b \wedge d$, 则视为 \emptyset), 所以 Π_1 为 π -类.

令

$$\mathcal{C} := \{B(x, \rho) : x \in \mathbb{Q}, \rho \in \mathbb{Q}_+\}.$$

因为 \mathbb{Q} 为可数集, 所以 \mathcal{C} 为可数集.

命题 4.1.4. 设 O 是开集, 则存在开球 $B_i, i \in I$, 其中 I 是可数集, 使得

$$O = \bigcup_{i \in I} B_i.$$

证明. $\forall x \in O$, 取 $\rho > 0$ 使得 $B(x, \rho) \subset O$.

取 $r \in B(x, \frac{\rho}{3}) \cap \mathbb{Q}$. 令

$$B_x := B\left(r, \frac{\rho}{2}\right).$$

则 $x \in B(r, \frac{\rho}{2})$, 且 $\forall y \in B(r, \frac{\rho}{2})$ 有

$$|y - x| < \rho.$$

所以 $B(r, \frac{\rho}{2}) \subset B(x, \rho) \subset O$, 且 $O = \bigcup_{x \in O} B_x$.

由于 $\forall x, B_x \in \mathcal{C}$, 而 \mathcal{C} 为可数集, 所以存在指标集 I 及开球 $B_i, i \in I$, 使得

$$\forall x \in O, \exists i \in I, \text{使得 } B_x = B_i.$$

于是

$$O = \bigcup_{i \in I} B_i.$$

□

由此命题, $\mathcal{O} \subset \sigma(\Pi_1)$, 因此我们有

$$\mathcal{B} = \sigma(\Pi_1).$$

仍然对 $\forall -\infty \leq a \leq b < \infty$, 以 $(a, b]$ 表示左开右闭的区间

$$(a, b] := \{x \in \mathbb{R} : a < x \leq b\}.$$

再约定

$$(a, \infty] := (a, \infty).$$

以 Π_2 表示这样的区间全体:

$$\Pi_2 := \{(a, b] : -\infty \leq a \leq b \leq \infty\}.$$

由于对任意 $a < b, c < d$

$$(a, b] \cap (c, d] = (a \vee c, b \wedge d],$$

所以 Π_2 也为 π 类.

我们来看看 $\sigma(\Pi_2)$ 是什么. 由于

$$(a, b] = (a, b) \cup \{b\},$$

所以 $(a, b] \in \mathcal{B}$, 故 $\sigma(\Pi_2) \subset \mathcal{B}$. 又由于

$$(a, b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right],$$

所以

$$\Pi_1 \subset \sigma(\Pi_2).$$

所以

$$\mathcal{B} = \sigma(\Pi_1) \subset \sigma(\Pi_2).$$

于是

$$\mathcal{B} = \sigma(\Pi_2).$$

令

$$\Pi_3 := \{[a, b] : -\infty \leq a \leq b \leq \infty\}.$$

这里同样约定 $[-\infty, \infty] = (-\infty, \infty)$, 等等. 由于 $[a, b]$ 是闭集, 所以

$$\sigma(\Pi_3) \subset \mathcal{B}.$$

又由于

$$(a, b) = \bigcap_{n=1}^{\infty} \left[a + \frac{1}{n}, b - \frac{1}{n} \right],$$

所以

$$\sigma(\Pi_3) \supset \sigma(\Pi_1) = \mathcal{B}.$$

所以

$$\sigma(\Pi_3) = \mathcal{B}.$$

再令

$$\Pi_4 := \{[a, b) : -\infty \leq a \leq b \leq \infty\}.$$

在 $\pm\infty$ 处做同样的约定. 和 Π_2 同样可证

$$\mathcal{B} = \sigma(\Pi_4).$$

令

$$\Pi_5 = \{(-\infty, a), -\infty < a \leq \infty\}.$$

由于 $\Pi_5 \subset \Pi_4$, 所以

$$\Pi_5 \subset \sigma(\Pi_4) = \mathcal{B}.$$

所以

$$\sigma(\Pi_5) \subset \mathcal{B}.$$

又

$$[a, b) = (-\infty, b) - (-\infty, a),$$

所以

$$\Pi_4 \subset \sigma(\Pi_5).$$

因此

$$\mathcal{B} = \sigma(\Pi_4) = \sigma(\Pi_5).$$

令

$$\Pi_6 = \{(-\infty, a], -\infty < a \leq \infty\},$$

由于 $\Pi_6 \subset \Pi_2$, 于是

$$\sigma(\Pi_6) \subset \mathcal{B}.$$

又由于

$$(-\infty, a) = \bigcup_{n=1}^{\infty} \left(-\infty, a - \frac{1}{n} \right],$$

所以

$$\mathcal{B} = \sigma(\Pi_5) \subset \sigma(\Pi_6).$$

最后, 令

$$\Pi_7 = \{(a, \infty), -\infty \leq a < \infty\},$$

$$\Pi_8 = \{[a, \infty), -\infty \leq a < \infty\}.$$

也可证明

$$\mathcal{B} = \sigma(\Pi_7) = \sigma(\Pi_8).$$

下面我们转向多维情形. 我们将主要讨论二维情况, 更高维情况类似. 在 \mathbb{R}^2 上, 定义两点 $x = (x_1, x_2)$ 与 $y = (y_1, y_2)$ 的距离:

$$d_1(x, y) := \sqrt{|x_1 - y_1|^2 + |x_2 - y_2|^2}$$

及

$$d_2(x, y) := |x_1 - y_1| \vee |x_2 - y_2|.$$

对 $x \in \mathbb{R}^2$, $\rho \geq 0$, 相应地可以定义开球

$$B_1(x, \rho) := \{y : d_1(x, y) < \rho\}, \quad B_2(x, \rho) := \{y : d_2(x, y) < \rho\}.$$

下面的 d 及 B , 既可以是 d_1 与 B_1 , 也可以是 d_2 与 B_2 .

设 $O \subset \mathbb{R}^2$. 若 $\forall x \in O, \exists \rho > 0$, 使得 $B(x, \rho) \subset O$, 则 O 称为开集. 显然, 用两种开球定义的开集是一样的.

以 \mathcal{O} 表示开集全体. 令

$$\mathcal{B}^2 := \sigma(\mathcal{O}),$$

称为二维Borel σ -代数.

设 $C \subset \mathbb{R}^2$, 若 C 包含了它全部的聚点, 则称为闭集.

用与一维情况完全一样的方法, 可以证明:

命题 4.1.5. C 为闭集 $\iff C^c$ 为开集.

因此, 所有闭集 $\in \mathcal{B}^2$. 特别地, $\forall a, b \in \mathbb{R}$,

$$\{(x, y) \in \mathbb{R}^2 : x = a, y \leq b\}, \quad \{(x, y) \in \mathbb{R}^2 : x \leq a, y = b\}$$

均是闭集, 所以也均是Borel集.

同样地, 用与一维情况一样的方法, 可以证明:

命题 4.1.6. 设 O 为开集, 则存在开球 $B_i, i \in I, I$ 为可数集, 使得

$$O = \bigcup_{i \in I} B_i.$$

令

$$\Pi := \{B : B \text{ 是开球}\}$$

则由上一命题有

$$\mathcal{B}^2 = \sigma(\Pi).$$

对 $A, B \subset \mathbb{R}$, 令

$$A \times B := \{(x, y) \in \mathbb{R}^2 : x \in A, y \in B\}.$$

$$\Pi_i^2 := \{A \times B : A, B \in \Pi_i, i = 1, 2, \dots, 8\}.$$

具体一点, 即

$$\Pi_1^2 := \{(a_1, b_1) \times (a_2, b_2), -\infty \leq a_i \leq b_i \leq \infty, i = 1, 2\},$$

$$\Pi_2^2 := \{(a_1, b_1] \times (a_2, b_2], -\infty \leq a_i \leq b_i \leq \infty, i = 1, 2\},$$

等等.

我们有

命题 4.1.7.

$$\mathcal{B}^2 = \sigma(\Pi_i^2), i = 1, 2, \dots, 8.$$

证明. (i) $i = 1$:

因为 $\forall A \in \Pi_1^2$, A 为开集, 所以 $\Pi_1^2 \subset \mathcal{B}^2$. 从而 $\sigma(\Pi_1^2) \subset \mathcal{B}^2$.

又由命题 4.1.6 后面的结论, 有

$$\mathcal{B}^2 = \sigma(d_2 \text{ 意义下的开球}) \subset \sigma(\Pi_1^2).$$

所以 $\mathcal{B}^2 = \sigma(\Pi_1^2)$.

(ii) $i = 2, 3, 4$: 按一维的方法如法炮制.

(iii) $i = 5$: 因为 $\Pi_5^2 \subset \Pi_4^2$, 所以 $\sigma(\Pi_5^2) \subset \sigma(\Pi_4^2) \subset \mathcal{B}^2$.

反之, $\forall a_1 \leq a_2, b_1 \leq b_2$, 有

$$\begin{aligned} [a_1, a_2) \times [b_1, b_2) &= [(-\infty, a_2) \times (-\infty, b_2) - (-\infty, a_1) \times (-\infty, b_2)] \\ &\quad - [(-\infty, a_2) \times (-\infty, b_1) - (-\infty, a_1) \times (-\infty, b_1)], \end{aligned}$$

所以 $\Pi_4^2 \subset \sigma(\Pi_5^2)$. 于是

$$\mathcal{B}^2 = \sigma(\Pi_4^2) \subset \sigma(\Pi_5^2).$$

剩下几种情况和一维相应情况的证明类似. □

一般地, 对任意 $n \in \mathbb{N}_{++}$, 以 \mathcal{B}^n 表示 n -维 Borel σ -代数. 有了 Borel σ -代数的概念, 就有了 Borel 可测函数的概念.

定义 4.1.8. 设 $f : \mathbb{R}^n \mapsto \mathbb{R}^m$. 若

$$f^{-1}(B) \in \mathcal{B}^n, \forall B \in \mathcal{B}^m,$$

则称 f 为 Borel 可测函数, 简称 Borel 函数.

上述定义中的条件常常简记为

$$f^{-1}(\mathcal{B}^m) \subset \mathcal{B}^n.$$

Borel函数复合Borel函数仍然是Borel函数, 即我们有下面的

命题 4.1.9. 设 $\varphi: \mathbb{R}^n \mapsto \mathbb{R}^m$ 及 $\psi: \mathbb{R}^m \mapsto \mathbb{R}^l$ 均为 Borel函数, 则 $\psi \circ \varphi: \mathbb{R}^n \mapsto \mathbb{R}^l$ 也为 Borel函数.

证明. 设 $A \in \mathcal{B}^l$, 则 $\psi^{-1}(A) \in \mathcal{B}^m$, 因此

$$(\psi \circ \varphi)^{-1}(A) = \varphi^{-1}(\psi^{-1}A) \in \mathcal{B}^n.$$

□

我们可以简化验证一个函数是Borel函数的手续. 首先我们有:

命题 4.1.10. 设 \mathcal{C} 是 \mathbb{R}^m 的某些子集构成的集类, 且 $\sigma(\mathcal{C}) = \mathcal{B}^m$. 则 $\varphi: \mathbb{R}^n \mapsto \mathbb{R}^m$ 是 Borel函数的充要条件是 $\varphi^{-1}(\mathcal{C}) \subset \mathcal{B}^n$.

我们知道有很多这样的集类. 例如可取 \mathcal{C} 为开集全体, 或者闭集全体, 或者以及上面所定义的 $\Pi_i^m, i = 1, \dots, 8$, 等等.

证明. 必要性显然, 往证充分性. 证法一. 令

$$\mathcal{G} := \{A \in \mathcal{B}^m : \varphi^{-1}(A) \in \mathcal{B}^n\}.$$

则 \mathcal{G} 为 σ -代数, 且 $\mathcal{G} \supset \mathcal{C}$. 所以

$$\mathcal{B}^m = \sigma(\mathcal{C}) \subset \mathcal{G} \subset \mathcal{B}^m.$$

证法2. 因为 $\varphi^{-1}(\mathcal{C}) \subset \mathcal{B}^n$, 因此 $\sigma(\varphi^{-1}(\mathcal{C})) \subset \mathcal{B}^n$. 由命题4.1.1有

$$\varphi^{-1}(\mathcal{B}^m) = \varphi^{-1}(\sigma(\mathcal{C})) = \sigma(\varphi^{-1}(\mathcal{C})) \subset \mathcal{B}^n.$$

□

由此我们推得:

推论 4.1.11. $\varphi = (\varphi_1, \dots, \varphi_m): \mathbb{R}^n \mapsto \mathbb{R}^m$ 为 Borel函数的充要条件是每个 $\varphi_i, i = 1, 2, \dots, m$, 都是 Borel函数.

证明. 必要性显然.(见第三章第二节习题5.) 至于充分性, 若每个 φ_i 都是 Borel函数, 则

$$\varphi_i^{-1}((-\infty, a_i]) \in \mathcal{B}^n, \quad \forall i = 1, 2, \dots, m.$$

所以,

$$\varphi^{-1}\left(\prod_{i=1}^m (-\infty, a_i]\right) = \bigcap_{i=1}^m \varphi_i^{-1}((-\infty, a_i]) \in \mathcal{B}^n.$$

再用上一命题即完成证明.

□

有了这个结果, 很多命题的证明往往退化到只需对数值Borel函数证明. 数值Borel函数包含了哪些函数呢? 首先, 它包含了连续函数, 即我们有:

命题 4.1.12. 设 $\varphi: \mathbb{R}^n \mapsto \mathbb{R}$ 连续, 则 φ 为 Borel 函数.

证明. 设 $A \subset \mathbb{R}$ 为开集, 则 $\varphi^{-1}(A)$ 也为开集, 因而为 Borel 集. 再用命题 4.1.10 即可. \square

其次, 对任何一个 \mathbb{R}^n 上的 Borel 集 $A \in \mathcal{B}^n$, 其示性函数

$$f(x) := 1_A(x)$$

显然也是 Borel 函数.

设 $a_1, \dots, a_k \in \mathbb{R}$, $A_1, \dots, A_k \in \mathcal{B}^n$. 由于 $(x_1, \dots, x_n) \mapsto a_1 x_1 + \dots + a_k x_k$ 为连续函数, 所以函数

$$\varphi(x_1, \dots, x_n) := \sum_{i=1}^k a_i 1_{A_i}$$

也为 Borel 函数. 这样的函数结构简单, 因此称为简单 Borel 函数.

并不是所有的 Borel 函数都是简单 Borel 函数, 当然. 不过, 简单 Borel 函数却可以逼近任何 Borel 函数. 这是一个非常有用的结果, 其精确叙述如下:

命题 4.1.13. 设 f 为 \mathbb{R}^n 上的数值 Borel 函数, 则存在一系列简单 Borel 函数 f_k , 使得

$$\lim_{k \rightarrow \infty} f_k(x) = f(x) \quad \forall x \in \mathbb{R}^n.$$

证明. 设 $f \geq 0$. 令

$$f_k(x) := \sum_{i=0}^{2^k-1} i 2^{-k} 1_{f \in [i 2^{-k}, (i+1) 2^{-k})} + 2^k 1_{f \geq 2^k}.$$

易证 f_k 满足要求. 一般地, 令

$$f_k := f_k^+ - f_k^-$$

即可, 这里 f^+ 、 f^- 分别表示 f 的正部与负部. \square

将定义 4.1.8 中的出发地 $(\mathbb{R}^m, \mathcal{B}^m)$ 换为概率空间 (Ω, \mathcal{F}, P) 中的 (Ω, \mathcal{F}) , 再取 $n = 1$, 就得到随机变量的定义. 不过为方便起见, 我们允许随机变量取 $\pm\infty$ 值.

定义 4.1.14. 函数 $\xi: \Omega \mapsto \bar{\mathbb{R}} := [-\infty, +\infty]$ 若满足

$$\xi^{-1}(A) := \{\omega: \xi(\omega) \in A\} \in \mathcal{F}, \quad \forall A \in \mathcal{B},$$

则称为随机变量; 若进一步有 $P(|\xi| = \infty) = 0$ (或等价地, $|\xi| < \infty$ a.s.), 则称为实值随机变量.

既然随机变量可以取 ∞ , 那么就不可避免地会碰到涉及到 ∞ 的运算. 在此我们规定:

$$(+\infty) + (+\infty) = +\infty, \quad (-\infty) + (-\infty) = -\infty, \quad (+\infty) - (+\infty) = +\infty, \quad (-\infty) - (+\infty) = -\infty,$$

对任何有限数 a ,

$$a + (\pm\infty) = \pm\infty, \quad a - (\mp\infty) = \pm\infty, \quad (\pm\infty) \pm a = \pm\infty$$

对任意 $a > 0$ 及 $a = +\infty$,

$$a(\pm\infty) = \pm\infty,$$

对任意 $a < 0$ 及 $a = -\infty$,

$$a(\pm\infty) = \mp\infty.$$

而下列运算被认为是无意义的:

$$\pm\infty + (\mp\infty), \pm\infty - (\pm\infty), \mp\infty - (\mp\infty), 0 \times (\mp\infty).$$

多维随机变量可类似定义:

定义 4.1.15. 函数 $\xi: \Omega \mapsto \bar{\mathbb{R}}^n$ 若满足

$$\xi^{-1}(A) := \{\omega: \xi(\omega) \in A\} \in \mathcal{F}, \quad \forall A \in \mathcal{B}^n,$$

则称为 n -维随机变量.

我们注意到, 在随机变量的定义中, P 是不起作用的. 此外, 我们有下面的等价性条件:

命题 4.1.16. 设 $\xi = (\xi_1, \dots, \xi_n)$ 为 $\Omega \mapsto \mathbb{R}^n$. 则下列条件等价:

- (i) ξ 为 n -维随机变量;
- (ii) \forall 开集 O , $\xi^{-1}(O) \in \mathcal{F}$;
- (iii) 存在 $i \in \{1, \dots, n\}$, 使得

$$\xi^{-1}(A) \in \mathcal{F}, \quad \forall A \in \Pi_i^n.$$

证明. 只需注意

$$\mathcal{B}^n = \sigma(\mathcal{A}_1) = \sigma(\mathcal{A}_2),$$

其中 $\mathcal{A}_1, \mathcal{A}_2$ 分别是 (ii), (iii) 中的集类 (见第三章第二节习题5), 由命题4.1.1立得. \square

有了这个结果, 我们就很容易地知道, 判断一个多维函数是不是随机变量, 只要看各个分量即可. 即我们有:

命题 4.1.17. $\xi = (\xi_1, \dots, \xi_n)$ 是 n -维随机变量的充要条件是: 对任意 i , ξ_i 是随机变量.

证明留作习题.

在数学分析里我们经常碰到复合函数, 即函数的函数. 同理, 在概率论中我们也需要考虑随机变量的函数. 设 ξ 是随机变量, $\varphi: \mathbb{R} \mapsto \mathbb{R}$, 比如考虑 $\varphi(\xi)$. 问题是: $\varphi(\xi)$ 是否还是一个随机变量? 下面的结果回答了这个问题.

命题 4.1.18. 设 ξ 是 n -维随机变量, $\varphi: \mathbb{R}^n \mapsto \mathbb{R}^m$ 为 Borel 函数, 则 $\varphi(\xi)$ 为 m -维随机变量.

证明. 设 $A \in \mathcal{B}(\mathbb{R}^m)$. 由于 φ 是 Borel 函数, 所以 $\varphi^{-1}(A) \in \mathcal{B}(\mathbb{R}^n)$. 因此

$$[\varphi(\xi)]^{-1}(A) = \xi^{-1}(\varphi^{-1}(A)) \in \mathcal{F}.$$

故 $\varphi(\xi)$ 是随机变量. \square

习题

1. 证明原像保持所有的运算关系不变, 即

(a)

$$f^{-1}\left(\bigcup_{i \in I} A_i\right) = \bigcup_{i \in I} f^{-1}(A_i);$$

(b)

$$f^{-1}\left(\bigcap_{i \in I} A_i\right) = \bigcap_{i \in I} f^{-1}(A_i);$$

(c)

$$f^{-1}(A \setminus B) = f^{-1}(A) \setminus f^{-1}(B);$$

(d)

$$A \subset B \implies f^{-1}(A) \subset f^{-1}(B),$$

等等.

2. 设 $\varphi: \mathbb{R}^m \mapsto \mathbb{R}^n$ 为连续函数. 证明 φ 为 Borel 函数.

3. 设 $\varphi: \mathbb{R} \mapsto \mathbb{R}$ 为单调函数. 证明 φ 为 Borel 函数.

4. 设 $\forall n = 1, 2, \dots$, ξ_n 是随机变量, $A_n \in \mathcal{F}$, 且 $\{A_n\}$ 构成 Ω 的一个分割. 令

$$\xi(\omega) := \sum_{n=1}^{\infty} \xi_n(\omega) 1_{A_n}(\omega).$$

证明 ξ 是随机变量.

5. 设 $f: \mathbb{R}^n \mapsto \mathbb{R}^m$. 证明下列两条件均为 f 是 Borel 函数的等价条件:

(a) 对任意开集 $O \subset \mathbb{R}^m$, $f^{-1}(O) \in \mathcal{B}^n$;

(b) 对任意闭集 $C \subset \mathbb{R}^m$, $f^{-1}(C) \in \mathcal{B}^n$.

6. 设 $f_n: \mathbb{R}^m \rightarrow \mathbb{R}$ 为 Borel 函数. 证明

(a)

$$\liminf_n f_n, \limsup_n f_n$$

均为 Borel 函数.

(b)

$$\{x: \lim_n f_n(x) \text{ 存在}\} \in \mathcal{B}^m.$$

7. 设 $f: \mathbb{R}^n \mapsto \mathbb{R}^m$ 为连续函数. 证明它为 Borel 函数.

8. 设 ξ_1, ξ_2, \dots 为随机变量, 证明下面的量均是随机变量:

(a) $\eta_n := \max_{1 \leq i \leq n} \xi_i$, $\eta := \sup_{i \geq 1} \xi_i$;

(b) $\zeta_n := \min_{1 \leq i \leq n} \xi_i$, $\zeta := \inf_{i \geq 1} \xi_i$;

(c) $\theta := \limsup_{n \rightarrow \infty} \xi_n$, $\gamma := \liminf_{n \rightarrow \infty} \xi_n$.

(d) $\alpha\xi_1 + \beta\xi_2, \xi_1\xi_2$.

(e)

$$\eta := \begin{cases} \xi_1/\xi_2, & \xi_2 \neq 0, \\ 0, & \xi_2 = 0. \end{cases}$$

4.2 分布函数

设 ξ 是实值随机变量. 对 $A \in \mathcal{B}$, 由于 $\xi^{-1}(A) \in \mathcal{F}$, 所以 $P(\xi^{-1}(A))$ 是有意义的. 因此可定义

$$\mu(A) := P(\xi^{-1}(A)).$$

易证 μ 是 $(\mathbb{R}, \mathcal{B})$ 上的概率, 因此 $(\mathbb{R}, \mathcal{B}, \mu)$ 是概率空间. μ 称为 ξ 的概率分布, 因为它反映了 ξ 的值是依什么概率分布的.

对随机变量 ξ , 定义其分布函数为

$$F_\xi(x) := P(\xi \leq x).$$

在不发生混淆的情况下, 往往简写为 F . 分布函数和概率分布的关系是:

$$F(x) = \mu((-\infty, x]).$$

我们有:

命题 4.2.1. 任意一个实值随机变量的分布函数 F 都具有下列性质:

- (i) $F: \mathbb{R} \mapsto [0, 1]$;
- (ii) F 单调上升且右连续;
- (iii)

$$F(-\infty) := \lim_{x \rightarrow -\infty} F(x) = 0,$$

$$F(\infty) := \lim_{x \rightarrow \infty} F(x) = 1.$$

证明. (i)显然.

(ii). 设 $x < y$, 则

$$F(y) - F(x) = P(\xi \leq y) - P(\xi \leq x) = P(x < \xi \leq y) \geq 0.$$

$\forall x, \forall n$, 当 $x \leq y < x + n^{-1}$ 时,

$$0 \leq F(y) - F(x) \leq F(x + n^{-1}) - F(x) = P(\xi \in (x, x + n^{-1}])$$

而

$$P(\xi \in (x, x + n^{-1}]) \downarrow P(\xi \in \emptyset) = 0, \quad n \rightarrow \infty.$$

所以

$$\lim_{y \downarrow x} F(y) = F(x).$$

而当 $y \uparrow x$ 时, $F(y) \uparrow$ 且 $F(y) \leq F(x)$. 所以 $F(x-) := \lim_{y \uparrow x} F(y)$ 存在.

(iii) 因为 F 单调, 所以 $F(-\infty)$ 与 $F(\infty)$ 均存在, 且

$$F(-\infty) = \lim_{n \rightarrow \infty} F(-n) = \lim_{n \rightarrow \infty} P(\xi \leq -n) = P(\xi = -\infty) = 0,$$

$$F(\infty) = \lim_{n \rightarrow \infty} F(n) = \lim_{n \rightarrow \infty} P(\xi \leq n) = P(\xi < \infty) = 1.$$

□

任何满足上面命题中这几条性质的函数, 可以脱离于随机变量, 也称为分布函数.

设 μ 是 $(\mathbb{R}, \mathcal{B})$ 上的概率, 定义

$$F(x) := \mu(-\infty, x],$$

则 F 是分布函数, 称为 μ 的分布函数; 反之, 设 F 是分布函数, 则有唯一一个概率 μ , 使得 F 是 μ 的分布函数. 这里 μ 的存在性我们就不证了, 可见于许多实变函数或实分析方面的教材; 可以证明 μ 是由 F 唯一决定的, 即我们有

命题 4.2.2. 设 μ, ν 是 $(\mathbb{R}^m, \mathcal{B}^m)$ 上的两个概率. 若

$$\mu((-\infty, a]) = \nu((-\infty, a]) \quad \forall a \in \mathbb{R}^m,$$

则 $\mu \equiv \nu$.

证明. 令

$$\mathcal{G} := \{A \in \mathcal{B} : \mu(A) = \nu(A)\}.$$

则易证 \mathcal{G} 为 λ -类; 但 \mathcal{G} 又包含了 π -类

$$\Pi := \{(-\infty, a], a \in \mathbb{R}^m\},$$

所以由 $\pi - \lambda$ 定理,

$$\mathcal{G} \supset \sigma(\Pi) = \mathcal{B}.$$

□

因此, $(\mathbb{R}, \mathcal{B})$ 上的概率和 \mathbb{R} 上的分布函数是一一对应的. 如果随机变量 ξ, η 的概率分布一样, 我们称 ξ, η 同分布. 显然, 这等价于 ξ, η 的分布函数是一样的.

习题

1. 设 ξ 是实值随机变量. 证明 ξ 的概率分布是 $(\mathbb{R}, \mathcal{B})$ 上的概率.
2. 设 ξ, η 是定义在概率空间 (Ω, \mathcal{F}, P) 上的随机变量. 证明 $\xi + \eta, -\xi, \xi^2, |\xi|, \xi^+, \xi^-, \xi\eta$ 也是随机变量.(稍后做)
3. 定义在同一概率空间上的两个随机变量 ξ, η 如果满足:

$$P(\omega : \xi(\omega) = \eta(\omega)) = 1,$$

则称为 ξ 与 η 几乎必然相等, 记为 $\xi = \eta$ a.s. 或 p.s. (a.s. 是英语almost surely的缩写, p.s. 是法语presque sûrement的缩写.) 证明: 若 $\xi = \eta$ a.s., 则 $F_\xi = F_\eta$.

4. 举例说明存在两个随机变量 ξ, η , 使得 $\forall \omega, \xi(\omega) \neq \eta(\omega)$, 但 $F_\xi = F_\eta$.
5. 设 ξ 为随机变量, F 为其分布函数.

(a) 令

$$A := \{x : P(\xi = x) > 0\}.$$

证明 A 最多有可数个元素.

(b) 令

$$B = \{x : F(x) - F(x-) > 0\}.$$

证明 $A = B$.

(c) 证明

$$P(\xi < x) = F(x-), \quad P(\xi = x) = F(x) - F(x-).$$

6. $\forall -\infty \leq y < x \leq \infty$,

$$F(x) - F(y) = P(\xi \in (y, x]),$$

$$F(x) - F(y-) = P(\xi \in [y, x]),$$

$$F(x-) - F(y-) = P(\xi \in [y, x)),$$

$$F(x-) - F(y) = P(\xi \in (y, x)).$$

7. 设 $A = \sum_{i=1}^{\infty} (y_i, x_i]$, 则

$$\sum_{i=1}^{\infty} [F(x_i) - F(y_i)] = P(\xi \in A).$$

4.3 分类

根据分布函数的性质, 可将随机变量分为离散型随机变量、连续型随机变量. 当然, 也有既非离散亦非连续的随机变量, 但本课程不涉及其具体例子.

如果随机变量 ξ 的值域为可数集 $\{x_1, x_2, \dots\}$, 则称为离散型随机变量, 其分布函数 $F(x)$ 称为离散型分布. 此时有 $\Delta F(x_i) := F(x_i) - F(x_i-) > 0$, 其中 $F(x-)$ 是 F 在 x 的左极限. 我们称 $\{(x_k, p_k), k = 1, 2, \dots\}$ 为 ξ 的分布列, 其中

$$P(\xi = x_k) = p_k.$$

注意! 我们这里认为 x_1, x_2, \dots 中没有相同的, 即

$$x_i \neq x_j, \quad \forall i \neq j.$$

如果出现了重复的, 则必须把它们看成是一个点, 且将相应的概率加起来. 比如说, 如果随机变量 ξ 的取值为 x_1, x_2, x_3 , 但 x_1, x_2, x_3 并不是两两互异的, 而是

$$x_1 = 0, \quad x_2 = 1, \quad x_3 = 0.$$

且

$$P(\xi = x_1) = \frac{1}{2}, \quad P(\xi = x_2) = \frac{1}{4}, \quad P(\xi = x_3) = \frac{1}{4},$$

则分布列似乎应为:

$$\left\{ \left(0, \frac{1}{2}\right), \left(1, \frac{1}{4}\right), \left(0, \frac{1}{4}\right) \right\}.$$

但这样写只会引起混乱, 让你不知道 $P(\xi = 0)$ 的概率到底是多少. 所以要合并同类项, 写成

$$\left\{ \left(0, \frac{3}{4}\right), \left(1, \frac{1}{4}\right) \right\}.$$

分布列包含 ξ 的可能取值 x_1, x_2, \dots 和分别取这些值的概率 p_1, p_2, \dots 两方面的信息. 显然,

$$p_k > 0, \quad \forall k \geq 1, \quad \text{且} \quad \sum_k p_k = 1.$$

反之, 对所有满足上面两个性质的 p_k , 结合一个可列点集 $\{x_1, x_2, \dots\}$, 都可以构造一个概率空间和定义在其上的随机变量 ξ , 使得它以 $\{(x_k, p_k), k = 1, 2, \dots\}$ 为分布列.

设 ξ 的分布列为 $\{(x_k, p_k), k = 1, 2, \dots\}$, 那么其分布函数为

$$F(x) = \sum_{k: x_k \leq x} p_k,$$

所以由分布列可以唯一地确定分布函数. 反之, 若 ξ 为离散分布, 值域为 $\{x_k\}$, 分布函数为 F . 则

$$p_k = P(\xi = x_k) = F(x_k) - F(x_k -).$$

因此 $\{(x_k, p_k)\}$ 为分布列. 所以由分布函数也可唯一地确定分布列.

因此我们得到:

命题 4.3.1. 对离散分布而言, 分布函数和分布列是相互唯一确定的.

直观上, 离散型分布函数是一个阶梯函数. 但因为可列集 $\{x_1, x_2, \dots\}$ 的结构可以非常复杂, 所以对应的分布函数也可以非常复杂, 见例2.9.

我们来看几个离散型分布的重要例子.

1. Bernoulli分布.

设 $p \in (0, 1)$. 若 ξ 的分布函数为

$$F(x) = \begin{cases} 0, & x < 0, \\ q, & 0 \leq x < 1, \\ 1, & x \geq 1, \end{cases}$$

$q := 1 - p$, 则称 ξ 服从参数为 p 的Bernoulli分布, 记为 $\xi \sim B(p)$. 此时 ξ 的分布列为

$$(0, q), (1, p).$$

2. 二项分布

设 $p \in (0, 1)$, $q := 1 - p$. 若 ξ 的分布列为 $\{(k, p_k), k = 0, \dots, n\}$, 其中

$$p_k := C_n^k p^k q^{n-k},$$

则称 ξ 服从二项分布, 记为 $\xi \sim B(n, p)$.

显然, $B(1, p) = B(p)$.

3. Poisson分布.

设 $\lambda > 0$. 若 ξ 的分布列为 $\{(k, p_k), k = 0, 1, \dots\}$, 其中

$$p_k := \frac{\lambda^k}{k!} e^{-\lambda},$$

则称 ξ 服从 Poisson 分布, 记为 $\xi \sim P(\lambda)$.

4. 几何分布.

设 $p \in (0, 1)$, $q := 1 - p$. 若 ξ 的分布列为 $\{(k, p_k), k = 1, 2, \dots\}$, 其中

$$p_k := pq^{k-1},$$

则称 ξ 服从几何分布, 记为 $\xi \sim Ge(p)$.

设 F 是随机变量 ξ 的分布函数. 令

$$A := \{x : \Delta F(x) := F(x) - F(x-) > 0\}.$$

由上一节的习题, ξ 是离散型随机变量的充要条件是

$$\sum_{x \in A} \Delta F(x) = 1.$$

此时 ξ 的值域即 A .

与此对应的情况是 A 为空集, 此时 F 是连续函数. 在这种情况下, 有一种更加特别也更加引人注目的情况, 即 F 是绝对连续函数, 亦即它可以表示为某个可积函数的不定积分. 我们给这些随机变量一个名字.

定义 4.3.2. 如果随机变量 ξ 的分布函数 $F(x)$ 满足

$$F(x) = \int_{-\infty}^x p(t) dt, \quad \forall x \in \mathbb{R}, \quad (3.1)$$

其中 p 是非负 Riemann¹ 可积函数,² 则称 $F(x)$ 为连续型分布, ξ 为连续型随机变量, p 为 ξ 的或 F 的密度函数(或分布密度), 简称为密度.

注1. 本书涉及到的密度函数都是分段连续函数, 所以都是 Riemann 可积的.

注2. 在 p 的连续点上, 有 $p(x) = F'(x)$. 所以在这些点上, p 由 F 唯一确定.

注3. 显然, 改变函数 p 在有限个点上的值并不会影响等式(3.1); 更进一步, 学了实变函数后就知道, 在任何一个 Lebesgue 零测集上改变 p 的值都不会影响这个等式, 所以密度函数 p 不是绝对“唯一的”, 而只是在几乎处处相等意义下的唯一.

注4. 密度函数的直观意义是: 当 $b - a > 0$ 很小, $x \in [a, b]$ 且 x 是 p 的连续点时, 有

$$P(\xi \in [a, b]) \approx p(x)(b - a).$$

这里之所以强调 x 是 p 的连续点, 是因为 p 在任意一个单点上的值都是可以随意改变的, 因此可以取一个与 ξ 毫无关系的值; 但如果限定 p 在 x 处连续, 则 $p(x)$ 便唯一确定, 无法改变了.

¹Georg Friedrich Bernhard Riemann, 1826-1866, 德国数学家.

²等你学了实变函数之后, 你可以把 Riemann 替换为 Lebesgue.

显然, 密度函数满足

$$\int_{-\infty}^{\infty} p(x) dx = 1;$$

而有必要时在有限个点上改变 p 的值, 我们还可以假定它满足

$$p(x) \geq 0, \quad \forall x \in \mathbb{R}.$$

反之, 给定满足上面两个性质的函数 p , 总可以通过等式(3.1)定义函数 F , 且 F 为一连续型分布函数.

所以, 如果哪个人声称 p 是某随机变量的密度函数, 那么他必须说明 p 满足上面两条性质, 否则就是瞎说.

我们来看几个连续型分布的例子.

1. 指数分布

小星星, 眨眼睛, 你是什么小精灵? 设小星星在遥远的深空眨着眼睛, 小朋友在地面捕捉它们. 假设他在两个不同的时间段内捕捉到多少颗星星的事件是相互独立的, 且捕捉到多少星星的概率只与时间的长度有关, 而与具体的起止点无关. 我们想知道他为捕捉到第一个星星需要等待的时间的概率分布.

若以 $\xi(t)$ 记在 $[0, t]$ 内捕捉到的星星颗数, 那么 $\xi(s+t) - \xi(t)$ 就是在 $[t, t+s]$ 内捕捉到的星星数. 注意这里假设的两个特征, 所以

(i) $\forall k \in \mathbb{N}_{++}, \forall 0 = t_0 < t_1 < t_2 < \cdots < t_k, n_1, n_2, \cdots, n_k \in \mathbb{N}_+$, 事件

$$\{\xi(t_i) - \xi(t_{i-1}) = n_i\}, \quad i = 1, 2, \cdots, k$$

相互独立.

(ii) $\forall m, P(\xi(s+t) - \xi(t) = m)$ 只与 s 有关而与 t 无关.

我们将第一个特征称为独立增量性, 而把第二个特征称为平稳性.

现在我们对找到第一颗星星的时间感兴趣. 以 $A(s, t)$ 记 $[s, t]$ 内没有捕捉到任何星星这个事件. 即

$$A(s, t) := \{\xi(t) - \xi(s) = 0\}.$$

则

$$A(0, s+t) = A(0, s)A(s, s+t), \quad \forall s, t \geq 0.$$

令

$$\varphi(s, t) := P(A(s, t)).$$

则由独立增量性有 $\varphi(0, s+t) = \varphi(0, s)\varphi(s, s+t)$, 而由平稳性有 $\varphi(s, s+t) = \varphi(t)$. 因此, 若令 $\varphi(t) := \varphi(0, t)$, 则有

$$\varphi(s+t) = \varphi(s)\varphi(t).$$

于是由后面的引理知存在 $\lambda \geq 0$ 使得

$$\varphi(t) = e^{-\lambda t}.$$

因此

$$P(A(0, t)) = e^{-\lambda t}.$$

这就是说, 若以 τ_1 表示捕捉到第一个星星的时刻, 那么

$$P(\tau_1 \geq t) = e^{-\lambda t}.$$

于是

$$P(\tau_1 > t) = \lim_{n \rightarrow \infty} P(\tau \geq t + n^{-1}) = e^{-\lambda t}.$$

从而

$$P(\tau_1 \leq t) = 1 - e^{-\lambda t}.$$

所以 τ_1 的分布是一个连续型分布, 相应的密度函数为

$$p(t) = \begin{cases} \lambda e^{-\lambda t}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

称为参数为 λ 的指数分布, 记为 $E(\lambda)$.

我们还要陈述以下事实:

记 τ_2 为捕捉到第二个星星的时刻. 由于是重新开始了捕捉, 所以 $\tau_2 - \tau_1$ 与 τ_1 独立, 且与 τ_1 同分布. 依次下去, 设 τ_n 时捕捉到第 n 个星星的时刻, 则 $\tau_n - \tau_{n-1}$ 与 τ_1 同分布, 且

$$\tau_1, \tau_2 - \tau_1, \dots, \tau_n - \tau_{n-1}, \dots$$

相互独立.

这些事实的严格证明需要用到更深刻随机过程理论, 超出了本课程的范围, 这里可先从直观上理解. 不过这里还是建议大家不妨思考一下, 到底需要什么样的理论才能给出它们的严格证明, 也许你能自己琢磨出些道道, 甚至建立一点理论, 创造出几个工具?——这比你不想直接去读别人的东西可能会有用一些, 你所受到的训练也可能会多一些, 毕竟任何理论和工具都是人创建的, 不是天上掉下来的.

那么, τ_n 服从什么分布? 这也是一个目前对我们稍难的问题, 但后面我们会知道, 通过以上几条性质的确是可以求出 τ_n 的分布的.

注意, 若天上根本没有星星, 那么 $\varphi \equiv 1$, 这对应着 $\lambda = 0$ 的情况.

作为推导指数分布的引子, 我们取了找星星的小朋友作为例子. 但事实上这个模型适用于一切与排队等待有关的问题. 比如说网店的客服平台接到第一个电话的时间, 地铁站第一个到来的乘客的时间, 保险公司收到的第一份保单的时间, 去银行排队需要等待的时间等等, 大体上都是服从指数分布的.

现在我们证明前面提到的引理.

引理 4.3.3. 设 $\varphi: \mathbb{R}_+ \mapsto \mathbb{R}_+$ 单调下降, 且满足 $\varphi(s+t) = \varphi(s)\varphi(t)$, $\forall s, t > 0$. 则存在 $\lambda \geq 0$ 使得

$$\varphi(t) = e^{-\lambda t}, \quad \forall t \geq 0.$$

证明. 对任意 $m, n \in \mathbb{N}_{++}$ 有

$$\varphi\left(\frac{n}{m}\right)^m = \varphi(n) = \varphi(1)^n.$$

所以

$$\varphi\left(\frac{n}{m}\right) = \varphi(1)^{\frac{n}{m}}.$$

因此, 令 $\varphi(1) = a$, 则有

$$\varphi(r) = a^r, \quad \forall r \in \mathbb{Q}_+.$$

由于 φ 单调下降, 故 $a \leq 1$, 因此有 $\lambda \geq 0$ 使 $a = e^{-\lambda}$. 再由 φ 的单调下降性有

$$\varphi(r_1) \leq \varphi(t) \leq \varphi(r_2), \quad \forall t \geq 0, \quad \forall r_1, r_2 \in \mathbb{Q}_+, r_1 \geq t \geq r_2.$$

令 $r_1 \uparrow t, r_2 \downarrow t$ 有

$$e^{-\lambda t} \leq \varphi(t) \leq e^{-\lambda t}.$$

所以 $\varphi(t) = e^{-\lambda t}, \forall t \geq 0$. □

2. 正态分布

设想你抛一枚硬币, 得到正面你记录下1, 得到反面你记录下-1. 进行100次后你将得到的数字加起来; 然后你重复这个试验1000次. 将最后得到的1000个数字放在一起观察一下, 你预料会看到什么情况? 你会看到大量的0? 大量的100, 还是大量的-100?

高尔顿做过这个试验.³ 当然他不是用抛硬币的方法, 而是从一块均匀地钉满了钉子的板子顶部中央放下一个小球, 小球下落时在每一层都会碰到一个钉子, 于是都有向左向右两种可能, 且可能性都是二分之一. 他想要观察的是, 放下大量的小球之后, 板子上的小球会堆积成什么形状?

现在在网上可用“高尔顿钉板”搜索到这个试验的动画演示. 小球最后会堆积成钟状样——不是现在的电子钟, 而是像电影《地道战》中高家庄里挂在村口的古老的大钟.

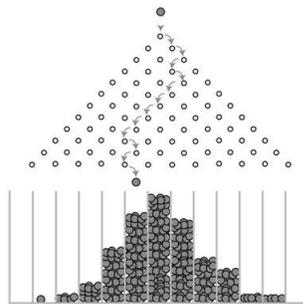


图 4.1: 高尔顿钉板试验示意图

现在我们要对这个试验进行严格的数学分析, 来说明形成这样的形状不是偶然的, 而是必然的.

我们需要用到下面的分析结果.

引理 4.3.4. 设 f 是定义在 \mathbb{R} 上的连续函数, 且

$$\int_{\mathbb{R}} |f(x)| dx < \infty,$$

³Francis Galton, 1822年2月16日-1911年1月17日, 英国生物统计学家, 达尔文表弟.

定义其Fourier变换:

$$\hat{f}(t) := \int_{\mathbb{R}} f(x)e^{itx} dx.$$

若

$$\int_{\mathbb{R}} |\hat{f}(t)| dt < \infty,$$

则

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{f}(t)e^{-itx} dt.$$

这个公式的证明在许多书中都可以找到, 例如[3, Ch. 2, Sect. 6], [11, Ch.III, Sect.2], [2, 第二章第四节].

注: 问题是, 这里的条件

$$\int_{\mathbb{R}} |\hat{f}(t)| dt < \infty$$

什么时候可以满足? 一个简单的充分条件是: 当 f 二次连续可微且在某有界集外恒为零时, 因为此时由分部积分易见

$$|\hat{f}(t)| \leq \frac{C}{1+t^2}.$$

下面这个漂亮的结果对整个概率论, 无论初等还是高等, 都非常重要. 它说的是, 忽略掉一个常数因子之后, $e^{-\frac{x^2}{2}}$ 的Fourier变换就是它自己——这在所有的函数中是独一无二的.

引理 4.3.5.

$$\sqrt{2\pi}e^{-\frac{t^2}{2}} = \int_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} dx, \forall t \in \mathbb{R}.$$

证明. 先证

$$\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}. \quad (3.2)$$

以 I 表示上述积分, 则

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2+y^2}{2}} dx dy \\ &= \int_{-\infty}^{\infty} dr \int_0^{2\pi} e^{-\frac{r^2}{2}} r d\theta \\ &= 2\pi \int_0^{\infty} e^{-r} dr \\ &= 2\pi. \end{aligned}$$

(3.2)得证.

于是, 对 $t \in \mathbb{R}$, 有

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} e^{tx} dx = e^{\frac{t^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}} dx = e^{\frac{t^2}{2}}.$$

由于上式两端都是 t 的解析函数, 故对任意 $z \in \mathbb{C}$ 均有

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} e^{zx} dx = e^{\frac{z^2}{2}}.$$

取 $z = it$ 即得结果. □

现在我们可以对高尔顿钉板试验做严格的数学分析了.

小球的每一次下落, 都相当于左移或右移了一步, 因此可理解为一个取值于 $\{1, -1\}$ 的随机变量. 这样, 以 ξ_n 表示小球在第 n 步的位移, 则 ξ_1, ξ_2, \dots 是独立同分布随机变量列, 且

$$P(\xi_1 = 1) = P(\xi_1 = -1) = \frac{1}{2}.$$

我们想看看对大的 n , 其和即小球现在的位置

$$S_n := \sum_{i=1}^n \xi_i$$

的分布呈何种形态.

直接计算可知

$$E[S_n^2] = n,$$

所以 S_n 的体量会越来越大. 由于只关心形态, 故我们先对它做一个相似变换, 即令

$$\zeta_n := \frac{1}{\sqrt{n}} S_n.$$

则 $E[\zeta_n^2] \equiv 1$. 这样 ζ_n 的体量就保持不变了, 我们也得以能集中研究其形态.

令

$$\varphi_n(t) := E[\exp\{it\zeta_n\}].$$

由独立性有

$$\begin{aligned} \varphi_n(t) &= \frac{1}{2^n} \left(\exp\left\{\frac{it}{\sqrt{n}}\right\} + \exp\left\{\frac{-it}{\sqrt{n}}\right\} \right)^n \\ &= \cos^n\left(\frac{t}{\sqrt{n}}\right). \end{aligned}$$

令 $n \rightarrow \infty$, 注意到

$$\cos^n\left(\frac{t}{\sqrt{n}}\right) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n,$$

得

$$\lim_{n \rightarrow \infty} \varphi_n(t) = \exp\left(-\frac{t^2}{2}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} dx.$$

设 H 为二次连续可微函数, 且在某个有界集外恒为零, 由引理4.3.4后面的注知

$$\int_{\mathbb{R}} |\hat{H}(t)| dt < \infty.$$

于是由引理4.3.4与引理4.3.5有

$$\begin{aligned} \lim_{n \rightarrow \infty} E[H(\zeta_n)] &= \frac{1}{2\pi} \lim_{n \rightarrow \infty} E\left[\int_{-\infty}^{\infty} \hat{H}(t) e^{-it\zeta_n} dt\right] \\ &= \frac{1}{2\pi} \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \hat{H}(t) \varphi_n(-t) dt \\ &= \frac{1}{(2\pi)^{3/2}} \int_{-\infty}^{\infty} \hat{H}(t) dt \int_{-\infty}^{\infty} e^{-itx} e^{-\frac{x^2}{2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} H(x) e^{-\frac{x^2}{2}} dx. \end{aligned}$$

这里我们用到了 E 和 $\int_{-\infty}^{\infty}$ 的交换次序. 由于这里的 E 是有限和, 所以实际上是有限和与积分的交换次序, 因而是不成问题的. 现在, 我们忽略 H 需要满足的条件, 形式地取

$$H(y) := 1_{(-\infty, x]}(y),$$

就有

$$P(\zeta_n \leq x) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy.$$

当然, 上面这样取 H 是不严格的, 但它所引导的方向是正确的, 过程可以严格化. 之所以可以严格化, 是因为等式最左边的项和最右边的项对于这样的 H 都是存在的, 因此可以通过用光滑函数逼近来实现. 事实上, 可取

$$H_m(y) = \begin{cases} 0, & y \leq -m-1, \\ y+m+1, & y \in (-m-1, -m], \\ 1, & y \in [-m, x], \\ -my+1+mx, & y \in (x, x+\frac{1}{m}], \\ 0, & y > x+\frac{1}{m}. \end{cases}$$

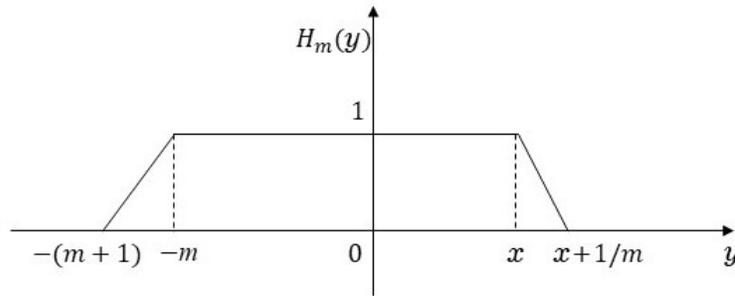


图 4.2: 函数 $H_m(y)$ 的图像

当然, 这样定义的 H_m 还是不满足需要的条件, 因为它只是一次可导的. 不过, 我们可以将它的不光滑处磨光, 就假装它是二次可导的(这道手续可以严格化, 见附录), 因此公式是成立的. 然后令 $m \rightarrow \infty$, 就得到等式对 H 成立. 至于为什么可以令 $m \rightarrow \infty$, 这是由后面将要证明的控制收敛定理保证的, 你先就这样用着. 所以当 n 很大时, 差不多有

$$P(\zeta_n \leq x) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du.$$

事实上, 使用更高超的技巧还可以证明, 不仅对以上特殊的独立随机变量列如此, 对更一般的独立随机变量列, 这个结论也是成立的.

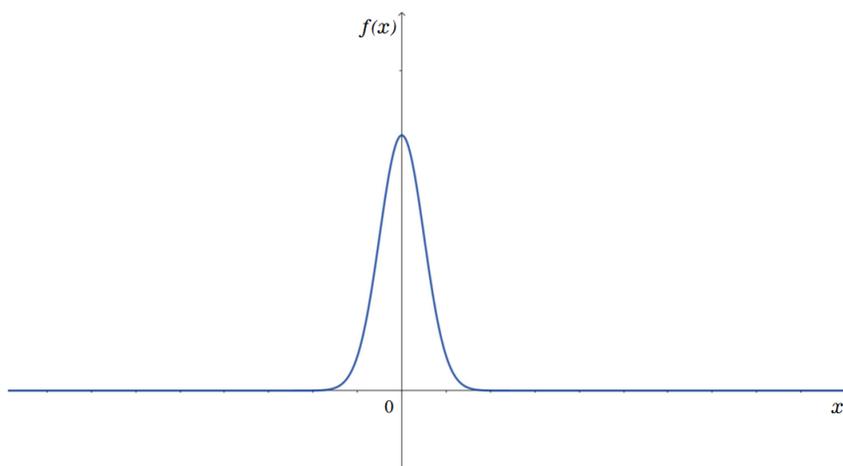
令

$$p(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

这个函数的图像正是大钟形的! (见图4.3)

由(3.2)知

$$\int_{-\infty}^{\infty} p(x) dx = 1.$$

图 4.3: 函数 $f(x)$ 的图像

所以 f 确定了一个以它为密度函数的分布 F , 而 ζ_n 的分布近似于这个 F , 当 n 很大时. 我们以后会知道, 这个 F 对相当大的一类随机变量列是其公共的不变的所谓弱极限, 所以人们认为只有这种情况是正常的, 其它情况都是异常的, 也因此这个分布也就自然应叫做正常分布, 即 normal distribution. 不过传统上, 其中译作为正式的学术名词, 它被书面化文雅化与故弄玄虚化了, 称为正态分布 (台湾称为常态分布).

一般地, 设 $\sigma \in \mathbb{R}$, $\sigma > 0$, 密度函数为

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

的分布称为参数为 μ, σ^2 的正态分布, 记为 $N(\mu, \sigma^2)$. 称 $N(0, 1)$ 为标准正态分布.

理论上我们可以并将证明任何一个分布函数都是某个随机变量的分布函数, 然而应用中最常见也最便于处理的要么是离散型的, 要么是连续型的.

4.4 多维随机变量的分布函数

对多维随机变量, 同样可以考虑其分布函数. 我们以二维为例, 更高维的情况类似.

定义 4.4.1. 设 ξ, η 均为实值随机变量. 令

$$F(x, y) := P(\xi \leq x, \eta \leq y),$$

称为 (ξ, η) 的分布函数. 必要时 F 可记为 $F_{\xi, \eta}$ 以宣示主权.

相应于一维时的 $F(\infty) = 1$, 我们有

$$\begin{aligned}
 F(\infty, \infty) &:= \lim_{x \rightarrow \infty, y \rightarrow \infty} F(x, y) \\
 &= \lim_{n \rightarrow \infty, m \rightarrow \infty} F(n, m) \\
 &= \lim_{n \rightarrow \infty, m \rightarrow \infty} P(\xi \leq n, \eta \leq m) \\
 &= P\left(\bigcup_{m, n=1}^{\infty} \{\xi \leq n, \eta \leq m\}\right) \\
 &= P((\xi, \eta) \in \mathbb{R}^2) \\
 &= 1.
 \end{aligned}$$

而 $\forall x \in \mathbb{R}$,

$$\begin{aligned}
 F(x, \infty) &:= \lim_{y \rightarrow \infty} F(x, y) \\
 &= \lim_{m \rightarrow \infty} F(x, m) \\
 &= \lim_{m \rightarrow \infty} P(\xi \leq x, \eta \leq m) \\
 &= P\left(\bigcup_{m=1}^{\infty} \{\xi \leq x, \eta \leq m\}\right) \\
 &= P(\xi \leq x, \eta \in \mathbb{R}) \\
 &= P(\xi \leq x) \\
 &= F_{\xi}(x).
 \end{aligned}$$

同理, $\forall y \in \mathbb{R}$,

$$F(\infty, y) = F_{\eta}(y).$$

以上这两个性质在直观上是明显的, 因为比如 $F(x, \infty)$ 意味着对 η 的取值没有任何限制, 而只是限制了 ξ 的值不能超过 x ; 另一方面, $F(x, -\infty)$ 则意味着 η 取任何值都不行, 是不可能事件, 所以其概率为零. 严格写出来就是: $\forall x \in \mathbb{R}$,

$$\begin{aligned}
 F(x, -\infty) &:= \lim_{y \rightarrow -\infty} F(x, y) \\
 &= \lim_{m \rightarrow -\infty} F(x, m) \\
 &= \lim_{m \rightarrow -\infty} P(\xi \leq x, \eta \leq m) \\
 &= P\left(\bigcap_{m=1}^{\infty} \{\xi \leq x, \eta \leq -m\}\right) \\
 &= P(\xi \leq x, \eta = -\infty) \\
 &\leq P(\eta = -\infty) \\
 &= 0.
 \end{aligned}$$

同理, $\forall y \in \mathbb{R}$,

$$F(-\infty, y) = 0.$$

由于 $\forall x$ 及 $\forall y_1 \leq y_2$,

$$F(x, y_2) - F(x, y_1) = P(\xi \leq x, \eta \in (y_1, y_2]) \geq 0,$$

所以对固定的 x , $y \mapsto F(x, y)$ 是单调上升函数, 且是右连左极的. 同理, 对固定的 y , 函数 $x \mapsto F(x, y)$ 也具有同样的性质.

不过这种单调性还不足以真正反映二维分布函数的特性: 它真正的特色单调性是, 对于任意 $x_1 \leq x_2, y_1 \leq y_2$, 有

$$\begin{aligned} & F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) \\ &= P(x_1 < \xi \leq x_2, y_1 < \eta \leq y_2) \geq 0. \end{aligned} \quad (4.3)$$

这个性质显然比固定任何一个变量时关于另一个变量的单调性更强.

多维离散型与连续型随机变量可类似定义. 我们还是以二维为例.

定义 4.4.2. 一个二维随机变量 (ξ, η) , 若其值域为 $\{(x_i, y_j), i, j = 1, 2, \dots\}$, 则称为离散型的. 此时, 令

$$p_{ij} := P(\xi = x_i, \eta = y_j).$$

则 $\{(x_i, y_j), p_{ij}\}$ 称为 (ξ, η) 的分布列.

例1. 多项分布

若每次试验的可能结果为 A_1, \dots, A_r , 而 $P(A_i) = p_i, i = 1, \dots, r, p_1 + \dots + p_r = 1$. 重复这个试验 n 次, 并设这 n 次试验间是相互独立的. 以 ξ_1, \dots, ξ_r 分别记 A_1, \dots, A_r 出现的次数, 则

$$P(\xi_1 = k_1, \xi_2 = k_2, \dots, \xi_r = k_r) = \frac{n!}{k_1! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r},$$

其中 $k_i \geq 0, k_1 + \dots + k_r = n$.

而多维连续性随机变量就是, 简言之, 其分布函数可以表示为另一可积函数之不定积分的随机变量. 不过, 由于现在是多元函数, 所以我们得加上一些技术性条件以保证将要涉及得各种运算的通畅性. 具体地说, 我们有:

定义 4.4.3. 若存在 $p: \mathbb{R}^2 \mapsto \mathbb{R}$ 满足

- (i) $p(x, y) \geq 0$;
- (ii) p 的间断点只出现在有限多条简单曲线上;⁴
- (iii) $\forall x, y \mapsto p(x, y)$ 分段连续, $\forall y, x \mapsto p(x, y)$ 分段连续;
- (iii) 除了有限个例外的 (u_0, v_0) 点外, 有

$$\lim_{u \rightarrow u_0} \int_{-\infty}^{\infty} |p(u, v) - p(u_0, v)| dv = 0,$$

$$\lim_{v \rightarrow v_0} \int_{-\infty}^{\infty} |p(u, v) - p(u, v_0)| du = 0,$$

使得

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y p(u, v) dudv,$$

则 F 称为连续型分布, (ξ, η) 称为连续型随机变量, p 称为 F 或 (ξ, η) 的密度函数(或分布密度).

⁴简单曲线是指有着参数方程 $x = x(t), y = y(t)(a \leq t \leq b)$ 的曲线, 其中 $x(t), y(t)$ 为连续可微函数. 见[6, p.180]

这里虽然有一些繁琐的技术性条件,但应用中 p 基本上都是连续的,即使不连续,也是很有规则地不连续,比如在一个矩形或者圆盘的边界上不连续.

当 (ξ, η) 的分布函数 F 给定时,我们有

$$F_{\xi}(x) = P(\xi \leq x, \eta \in \mathbb{R}) = F(x, \infty),$$

$$F_{\eta}(y) = P(\xi \in \mathbb{R}, \eta \leq y) = F(\infty, y).$$

因此 ξ 与 η 的分布也都确定下来了. F_{ξ} 与 F_{η} 于是称为 F 的两个边沿分布.相对应的, (ξ, η) 本来的分布 F 也往往称为联合分布,以示强调 ξ 与 η 是放在一起考虑的.

若 (ξ, η) 是离散型的,分布列为 $\{(x_i, y_j), p_{ij}\}$,那么 ξ 的值域就是 $\{x_i\}$,且

$$p_i^{\xi} := P(\xi = x_i) = \sum_j P(\xi = x_i, \eta = y_j) = \sum_j p_{ij}.$$

所以 ξ 也是离散型的,且分布列为 $\{(x_i, p_i^{\xi})\}$.同理 η 也是离散型的,且分布列为 $\{(y_j, p_j^{\eta})\}$,其中

$$p_j^{\eta} := \sum_i p_{ij}.$$

若 (ξ, η) 为连续型的,密度函数为 f .则 $\forall x$,

$$\begin{aligned} F_{\xi}(x) &= P(\xi \leq x) \\ &= P(\xi \leq x, \eta < \infty) \\ &= \int_{-\infty}^x \int_{-\infty}^{\infty} p(u, v) du dv \\ &= \int_{-\infty}^x du \int_{-\infty}^{\infty} p(u, v) dv. \end{aligned}$$

所以 ξ 也是连续型的,且密度函数为

$$p_{\xi}(x) = \int_{-\infty}^{\infty} p(x, v) dv.$$

同理, η 也是连续型的,且密度函数为

$$p_{\eta}(y) = \int_{-\infty}^{\infty} p(u, y) du.$$

下面看一些例子.

例2. 多维正态分布

这是最重要的多维连续型分布,没有之一.

我们知道,一维正态分布有两个参数, σ^2 和 μ .而在多维时,代替正数 σ^2 的是一个正定对称方阵,代替 μ 的是一个向量.

设 $\Sigma = (\sigma_{ij})$ 为 n 阶正定对称方阵, $\mu = (\mu_1, \dots, \mu_n)$ 为 n -维行向量.记

$$\Sigma^{-1} = (\gamma_{ij}).$$

则以

$$\begin{aligned} p(x_1, \dots, x_n) &= \frac{1}{(2\pi)^{n/2} (\det \Sigma)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} \sum_{j,k=1}^n \gamma_{jk} (x_j - \mu_j)(x_k - \mu_k) \right\} \\ &= \frac{1}{(2\pi)^{n/2} (\det \Sigma)^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (x - \mu) \Sigma^{-1} (x - \mu)' \right\} \end{aligned}$$

为密度的分布称为 n 维正态分布 $N(\mu, \Sigma)$, 其中“ $'$ ”表示向量或矩阵的转置. $N(0, I)$ 称为 n 维标准正态分布, 其中 I 为 n 阶单位矩阵.

将 $N(0, I)$ 的密度函数记为 φ , 则

$$\varphi(x_1, \dots, x_n) = \tilde{\varphi}(x_1) \cdots \tilde{\varphi}(x_n).$$

其中 $\tilde{\varphi}$ 是一维标准正态分布 $N(0, 1)$ 的密度函数. 因此有

$$\int_{\mathbb{R}^n} \varphi(x_1, \dots, x_n) dx_1 \cdots dx_n = \left(\int_{-\infty}^{\infty} \tilde{\varphi}(t) dt \right)^n = 1.$$

对一般的 $N(\mu, \Sigma)$, 是否也有

$$\int_{\mathbb{R}^n} p(x) dx = 1?$$

答案是肯定的. 和一维时一样, 也是将一般情形转化为标准情形证明, 兹给出如下:
设 $\lambda_i^2 > 0, i = 1, 2, \dots, n$ 为 Σ 的特征根, 而 A 为正交阵, 使得

$$\Sigma = A \begin{pmatrix} \lambda_1^2 & 0 & \cdots & 0 \\ 0 & \lambda_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n^2 \end{pmatrix} A'.$$

令

$$y := (x - \mu)A \begin{pmatrix} \lambda_1^{-1} & 0 & \cdots & 0 \\ 0 & \lambda_2^{-1} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n^{-1} \end{pmatrix}.$$

得

$$\int_{\mathbb{R}^n} p(x) dx = \int_{\mathbb{R}^n} \varphi(y) dy = 1,$$

其中 φ 是 $N(0, I)$ 的密度.

特别地, 设 (ξ, η) 服从二维正态分布 $N(\mu, \Sigma)$, 其中 $\mu = (\mu_1, \mu_2)$,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},$$

其中 $\sigma_1, \sigma_2 > 0, \rho \in (-1, 1)$. 则 (ξ, η) 的密度函数写成分量形式为

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \cdot \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right] \right\}.$$

此时, 也记为 $(\xi, \eta) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. 图4.4是 $N(5, 5, 1, 1, 0.5)$ 的密度函数图像.

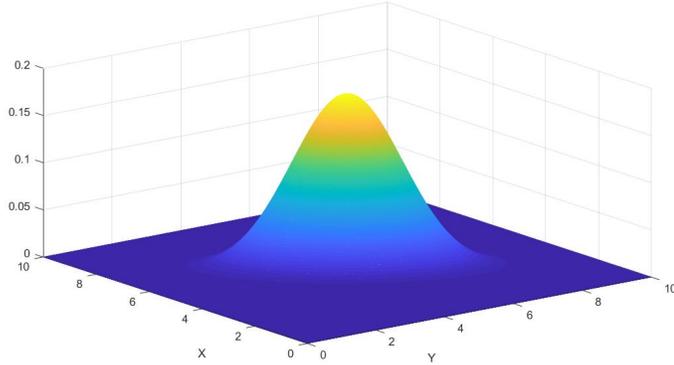


图 4.4: 二维正态分布密度函数的图像

现在求其边沿分布. 为此将 $p(x, y)$ 改写为

$$\begin{aligned}
 p(x, y) &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\} \\
 &\quad \cdot \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_2} \exp\left\{-\frac{\left[y - \left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right)\right]^2}{2\sigma_2^2(1-\rho^2)}\right\} \\
 &= \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\} \\
 &\quad \cdot \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_1} \exp\left\{-\frac{\left[x - \left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)\right)\right]^2}{2\sigma_1^2(1-\rho^2)}\right\}. \tag{4.4}
 \end{aligned}$$

由此易得

$$\begin{aligned}
 p_\xi(x) &= \int_{\mathbb{R}} p(x, y) dy = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, \\
 p_\eta(y) &= \int_{\mathbb{R}} p(x, y) dx = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}.
 \end{aligned}$$

这就是说, 正态分布的边沿分布仍为正态分布, $\xi \sim N(\mu_1, \sigma_1^2)$ 和 $\eta \sim N(\mu_2, \sigma_2^2)$. 注意这两个边沿分布都与 ρ 无关, 这就顺带举出了一个不同的联合分布有相同的边沿分布的例子.

然而, ρ 的意义呢? (4.4)中的另一项是什么呢? 我们要等等才能知道.

类似地, 多维正态分布的边沿分布也是正态分布. 但若还用上面的计算就比较复杂, 从后面的特征函数角度来看这就是显然的结论.

习题

1. 设 τ_1, \dots, τ_n 为独立同分布随机变量, 均服从参数为 λ 的指数分布. 证明 $\xi_n := \tau_1 + \dots + \tau_n$ 为连续型, 密度函数为

$$\frac{\lambda^n t^{n-1}}{(n-1)!} e^{-\lambda t}.$$

2. 证明:

$$\lim_{n \rightarrow \infty} \cos^n\left(\frac{t}{\sqrt{n}}\right) = e^{-\frac{t^2}{2}}.$$

3. 证明:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} dx = e^{-\frac{t^2}{2}}.$$

4. 举例说明, 当 (ξ, η) 与 (α, β) 有相同的边沿分布时, 不一定会有相同的联合分布.

5. 证明: 当联合分布是离散时, 边沿分布也是离散的; 当联合分布是连续时, 边沿分布也是连续的.

6. 设 ξ 服从 $[0, 1]$ 上的均匀分布, $\eta := \xi$.

(a) 求 (ξ, η) 的联合分布 $F(x, y)$, 并证明 F 是 (x, y) 的连续函数;

(b) 证明 (ξ, η) 不是连续型随机变量.

7. 写出对应于(4.3)的多维情形的公式并证明之.

8. 写出多维离散型与连续型随机变量的定义.

9. 设 ξ 服从 n -维正态分布, A 为 $n \times n$ 矩阵. 证明 ξA 也服从 n -维正态分布.

10. 对任意 $x \in \mathbb{R}^n$, 以 $(r(x), \theta_1(x), \dots, \theta_{n-1}(x))$ 表示 x 的球坐标. 设 ξ 服从 n -维标准正态分布, 求 $(r(\xi), \theta_1(\xi), \dots, \theta_{n-1}(\xi))$ 的分布. (建议: 先考虑 $n = 2$ 的情形?)

11. 设 ξ 为随机变量, $c \geq 0$ 为常数且 $c \neq 1$. 证明: 若 ξ 与 $c\xi$ 同分布, 则 $\xi \equiv 0$.

12. 设 F 是 n -维分布函数, F_k 是其第 k 个边沿分布.

(a) 设 a_k 是 F_k 的连续点, 证明: 对任意固定的 $x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_n$, a_k 是函数

$$x_k \mapsto F(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_n)$$

的连续点;

(b) 若 $a = (a_1, \dots, a_n)$, 且 a_k 为 F_k 的连续点, $\forall k = 1, \dots, n$, 则称 a 为 F 的连续点. 证明: 若 a 为 F 的连续点, 则

$$F(a) = P(\xi_1 \prec a_1, \dots, \xi_n \prec a_n),$$

其中 \prec 可随意取为 $<$ 或 \leq ;

(c) 若 $a = (a_1, \dots, a_n)$, $b = (b_1, \dots, b_n)$, $a \leq b$ (即 $a_k \leq b_k, \forall k$), 且 a, b 为 F 的连续点, 证明

$$P(a_1 \prec \xi_1 \prec b_1, \dots, a_n \prec \xi_n \prec b_n) = F((a, b]) := P(\in (\xi_1, \dots, \xi_n)(a, b]),$$

其中 \prec 可随意取为 $<$ 或 \leq .

4.5 条件分布

设 η 是随机变量, A 是事件, 则比对着条件概率, 给定 A 时 η 的条件分布很自然定义为

$$P(\eta \leq y|A) := \frac{P(\{\eta \leq y\} \cap A)}{P(A)},$$

只要 $P(A) > 0$. 受此启发, 假设 ξ 是另一个随机变量, 取 $A = \{\xi = x\}$, 则给定 $\xi = x$ 时 η 的条件分布理应定义为

$$P(\eta \leq y|\xi = x) = \frac{P(\{\eta \leq y\} \cap \{\xi = x\})}{P(\xi = x)}. \quad (5.5)$$

然而 $P(\xi = x)$ 极有可能等于零(例如当 ξ 为连续型时), 所以不能一概笼统地这样定义, 需要区分情况讨论.

一. 离散情形.

设 (ξ, η) 是离散型的, 分布列为 $\{(x_i, y_j), p_{ij}\}$. 则定义

$$p_{ji}^{\eta|\xi} := P(\eta = y_j|\xi = x_i) = \frac{P(\eta = y_j, \xi = x_i)}{P(\xi = x_i)} = p_{ij} (p_i^\xi)^{-1},$$

称为给定 $\xi = x_i$ 时 η 的条件分布列. 类似地,

$$p_{ij}^{\xi|\eta} := P(\xi = x_i|\eta = y_j) = \frac{P(\xi = x_i, \eta = y_j)}{P(\eta = y_j)} = p_{ij} (p_j^\eta)^{-1}$$

称为给定 $\eta = y_j$ 时 ξ 的条件分布列.

此时皆无分式的分母有可能为零的问题.

二. 连续情形.

设 (ξ, η) 是连续型随机变量, 分布密度为 $p(x, y)$. 此时(5.5)中右端分式的分母为零, 因此需要通过极限定义其值.

$$\begin{aligned} P(\eta \leq y|\xi = x) &:= \lim_{\varepsilon \downarrow 0} P(\eta \leq y|x - \varepsilon \leq \xi < x + \varepsilon) \\ &= \lim_{\varepsilon \downarrow 0} \frac{P(\eta \leq y, x - \varepsilon < \xi < x + \varepsilon)}{P(x - \varepsilon < \xi < x + \varepsilon)} \\ &= \lim_{\varepsilon \downarrow 0} \frac{\int_{x-\varepsilon}^{x+\varepsilon} du \int_{-\infty}^y p(u, v) dv}{\int_{x-\varepsilon}^{x+\varepsilon} du \int_{-\infty}^{\infty} p(u, v) dv} \\ &= \lim_{\varepsilon \downarrow 0} \frac{(2\varepsilon)^{-1} \int_{x-\varepsilon}^{x+\varepsilon} du \int_{-\infty}^y p(u, v) dv}{(2\varepsilon)^{-1} \int_{x-\varepsilon}^{x+\varepsilon} du \int_{-\infty}^{\infty} p(u, v) dv}. \end{aligned}$$

当 x 满足 $p_\xi(x) > 0$ 且

$$\lim_{u \rightarrow x} \int_{-\infty}^{\infty} |p(u, v) - p(x, v)| dv = 0$$

时(只有有限个点不满足此式), 上述极限等于

$$\frac{\int_{-\infty}^y p(x, v) dv}{p_\xi(x)},$$

称为给定 $\xi = x$ 时, η 的条件分布函数. 所以, 给定 $\xi = x$ 时, η 的条件密度定义为

$$p_{\eta|\xi}(y|x) := \begin{cases} \frac{p(x,y)}{p_\xi(x)}, & p_\xi(x) \neq 0, \\ 0, & p_\xi(x) = 0. \end{cases}$$

其中 p_ξ 为 ξ 的边沿密度.

细心的读者会产生问题: 为什么在 $p_\xi(x) = 0$ 时, 要将 $p_{\eta|\xi}(y|x)$ 定义为恒等于零? 这是因为我们假定了 $y \mapsto p(x, y)$ 是分段连续的, 而

$$p_\xi(x) = \int_{-\infty}^{\infty} p(x, y) dy,$$

所以若 $p_\xi(x) = 0$, 那么当 y 是 $p(x, \cdot)$ 的连续点时, 必有 $p(x, y) = 0$. 但函数 $p(x, \cdot)$ 只有有限个不连续点, 所以它只可能在有限个点上不等于零. 再注意到在有限个点上改变 p 的值是不会改变它密度函数的身份的, 所以我们可以认为对所有的 y 均有 $p(x, y) = 0$. 而对条件密度而言, 真正重要的是等式

$$p(x, y) = p_{\eta|\xi}(y|x)p_\xi(x)$$

要成立. 既然 $p(x, y)$ 与 $p_\xi(x)$ 都等于零了, 所以实际上 $p_{\eta|\xi}(y|x)$ 定义为任何数都是可以的, 但为了方便起见, 就让它等于零好了.

注. 从逻辑的角度看, 条件分布是通过联合分布定义的, 因此是先有联合分布再有条件分布. 但实际情况是复杂的, 条件分布的定义本身的合理性在于它正确地反映了联合分布和条件分布之间应该满足的关系, 而很多时候, 尤其是在建立具体问题的概率模型时候, 是先确定了条件分布列, 然后依据这个关系确定联合分布的, 即通过

$$p_{ij} = p_{ij}^{\eta|\xi} p_i^\xi \quad \text{或} \quad p_{ij} = p_{ji}^{\xi|\eta} p_j^\eta$$

得到 p_{ij} , 或者是通过

$$p(x, y) = p_{\eta|\xi}(y|x)p_\xi(x)$$

得到 $p(x, y)$ 的. 而有了联合分布之后, 就可以确定更多的条件分布.

例1. 从装有黑白两球的袋中任取一只, 取黑球与白球的概率均为 $\frac{1}{2}$. 若第一次取到白球, 则终止; 若取到黑球, 则再来一次; 如此下去, 直到取到白球为止. 以 ξ 表示取到白球时所用的次数. 当 $\xi = n$ 时, 在装有 n 只黑球与 m 只白球的袋子里任取一只, 取每球的概率均等. 若取白球, 则试验中止; 若取到黑球, 则将球放回再行取球; 如此下去, 直到取到白球为止. 以 η 表示试验的第二阶段取到白球所用次数. 求 (ξ, η) 的联合分布.

解. 显然

$$p_i^\xi := P(\xi = i) = \frac{1}{2^i},$$

$$p_{ji}^{\eta|\xi} := P(\eta = j | \xi = i) = \left(\frac{i}{m+i} \right)^{j-1} \frac{m}{m+i}.$$

所以

$$p_{ij} = P(\xi = i, \eta = j) = \frac{1}{2^i} \left(\frac{i}{m+i} \right)^{j-1} \frac{m}{m+i}.$$

例2. 设 ξ, η 独立, 分布列分别为 (x_i, p_i) 与 (y_j, q_j) . 求给定 η 时 $\xi + \eta$ 的条件分布列.

解. $\xi + \eta$ 的值域显然为 $\{x_i + y_j\}$, 而

$$\begin{aligned} P(\xi + \eta = x_i + y_j | \eta = y_j) &= \frac{P(\xi + \eta = x_i + y_j, \eta = y_j)}{P(\eta = y_j)} \\ &= \frac{P(\xi = x_i, \eta = y_j)}{P(\eta = y_j)} \\ &= \frac{P(\xi = x_i)P(\eta = y_j)}{P(\eta = y_j)} \\ &= P(\xi = x_i). \end{aligned}$$

所以, 当 $z \in \{x_i + y_j, i, j = 1, 2, \dots\}$ 时

$$P(\xi + \eta = z | \eta = y_j) = P(\xi = z - y_j).$$

例3 设 $(\xi, \eta) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. 上节已经知道, $\xi \sim N(\mu_1, \sigma_1^2)$ 和 $\eta \sim N(\mu_2, \sigma_2^2)$. 回顾

$$\begin{aligned} p(x, y) &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{(x - \mu_1)^2}{2\sigma_1^2}\right\} \\ &\quad \cdot \frac{1}{\sqrt{2\pi(1 - \rho^2)}\sigma_2} \exp\left\{-\frac{\left[y - \left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)\right)\right]^2}{2\sigma_2^2(1 - \rho^2)}\right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{(y - \mu_2)^2}{2\sigma_2^2}\right\} \\ &\quad \cdot \frac{1}{\sqrt{2\pi(1 - \rho^2)}\sigma_1} \exp\left\{-\frac{\left[x - \left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2)\right)\right]^2}{2\sigma_1^2(1 - \rho^2)}\right\}. \end{aligned}$$

因此两个条件密度为

$$\begin{aligned} p_{\eta|\xi}(y|x) &= \frac{1}{\sqrt{2\pi(1 - \rho^2)}\sigma_2} \exp\left\{-\frac{\left[y - \left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1)\right)\right]^2}{2\sigma_2^2(1 - \rho^2)}\right\}, \\ p_{\xi|\eta}(x|y) &= \frac{1}{\sqrt{2\pi(1 - \rho^2)}\sigma_1} \exp\left\{-\frac{\left[x - \left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2)\right)\right]^2}{2\sigma_1^2(1 - \rho^2)}\right\}, \end{aligned}$$

即两个边沿分布也是正态的. 具体地, 给定 $\xi = x$ 时, η 的条件分布为 $N(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x - \mu_1), \sigma_2^2(1 - \rho^2))$; 给定 $\eta = y$ 时, ξ 的条件分布为 $N(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2), \sigma_1^2(1 - \rho^2))$.

习题

1. 设 ξ, η 独立, 分布列分别为 (x_i, p_i) 与 (y_j, q_j) . 设 $f, g: \mathbb{R} \mapsto \mathbb{R}$, 且 f 是单射. 求 $(f(\xi) + g(\eta), \eta)$ 的联合分布列.

4.6 随机变量的存在性

在实际问题中,看得见摸得着的往往是分布函数,概率空间及随机变量则是在建立实际问题的概率模型时,为了数学处理的方便而出现的.所以,实际工作中一个很重要的问题是:是否对任意一个分布函数,都存在一个概率空间和定义在其上的随机变量,使得该分布正是此随机变量的分布函数?在理论上,这个问题更加重要,比如一系列随机变量的分布函数若在某种意义上收敛到一分布函数 F ,那么有没有一个概率空间足以支撑起一个随机变量,使得其分布函数就是 F ?

如果问题只是这样提,回答倒是比较简单的,因为分布函数 F 会在 $(\mathbb{R}, \mathcal{B})$ 上生成一个 Lebesgue-Stieltjes⁵ 测度 μ_F ,于是 $(\mathbb{R}, \mathcal{B}, \mu_F)$ 就是一个概率空间,且易证这上面的随机变量 $\xi(\omega) := \omega$ 的分布就是 F .

我们真正想问的是下面的问题:是否存在一个公共的概率空间,使得对任何分布函数 F ,都存在定义在该空间上以 F 为分布函数的随机变量?

问题的答案有些出人意料,即不但有这样的概率空间,而且此概率空间可以统一地取为 $[0, 1]$ 及其上的Lebesgue测度.更有甚者,不但对 \mathbb{R} 上的分布函数如此,对 \mathbb{R}^n 乃至抽象空间上的分布函数也是如此.对抽象空间我们就不证了,可见[7];对 \mathbb{R}^n 留做习题.下面我们仅对 \mathbb{R} 的情况证明.

首先我们回忆一下分布函数的定义.一个单调上升的函数 $F: \mathbb{R} \mapsto [0, 1]$,若满足

- (i) $\forall x, F$ 在 x 处右连续;
- (ii) $F(\infty) = 1, F(-\infty) = 0$,

则称为分布函数.

我们现在证明:

定理 4.6.1. 设 F 是分布函数,则在 $([0, 1], \mathcal{B}, dx)$ 上存在随机变量 ξ ,使得 ξ 的分布函数恰为 F .

证明. 我们从最简单的情形开始.

- (1) 设存在 $-\infty < x_1 < x_2 < \cdots < x_n < \infty$,及 $p_i > 0, i = 1, \cdots, n$,使得

$$\sum_{i=1}^n p_i = 1,$$

$$F(x) = \sum_{i: x_i \leq x} p_i.$$

因此所对应的随机变量——如果有这样的随机变量的话——应该是离散型的,即

$$P(\xi = x_i) = p_i.$$

这样的随机变量很容易构造,即令

$$A_1 := [0, p_1), \cdots, A_k := [p_1 + \cdots + p_{k-1}, p_1 + \cdots + p_k), 2 \leq k \leq n.$$

$$\xi(\omega) := x_i, \omega \in A_i.$$

⁵Thomas Jan Stieltjes, 1856-1894, 荷兰-法国数学家.

则易见 ξ 的分布函数为 F .

(2) 设存在 $\{x_i, i = 0, \pm 1, \pm 2, \dots\}$, $x_i \in \mathbb{R}$, 及 $\{p_i, i = 0, \pm 1, \pm 2, \dots\}$, $p_i > 0$, 使得

$$-\infty < \dots < x_{-n-1} < x_{-n} < \dots < x_{-1} < x_0 < x_1 < \dots < x_n < x_{n+1} < \dots < \infty,$$

$$\sum_{i=-\infty}^{\infty} p_i = 1,$$

$$F(x) = \sum_{x_i \leq x} p_i.$$

令

$$A_i := \left[\sum_{k=-\infty}^{i-1} p_k, \sum_{k=-\infty}^i p_k \right),$$

$$\xi(\omega) := x_i, \omega \in A_i.$$

则 ξ 的分布函数为 F .

(3) 一般地, 设 F 为分布函数. $\forall n$, 定义分割

$$\mathcal{P}_n = \{r_k^n : k = 0, \pm 1, \pm 2, \dots\},$$

使之满足:

(i)

$$r_k^n < r_{k+1}^n, \quad \forall n, k;$$

(ii)

$$\sup_k |r_{k+1}^n - r_k^n| \leq \frac{1}{2^n};$$

(iii) \mathcal{P}_n 是逐步加细的, 即 $\forall n, k, \exists k'$, 使得 $r_{k'}^{n+1} = r_k^n$. (这样的分割是存在的, 比如可以取 $r_k^n := k2^{-n}$.)

定义

$$\xi_n(\omega) := r_k^n, \quad \omega \in [F(r_k^n), F(r_{k+1}^n)).$$

令 $h_n(x)$ 是诸 $\{r_k^n, k \in \mathbb{N}\}$ 里面大于 x 中的最小者, 即

$$h_n(x) := \min\{r_k^n : r_k^n > x\}.$$

则

$$\xi_n(\omega) \leq x \iff \omega \in [0, F(h_n(x))).$$

因此 ξ_n 的分布函数为

$$F_n(x) = P(\xi_n \leq x) = F(h_n(x)).$$

又显然有

$$0 \leq \xi_{n+1}(\omega) - \xi_n(\omega) \leq 2^{-n}, \quad \forall \omega.$$

因此 ξ_n (对 n)单调上升(对 ω)一致收敛. 设其极限为 ξ . $\forall x \in \mathbb{R}$, 则

$$\bigcap_{n=1}^{\infty} \{\xi_n \leq x\} = \{\xi \leq x\},$$

注意 $h_n(x) \downarrow x$, 所以, 由 F 的右连续性, 有

$$\begin{aligned} P(\xi \leq x) &= \lim_{n \rightarrow \infty} P(\xi_n \leq x) \\ &= \lim_{n \rightarrow \infty} F(h_n(x)) \\ &= F(x). \end{aligned}$$

□

注. 在一维情况, 随机变量的存在性在大部分书中是用广义反函数证明的. 这种证明方法的优点是比较简短(不过, 如果要算上在广义反函数方面的准备工作, 实际上也简短不了多少. [12] 中这个准备是不完备的, 证明也是有瑕疵的), 缺点是不能应用到其它场合. 上述证明原则上可应用于其它场合.

习题

1. 设 $F: \mathbb{R}^2 \mapsto [0, 1]$ 满足:

(a) $\forall x_1 \leq y_1, x_2 \leq y_2,$

$$F(y_1, y_2) - F(x_1, y_2) - F(y_1, x_2) + F(x_1, x_2) \geq 0;$$

(b)

$$\lim_{x_1, x_2 \rightarrow \infty} F(x_1, x_2) = 1;$$

(c)

$$\begin{aligned} \lim_{x_2 \rightarrow -\infty} F(x_1, x_2) &= 0, \quad \forall x_1, \\ \lim_{x_1 \rightarrow -\infty} F(x_1, x_2) &= 0, \quad \forall x_2; \end{aligned}$$

(d) $\forall x_1, x_2,$

$$\lim_{y_1 \downarrow x_1, y_2 \downarrow x_2} F(y_1, y_2) = F(x_1, x_2).$$

证明: 存在概率空间 (Ω, \mathcal{F}, P) 及其上的随机变量 (ξ, η) , 使得其分布函数恰为 F . (建议: 也许先考虑 F 连续的情况会容易一点?)

4.7 随机变量的函数

设 ξ 是随机变量, $\varphi: \mathbb{R} \mapsto \mathbb{R}$. 由命题 4.1.18, $\varphi(\xi)$ 仍是一个随机变量. 本节要考虑的问题是, 如果知道 ξ 的分布, 能决定 $\varphi(\xi)$ 的分布吗, 其分布该怎么来求?

理论上说, $\varphi(\xi)$ 的分布是由 ξ 的分布唯一确定的. 因为

$$\{\omega: \varphi(\xi) \in B\} = \{\xi \in \varphi^{-1}(B)\},$$

所以

$$P(\{\omega: \varphi(\xi) \in B\}) = P(\{\xi \in \varphi^{-1}(B)\}).$$

但实际上一般很难直接由 ξ 的分布函数求出 $\varphi(\xi)$ 的分布函数. 不过在一些简单情况, 是有办法算出来的.

首先, 如果 ξ 是离散型的, 那么 $\varphi(\xi)$ 也是离散型的. 以 S 记 ξ 的值域, 那么 $\varphi(\xi)$ 的值域就是 $\varphi(S)$. 对任意 $y \in \varphi(S)$,

$$P(\varphi(\xi) = y) = P(\xi \in \varphi^{-1}(y)) = \sum_{x \in S: \varphi(x)=y} P(\xi = x).$$

剩下的就是具体计算问题了.

ξ 是连续型时事情会麻烦一些, 也许是麻烦很多. 这时, 首先 $\eta := \varphi(\xi)$ 有可能是连续型的, 也有可能是离散型的, 也可能两者都不是 (请自己举个例子). 设 ξ 的密度函数是 p , 我们来看两种情况及相应的具体例子.

1. 设 $\varphi(x) = -x$, 则

$$P(\eta \leq x) = P(\xi \geq -x) = \int_{-x}^{\infty} p(u) du = \int_{-\infty}^x p(-u) du.$$

所以 η 也是连续型的, 其分布密度为 $p(-x)$.

2. 设 ξ 的值域为一区间 I , 有限或无限, 开或闭, 半开或半闭, 均不限. 设 φ 是 I 上单调函数, 我们假设单调上升好了. 如果单调下降的话, 可先考虑 $-\varphi$, 然后用情形1的结果. 这时,

$$P(\eta \leq x) = \begin{cases} 1, & \text{当 } x \text{ 在 } I \text{ 右边,} \\ 0, & \text{当 } x \text{ 在 } I \text{ 左边.} \end{cases}$$

而当 $x \in I$ 时

$$\begin{aligned} P(\eta \leq x) &= P(\xi \leq \varphi^{-1}(x)) \\ &= \int_{-\infty}^{\varphi^{-1}(x)} f(u) du. \end{aligned}$$

一般情况下也只能到此而止了. 但如果 φ 连续可导, 则上式等于

$$\int_{-\infty}^x f(\varphi^{-1}(u)) (\varphi^{-1}(u))' du.$$

所以 η 也是连续型的, 其分布密度为

$$g(x) = \begin{cases} 0 & x \notin \varphi(I), \\ f(\varphi^{-1}(x)) (\varphi^{-1}(x))' & x \in \varphi(I). \end{cases}$$

3. 若 ξ 的值域可分解为可数个不相交的区间 $\{I_i, i = 1, 2, \dots\}$ 的并, 而 φ 在每个 I_i 上严格单调 (可在不同的 I_i 上有不同的单调性), 连续可微. 则利用

$$P(\eta \leq x) = \sum_{i=1}^{\infty} P(\varphi(\xi) \leq x, \xi \in I_i).$$

若 I_i 的左端点为 a_i , 右端点为 b_i (开或闭均可), φ 在 I_i 上单调上升, 则

$$P(\varphi(\xi) \leq x, \xi \in I_i) = \int_{a_i}^{(\varphi^{-1}(x) \vee a_i) \wedge b_i} f(y) dy.$$

若 φ 在 I_i 上单调下降, 则

$$P(\varphi(\xi) \leq x, \xi \in I_i) = \int_{(\varphi^{-1}(x) \vee a_i) \wedge b_i}^{b_i} f(y) dy.$$

如此这般可求出密度. 但一般的公式写出来太复杂, 不如按此原则具体问题具体算, case by case.

习题

1. 设 ξ 是连续型随机变量, 分布密度是 f . 证明下面的随机变量也是连续型的, 并求出分布密度.
 - (a) $\xi_1 := \xi^2$;
 - (b) $\xi_2 := \sin \xi$;
 - (c) $\xi_3 := \exp(\xi)$;
 - (d) $\xi_4 := |\xi|$;
 - (e) $\xi_5 := |\xi|^\alpha, \alpha > 0$.
2. 设 $\xi \sim N(0, 1)$. 证明 $-\xi \sim N(0, 1)$.
3. 设 $\xi \sim N(\mu, \sigma^2)$. 证明 $\frac{\xi - \mu}{\sigma} \sim N(0, 1)$.
4. 设 ξ 的分布函数为 $F, a < b$ 为常数. 令 $\eta := \xi 1_{(a,b)}(\xi)$. 求 η 的分布函数.

4.8 多维随机变量的函数

以上我们考虑的随机变量是一维的, 函数 φ 也是 \mathbb{R}^1 到 \mathbb{R}^1 的. 本节我们考虑多维情形.

设 ξ 是 \mathbb{R}^n -值随机变量, $\varphi: \mathbb{R}^n \mapsto \mathbb{R}^m$ 是Borel函数. 由命题4.1.18, 则 $\eta := \varphi(\xi)$ 是 \mathbb{R}^m -值随机变量.

理论上说, η 的分布是由 ξ 的分布和 φ 唯一确定的. 不过理论是一回事, 实际计算则是另一回事. 要给出 η 分布的解析表达式, 只有在零星几种情况是可能的. 不过幸运的是, 这零星的几种情况是非常重要的. 其中最重要的是:

命题 4.8.1. 设 (ξ, η) 是二维离散型随机变量. 分布列为 $\{(x_i, y_j), p_{ij}, i, j = 1, 2, \dots\}$. 令 $\zeta := \xi + \eta$. 则 ζ 也是离散型的, 值域为 $\{x_i + y_j, i, j = 1, 2, \dots\}$, 且对于 $z \in \{x_i + y_j, i, j = 1, 2, \dots\}$ 有

$$P(\zeta = z) = \sum_{x_i + y_j = z} p_{ij}.$$

证明. 我们有

$$P(\xi + \eta = z) = \sum_{i,j: x_i + y_j = z} P(\xi = x_i, \eta = y_j) = \sum_{i,j: x_i + y_j = z} p_{ij}.$$

□

现在考虑另一种情况. 设 ξ 是 n 维连续型随机变量, 值域为区域 E_1 , 密度为 p (因此在 E_1^c 上 p 为零). $\varphi: E_1 \rightarrow E_2$, $E_2 \subset \mathbb{R}^n$, 且 $\varphi: E_1 \rightarrow E_2$ 为连续可微的双射. 设Jacobi行列式

$$\frac{\partial \varphi(u)}{\partial u} \neq 0, \quad \forall u \in E_1.$$

我们来求 $\eta := \varphi(\xi)$ 的分布.

为此, 我们需要下面的结果.

定理 4.8.2. 设 ξ 为 n 维连续型随机变量, 密度函数为 p , 且 $D \subset \mathbb{R}^n$ 为有限条简单曲面(曲线)围成的区域(开或闭均可). 则

$$P(\xi \in D) = \int_D p(x) dx,$$

上式右边理解为Riemman积分.

这个结果成立的直观理由是: 为求 $P(\xi \in D)$, 将 D 分为若干小块:

$$D = \sum_{i=1}^k D_i.$$

则

$$P(\xi \in D) = \sum_{i=1}^k P(\xi \in D_i) \approx \sum_{i=1}^k p(x_i) |D_i|,$$

其中 $x_k \in D_k$, $|D_k|$ 表示 D_k 的体积. 当小块的直径趋于0时, 就得到所要的公式. 详细证明见后面的定理5.9.2.

对于连续型随机变量, 命题4.8.1变为

命题 4.8.3. 设 (ξ, η) 是二维连续型随机变量, 密度函数为 $p(x, y)$, 则 $\zeta = \xi + \eta$ 也是连续型的, 且密度函数为

$$h(z) := \int_{-\infty}^{\infty} p(z-x, x) dx = \int_{-\infty}^{\infty} p(x, z-x) dx.$$

证明. 由定理4.8.2, 我们有

$$\begin{aligned} P(\xi + \eta \leq z) &= \iint_{x+y \leq z} p(x, y) dx dy \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^{z-x} p(x, y) dy \\ &= \int_{-\infty}^{\infty} dx \int_{-\infty}^z p(x, y-x) dy \\ &= \int_{-\infty}^z dy \int_{-\infty}^{\infty} p(x, y-x) dx. \end{aligned}$$

所以 $\xi + \eta$ 是连续型的, 且密度函数为

$$\int_{-\infty}^{\infty} p(x, z-x) dx$$

另一表达式可通过变量代换得到. □

现在回到一般的 $\eta := \varphi(\xi)$ 的密度函数的计算. $\forall x = (x_1, \dots, x_n) \in E_2$, 令

$$E_{2,x} := \{y = (y_1, \dots, y_n) \in E_2 : y_1 \leq x_1, \dots, y_n \leq x_n\}.$$

因此

$$\begin{aligned} P(\eta_1 \leq x_1, \dots, \eta_n \leq x_n) &= \int_{\varphi^{-1}(E_{2,x})} p(u) du \\ &= \int_{E_{2,x}} p(\varphi^{-1}(y)) \left| \frac{\partial \varphi^{-1}(y)}{\partial y} \right| dy. \end{aligned}$$

由此得到 η 也是连续型的, 且密度为

$$q(y) := \begin{cases} p(\varphi^{-1}(y)) \left| \frac{\partial \varphi^{-1}(y)}{\partial y} \right| & y \in E_2, \\ 0 & y \notin E_2. \end{cases}$$

习题

1. 设 ξ 为 n 维随机变量, $f: \mathbb{R}^n \mapsto \mathbb{R}^m$ 为Borel函数. 证明 $f(\xi)$ 为 m 维随机变量.
2. 设 ξ 为随机变量(可能是多维). 证明 $\{\omega : |\xi(\omega)| = \infty\} \in \mathcal{F}$.
3. 设 ξ_1 与 ξ_2 同分布, η_1 与 η_2 同分布. 问 $\xi_1 + \eta_1$ 是否与 $\xi_2 + \eta_2$ 同分布? 证明或举出反例.
4. 设 ξ 是 n 维连续型随机变量, 密度函数形式为 $p(x) = f(|x|)$. $(r(x), \theta_1(x), \dots, \theta_{n-1}(x))$ 表示 x 的极坐标. 证明: $r(\xi), \theta_1(\xi), \dots, \theta_{n-1}(\xi)$ 独立.

5 期望与积分

将一组数据相加, 然后除以数据个数, 就得到这组数据的平均值; 将一个函数求积分并除以积分区间的长度, 就得到这个函数的平均值. 同理, 将一个随机变量求相对于概率的加权和, 就得到这个随机变量的平均值, 通称数学期望, 简称期望. 期望现在一律用 E 表示, 代表英语的Expectation, 也可认为是法语的Espérance.

Lévy在[10]中, 将现在我们称之为期望的东西按本意叫做“可能值的平均”(la moyenne des valeurs possibles), 并用 M 表示. 所以, 在早期的文献(尤其是俄文文献及俄文文献的其他语言译本)中, 也常见用 M 表示期望的, 代表Moyenne, 或者Mean.

期望除了代表平均值这个功能外, 还是研究概率论的最基本的工具. 它在概率论中的地位就相当于积分在数学分析中的地位, 没有它概率论就寸步难行, 就没有太多的“论——Theory”可言.

5.1 简单随机变量情形

在离散概型下, 对随机变量 ξ , 我们曾定义了其期望 $E[\xi]$. 即

$$E[\xi] := E[\xi^+] - E[\xi^-],$$

如果等式右边的两项中至少一项有限.

在公理化框架下, 对只取可数个值的随机变量, 我们可以类似地定义期望. 给定概率空间 (Ω, \mathcal{F}, P) , 我们先引进一个名词.

定义 5.1.1. 设随机变量 ξ 可表示为

$$\xi = \sum_{n=1}^{\infty} x_n 1_{A_n}, \quad x_n \in \bar{\mathbb{R}}, A_n \in \mathcal{F}, \forall n,$$

其中 $\{A_n, n = 1, 2, \dots\}$ 为 Ω 的一个分割, 则称 ξ 为简单随机变量. 若还有 $x_n \geq 0, \forall n$, 则称 ξ 为非负简单随机变量.

我们先对非负简单随机变量定义期望.

定义 5.1.2. 设 ξ 为非负简单随机变量:

$$\xi = \sum_{n=1}^{\infty} x_n 1_{A_n}, \quad x_n \geq 0, \forall n.$$

定义

$$E[\xi] := \sum_{n=1}^{\infty} x_n P(A_n).$$

提醒一下, 这里有可能出现 $x_n = \infty$ 的情况. 此时, 记住我们的约定:

$$0 \times \pm\infty = 0, \quad x \times \pm\infty = \pm\infty \quad (x > 0).$$

注意在上述定义中, 等式右边是非负项级数, 所以收敛是肯定的, 无非收敛到是无限或有限, 且其值与各项的排序无关. 若收敛到有限数, 则称 ξ 可积.

我们必须说明这个定义的合理性, 即若还有另外一个分割 $\{B_m, m = 1, 2, \dots\}$ 使得 ξ 也可以表示为

$$\xi = \sum_{m=1}^{\infty} y_m 1_{B_m}, \quad y_m \geq 0.$$

则一定有

$$\sum_{n=1}^{\infty} x_n P(A_n) = \sum_{m=1}^{\infty} y_m P(B_m).$$

事实上, 若 $A_n B_m \neq \emptyset$, 则必在其上有 $x_n = y_m$. 因此由概率的可列可加性有

$$\begin{aligned} \sum_{n=1}^{\infty} x_n P(A_n) &= \sum_{n=1}^{\infty} x_n \sum_{m=1}^{\infty} x_n P(A_n B_m) \\ &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} x_n P(A_n B_m) \\ &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} y_m P(A_n B_m) \\ &= \sum_{m=1}^{\infty} y_m P(B_m), \end{aligned}$$

其中两个求和号能交换次序是因为现在是正项级数.

对可正可负的随机变量, 期望是将其分解为正部负部分别定义期望. 具体地说, 设

$$\xi := \sum_{n=1}^{\infty} x_n 1_{A_n}.$$

则

$$\xi = \xi^+ - \xi^-,$$

其中

$$\xi^+ := \sum_{n=1}^{\infty} x_n^+ 1_{A_n}, \quad \xi^- := \sum_{n=1}^{\infty} x_n^- 1_{A_n}.$$

由于 ξ^+ 与 ξ^- 均为非负简单随机变量, 所以其期望均有定义. 我们有

定义 5.1.3. 设 ξ 是简单随机变量. 若 $E[\xi^+]$ 与 $E[\xi^-]$ 中至少一个有限, 则称 $E[\xi]$ 存在, 且定义为

$$E[\xi] := E[\xi^+] - E[\xi^-].$$

若两者均有限, 则称 ξ 可积.

由定义直接得到:

命题 5.1.4. 简单随机变量

$$\xi := \sum_{n=1}^{\infty} x_n 1_{A_n}$$

可积的充要条件是

$$\sum_{n=1}^{\infty} |x_n| P(A_n) < \infty,$$

且此时

$$E[\xi] = \sum_{n=1}^{\infty} x_n P(A_n).$$

简单随机变量的期望有以下性质:

命题 5.1.5. (i) 线性性: 设 $a, b \in \mathbb{R}$, ξ, η 可积, 则 $a\xi + b\eta$ 也可积, 且

$$E[a\xi + b\eta] = aE[\xi] + bE[\eta];$$

(ii) 单调性: 设 ξ, η 可积, 则:

$$\xi \leq \eta \implies E[\xi] \leq E[\eta];$$

(iii) 设 ξ 可积, 则

$$|E[\xi]| \leq E[|\xi|].$$

证明. (i) 设

$$\xi = \sum_{n=1}^{\infty} c_n 1_{A_n},$$

$$\eta = \sum_{n=1}^{\infty} d_n 1_{B_n},$$

其中 $\sum_{n=1}^{\infty} A_n = \sum_{n=1}^{\infty} B_n = \Omega$. 将它们各自改写为

$$\xi = \sum_{m,n=1}^{\infty} c_n 1_{B_m A_n},$$

$$\eta = \sum_{m,n=1}^{\infty} d_m 1_{B_m A_n}.$$

则

$$a\xi + b\eta = \sum_{m,n=1}^{\infty} (ac_n + bd_m) 1_{B_m A_n}.$$

因此 $a\xi + b\eta$ 也是简单随机变量, 且

$$\begin{aligned}
 E[a\xi + b\eta] &= \sum_{m,n=1}^{\infty} (ac_n + bd_m)P(B_m A_n) \\
 &= a \sum_{m,n=1}^{\infty} c_n P(B_m A_n) + b \sum_{m,n=1}^{\infty} d_m P(B_m A_n) \\
 &= a \sum_{n=1}^{\infty} c_n \sum_{m=1}^{\infty} P(B_m A_n) + b \sum_{m=1}^{\infty} d_m \sum_{n=1}^{\infty} P(B_m A_n) \\
 &= a \sum_{n=1}^{\infty} c_n P\left(\left(\bigcup_{m=1}^{\infty} B_m\right) A_n\right) + b \sum_{m=1}^{\infty} d_m P\left(B_m \left(\bigcup_{n=1}^{\infty} A_n\right)\right) \\
 &= a \sum_{n=1}^{\infty} c_n P(A_n) + b \sum_{m=1}^{\infty} d_m P(B_m) \\
 &= aE[\xi] + bE[\eta].
 \end{aligned}$$

注意上述运算之所以畅通无阻, 是因为我们假定了可积性, 即

$$\sum_{n=1}^{\infty} |c_n|P(A_n) < \infty, \quad \sum_{n=1}^{\infty} |d_n|P(B_n) < \infty.$$

(见习题2.)

(ii) 首先注意若 $\xi \geq 0$, 则 $E[\xi] \geq 0$. 其次, 因 $\eta - \xi \geq 0$, 由(i),

$$E[\eta] - E[\xi] = E[\eta - \xi] \geq 0.$$

(iii) 因 $-|\xi| \leq \xi \leq |\xi|$, 由(1)和(ii),

$$-E[|\xi|] \leq E[\xi] \leq E[|\xi|].$$

□

习题

1. 设 $\xi_i, i = 1, \dots, n$ 是简单随机变量, $f: \mathbb{R}^n \mapsto \mathbb{R}$ 是函数. 证明 $f(\xi_1, \dots, \xi_n)$ 仍然是简单随机变量.

2. 证明简单随机变量 $\xi := \sum_{n=1}^{\infty} x_n 1_{A_n}$ 可积的充要条件是

$$\sum_{n=1}^{\infty} |x_n|P(A_n) < \infty.$$

3. 设简单随机变量 $\xi := \sum_{n=1}^{\infty} x_n 1_{A_n}$ 可积. 证明

$$E[\xi] := \sum_{n=1}^{\infty} x_n P(A_n).$$

5.2 一般情形

现在我们要对一般的随机变量也定义期望. 定义的方法是通过离散逼近. 具体地说就是:

定义 5.2.1. (i) 设 ξ 是非负实值随机变量, ξ_n 是简单随机变量列, 且

$$\xi_n(\omega) \Rightarrow \xi(\omega) \text{ a.s.},$$

即 ξ_n 在一个可略集外, 关于 ω 一致收敛于 ξ , 则定义

$$E[\xi] := \lim_{n \rightarrow \infty} E[\xi_n].$$

(ii) 设 ξ 是非负随机变量, 且 $P(\xi = \infty) > 0$, 则定义

$$E[\xi] = \infty.$$

为使这个定义合理, 我们要说明两件事情. 第一是对任意非负实值随机变量, 至少有一个简单随机变量序列 ξ_n 使得 $\xi_n \Rightarrow \xi$ a.s.. 这的确是可以做到的, 例如可令

$$j_n(\xi) = \sum_{k=0}^{\infty} k 2^{-n} 1_{[k 2^{-n}, (k+1) 2^{-n})}(\xi).$$

则除非 $\xi(\omega) = \infty$ (这样的 ω 全体构成一个可略集), 否则均有

$$|j_n(\xi)(\omega) - \xi(\omega)| < 2^{-n}.$$

所以 $j_n(\xi) \Rightarrow \xi$ a.s..

第二是对一致收敛于 ξ 的 ξ_n , 极限 $\lim_{n \rightarrow \infty} E[\xi_n]$ 的确存在且不依赖于 ξ_n 的选择.

因为 $\xi_n \Rightarrow \xi$ a.s., 所以 $\forall \varepsilon > 0, \exists N_\varepsilon$, 当 $m, n \geq N_\varepsilon$ 有

$$|\xi_n - \xi_m| \leq \varepsilon, \text{ a.s. } \omega.$$

故

$$E[|\xi_n - \xi_m|] \leq \varepsilon.$$

因此当 $n \geq N_\varepsilon$ 时, 所有的 ξ_n 要么同时可积, 要么同时不可积, 且当它们可积时, 对 $m, n > N_\varepsilon$ 有

$$|E[\xi_n] - E[\xi_m]| \leq E[|\xi_n - \xi_m|] \leq \varepsilon.$$

所以极限的确存在.

若 $\xi_n \Rightarrow \xi$ a.s., $\eta_n \Rightarrow \xi$ a.s., 令

$$\zeta_n = \begin{cases} \xi_k & n = 2k - 1 \\ \eta_k & n = 2k. \end{cases}$$

则 $\zeta_n \Rightarrow \xi$ a.s.. 所以 $\lim_{n \rightarrow \infty} E[\zeta_n]$ 存在. 因此

$$\lim_{n \rightarrow \infty} E[\xi_n] = \lim_{n \rightarrow \infty} E[\eta_n].$$

所以这个极限不依赖于 ξ_n 的选择.

现在我们可以对一般的随机变量定义期望.

定义 5.2.2. 对一般的 ξ , 若下面两式中至少之一成立:

$$E[\xi^+] < \infty, \quad E[\xi^-] < \infty,$$

则定义

$$E[\xi] := E[\xi^+] - E[\xi^-].$$

若 $E[\xi^+] + E[\xi^-] < \infty$, 则称 ξ 可积.

注. 1. 按定义, 非负实值随机变量的数学期望总是存在的, 它要么是实数, 要么是无穷大. 并且, 因为

$$E[j_n(\xi)] \leq E[j_{n+1}(\xi)] \leq E[j_n(\xi)] + 2^{-n},$$

所以 $E[\xi]$ 是无穷大的充分必要条件是每个 $E[j_n(\xi)]$ 都是无穷大.

若 ξ 为可积随机变量, 定义

$$j_n(\xi) := j_n(\xi^+) - j_n(\xi^-).$$

因 $j_n(\xi^+) \Rightarrow \xi^+ \text{ a.s.}$, $j_n(\xi^-) \Rightarrow \xi^- \text{ a.s.}$, 故 $j_n(\xi) \Rightarrow \xi^+ - \xi^- = \xi \text{ a.s.}$, 且 $\lim_{n \rightarrow \infty} E[j_n(\xi)] = E[\xi]$.

而且, 若 ξ 可积且

$$\xi_n(\omega) \Rightarrow \xi(\omega) \text{ a.s.},$$

则总有

$$E[\xi] = \lim_{n \rightarrow \infty} E[\xi_n].$$

特别地,

$$\xi = \eta \text{ a.s.} \implies E[\xi] = E[\eta].$$

2. 同样是按定义, 当 ξ 可积时, 必有 $P(\xi = \infty) = 0$.

3. 对一般的随机变量 ξ , 之所以要求 $E[\xi^+]$ 与 $E[\xi^-]$ 中至少一个是有限的, 是因为要避免出现 $\infty - \infty$ 这样一个没有意义的情况.

4. ξ 可积时, $|E[\xi]| \leq E[|\xi|] < \infty$.

5. 由定义, 期望是一种和的极限, 因此是积分, 故也常常用积分符号表示期望, 即用

$$\int_{\Omega} \xi(\omega) P(d\omega), \quad \int_{\Omega} \xi(\omega) dP(\omega), \quad \text{或} \quad \int \xi dP$$

表示 $E[\xi]$.

注意, 如果取 $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0, 1])$, $P = \text{Lebesgue测度}$, 则上面定义的积分就是所谓的Lebesgue积分. 而将所有区间 $[n, n+1)$ 合并起来, 则得到 \mathbb{R} 上的Lebesgue积分. 根据实变函数论, 一个非负函数如果是Riemann可积的, 则一定是Lebesgue可积的, 且两个积分相等. 所以, 此后我们将不加区别地使用

$$\int_{\mathbb{R}} f(t) dt$$

代表这个积分, 并默认它首先是Lebesgue积分, 但如果它是Riemann可积的(比如说 f 分段连续时), 也可理解为Riemann积分.

根据定义易见, 当 ξ 是离散型的时候, $E[\xi]$ 可以通过 ξ 的分布列进行计算. 自然的期待是, 当 ξ 为连续型时, $E[\xi]$ 可以通过 ξ 的密度函数来计算. 这就是下面的:

命题 5.2.3. (i) 设 ξ 是离散型随机变量, 分布列为 (x_n, p_n) . 若 $\sum_n |x_n|p_n < \infty$, 则

$$E[\xi] = \sum_n x_n p_n.$$

(ii) 设 ξ 为连续型随机变量, 密度函数为 $p(x)$. 若

$$\int_{-\infty}^{\infty} |x|p(x)dx < \infty,$$

则 $E[\xi]$ 存在且

$$E[\xi] = \int_{-\infty}^{\infty} x f(x) dx.$$

证明. (i) 此时 ξ 可表示为

$$\xi = \sum_n x_n 1_{A_n},$$

其中 $A_n = \{\xi = x_n\}$. 因此 $P(A_n) = p_n$. 于是 $\sum_n |x_n|p_n < \infty$ 时有

$$E[\xi] = \sum_n x_n p_n.$$

(ii) 只要注意

$$\begin{aligned} \left| E[j_n(\xi^+)] - \int_0^{\infty} x f(x) dx \right| &= \left| \sum_{k=0}^{\infty} k 2^{-n} P(j_n(\xi) = k 2^{-n}) - \int_0^{\infty} x f(x) dx \right| \\ &= \left| \sum_{k=0}^{\infty} \int_{k 2^{-n}}^{(k+1) 2^{-n}} (k 2^{-n} - x) f(x) dx \right| \\ &\leq 2^{-n} \int_0^{\infty} f(x) dx \\ &\leq 2^{-n} \end{aligned}$$

和

$$\left| E[j_n(\xi^-)] + \int_{-\infty}^0 x f(x) dx \right| \leq 2^{-n}.$$

□

现在我们研究期望的基本性质. 首先我们有:

引理 5.2.4. 设 ξ, η 可积, a, b 是常数, 则

$$E[a\xi + b\eta] = aE[\xi] + bE[\eta].$$

这个结果称为期望的线性性.

证明. 设 ξ_n, η_n 是简单随机变量, 且

$$\xi_n \Rightarrow \xi \text{ a.s.}, \quad \eta_n \Rightarrow \eta \text{ a.s.}$$

则 $a\xi_n + b\eta_n$ 也是简单随机变量且

$$a\xi_n + b\eta_n \Rightarrow a\xi + b\eta \text{ a.s.}$$

于是, 用命题5.1.5之(i), 我们有

$$\begin{aligned} E[a\xi + b\eta] &= \lim_{n \rightarrow \infty} E[a\xi_n + b\eta_n] \\ &= a \lim_{n \rightarrow \infty} E[\xi_n] + b \lim_{n \rightarrow \infty} E[\eta_n] \\ &= aE[\xi] + bE[\eta]. \end{aligned}$$

□

推论 5.2.5. 设 ξ 与 η 期望均存在, 则 (i)

$$\xi \geq 0 \implies E[\xi] \geq 0;$$

(ii)

$$\xi \leq \eta \implies E[\xi] \leq E[\eta];$$

(iii)

$$|E[\xi]| \leq E[|\xi|].$$

证明. (i) 直接由期望的定义.

(ii) 若 ξ, η 可积, 直接由(i)及线性性有

$$E[\eta] - E[\xi] = E[\eta - \xi] \geq 0.$$

若 ξ 不可积, 因为 $j_n(\xi) \leq j_n(\eta)$, 则 η 也不可积, 结论也成立.

(iii) 因为

$$-|\xi| \leq \xi \leq |\xi|,$$

所以

$$-E[|\xi|] \leq E[\xi] \leq E[|\xi|].$$

□

我们由此可以得到Chebyshev¹不等式:

定理 5.2.6. (Chebyshev不等式) 设 $\xi \geq 0$ 是随机变量, $a > 0$ 是常数. 则

$$P(\xi \geq a) \leq \frac{E[\xi]}{a}.$$

证明. 我们有

$$aP(\xi \geq a) = aE[1_{\xi \geq a}] \leq E[\xi 1_{\xi \geq a}] \leq E[\xi].$$

□

¹Pafnuty Lvovich Chebyshev, 1821-1894, 俄罗斯数学家.

此不等式很少有直接使用的, 使用得更多的是它的一些变体. 这里的总体想法是, 如果 f 是定义在 $[0, \infty)$ 上的单调上升函数, 且 $x > 0$ 时 $f(x) > 0$, 那么对任意(即不管是否非负) ξ 有

$$P(|\xi| \geq a) = P(f(|\xi|) \geq f(a)) \leq \frac{E[f(|\xi|)]}{f(a)}.$$

以下是几种常用的取法: 取 $f(x) = x^\alpha$, $\alpha > 0$ 就有

$$P(|\xi| \geq a) \leq \frac{E[|\xi|^\alpha]}{a^\alpha};$$

对 $|\xi - E[\xi]|$ 用上式则有

$$P(|\xi - E[\xi]| \geq a) \leq \frac{E[|\xi - E[\xi]|^\alpha]}{a^\alpha};$$

取 $f(x) = \exp\{\lambda x\}$, $\lambda \in I$, 其中 $I \subset \mathbb{R}_+$, 则有

$$P(|\xi| \geq a) \leq e^{-\lambda a} E[\exp\{\lambda|\xi|\}];$$

等等. 在最后一个不等式中, 因为 λ 可以跑遍 I , 故

$$P(|\xi| \geq a) \leq \inf_{\lambda \in I} e^{-\lambda a} E[\exp\{\lambda|\xi|\}].$$

这么做的好处是对不同的 a , 使得右边接近最小值的 λ 可能不一样的, 因此需要对不同的 a 取最合适的 λ , 以得到最佳估计.

由此可以得到:

推论 5.2.7. $E[|\xi|] = 0$ 当且仅当 $P(\xi = 0) = 1$.

证明. $\forall n$, 由 Chebyshev 不等式有

$$P(|\xi| \geq \frac{1}{n}) \leq nE[|\xi|] = 0.$$

由于

$$\{\xi \neq 0\} = \bigcup_{n=1}^{\infty} \left\{ |\xi| \geq \frac{1}{n} \right\},$$

所以由概率的下连续性有

$$P(\xi \neq 0) = \lim_{n \rightarrow \infty} P\left(\left\{ |\xi| \geq \frac{1}{n} \right\}\right) = 0.$$

□

我们还可以得到:

命题 5.2.8. 设 ξ^2, η^2 可积, 则成立 Cauchy²-Schwarz³ 不等式

$$(E[\xi\eta])^2 \leq E[\xi^2]E[\eta^2].$$

当 $E[\xi^2]$ 和 $E[\eta^2]$ 都为零时, 显然等式成立; 当 $E[\xi^2]$ 和 $E[\eta^2]$ 有一个不为零, 比如当 $E[\xi^2] > 0$ 时, 等式成立的充要条件是存在 $t_0 \in \mathbb{R}$, 使 $P(\eta = t_0\xi) = 1$.

²Augustin Louis Cauchy, 1789-1857, 法国数学家.

³Hermann Amandus Schwarz, 1843-1921, 波兰-德国数学家.

证明. 因为 $2\xi\eta \leq \xi^2 + \eta^2$, 显然 $\xi\eta$ 可积. 记

$$f(t) := E[(t\xi - \eta)^2],$$

则 $f(t) \geq 0$, 且

$$f(t) = t^2 E[\xi^2] - 2t E[\xi\eta] + E[\eta^2].$$

当 $E[\xi^2] = 0$, 则 $P(\xi = 0) = 1$. 因此不等式显然成立. 当 $E[\xi^2] > 0$ 时, $f(t)$ 为关于 t 的非负一元二次函数, 因而其判别式

$$\Delta = 4(E[\xi\eta])^2 - 4E[\xi^2]E[\eta^2] \leq 0,$$

即 Cauchy 不等式成立.

当 $E[\xi^2] > 0$ 时, 若等式成立, 则存在 $t_0 \in \mathbb{R}$, 使 $f(t_0) = E[(t_0\xi - \eta)^2] = 0$. 由推论 5.2.7, 因此 $P(\eta = t_0\xi) = 1$. 若 $P(\eta = t_0\xi) = 1$, 显然等式成立. \square

习题

1. 将引理 5.2.4 推广到期望存在的情景, 即证明: 当 $aE[\xi] + bE[\eta]$ 有意义时, $E[a\xi + b\eta]$ 也存在且两者相等; 反之, 当 $E[a\xi + b\eta]$ 存在时, $aE[\xi] + bE[\eta]$ 有意义且两者相等.
2. 证明: ξ 可积的充要条件为

$$\sum_{n=0}^{\infty} P(|\xi| \geq n) < \infty.$$

5.3 计算期望的例子

前面我们定义了随机变量的期望并研究了它的一些基本性质. 至于它的计算, 在有些情况下该随机变量的分布是已知的, 此时期望就是一个级数或者是一个积分, 因此问题是比较明确的(明确地难或者明确地易); 另外一些情况则是其分布并不清楚, 此时就要利用期望的性质, 而不是死算; 还有一些情况应用中的问题, 这样的问题更具有挑战性, 当然也就更有趣.

我们来看一些例子.

1. 设 $\xi \sim B(p)$, 即参数为 p 的 Bernoulli 分布.

我们有

$$E[\xi] = 0 \cdot q + 1 \cdot p = p.$$

2. 设 $\xi_i \sim B(p_i)$, $i = 1, \dots, n$. 令

$$\xi := \sum_{i=1}^n \xi_i.$$

求 $E[\xi]$.

ξ 显然是值域为 $\{0, 1, \dots, n\}$ 的离散型随机变量. 如果直接按定义计算, 为求 $E[\xi]$, 得知道 ξ 的分布列, 但不幸的是我们并不知道它. 注意, 我们并没有假定 ξ_i 们独立(这是下一章将介绍的概念). 当 ξ_i 们独立时倒是能计算出 ξ 的分布列, 但计算会比较繁琐, 因而容易出错, 并且是毫无必要的.

好在我们注意到, 在引理5.2.4中, 并不要求 ξ 和 η 独立. 所以有

$$E[\xi] = \sum_{i=1}^n E[\xi_i] = \sum_{i=1}^n p_i.$$

特别地, 当 $p_i = p, \forall i = 1, 2, \dots, n$ 时, 上式等于 np .

所以你一定要记牢了, 在公式

$$E[\xi + \eta] = E[\xi] + E[\eta]$$

中, 不需要 ξ 与 η 独立, 这会省掉很多麻烦.

把这个例子应用到二项分布. 设 $\xi \sim B(n, p)$, 由2.3节例2, 则

$$\xi = \sum_{i=1}^n \xi_i,$$

其中 $\xi_i \sim B(p)$. 所以 $E[\xi] = np$. 注意我们这里并不需要用到 ξ_1, \dots, ξ_n 的独立性, 尽管它们的确是独立的.

再看一个例子:

3. 在 $\{1, 2, \dots, n\}$ 中无放回地取 k 个数($k \leq n$), 求取出的数字之和 ξ 的期望.

设第 i 次取出的是 ξ_i , 显然 ξ_i 与 ξ_1 同分布, $\forall i$. 因此由期望的线性性有

$$E[\xi] = \sum_{i=1}^k E[\xi_i] = kE[\xi_1].$$

因为

$$P(\xi_1 = j) = \frac{1}{n}, \quad \forall j = 1, \dots, n,$$

所以

$$E[\xi_1] = \frac{1}{n} \sum_{j=1}^n j = \frac{n+1}{2}.$$

由此

$$E[\xi] = \frac{k(n+1)}{2}.$$

我们注意到, 当 $k = n$ 时, $E[\xi] = \frac{n(n+1)}{2}$. 这当然是必须的, 否则公式就错了.

4. 设 $\xi \sim P(\lambda)$, 即参数为 λ 的Poisson分布.

我们有

$$\begin{aligned} E[\xi] &= \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} \\ &= \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = \lambda. \end{aligned}$$

5. 设 ξ 服从几何分布.

此时

$$P(\xi = k) = q^{k-1} p, \quad k = 1, 2, \dots, q = 1 - p, \quad 0 < p < 1.$$

所以

$$\begin{aligned} E[\xi] &= \sum_{k=1}^{\infty} kq^{k-1}p = p \frac{d}{dq} \sum_{k=1}^{\infty} q^k \\ &= p \frac{d}{dq} \left[\frac{q}{1-q} \right] = p \frac{1}{(1-q)^2} = \frac{1}{p}, \end{aligned}$$

其中第二个等式是因为幂级数在收敛半径内求导和级数可以换序.

6. 设天上有 n 架飞机, 地下有 m 架高炮, 每门高炮独立瞄准射击. 某门高炮瞄准了某架飞机时命中率是 α ; 没有瞄准时, 命中率是0(即瞎撞上的概率是0). 求一次齐射之后, 命中飞机的平均数.

V.I.Arnold⁴ 讲过一个故事: 二战期间, 德军兵临莫斯科城下时, 斯大林为城市防空问题请教过Kolmogorov, 问怎样协调各门高炮瞄准的目标才能达到最好效果. Kolmogorov的回答是: 乱打(即各打各的). 本题试图给Kolmogorov的回答一个解释.

令 C_i 表示第 i 架飞机被击中,

$$\xi_i := 1_{C_i}.$$

则被击中的飞机数为 $\xi := \sum_{i=1}^n \xi_i$. 所以

$$E[\xi] = E\left[\sum_{i=1}^n \xi_i\right] = \sum_{i=1}^n E[\xi_i].$$

由于每架飞机的地位相同, 所以 $E[\xi_1] = \cdots = E[\xi_n]$. 因此 $E[\xi] = nE[\xi_1]$.

下面求 $E[\xi_1]$. 令

$$A_k := \{\text{第}k\text{门炮瞄准第1架飞机}\},$$

$$B_k := \{\text{第}k\text{门炮命中第1架飞机}\},$$

则

$$P(A_k) = \frac{1}{n}, \quad P(B_k|A_k) = \alpha, \quad k = 1, 2, \dots, m.$$

注意 $A_1B_1, A_2B_2, \dots, A_mB_m$ 独立, 因此 $(A_1B_1)^c, (A_2B_2)^c, \dots, (A_mB_m)^c$ 也独立. 且

$$C_1 = \bigcup_{k=1}^m A_k B_k, \quad P(C_1^c) = \prod_{k=1}^m (A_k B_k)^c$$

所以

$$\begin{aligned} P(C_1) &= 1 - P(C_1^c) \\ &= 1 - P\left(\bigcap_{k=1}^m (A_k B_k)^c\right) \\ &= 1 - \prod_{k=1}^m P((A_k B_k)^c) \\ &= 1 - \prod_{k=1}^m (1 - P(A_k B_k)) \\ &= 1 - \prod_{k=1}^m (1 - P(A_k)P(B_k|A_k)) \\ &= 1 - \left(1 - \frac{\alpha}{n}\right)^m. \end{aligned}$$

⁴Vladimir Igorevich Arnold, 1937-2010, 苏联-俄罗斯数学家, Kolmogorov的学生.

所以

$$E[\xi_1] = 1 - \left(1 - \frac{\alpha}{n}\right)^m.$$

于是

$$E[\xi] = n \left(1 - \left(1 - \frac{\alpha}{n}\right)^m\right).$$

因为当 n 大时

$$\left(1 - \frac{\alpha}{n}\right)^m \approx 1 - m\frac{\alpha}{n},$$

所以当 n 大时,

$$E[\xi] \approx m\alpha.$$

而理想的情况是, 通过相互协调, 安排每门高炮瞄准不同的飞机, 击中飞机数的期望最好也是 $m\alpha$ 架. 所以各自为战和相互协调的最好效果差不多. 况且相互协调是需要时间的, 因此协调过程中很可能贻误战机. 所以Kolmogorov的答案是各自为战.

7. 设一个班有两个组, A 组有 n 个学生, B 组有 m 个学生. 教师每次从学生中随机抽查一个人的作业. 问: 若要抽查到 A 组中 r 个不同学生的作业, 需要抽查的总次数的期望是多少?

以 $\xi_1, \xi_2, \dots, \xi_r$ 记依次抽查到 A 组中第 $1, 2, \dots, r$ 新学生需要的次数 (即: ξ_i 表示在 A 组中抽到第 $i-1$ 个学生开始算起, 再抽到第 i 个新学生需要的次数). 则抽查出 r 个学生需要的次数为

$$\xi = \xi_1 + \xi_2 + \dots + \xi_r.$$

令

$$p_i := \frac{n-i+1}{m+n}, \quad q_i := 1-p_i, \quad i=1, 2, \dots, r,$$

则 p_i 表示 A 组中抽到第 $i-1$ 个学生后, 下次再从 A 组中抽第 i 个学生的概率. 因此

$$P(\xi_1 = k) = q_1^{k-1}p_1, \quad k=1, 2, \dots,$$

一般地有

$$P(\xi_i = k) = q_i^{k-1}p_i, \quad k=1, 2, \dots.$$

所以 ξ_i 服从参数为 p_i 的几何分布, 且

$$E[\xi_i] = \frac{1}{p_i}.$$

所以

$$\begin{aligned} E[\xi] &= E[\xi_1 + \dots + \xi_r] \\ &= \sum_{i=1}^r \frac{m+n}{n-i+1} \\ &= (m+n) \sum_{i=1}^r \frac{1}{n-i+1}. \end{aligned}$$

我们看一个特殊情况.

若 $r = n$ (即要抽查到 A 组所有同学), 则

$$E[\xi] = (m+n) \sum_{i=1}^n \frac{1}{i}.$$

由于

$$1 + \frac{1}{2} + \cdots + \frac{1}{n} = \ln n + C + \varepsilon_n,$$

其中 C 是Euler常数:

$$C = 0.5772 \dots \dots ,$$

而

$$\lim_{n \rightarrow \infty} \varepsilon_n = 0.$$

特别地

$$\lim_{n \rightarrow \infty} \frac{1}{\ln n} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n} \right) = 1.$$

所以对大的 n ,

$$E[\xi] \approx (m+n) \ln n.$$

若 $m = n, r = n$, 则对大 n ,

$$E[\xi] \approx 2n \cdot \ln n.$$

8. 设 ξ 的分布列为

$$P\left(\xi = (-1)^k \frac{2^k}{k}\right) = \frac{1}{2^k}, \quad k = 1, 2, \dots,$$

求 $E[\xi]$.

我们有

$$P\left(\xi^+ = \frac{2^{2k}}{2k}\right) = \frac{1}{2^{2k}},$$

$$P\left(\xi^- = \frac{2^{2k+1}}{2k+1}\right) = \frac{1}{2^{2k+1}}.$$

虽然

$$\sum_{k=1}^{\infty} (-1)^k \frac{2^k}{k} \frac{1}{2^k} = \sum_{k=1}^{\infty} (-1)^k \frac{1}{k} = -\ln 2,$$

但由于

$$E[\xi^+] = \sum_{k=1}^{\infty} \frac{1}{2k} = \infty,$$

$$E[\xi^-] = \sum_{k=1}^{\infty} \frac{1}{2k+1} = \infty,$$

故 $E[\xi]$ 不存在.

9. 设 $x \sim U(a, b)$ 为均匀分布, 密度函数为

$$p(x) = \frac{1}{b-a} 1_{[a,b]}(x).$$

此时

$$E[\xi] = \int_a^b \frac{1}{b-a} x dx = \frac{a+b}{2}.$$

10. 设 $\xi \sim N(\mu, \sigma^2)$ 为正态分布.

我们知道, 此时密度函数为

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

所以

$$\begin{aligned} E[\xi] &= \int_{-\infty}^{\infty} xp(x)dx \\ &= \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z + \mu) e^{-\frac{z^2}{2}} dz \\ &= \mu \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \\ &= \mu, \end{aligned}$$

亦即 $N(a, \sigma^2)$ 中的 μ 就是期望.

11. 设 $\xi \sim N(0, 1)$. 对任意 $t \in \mathbb{R}$, 求 $E[\xi^n]$.

我们有

$$\begin{aligned} E[\exp(t\xi)] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{x^2}{2}} dx \\ &= e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-t)^2} dx \\ &= e^{\frac{t^2}{2}} = \sum_{n=0}^{\infty} \frac{t^{2n}}{2^n n!}. \end{aligned}$$

由于

$$E[\exp(t\xi)] = \sum_{n=0}^{\infty} \frac{t^n}{n!} E[\xi^n],$$

比较 t^n 的系数即得

$$E[\xi^n] = \begin{cases} 0, & n \text{ 为奇数} \\ (2n-1)!!, & n \text{ 为偶数.} \end{cases}$$

12. $\xi \sim E(\lambda)$ 为指数分布, 密度函数为

$$p(x) = \lambda e^{-\lambda x} 1_{x \geq 0}.$$

所以

$$E[\xi] = \int_0^{\infty} \lambda x e^{-\lambda x} dx = - \int_0^{\infty} x d e^{-\lambda x} = \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda}.$$

而对 $t \in (0, \lambda)$,

$$E[e^{t\xi}] = \int_0^{\infty} \lambda e^{-(\lambda-t)x} dx = \frac{\lambda}{\lambda-t}.$$

同样比较系数即得

$$E[\xi^n] = \lambda^{-n}.$$

13. ξ 服从 *Cauchy* 分布, 密度函数为

$$p(x) = \frac{1}{\pi} \frac{1}{1+x^2}.$$

因为

$$E[\xi^+] = E[\xi^-] = \int_0^{\infty} |x| \frac{1}{\pi(1+x^2)} dx = \infty.$$

所以 ξ 的期望不存在.

14. 设 ξ 的分布密度为

$$p(x) = \begin{cases} \frac{1}{\pi(1+x^2)} & x \geq 0, \\ \frac{1}{2}e^x & x < 0, \end{cases}$$

求 $E[\xi]$

此时,

$$E[\xi^+] = \infty, \quad E[\xi^-] = \frac{1}{2}.$$

所以

$$E[\xi] = \infty.$$

习题

1. 证明: 对任意 n ,

$$\begin{aligned} & \sum_{k=-\infty}^{\infty} k2^{-n} P(\xi \in [k2^{-n}, (k+1)2^{-n})) \\ & \leq E[\xi] \leq \sum_{k=-\infty}^{\infty} (k+1)2^{-n} P(\xi \in [k2^{-n}, (k+1)2^{-n})), \end{aligned}$$

且

$$\left| E[\xi] - \sum_{k=-\infty}^{\infty} k2^{-n} P(\xi \in [k2^{-n}, (k+1)2^{-n})) \right| \leq 2^{-n},$$

$$\left| E[\xi] - \sum_{k=-\infty}^{\infty} (k+1)2^{-n} P(\xi \in [k2^{-n}, (k+1)2^{-n})) \right| \leq 2^{-n}.$$

2. 一个班级有50个学生, 某次课后依次随机地取回作业本.

(a) 一共会有多少种不同的结果?

(b) 假设所有结果都是等可能的. 如果最后一个学生发现取回的作业本不是自己的, 他或她就设法找到那个拿了他或她作业本的人, 要求交换; 如果这个被要求交换的人发现交换到的作业本不是自己的, 也同样设法找到拿了他或她作业本的人, 要求交换. 这样逐次进行下去, 直到某次被要求交换的人刚好拿到了自己的作业本为止. 求交换总次数的数学期望.

3. 有两桶酒一样多, 一桶白酒, 一桶红酒. 一个大妈从红酒桶里舀出一满瓢倒入白酒桶中, 用瓢在桶里一转悠, 然后再从白酒桶里舀出一满瓢倒入红酒桶中, 再一转悠. 最后, 从两只酒桶里各舀出一瓢. 问: 红酒桶里舀出的那一瓢中的白酒成分的期望值和白酒桶里舀出的那一瓢中的红酒成分的期望值谁大谁小?

4. 设 $p_n \geq 0$. 令 $r_n := \sum_{k=n}^{\infty} p_k$. 证明:

$$\sum_{n=1}^{\infty} np_n = \sum_{n=1}^{\infty} r_n,$$

即两边同时收敛或同时发散, 且收敛时收敛到同一极限. 由此证明若 ξ 是取非负整数值的随机变量, 则

$$E[\xi] = \sum_{n=1}^{\infty} P(\xi \geq n).$$

5. 设 $\{c_n\}$ 是单调上升到正无穷的正数列, 且存在 c 使得 $c_{n+1} \leq cc_n, \forall n$. 设 $\xi \geq 0$, 证明 $E[\xi]$ 有限等价于

$$\sum_{n=1}^{\infty} c_n P(\xi \in [c_n, c_{n+1})) < \infty,$$

又等价于

$$\sum_{n=1}^{\infty} P(\xi \geq c_n) < \infty.$$

6. 设 $p > 0$ 且

$$\lim_{x \rightarrow +\infty} x^p P(|\xi| > x) = 0.$$

证明 $E[|\xi|^r] < \infty, \forall r \in [0, p)$, 并举例说明 r 不能等于 p .

7. 赌注加倍问题. 一个人参加一场赌博, 他的计划是: 赢了就离开, 输了就将赌注加倍继续赌, 直到赌资不够用(即不到上一盘赌资的2倍)为止. 如果他手上有250元钱, 从1元钱起赌. 假设每盘他赢的概率都是 $\frac{1}{2}$. 问他在这场赌博中的预期收益是多少?

8. 从 $0, 1, \dots, n$ 中随机取一个数, 每数被取到的概率依次为 p_0, p_1, \dots, p_n . 不放回地连续取 k 次. 求取到的数之和的期望.

9. 设一个城市有1千万人. 现在怀疑其中有1000人感染了新冠而进行全员核酸. 问: 在第一次查到感染者之前, 预期要做多少人的核酸? 为查出所有的感染者, 预期有要做多少人的核酸?

10. 袋中装了 N 个编号为 $1, 2, \dots, N$ 的球, 摸出一个, 记下号码, 再放回去. 这样进行下去. 求:

- (a) 第一次取出的那个球, 预期还需要摸多少次才能再次取出?
 (b) 预期需要摸多少次, 才会出现一个球被摸到了两次的情况?
 (c) 设摸了 n 次. 求摸到的最大号码和最小号码的联合概率分布.
 (d) 改为无放回地摸了 n 次, 求摸到的最大号码和最小号码的联合概率分布.

5.4 随机变量列的收敛性

下节我们将研究积分号下取极限, 即极限和期望的交换顺序问题. 由于期望就是积分, 所以这个问题我们在数学分析里已经有所接触, 但那时的条件一般都是被积函数一致收敛, 这无疑是太强了. 实际上对随机变量来说, 逐点收敛都太强了, 所以我们将在本节寻找一些较弱的条件. 考察一下随机变量序列的各种不同的收敛性.

设 $\{\xi_n\}$ 是定义在概率空间 (Ω, \mathcal{F}, P) 上的随机变量列. 在考察其收敛性时, 最容易想到的收敛性就是移植数学分析里面的逐点收敛, 即对每个 ω , $\xi_n(\omega)$ 都收敛. 但由于在概率论中, 我们可以忽略概率为零的事件(可略集), 所以可以允许在某个可略集上不收敛. 因此, 我们有下面的定义:

定义 5.4.1. 设 $\xi_n, \xi, n = 1, 2, \dots$ 为随机变量. 若存在 $A \in \mathcal{F}, P(A) = 0$, 使得

$$\lim_{n \rightarrow \infty} \xi_n(\omega) = \xi(\omega), \quad \forall \omega \notin A,$$

则称 ξ_n 几乎必然收敛于 ξ , 记为 $\lim_{n \rightarrow \infty} \xi_n = \xi$ a.s., 或 $\xi_n \xrightarrow{a.s.} \xi$.

几乎必然收敛当然是一件很美好的事情, 不过这个要求非常苛刻, 往往难以达到. 再说概率论里面关心的往往不是每个样本的性质, 而是全局的性质. 这样, 就出现了下面的概念.

定义 5.4.2. 设 $\xi_n, \xi, n = 1, 2, \dots$ 为随机变量. 若 $\forall \varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P(|\xi_n - \xi| > \varepsilon) = 0,$$

则称 ξ_n 依概率收敛于 ξ , 记为 $\lim_{n \rightarrow \infty} \xi_n = \xi(P)$, 或 $\xi_n \xrightarrow{P} \xi$.

几乎必然收敛序列的极限无疑是几乎必然唯一的. 即若同时有 $\xi_n \xrightarrow{a.s.} \xi, \xi_n \xrightarrow{a.s.} \xi'$, 则 $\xi = \xi'$ a.s., 这是因为

$$\{\omega : \xi(\omega) \neq \xi'(\omega)\} \subset \{\omega : \xi_n(\omega) \not\rightarrow \xi(\omega)\} \cup \{\omega : \xi_n(\omega) \not\rightarrow \xi'(\omega)\},$$

所以

$$P(\{\omega : \xi(\omega) \neq \xi'(\omega)\}) \leq P(\{\omega : \xi_n(\omega) \not\rightarrow \xi(\omega)\}) + P(\{\omega : \xi_n(\omega) \not\rightarrow \xi'(\omega)\}) = 0.$$

那么, 依概率收敛的极限是否也是几乎必然唯一的呢? 答案也是肯定的, 因为若同时有 $\xi_n \xrightarrow{P} \xi, \xi_n \xrightarrow{P} \xi'$, 则 $\forall n \in \mathbb{N}_{++}, \varepsilon > 0$ 有

$$\begin{aligned} \{|\xi - \xi'| > \varepsilon\} &\subset \{|\xi - \xi_n| + |\xi' - \xi_n| > \varepsilon\} \\ &\subset \left\{|\xi - \xi_n| > \frac{\varepsilon}{2}\right\} \cup \left\{|\xi' - \xi_n| > \frac{\varepsilon}{2}\right\}. \end{aligned}$$

所以

$$P(|\xi - \xi'| > \varepsilon) \leq P\left(|\xi - \xi_n| > \frac{\varepsilon}{2}\right) + P\left(|\xi' - \xi_n| > \frac{\varepsilon}{2}\right).$$

因此

$$P(|\xi - \xi'| > \varepsilon) \leq \lim_{n \rightarrow \infty} P\left(|\xi - \xi_n| > \frac{\varepsilon}{2}\right) + \lim_{n \rightarrow \infty} P\left(|\xi' - \xi_n| > \frac{\varepsilon}{2}\right) = 0.$$

这样就有

$$P(|\xi - \xi'| \neq 0) = P\left(\bigcup_{k=1}^{\infty} \{|\xi - \xi'| > k^{-1}\}\right) \leq \sum_{k=1}^{\infty} P(|\xi - \xi'| > k^{-1}) = 0.$$

这两种收敛性是什么关系呢?

ω 固定时, $\xi_n(\omega)$ 收敛到 $\xi(\omega)$, 用 $\varepsilon - \delta$ 语言来描述, 就是说: $\forall m, \exists n$, 使得 $\forall k \geq n$ 均有 $|\xi_n(\omega) - \xi(\omega)| \leq \frac{1}{m}$. 所以

$$\{\omega : \xi_n(\omega) \rightarrow \xi(\omega)\} = \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \left\{ \omega : |\xi_k(\omega) - \xi(\omega)| \leq \frac{1}{m} \right\}.$$

于是由De Morgan原理

$$\{\omega : \xi_n(\omega) \not\rightarrow \xi(\omega)\} = \bigcup_{m=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \left\{ \omega : |\xi_k(\omega) - \xi(\omega)| > \frac{1}{m} \right\}.$$

所以由概率的下连续性有

$$P(\{\omega : \xi_n(\omega) \not\rightarrow \xi(\omega)\}) = \lim_{m \rightarrow \infty} P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \left\{ \omega : |\xi_k(\omega) - \xi(\omega)| > \frac{1}{m} \right\}\right).$$

由于

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \left\{ \omega : |\xi_k(\omega) - \xi(\omega)| > \frac{1}{m} \right\}\right)$$

随 m 递增, 故 $P(\{\omega : \xi_n(\omega) \not\rightarrow \xi(\omega)\}) = 0$ 就等价于

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \left\{ \omega : |\xi_k(\omega) - \xi(\omega)| > \frac{1}{m} \right\}\right) = 0, \quad \forall m,$$

再由概率的上连续性, 这又等价于

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{k=n}^{\infty} \left\{ \omega : |\xi_k(\omega) - \xi(\omega)| > \frac{1}{m} \right\}\right) = 0, \quad \forall m. \quad (4.1)$$

这个当然意味着

$$\lim_{n \rightarrow \infty} P\left(\omega : |\xi_n(\omega) - \xi(\omega)| > \frac{1}{m}\right) = 0, \quad \forall m, \quad (4.2)$$

而这就是 $\xi_n \xrightarrow{P} \xi$. 所以我们证明了:

命题 5.4.3.

$$\xi_n \xrightarrow{a.s.} \xi \implies \xi_n \xrightarrow{P} \xi.$$

那么这两者是否实际上是等价的呢? 答案是否定的, 例子如下.

例1. 设 $\Omega = [0, 1)$, $\mathcal{F} = \text{Borel}\sigma\text{-代数}$, $P = \text{Lebesgue测度}$. 令

$$\begin{aligned}\xi_{1,1} &\equiv 1; \\ \xi_{2,k} &= 1_{[k2^{-1}, (k+1)2^{-1})}, \quad k = 0, 1; \\ \xi_{n,k} &= 1_{[k2^{-n+1}, (k+1)2^{-n+1})}, \quad k = 0, 1, \dots, 2^n - 1; \\ &\dots\dots\dots\end{aligned}$$

再将这些随机变量按从上到下, 从左到右的次序排成一列, 记为 ξ_n . 则 ξ_n 对每个 ω 都有收敛到0和1的两个不同子列, 因此不收敛. 但因为 $\forall 0 < \delta < 1$, $P(|\xi_{n,k} - 0| > \delta) \leq 2^{-n}$, 因此却是依概率收敛于零的.

(4.2)与(4.1)差多远呢?

首先注意, 如果将(4.2)加强为: $\forall m \geq 1$, 有

$$\sum_{n=1}^{\infty} P\left(|\xi_n - \xi| > \frac{1}{m}\right) < \infty, \quad (4.3)$$

那么

$$P\left(\bigcup_{k=n}^{\infty} \left\{\omega : |\xi_k(\omega) - \xi(\omega)| > \frac{1}{m}\right\}\right) \leq \sum_{k=n}^{\infty} P\left(|\xi_k - \xi| > \frac{1}{m}\right) < \infty.$$

于是(4.1)满足. 所以, 如果一个依概率收敛的随机变量列收敛的速度足够快, 那么它就是几乎必然收敛的.

现在, 若 $\xi_n \xrightarrow{P} \xi$, 那么由(4.2), $\forall k \geq 1$, 可找到 n_k , 且 n_k 随 k 而递增, 使得

$$P\left(|\xi_{n_k} - \xi| > \frac{1}{k}\right) < 2^{-k}.$$

因此, $\forall m \geq 1$,

$$\sum_{k=m}^{\infty} P\left(|\xi_{n_k} - \xi| > \frac{1}{m}\right) \leq \sum_{k=m}^{\infty} P\left(|\xi_{n_k} - \xi| > \frac{1}{k}\right) < \infty.$$

于是 ξ_{n_k} 几乎必然收敛于 ξ . 这样我们就证明了:

定理 5.4.4. 若 $\xi_n \xrightarrow{P} \xi$, 那么一定存在子列 ξ_{n_k} , 使得 $\xi_{n_k} \xrightarrow{a.s.} \xi$.

我们顺便得到:

推论 5.4.5. 设 $\xi_n \xrightarrow{P} \xi$, 且存在随机变量 η 使得 $|\xi_n| \leq \eta$ a.s., $\forall n$. 则 $|\xi| \leq \eta$ a.s.. 特别地, 当 $\eta \equiv C$, 则 $\xi \leq C$.

证明. 因 $\xi_n \xrightarrow{P} \xi$ 和定理5.4.4, 存在子列 ξ_{n_k} , 使得 $\xi_{n_k} \xrightarrow{a.s.} \xi$. 又因为 $|\xi_{n_k}| \leq \eta$ a.s., $\forall k$, 因此 $|\xi| \leq \eta$ a.s.. \square

显然, (4.1)等价于

$$\lim_{N \rightarrow \infty} P\left(\sup_{n \geq N} |\xi_n - \xi| > \varepsilon\right) = 0, \quad \forall \varepsilon > 0. \quad (4.4)$$

类似于(4.1), 我们给出一个几乎必然收敛的Cauchy准则.

引理 5.4.6. 设 $\{\xi_i\}$, ξ 是随机变量, 则 $\xi_i \xrightarrow{a.s.} \xi$ 的充要条件是

$$P\left(\bigcap_{N=1}^{\infty} \bigcup_{m,n=N}^{\infty} \{|\xi_m - \xi_n| > \varepsilon\}\right) = 0, \quad \forall \varepsilon > 0. \quad (4.5)$$

这又等价于

$$\lim_{N \rightarrow \infty} P\left(\sup_{m,n \geq N} |\xi_m - \xi_n| > \varepsilon\right) = 0, \quad \forall \varepsilon > 0; \quad (4.6)$$

证明. 因为

$$\begin{aligned} \{\omega : \xi_n(\omega) \rightarrow \xi(\omega)\} &= \{\omega : \forall k > 0, \exists N, \text{使得 } |\xi_m(\omega) - \xi_n(\omega)| \leq k^{-1}, \forall m, n \geq N\} \\ &= \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{m,n=N}^{\infty} \{|\xi_m(\omega) - \xi_n(\omega)| \leq k^{-1}\}, \end{aligned}$$

所以, $\xi_n \xrightarrow{a.s.} \xi$ 的充要条件是

$$P\left(\bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{m,n=N}^{\infty} \{|\xi_m(\omega) - \xi_n(\omega)| \leq k^{-1}\}\right) = 1,$$

即

$$P\left(\bigcup_{k=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{m,n=N}^{\infty} \{|\xi_m(\omega) - \xi_n(\omega)| > k^{-1}\}\right) = 0.$$

显然,此式满足的充要条件是

$$P\left(\bigcap_{N=1}^{\infty} \bigcup_{m,n=N}^{\infty} \{|\xi_m(\omega) - \xi_n(\omega)| > k^{-1}\}\right) = 0, \quad \forall k > 0.$$

这又等价于

$$P\left(\bigcap_{N=1}^{\infty} \bigcup_{m,n=N}^{\infty} \{|\xi_m(\omega) - \xi_n(\omega)| > \varepsilon\}\right) = 0, \quad \forall \varepsilon > 0.$$

□

5.5 积分(期望)号下取极限

现在我们就开始考虑积分号和极限的交换问题. 我们固定一个概率空间 (Ω, \mathcal{F}, P) , 所有的随机变量均定义在这个空间上.

首先我们证明一个预备结果.

引理 5.5.1. 若 $\xi \geq 0$, 则

$$\lim_{M \uparrow \infty} E[\xi \wedge M] = E[\xi].$$

证明. 若 $P(\xi = \infty) > 0$, 则 $E[\xi] = \infty$, 而

$$E[\xi \wedge M] \geq MP(\xi = \infty) \rightarrow \infty,$$

所以结论成立.

下面设 $P(\xi = \infty) = 0$. 由于 $E[\xi \wedge M]$ 是 M 的单调上升函数, 故

$$\lim_{M \uparrow \infty} E[\xi \wedge M] = \lim_{m \uparrow \infty} E[\xi \wedge m] \leq E[\xi],$$

这里 m 取整数. 因为 $\forall m \geq 1$,

$$\sum_{k=0}^m k 1_{[k, k+1)}(\xi) \leq \xi \wedge m \leq \sum_{k=0}^m (k+1) 1_{[k, k+1)}(\xi),$$

所以

$$\sum_{k=0}^m k P(\xi \in [k, k+1)) \leq E[\xi \wedge m] \leq \sum_{k=0}^m k P(\xi \in [k, k+1)) + 1.$$

因此

$$\sum_{k=0}^{\infty} k P(\xi \in [k, k+1)) \leq \lim_{m \rightarrow \infty} E[\xi \wedge m] \leq \sum_{k=0}^{\infty} k P(\xi \in [k, k+1)) + 1.$$

再注意

$$\sum_{k=0}^{\infty} k 1_{[k, k+1)}(\xi) \leq \xi \leq \sum_{k=0}^{\infty} (k+1) 1_{[k, k+1)}(\xi),$$

所以

$$\sum_{k=0}^{\infty} k P(\xi \in [k, k+1)) \leq E[\xi] \leq \sum_{k=0}^{\infty} k P(\xi \in [k, k+1)) + 1,$$

若 $E[\xi] = \infty$, 则

$$\sum_{k=0}^{\infty} k P(\xi \in [k, k+1)) = \infty.$$

从而

$$\lim_{m \rightarrow \infty} E[\xi \wedge m] = \infty.$$

若 $E[\xi] < \infty$, 则

$$\sum_{k=0}^{\infty} (k+1) P(\xi \in [k, k+1)) \leq \sum_{k=0}^{\infty} k P(\xi \in [k, k+1)) + 1 < \infty.$$

用

$$0 \leq \xi - \xi \wedge m \leq \sum_{k=m}^{\infty} (k+1-m) 1_{\xi \in [k, k+1)},$$

得到

$$\lim_{m \rightarrow \infty} (E[\xi] - E[\xi \wedge m]) \leq \lim_{m \rightarrow \infty} \sum_{k=m}^{\infty} k P(\xi \in [k, k+1)) = 0.$$

□

为书写方便记, 以后将使用记号:

$$E[\xi; A] := E[\xi 1_A].$$

首先我们可证明下面简单的

命题 5.5.2 (有界收敛定理). 设 ξ_n 依概率收敛到 ξ , 且存在常数 C 使得

$$|\xi_n| \leq C \text{ a.s.}, \forall n.$$

则

$$E[\xi_n] \rightarrow E[\xi].$$

证明. 由推论5.4.5, $|\xi_n - \xi| \leq |\xi_n| + |\xi| \leq 2C \text{ a.s.}$. 因此 $\forall \varepsilon > 0$,

$$\begin{aligned} |E[\xi_n] - E[\xi]| &\leq E[|\xi_n - \xi|] \\ &= E[|\xi_n - \xi|; |\xi_n - \xi| > \varepsilon] + E[|\xi_n - \xi|; |\xi_n - \xi| \leq \varepsilon] \\ &\leq 2CP(|\xi_n - \xi| > \varepsilon) + \varepsilon. \end{aligned}$$

所以

$$\limsup_{n \rightarrow \infty} |E[\xi_n] - E[\xi]| \leq \varepsilon.$$

由 ε 的任意性结论即得结论. □

另外一种简单情况是关于非负单调上升序列的

命题 5.5.3 (单调收敛定理). 设 $0 \leq \xi_n \uparrow \xi \text{ a.s.}$, 则 $E[\xi] = \lim_{n \rightarrow \infty} E[\xi_n]$.

证明. 若 $\sup_n E[\xi_n] = \infty$. 由于 $E[\xi_n]$ 单调上升, 所以必有

$$\lim_{n \rightarrow \infty} E[\xi_n] = \infty.$$

因 $\xi_n \leq \xi$, 所以 $E[\xi] = \infty$.

下面设 $\sup_n E[\xi_n] < \infty$. 先证

$$E[\xi] < \infty.$$

若否, 则由引理5.5.1有

$$\lim_{m \rightarrow \infty} E[\xi \wedge m] = \infty.$$

但由有界收敛定理, $\forall m \geq 1$,

$$\lim_{n \rightarrow \infty} E[\xi_n \wedge m] = E[\xi \wedge m].$$

所以必有 n 使得

$$E[\xi_n] \geq E[\xi_n \wedge m] > E[\xi \wedge m] - 1.$$

由于 m 是任意的, 这与 $\sup_n E[\xi_n] < \infty$ 矛盾, 于是 $E[\xi] < \infty$.

由于 $E[\xi_n \wedge m]$ 关于 m, n 都是单调上升的, 且有上界, 故全面极限

$$\lim_{m, n \rightarrow \infty} E[\xi_n \wedge m]$$

存在且有限. 于是, 由引理5.5.1, 两个累次极限可换序以及有界收敛定理有

$$\begin{aligned} \lim_{n \rightarrow \infty} E[\xi_n] &= \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} E[\xi_n \wedge m] \\ &= \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} E[\xi_n \wedge m] \\ &= \lim_{m \rightarrow \infty} E[\xi \wedge m] \\ &= E[\xi]. \end{aligned}$$

□

由此可以立即得到:

推论 5.5.4. 设 $\{\xi_n\}$ 单调, ξ_1 可积. 令 $\xi = \lim_{n \rightarrow \infty} \xi_n$. 则

$$E[\xi] = \lim_{n \rightarrow \infty} E[\xi_n].$$

证明. 不妨设 $\{\xi_n\}$ 单调上升(单调下降时考虑 $\{-\xi_n\}$ 即可).

因为 $\xi_n - \xi_1 \geq 0$, 所以 $E[\xi_n - \xi_1]$ 存在. 又因为 ξ_1 可积, $\xi_n = (\xi_n - \xi_1) + \xi_1$, 所以 $E[\xi_n]$ 也存在, 且

$$E[\xi_n - \xi_1] = E[\xi_n] - E[\xi_1].$$

于是

$$\begin{aligned} \lim_{n \rightarrow \infty} E[\xi_n] - E[\xi_1] &= \lim_{n \rightarrow \infty} E[\xi_n - \xi_1] \\ &= E[\xi - \xi_1] \\ &= E[\xi] - E[\xi_1]. \end{aligned}$$

所以 $\lim_{n \rightarrow \infty} E[\xi_n] = E[\xi]$.

□

由这个结果可以推出许多有用的结果.

推论 5.5.5 (Lebesgue基本定理). 设 $\xi_n, n = 1, 2, \dots$ 是随机变量, 且

$$\sum_{n=1}^{\infty} E[|\xi_n|] < \infty.$$

则 $\xi := \sum_{n=1}^{\infty} \xi_n$ 几乎必然收敛, ξ 可积, 且

$$\lim_{n \rightarrow \infty} E \left[\left| \xi - \sum_{i=1}^n \xi_i \right| \right] = 0.$$

特别地,

$$E[\xi] = \sum_{n=1}^{\infty} E[\xi_n].$$

证明. 令

$$\eta_n := \sum_{i=1}^n |\xi_i|, \quad \eta := \sum_{i=1}^{\infty} |\xi_i|.$$

则 η 可积且 $\eta_n \uparrow \eta$. 因此

$$E \left[\sum_{n=1}^{\infty} |\xi_n| \right] = \sum_{n=1}^{\infty} E[|\xi_n|].$$

所以

$$\sum_{n=1}^{\infty} |\xi_n| < \infty \text{ a.s.}$$

于是 $\xi := \sum_{n=1}^{\infty} \xi_n$ 几乎必然收敛且可积, 满足

$$\begin{aligned} E \left[\left| \xi - \sum_{i=1}^n \xi_i \right| \right] &\leq E \left[\left| \sum_{i=n+1}^{\infty} \xi_i \right| \right] \\ &\leq E \left[\sum_{i=n+1}^{\infty} |\xi_i| \right] \\ &= \sum_{i=n+1}^{\infty} E[|\xi_i|] \rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

□

我们曾经证明过, 若 $\{A_n, n = 1, 2, \dots\}$ 互不相交,

$$\xi := \sum_{n=1}^{\infty} x_n 1(A_n),$$

且 $\sum_{n=1}^{\infty} |x_n| P(A_n) < \infty$, 则

$$E[\xi] = \sum_{n=1}^{\infty} x_n P(A_n).$$

然而, 运用上面的结果, 现在易见 $\{A_n, n = 1, 2, \dots\}$ 不必互不相交, 结论依然成立.

我们还可以得到:

推论 5.5.6 (Fatou引理). (i) 设 η 可积, 且 $\forall n, \xi_n \geq \eta$ a.s.. 则

$$E \left[\liminf_{n \rightarrow \infty} \xi_n \right] \leq \liminf_{n \rightarrow \infty} E[\xi_n].$$

(ii) 设 η 可积, 且 $\forall n, \xi_n \leq \eta$ a.s.. 则

$$E \left[\limsup_{n \rightarrow \infty} \xi_n \right] \geq \limsup_{n \rightarrow \infty} E[\xi_n].$$

证明. 我们只证(i), (ii)的证明是类似的.

令 $\eta_1 = \eta$, $\eta_n := \inf_{k \geq n} \xi_k$, $n \geq 2$. 则 $\{\eta_n\}$ 单调上升. 于是

$$\begin{aligned} E[\liminf_n \xi_n] &= E[\lim_n \eta_n] = \lim_n E[\eta_n] \\ &\leq \lim_n \inf_{k \geq n} E[\xi_k] = \liminf_n E[\xi_n]. \end{aligned}$$

□

使用这个推论的第二个结论就可以得到下面十分重要的

推论 5.5.7 (控制收敛定理). 设 $\xi_n \xrightarrow{P} \xi$, 且存在 $\eta \geq 0$, $E[\eta] < \infty$, 使得对任意 n , $|\xi_n| \leq \eta$ a.s.. 则

$$\lim_{n \rightarrow \infty} E[|\xi - \xi_n|] = 0.$$

特别地, 若 $\xi_n \xrightarrow{P} \xi$ 且存在常数 $C > 0$ 使得 $|\xi_n| \leq C$ a.s., $\forall n$, 则

$$\lim_{n \rightarrow \infty} E[|\xi - \xi_n|] = 0.$$

证明. 先设 $\lim_{n \rightarrow \infty} \xi_n = \xi$ a.s.. 因为

$$|\xi_n - \xi| \leq 2\eta.$$

所以由Fatou引理

$$\limsup_{n \rightarrow \infty} E[|\xi_n - \xi|] \leq E \left[\limsup_{n \rightarrow \infty} |\xi_n - \xi| \right] = 0.$$

再看一般情况. 用反证法. 设有子列 $\{n_k\}$ 及正常数 ε 使得

$$E[|\xi - \xi_{n_k}|] \geq \varepsilon, \quad \forall k. \quad (5.7)$$

在 $\{n_k\}$ 中再取子列 $\{m_k\}$ 使得 $\lim_{k \rightarrow \infty} \xi_{m_k} = \xi$ a.s.. 则

$$\lim_{k \rightarrow \infty} E[|\xi - \xi_{m_k}|] = 0,$$

而这与(5.5.8)矛盾. □

从上面的证明中可以看出, 当条件是 $\xi_n \xrightarrow{a.s.} \xi$ 时, 控制收敛定理就是Fatou引理第二个结论的直接结果. 但因为给的条件是 $\xi_n \xrightarrow{P} \xi$, 因此需要通过反证法选取矛盾子列过渡一下. 同理, 一样的技巧用于Fatou引理的第一个结论, 就可以得到下面常常是相当有用的推论.

推论 5.5.8. 设 $\xi_n \geq 0$, $\xi_n \xrightarrow{P} \xi$, 且

$$\liminf_{n \rightarrow \infty} E[\xi_n] < \infty,$$

则 ξ 可积且

$$E[\xi] \leq \liminf_{n \rightarrow \infty} E[\xi_n].$$

习题

1. 设 $E[|\xi|] < \infty$. 证明:

$$\lim_{x \rightarrow \infty} xP(|\xi| \geq x) = 0.$$

2. 设 $\xi_n \xrightarrow{P} \xi$ 且

$$\lim_{m, n \rightarrow \infty} E[|\xi_n - \xi_m|] = 0.$$

证明

$$\lim_{n \rightarrow \infty} E[|\xi_n - \xi|] = 0.$$

3. 设 ξ 可积. 证明: $\forall \varepsilon > 0, \exists \delta > 0$, 使得

$$P(A) < \delta \Rightarrow E[|\xi|1_A] < \varepsilon.$$

4. 举例说明, 在推论5.5.8中, $\xi_n \geq 0$ 这个条件不能去掉.

5. 设 $\{A_n\}$ 是一列事件. 证明:

$$P(\liminf_{n \rightarrow \infty} A_n) \leq \liminf_{n \rightarrow \infty} P(A_n) \leq \limsup_{n \rightarrow \infty} P(A_n) \leq P(\limsup_{n \rightarrow \infty} A_n).$$

6. 设 $\xi, \xi_n, n = 1, 2, \dots$ 是随机变量, $\varepsilon_n, n = 1, 2, \dots$ 是正数, 使得

$$\lim_{n \rightarrow \infty} \varepsilon_n = 0,$$

$$\sum_{n=1}^{\infty} P(|\xi_n - \xi| \geq \varepsilon_n) < \infty.$$

证明: $\xi_n \xrightarrow{a.s.} \xi$.

7. 设 $\xi_n \xrightarrow{P} \xi$. 证明存在子列 $\{\xi_{n_k}\}$ 使得 $\xi_{n_k} \xrightarrow{a.s.} \xi$.

8. 设 ξ_n, ξ 为 (Ω, \mathcal{F}, P) 上的非负可积随机变量. 利用

$$(\xi - \xi_n)^+ \leq \xi,$$

证明: $E|\xi_n - \xi| \rightarrow 0$ 当且仅当

$$\xi_n \xrightarrow{P} \xi \quad \text{且} \quad E\xi_n \rightarrow E\xi.$$

9. 证明如下的Pratt引理: 设 $\xi_n, \eta_n, \zeta_n, \xi, \eta, \zeta$ 均为可积随机变量. 若

$$\xi_n \xrightarrow{P} \xi, \eta_n \xrightarrow{P} \eta, \zeta_n \xrightarrow{P} \zeta,$$

$$\xi_n \leq \eta_n \leq \zeta_n,$$

$$E\xi_n \rightarrow E\xi, E\zeta_n \rightarrow E\zeta,$$

则

(a)

$$E\eta_n \rightarrow E\eta.$$

(b)

$$\xi_n \leq 0 \leq \zeta_n \implies E|\eta_n - \eta| \rightarrow 0.$$

10. 对每个定理, 逐一构造反例, 说明定理的条件遭到破坏时, 结论不再成立.

11. 构造非负随机变量列 ξ_n , 使得

$$E[\liminf_{n \rightarrow \infty} \xi_n] < \liminf_{n \rightarrow \infty} E[\xi_n].$$

12. 设 ξ_n 是非负随机变量, $\xi_n \xrightarrow{P} \xi$, $E[\xi_n] \rightarrow E[\xi]$.

(a) 证明:

$$E[\xi_n \wedge \xi] \rightarrow E[\xi].$$

(b) 证明:

$$E[\xi_n \vee \xi] \rightarrow E[\xi].$$

(c) 证明:

$$E[|\xi_n - \xi|] \rightarrow 0.$$

13. 设 $\{\xi_n\}$ 一致可积的随机变量, 即满足:

$$\lim_{c \rightarrow \infty} \sup_n E[|\xi_n| 1_{|\xi_n| \geq c}] = 0.$$

再设 $\xi_n \xrightarrow{P} \xi$. 证明:

(a)

$$\sup_n E[|\xi_n|] < \infty;$$

(b)

$$E[\liminf_n \xi_n] \leq \liminf_n E[\xi_n] \leq \limsup_n E[\xi_n] \leq E[\limsup_n \xi_n];$$

(c)

$$E[|\xi|] < \infty;$$

(d) $\forall \varepsilon > 0, \exists \delta > 0$, 使得

$$P(A) < \delta \implies \sup_n E[|\xi_n| 1_A] < \varepsilon;$$

(e) $\{\xi_n - \xi\}$ 一致可积;

(f)

$$\lim_n E[|\xi_n - \xi|] = 0.$$

14. 设 $\varphi: \mathbb{R}_+ \mapsto \mathbb{R}_+$, 且 $\lim_{x \rightarrow \infty} x^{-1}\varphi(x) = \infty$. 证明: 若

$$\sup_n E[\varphi(|\xi_n|)] < \infty,$$

则 $\{\xi_n\}$ 一致可积.

5.6 Fubini定理

设 $(\Omega_i, \mathcal{F}_i, P_i)$ 为概率空间, $i = 1, 2$. 令

$$(\Omega, \mathcal{F}, P) = (\Omega_1, \mathcal{F}_1, P_1) \times (\Omega_2, \mathcal{F}_2, P_2).$$

就像数学分析里重积分可以化为累次积分进行计算一样, 期望也有这样的性质. 即我们有下面的

定理 5.6.1 (Fubini⁵定理). 设 f 是 (Ω, \mathcal{F}, P) 上的可积随机变量. 则

$$E[f(\omega_1, \omega_2)] = E[E[f(x, \omega_2)]|_{x=\omega_1}]. \quad (6.8)$$

证明. 只需对非负随机变量证明, 一般情形可分别考虑正负部.

(i) 若

$$f = 1_{A_1 \times A_2},$$

其中 $A_i \in \mathcal{F}_i$. 则

$$\begin{aligned} (6.8) \text{的左端} &= P(A_1 \times A_2) = P_1(A_1) \times P_2(A_2) \\ &= E[1_{A_1}(\omega_1)]P(A_2) = E[1_{A_1}(\omega_1)P(A_2)] \\ &= E[E[1_{A_1 \times A_2}(x, \omega_2)]|_{x=\omega_1}] = \text{右端}. \end{aligned}$$

(ii) 令

$$\mathcal{G} := \{B \in \mathcal{F} : (6.8) \text{对 } f := 1_B \text{ 成立.}\}$$

则由(i), \mathcal{G} 包含所有形如 $A_1 \times A_2$ 的集合, 而这些集合全体构成 π -类. 由期望的线性性知 \mathcal{G} 对真差封闭, 由单调收敛定理知 \mathcal{G} 对单调上升的序列封闭. 因此 \mathcal{G} 为 λ -类. 于是由 $\pi - \lambda$ -定理, $\mathcal{G} = \mathcal{F}$.

(iii) 由期望的线性性, 知对任何有限个取值的简单随机变量 f , (6.8) 成立.

(iv) 由于任何非负随机变量可由非负简单随机变量单调上升地逼近, 由单调收敛定理知(6.8)对任意非负简单随机变量 f 成立. \square

特别地, 设 (Ω, \mathcal{F}, P) 为任一概率空间, $\Omega' = [0, 1]$, $\mathcal{F}' = \mathcal{B}([0, 1])$, $P' = \text{Lebesgue测度}$. 在此两空间的乘积空间上用上述定理, 就有

$$\int_0^1 E[\xi(t)]dt = E \int_0^1 \xi(t)dt.$$

完全类似地, 考虑整个 \mathbb{R} 上的 Lebesgue 测度, 就有

定理 5.6.2. 设 $\xi(t, \omega)$ 为 (t, ω) 的非负二元可测函数, 则

$$\int_{\mathbb{R}} E[\xi(t)]dt = E \left[\int_{\mathbb{R}} \xi(t)dt \right].$$

习题

1. 设 (ξ, η) 是二维连续型随机变量, 密度函数为 $p(x, y)$. 用 Fubini 定理重新证明: ξ 与 η 的也为连续型随机变量, 密度函数分别为

$$p_1(x) := \int_{-\infty}^{\infty} p(x, y)dy,$$

与

$$p_2(y) := \int_{-\infty}^{\infty} p(x, y)dx.$$

2. 设数列 $\{x_{ij}\}$ 满足 $\sum_{i,j} |x_{ij}| < \infty$. 用 Fubini 定理证明

$$\sum_{i,j} x_{ij} = \sum_i \sum_j x_{ij} = \sum_j \sum_i x_{ij}.$$

并举例说明 $\sum_{i,j} |x_{ij}| = \infty$ 结论不成立.

⁵Guido Fubini, 1879-1943, 意大利-美国数学家.

5.7 带参数的期望

设 $I = (t_0 - \varepsilon, t_0 + \varepsilon)$ 是 \mathbb{R} 上的开区间, 而对任意 $t \in I$, $\xi(t)$ 是随机变量, 且

$$E[|\xi(t)|] < \infty, \quad \forall t \in I. \quad (7.9)$$

我们将研究函数

$$f(t) := E[\xi(t)]$$

在 t_0 处的连续性与可微性问题.

我们先看连续性.

定理 5.7.1. 设

- (i) 对任意 $t_n \rightarrow t_0$, $\xi(t_n) \xrightarrow{P} \xi(t_0)$;
- (ii) 存在可积随机变量 η 使得

$$|\xi(t)| \leq \eta \quad a.s., \quad \forall t \in I.$$

则 $f(t)$ 在 t_0 处连续.

证明. 为证 $f(t)$ 在 t_0 处连续, 由 Heine⁶ 定理, 只需证明对任意 $t_n \rightarrow t_0$, $f(t_n) \rightarrow f(t_0)$. 但这可以直接从控制收敛定理得到. \square

由此可以得到积分与期望交换顺序的条件, 即 Fubini 定理的另一种形式. 注意这里涉及的对 t 的积分为普通的 Riemann 积分.

定理 5.7.2 (Fubini 定理). 设 $-\infty < a < b < \infty$, 且

- (i) 对几乎所有的 ω , $[a, b] \ni t \mapsto \xi(t)$ 连续;
- (ii) 存在可积随机变量 η 使得

$$|\xi(t)| \leq \eta \quad a.s., \quad \forall t \in [a, b].$$

则

$$E \int_a^b \xi(t) dt = \int_a^b f(t) dt.$$

证明. 为简化记号, 不妨设 $[a, b] = [0, 1]$. 令

$$\xi_n(t) = \xi(2^{-n}[2^n t]).$$

由 Riemann 积分之定义, 有

$$\int_0^1 \xi(t) dt = \lim_{n \rightarrow \infty} \int_0^1 \xi_n(t) dt \quad a.s.,$$

所以 $\int_0^1 \xi(t) dt$ 为随机变量. 而

$$\left| \int_0^1 \xi_n(t) dt \right| \leq \int_0^1 |\xi_n(t)| dt \leq \eta.$$

⁶Heinrich Eduard Heine, 1821-1881, 德国数学家, 高斯、狄利克雷的学生.

所以由控制收敛定理,

$$\begin{aligned} E \left[\int_0^1 \xi(t) dt \right] &= \lim_{n \rightarrow \infty} E \left[\int_0^1 \xi_n(t) dt \right] \\ &= \lim_{n \rightarrow \infty} \int_0^1 E[\xi_n(t)] dt \\ &= \lim_{n \rightarrow \infty} \int_0^1 f(2^{-n} [2^n t]) dt \\ &= \int_0^1 f(t) dt. \end{aligned}$$

这里最后一个等式成立是因为 f 连续(由上一定理) □

如果 $\xi(t)$ 具有更好的光滑性, 那么 $f(t)$ 也具有更好的光滑性:

定理 5.7.3. 设

- (i) 对几乎所有的 ω , $\xi'(t) := \frac{\partial \xi(t)}{\partial t}$ 存在, $\forall t \in I$;
- (ii) 存在可积随机变量 η , 使得

$$\sup_{t \in I} |\xi'(t)| \leq \eta \text{ a.s.}$$

则 f 在 t_0 处可微, 且

$$f'(t_0) = E[\xi'(t_0)].$$

证明. 由Heine定理, 只需证明: 任取 $\varepsilon_n \rightarrow 0$,

$$\lim_{n \rightarrow \infty} \frac{f(t_0 + \varepsilon_n) - f(t_0)}{\varepsilon_n} = E[\xi'(t_0)]. \quad (7.10)$$

注意

$$\frac{1}{\varepsilon_n} (f(t_0 + \varepsilon_n) - f(t_0)) = E \left[\frac{1}{\varepsilon_n} (\xi(t_0 + \varepsilon_n) - \xi(t_0)) \right],$$

由(i), 对几乎所有的 ω ,

$$\frac{1}{\varepsilon_n} (\xi(t_0 + \varepsilon_n) - \xi(t_0)) \rightarrow \xi'(t_0).$$

又因为对几乎所有的 ω , 由微分中值定理,

$$\frac{1}{\varepsilon_n} (\xi(t_0 + \varepsilon_n) - \xi(t_0)) = \xi'(t_0 + \delta_n(\omega)), \quad 0 \leq \delta_n(\omega) \leq \varepsilon_n,$$

所以

$$\left| \frac{1}{\varepsilon_n} (\xi(t_0 + \varepsilon_n) - \xi(t_0)) \right| = |\xi'(t_0 + \delta_n(\omega))| \leq \eta,$$

所以由控制收敛定理即得(7.10). □

习题

1. 证明定理5.7.2中, 条件(ii)可用 $\{\xi(t), t \in I\}$ 一致可积代替.

5.8 Riemann-Stieltjes积分

设 F 是 \mathbb{R} 上的单调上升右连续函数, φ 是 \mathbb{R} 上的函数. 我们将仿照Riemann积分定义所谓Riemann-Stieltjes积分. 这个积分粗略地说, 是以Riemann-Stieltjes和

$$I_n(\mathcal{P}, \Theta) := \sum \varphi(\theta_i)(F(x_i) - F(x_{i-1}))$$

代替Riemann和

$$\sum \varphi(\theta_i)(x_i - x_{i-1}),$$

然后取极限得到的量, 其中 $\theta_i \in [x_{i-1}, x_i]$, 且这个极限必须不依赖于 θ_i 和 x_i 的具体选择.

为了不陷入复杂的可积性问题, 我们假设 φ 是连续函数.

先看有限区间上的积分.

设 $-\infty < a < b < \infty$. 考虑 $(a, b]$ 的分割 $\mathcal{P} := \{x_0, x_1, \dots, x_n\}$, 即

$$a = x_0 < x_1 < \dots < x_n = b.$$

以 $\Delta = \max_{i=1, \dots, n} |x_i - x_{i-1}|$ 表示 \mathcal{P} 的步长, $\Theta = \{\{\theta_1, \dots, \theta_n\}, \theta_i \in [x_{i-1}, x_i]\}$ 为介点集. 做Riemann-Stieltjes和为

$$I_n(\mathcal{P}, \Theta) := \sum \varphi(\theta_i)(F(x_i) - F(x_{i-1})),$$

Darboux上和与下和分别为

$$\bar{S}_{\mathcal{P}} := \sum_{i=1}^n \max_{x_{i-1} \leq x \leq x_i} \varphi(x)(F(x_i) - F(x_{i-1})),$$

$$\underline{S}_{\mathcal{P}} := \sum_{i=1}^n \min_{x_{i-1} \leq x \leq x_i} \varphi(x)(F(x_i) - F(x_{i-1})).$$

因为 φ 在 $[a, b]$ 上一致连续, 故 $\forall \varepsilon > 0$, 当 Δ 充分小时就有

$$0 \leq \bar{S}_{\mathcal{P}} - \underline{S}_{\mathcal{P}} < \varepsilon.$$

所以下列极限存在

$$\int_{(a,b]} \varphi(x)dF(x) := \lim_{\Delta \rightarrow 0} \bar{S}_{\mathcal{P}} = \lim_{\Delta \rightarrow 0} I_n(\mathcal{P}, \theta) = \lim_{\Delta \rightarrow 0} \underline{S}_{\mathcal{P}},$$

称为 φ 对 F 的Riemann-Stieltjes积分. 注意我们这里用的记号是

$$\int_{(a,b]} \varphi(x)dF(x),$$

而非

$$\int_a^b \varphi(x)dF(x),$$

是因为在单点集 $\{a\}$ 上的积分可能并不为零. 事实上, 在单点集 $\{a\}$ 上的积分定义为

$$\int_{\{a\}} \varphi(x)dF(x) := \lim_{\varepsilon \rightarrow 0} \int_{(a-\varepsilon, a]} \varphi(x)dF(x) = \varphi(a)(F(a) - F(a-)),$$

所以只有当 F 在 a 连续时, 在 $\{a\}$ 上的积分才为零. 而定义

$$\int_{(a,b)} \varphi(x)dF(x) := \int_{(a,b]} \varphi(x)dF(x) - \int_{\{b\}} \varphi(x)dF(x) = \lim_{\varepsilon \rightarrow 0^+} \int_{(a,b-\varepsilon]} \varphi(x)dF(x).$$

在整个 \mathbb{R} 上的积分则定义为

$$\int_{\mathbb{R}} \varphi(x)dF(x) := \lim_{\substack{x \rightarrow -\infty \\ y \rightarrow \infty}} \int_{(x,y]} \varphi(u)dF(u),$$

只要右边的极限存在.

我们有:

命题 5.8.1. 设 ξ 为随机变量, F 为其分布函数, $\varphi: \mathbb{R} \mapsto \mathbb{R}$ 为连续函数. 则 $E[\varphi(\xi)]$ 存在的充要条件是 φ 对 F 的Riemann-Stieltjes积分存在, 且此时有

$$E[\varphi(\xi)] = \int_{\mathbb{R}} \varphi(x)dF(x).$$

证明. 不妨设 $\varphi \geq 0$, 一般情况可分解为正负部分开讨论.

令

$$\varphi_n(x) := \varphi(x)1(-n < x \leq n),$$

由单调收敛定理有

$$E[\varphi(\xi)] = \lim_{n \rightarrow \infty} E[\varphi_n(\xi)].$$

令

$$\varphi_{n,k}(x) := \varphi_n(2^{-k}([2^k x])).$$

故

$$0 \leq x - 2^{-k}[2^k x] \leq 2^{-k}.$$

从而由 φ 的连续性有

$$\lim_{k \rightarrow \infty} \varphi_{n,k}(x) = \varphi_n(x).$$

于是由控制收敛定理,

$$E[\varphi_n(\xi)] = \lim_{k \rightarrow \infty} E[\varphi_{n,k}(\xi)].$$

但

$$E[\varphi_{n,k}(\xi)] = \sum_{i=-n2^k}^{n2^k-1} \varphi(i2^{-k})[F((i+1)2^{-k}) - F(i2^{-k})],$$

所以

$$E[\varphi_n(\xi)] = \lim_{k \rightarrow \infty} E[\varphi_{n,k}(\xi)] = \int_{(-n,n]} \varphi(x)dF(x).$$

于是,

$$\lim_{n \rightarrow \infty} \int_{(-n,n]} \varphi(x)dF(x) = \lim_{n \rightarrow \infty} E[\varphi_n(\xi)] = E[\varphi(\xi)].$$

但 $\int_{(a,b]} \varphi(x)dF(x)$ 无论是作为 a 的函数还是 b 的函数都是单调的, 所以

$$\lim_{a \rightarrow -\infty, b \rightarrow \infty} \int_{(a,b]} \varphi(x)dF(x) = E[\varphi(\xi)],$$

即

$$\int_{\mathbb{R}} \varphi(x)dF(x) = E[\varphi(\xi)].$$

□

在上面这个命题中, 我们假定了 φ 连续. 这是由 Riemann-Stieltjes 积分的定义决定的, 因为 φ 不连续时, $\int_{\mathbb{R}} \varphi(x)dF(x)$ 有可能不存在. 例如, 若 ξ 服从集中于 0 点的单点分布, 即

$$F(x) = \begin{cases} 0, & x < 0, \\ 1, & x \geq 0, \end{cases}$$

若 $\varphi \equiv F$, 则易见 $\int_{\mathbb{R}} \varphi(x)dF(x)$ 不存在. 这时如果把这个积分理解为 Lebesgue-Stieltjes 积分, 即理解为

$$\int_{\mathbb{R}} \varphi(x)\mu(dx),$$

其中 μ 是 ξ 的分布, 命题对任意非负 Borel 可测函数就仍然是成立的, 见下节的讨论. 不过从计算的角度看, 至少当 φ 分段连续时, $E[\varphi(\xi)]$ 有更直接因而更实用的公式, 见习题 4.

这个命题是非常有用的. 它告诉我们如果只是计算 $\varphi(\xi)$ 的期望, 可不计算 $\varphi(\xi)$ 的分布, 而直接由 ξ 的分布来得到, 这会使计算量大大减少.

我们看两种特殊情况.

一. 离散型分布.

设 ξ 是离散型分布, 分布列为 $\{(x_i, p_i), i = 1, 2, \dots\}$. 设

$$\sum_{i=1}^{\infty} |\varphi(x_i)|p_i < \infty.$$

任意固定 n . 令 $I_n := \{i : x_i \in (-n, n]\}$. $\forall \varepsilon > 0$, 由于 φ 在 $[-n, n]$ 上一致连续, 所以当 m 充分大时,

$$\max_{x \in [k2^{-m}, (k+1)2^{-m}]} |\varphi(k2^{-m}) - \varphi(x)| < \varepsilon, \quad \forall k = -n2^m, \dots, n2^m - 1.$$

记 $\Delta_k^m := (k2^{-m}, (k+1)2^{-m}]$, 因此

$$\left| \sum_{x_i \in \Delta_k^m} \varphi(x_i)p_i - \varphi(k2^{-m})(F((k+1)2^{-m}) - F(k2^{-m})) \right| < \varepsilon \sum_{x_i \in \Delta_k^m} p_i.$$

于是

$$\left| \sum_{i \in I_n} \varphi(x_i)p_i - \int_{(-n,n]} \varphi(2^{-m}[2^m x])dF(x) \right| < \varepsilon \sum_{i \in I_n} p_i \leq \varepsilon.$$

令 $m \rightarrow \infty$ 得

$$\left| \sum_{i \in I_n} \varphi(x_i)p_i - \int_{(-n,n]} \varphi(x)dF(x) \right| \leq \varepsilon$$

再令 $n \rightarrow \infty$ 得

$$\left| \sum_{i=1}^{\infty} \varphi(x_i) p_i - \int_{\mathbb{R}} \varphi(x) dF(x) \right| \leq \varepsilon.$$

由 ε 的任意性得

$$\int_{\mathbb{R}} \varphi(x) dF(x) = \sum_{i=1}^{\infty} \varphi(x_i) p_i.$$

二. 连续型分布.

设 ξ 是连续型的, 密度函数为 f , φ 为一连续函数, 且 $\int_{-\infty}^{\infty} |\varphi(x)| f(x) dx < \infty$. $\forall a < b$, 令

$$x_i^n := a + \frac{i}{n}(b-a), \quad i = 0, 1, \dots, n.$$

则

$$\sum_{i=0}^{n-1} \varphi(x_i^n) (F(x_{i+1}^n) - F(x_i^n)) = \sum_{i=0}^{n-1} \varphi(x_i^n) \int_{x_i^n}^{x_{i+1}^n} f(x) dx$$

$\forall \varepsilon > 0$, 当 n 足够大时有

$$|\varphi(x_i^n) - \varphi(x)| < \varepsilon, \quad \forall x \in [x_i^n, x_{i+1}^n].$$

对 $x \in [x_i^n, x_{i+1}^n)$, 令 $h_n(x) = x_i^n$, 则有

$$\begin{aligned} & \left| \int_a^b \varphi(x) f(x) dx - \sum_{i=0}^{n-1} \varphi(x_i^n) (F(x_{i+1}^n) - F(x_i^n)) \right| \\ &= \left| \int_a^b \varphi(x) f(x) dx - \sum_{i=0}^{n-1} \varphi(x_i^n) \int_{x_i^n}^{x_{i+1}^n} f(x) dx \right| \\ &\leq \int_a^b |\varphi(x) - \varphi(h_n(x))| f(x) dx \\ &\leq \varepsilon \int_a^b f(x) dx \\ &\leq \varepsilon. \end{aligned}$$

令 $n \rightarrow \infty$ 得

$$\left| \int_a^b \varphi(x) f(x) dx - \int_{(a,b]} \varphi(x) dF(x) \right| \leq \varepsilon.$$

所以由 ε 的任意性有

$$\int_{(a,b]} \varphi(x) dF(x) = \int_a^b \varphi(x) f(x) dx.$$

最后, 令 $a \rightarrow -\infty$, $b \rightarrow \infty$ 得

$$\int_{\mathbb{R}} \varphi(x) dF(x) = \int_{-\infty}^{\infty} \varphi(x) f(x) dx.$$

也就是说, 此时 φ 对 F 的 Riemann-Stieltjes 积分可转化为 φ 的普通 Riemann 积分.

多维情况时有类似的理论. 首先, 若 F 是 $(\xi_1, \xi_2, \dots, \xi_n)$ 的分布函数, φ 是 \mathbb{R}^n 上的连续函数, 则可定义Riemann-Stieltjes积分:

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi(x_1, x_2, \dots, x_n) dF(x_1, x_2, \dots, x_n).$$

类似可证明, 当 $E[\varphi(\xi_1, \dots, \xi_n)]$ 存在的充要条件是上述积分也存在, 且此时两者相等:

$$E[\varphi(\xi_1, \dots, \xi_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi(x_1, x_2, \dots, x_n) dF(x_1, x_2, \dots, x_n).$$

特别地, 当 (ξ_1, \dots, ξ_n) 是离散型且分布列为 $\{(x_{i_1}^1, x_{i_2}^2, \dots, x_{i_n}^n), p_{i_1, \dots, i_n}\}$ 时,

$$E[\varphi(\xi_1, \dots, \xi_n)] = \sum_{i_1, \dots, i_n} \varphi(x_{i_1}^1, \dots, x_{i_n}^n) p_{i_1, \dots, i_n},$$

当然, 此时不需要 φ 连续, 而只要满足

$$\sum_{i_1, \dots, i_n} |\varphi(x_{i_1}^1, \dots, x_{i_n}^n)| p_{i_1, \dots, i_n} < \infty$$

便足矣. 而当 (ξ_1, \dots, ξ_n) 是连续型且有密度函数 f 时, 我们有

$$E[\varphi(\xi_1, \dots, \xi_n)] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \varphi(x_1, x_2, \dots, x_n) f(x_1, x_2, \dots, x_n) dx_1 \cdots dx_n.$$

从上面的结果可以看出, 如果两个随机变量 ξ 和 η 同分布, 那么对任意连续函数 φ , $\varphi(\xi)$ 与 $\varphi(\eta)$ 都有相同的期望. 反过来, 我们也可以证明, 如果对足够多的 f , $f(\xi)$ 与 $f(\eta)$ 的期望仍然相等, 那么 ξ 和 η 就没得选择, 必须同分布. 为精确地阐述这一结果, 我们引进函数类:

$$C_0^\infty(\mathbb{R}^m) := \{f : \mathbb{R}^m \mapsto \mathbb{R} : f \text{ 无穷次可微且在某个有界集外恒等于零}\},$$

$m = 1$ 时, 简记 $C_0^\infty := C_0^\infty(\mathbb{R})$. 同时, 令

$$C_0(\mathbb{R}^m) := \{f : \mathbb{R}^m \mapsto \mathbb{R} : f \text{ 连续且在某个有界集外恒等于零}\},$$

$$C_b(\mathbb{R}^m) := \{f : \mathbb{R}^m \mapsto \mathbb{R} : f \text{ 有界连续}\}.$$

无混淆危险时, 我们常常省略 \mathbb{R}^m . 显然

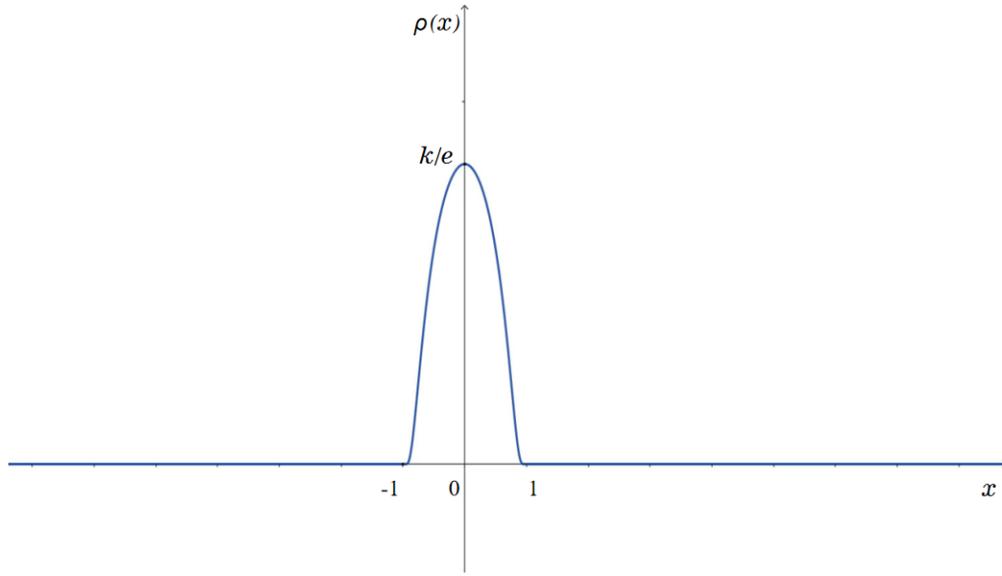
$$C_0^\infty \subset C_0 \subset C_b.$$

这样的函数是存在的, 很多的. 例如, 令

$$\rho(x) := \begin{cases} k \exp\{-\frac{1}{1-|x|^2}\} & |x| < 1, \\ 0 & |x| \geq 1, \end{cases}$$

其中 k 是常数. 则 $\rho \in C_0^\infty(\mathbb{R}^m)$. 下面我们固定 k 使得

$$\int_{\mathbb{R}^m} \rho(x) dx = 1.$$

图 5.1: 函数 $\varphi(x)$ 的图像

对任意 n , 令

$$\rho_n(x) = n^m \rho(nx).$$

则仍有 $\rho_n \in C_0^\infty(\mathbb{R}^m)$.

$\forall f \in C_0(\mathbb{R}^m)$, 令

$$f_n(x) := f * \rho_n(x) := \int_{\mathbb{R}^m} \rho_n(x-y) f(y) dy,$$

一般称 $f * g$ 为 f 与 g 的卷积. 显然 $f * g = g * f$.

由习题3, $f_n \in C_0^\infty(\mathbb{R}^m)$ 且 $f_n \rightrightarrows f$.

我们有:

定理 5.8.2. 设对任意 $f \in C_0^\infty(\mathbb{R}^m)$, $E[f(\xi)] = E[f(\eta)]$, 那么 ξ 与 η 同分布.

证明. 设 $f \in C_0$, 令

$$f_n := f * \rho_n.$$

则 $f_n \in C_0^\infty$ 且 $f_n(x) \rightarrow f(x), \forall x$. 因此由控制收敛定理,

$$E[f(\xi)] = \lim_{n \rightarrow \infty} E[f_n(\xi)] = \lim_{n \rightarrow \infty} E[f_n(\eta)] = E[f(\eta)].$$

设 $a, b \in \mathbb{R}^m$ 且 $a < b$ (即 $a_i < b_i, \forall i = 1, \dots, m$), 取一系列 $f_n \in C_0$ 使得 $f_n(x) \rightarrow 1_{(a,b]}(x), \forall x$, 且 $|f_n(x)| \leq 1, \forall n \geq 1$. 以 F 和 G 分别记 ξ 和 η 的分布函数, 则由控制收敛定理

$$\begin{aligned} F(b) - F(a) &= E[1_{(a,b]}(\xi)] \\ &= \lim_{n \rightarrow \infty} E[f_n(\xi)] \\ &= \lim_{n \rightarrow \infty} E[f_n(\eta)] \\ &= E[1_{(a,b]}(\eta)] \\ &= G(b) - G(a). \end{aligned}$$

再令 $a \rightarrow -\infty$ 即可. □

习题

1. 设 ξ, η, ζ 是随机变量, 且绝对值均小于 1. 证明:

$$|E[\xi\zeta - \eta\zeta]| \leq 1 - E[\xi\eta].$$

2. 令

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$$

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy.$$

证明:

$$\frac{x}{1+x^2} \varphi(x) < 1 - \Phi(x) < \frac{1}{x} \varphi(x), \quad \forall x > 0.$$

3. $\forall f \in C_0(\mathbb{R}^m)$, 令

$$f_n(x) := \rho_n * f(x) := \int_{\mathbb{R}^m} \rho_n(x-y) f(y) dy.$$

证明 $f_n \in C_0^\infty(\mathbb{R}^m)$ 且 $f_n \rightrightarrows f$.

4. 设 ξ 的分布函数为 F , φ 为 \mathbb{R} 上的非负 Borel 函数. 设有

$$-\infty < \cdots < x_{-n} < \cdots < x_{-1} < x_0 < x_1 < \cdots < x_n < \cdots < \infty$$

使得 φ 与 F 在所有 (x_n, x_{n+1}) 内都是连续的. 证明:

$$E[\varphi(\xi)] = \sum_{n=-\infty}^{\infty} \varphi(x_n) (F(x_n) - F(x_{n-})) + \sum_{n=-\infty}^{\infty} \int_{(x_n, x_{n+1})} \varphi(x) dF(x).$$

5. 设 ξ, η 是 m -维随机变量, 且

$$E[f(\xi)] = E[f(\eta)] \quad \forall f \in C_0^\infty(\mathbb{R}^m).$$

证明 ξ 与 η 同分布.

6. 设 ξ, η 的期望存在. 证明:

$$E[\xi - \eta] = \int_{-\infty}^{\infty} [P(\eta < x \leq \xi) - P(\xi < x \leq \eta)] dx.$$

7. 设 ξ 是可积随机变量. 证明:

(a) $\forall a \geq 0$,

$$E[\xi 1_{\xi \geq a}] = aP(\xi \geq a) + \int_a^\infty P(\xi \geq x) dx;$$

(b) $\forall a \leq 0$,

$$E[\xi 1_{\xi \leq a}] = aP(\xi \leq a) + \int_{-\infty}^a P(\xi \leq x) dx;$$

(c)

$$\lim_{|a| \rightarrow \infty} \left(|a|P(|\xi| \geq |a|) + \int_{|a|}^{\infty} P(|\xi| \geq x) dx \right) = 0.$$

8. 设 ξ 是可积随机变量, F 是其分布函数, φ 是 \mathbb{R} 上有界的分段连续可微函数. 证明

$$E[\varphi(\xi)] = - \int_{\mathbb{R}} \varphi'(x)F(x)dx.$$

5.9 Lebesgue-Stieltjes积分

回顾4.2节中分布的概念. 设 (Ω, \mathcal{F}, P) 为概率空间, ξ 为其上的 \mathbb{R}^m -值随机变量. 对 $A \in \mathcal{B}$, 令

$$\mu(A) := P(\xi^{-1}(A)),$$

则容易验证, μ 为 $(\mathbb{R}^m, \mathcal{B}^m)$ 上的概率, 因此 $(\mathbb{R}^m, \mathcal{B}^m, \mu)$ 为概率空间.

在概率空间 $(\mathbb{R}^m, \mathcal{B}^m, \mu)$ 上, 随机变量就是Borel函数. 因此对 $f \in \mathcal{B}^m$, 可定义

$$\int_{\mathbb{R}} f(x)d\mu := E^0[f].$$

这里 E^0 表示概率空间 $(\mathbb{R}^m, \mathcal{B}^m, \mu)$ 上的期望. 这样定义出来的积分称为 F 的Lebesgue-Stieltjes积分.

我们有

命题 5.9.1.

$$E[f(\xi)] = E^0[f], \quad \forall f \in \mathcal{B}^m,$$

即上述等式两边要么都有意义, 要么都没意义, 且有意义时两者相等.

证明. 第一步, 设 $f = 1_A$, $A \in \mathcal{B}^m$, 则由定义, 等式成立.

第二步, 对 $f = \sum_{i=1}^n a_i 1_{A_i}$, $A_i \in \mathcal{B}^m$, 由期望的线性和第一步, 知等式仍然成立.

第三步, 对 $f \in \mathcal{B}_+^m$, 取一列有限取值的简单函数 $f_n \uparrow f$. 则由第二步和单调收敛定理, 知等式成立.

第四步, 对一般的 $f \in \mathcal{B}^m$, 有 $f = f^+ - f^-$. 所以由上一步,

$$E[f(\xi)] = E[f^+(\xi)] - E[f^-(\xi)] = E^0[f^+] - E^0[f^-] = E^0[f].$$

□

因为 \mathbb{R}^m 上的连续函数是Borel函数, 所以是 $(\mathbb{R}^m, \mathcal{B}^m, \mu)$ 上的随机变量. 因此上面的命题对连续函数成立. 但对连续函数 f , 由命题5.8.1又有

$$E[f(\xi)] = \int_{\mathbb{R}^m} f(x)dF(x),$$

所以

$$\int_{\mathbb{R}^m} f(x)d\mu = \int_{\mathbb{R}^m} f(x)dF(x).$$

进一步可以证明, 若 f 为Borel函数, 且 $|f|$ 为Riemann-Stieltjes可积, 则 f 为Lebesgue-Stieltjes可积(这是实变函数的重要结论, 例如可见[18, 定理4.7.3]). 因此为方便及记号简单起见, 我

们以后对Borel函数 f , 把上式左边的积分通通记为右边, 并约定: 当Riemman-Stieltjes积分存在时, 它就是Riemman-Stieltjes积分; 当Riemman-Stieltjes积分不存在时, 它就是Lebesgue-Stieltjes积分.

为什么要在Riemman-Stieltjes积分和Lebesgue-Stieltjes积分间换来换去?这难道不是多余的吗?

不, 不是多余的. 这是因为, Riemman-Stieltjes积分里可积性的条件比较苛刻, 因而可积函数比较少, 且积分号下取极限的条件比较苛刻, 因此运算起来受到很多限制, 很不方便. 这时候Lebesgue-Stieltjes积分就显示出它的优势了——我们看到过的, 此时任何非负Borel函数的积分都是存在的, 并且积分号和极限号交换的条件要宽松许多. 而一旦交换成功后, 我们又往往可以用Riemman-Stieltjes积分进行实际计算(如果只从定义出发, Lebesgue-Stieltjes积分除了平凡的情况外, 几乎都是不能算的). 所以Riemman-Stieltjes积分与Lebesgue-Stieltjes积分的关系就像普通公路与高速公路的关系: 你离不开高速公路跨越千山万水, 但也离不开普通公路开回家门口.

为用说明实例说明这一点, 也因为其本身就是一个重要的结果, 我们来证明:

定理 5.9.2. 设 ξ 为 n 维随机变量, 分布函数为 F , $D \in \mathcal{B}^n$. 则

$$P(\xi \in D) = \int_D dF := \int 1_D(x) dF(x).$$

当 ξ 是连续型随机变量, 密度函数为 p , 且 D 为有限条简单曲面围成的区域(开或闭均可)时,⁷ 上式右边可理解为Riemman积分 $\int_D p(x) dx$.

我们先做一点准备工作. 设 D 为 \mathbb{R}^n 中一区域. 令

$$d(x, D) := \inf\{d(x, y) : y \in D\},$$

其中

$$d(x, y) := \sqrt{\sum_{i=1}^n (x_i - y_i)^2}.$$

显然有

$$|d(x, D) - d(y, D)| \leq |x - y|.$$

特别地, d 是连续函数.

证明. 在上一命题中取 $f = 1_D$ 即得到第一个结论.

设 ξ 是连续型的, 密度函数为 p . 因为 $e^{-nd(x, D)}$ 是连续函数, 且收到 1_D , a.s. dx , 则

$$\int e^{-nd(x, D)} dF(x) = \int e^{-nd((x_1, \dots, x_n), D)} p(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

因此, 由控制收敛定理,

$$\int 1_D(x) dF(x) = \int 1_D(x_1, \dots, x_n) p(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

当 D 为满足条件的区域时, 上述右端的被积函数只在有限条简单曲面上不连续(原来 p 的不连续点加上 D 的边界), 因而是Riemman可积的, 所以第二个结论成立. \square

⁷简单曲面是指有着参数方程 $x_1 = x_1(t_1, \dots, t_{n-1}), \dots, x_n = x_n(t_1, \dots, t_{n-1}) (a_i \leq t_i \leq b_i)$ 的曲线, 其中 $x_i(t_1, \dots, t_{n-1})$ 为连续可微函数, $\forall i = 1, \dots, n$.

理论上, 设 $\varphi: \mathbb{R}^n \mapsto \mathbb{R}$ 是连续函数, 且 $\forall x$, 区域 $D_x := \{\varphi \leq x\}$ 的边界是由有限条简单曲面构成的, 那么可以通过这个定理求出 $\varphi(\xi)$ 的分布, 因为

$$P(\varphi(\xi) \leq x) = \int_{y \in D_x} dF(y) = \int 1_{D_x}(y_1, \dots, y_n) p(y_1, \dots, y_n) dy_1 \cdots dy_n,$$

其中右边的积分为Riemman积分. 这个公式是计算随机变量的函数的分布的理论依据.

习题

1. 设一维分布函数 F 对应的概率为 μ . 证明:

$$\mu(\{x\}) = F(x) - F(x-),$$

$$\mu((x, y]) = F(y) - F(x),$$

$$\mu((x, y)) = F(y-) - F(x),$$

$$\mu([x, y]) = F(y) - F(x-),$$

$$\mu([x, y)) = F(y-) - F(x-).$$

2. 设 ξ, η 是随机变量, $F(x, y)$ 是其联合分布函数, F_1 与 F_2 是其两个边沿分布函数. 判断下列命题的正确性, 正确的给予证明, 不正确的给予反例.

- (a) 设 $(x, y) \in \mathbb{R}^2$. $P((\xi, \eta) = (x, y)) = 0$ 的充分必要条件是 F 在 (x, y) 处连续;
 (b) F 在 (x, y) 处连续时, F_1 在 x 处连续, F_2 在 y 处连续;
 (c) F_1 在 x 处连续, F_2 在 y 处连续时, F 在 (x, y) 处连续;
 (d) F 在 (x, y) 处连续的充分必要条件是: F_1 在 x 处连续与 F_2 在 y 处连续两者中至少有一个成立.

3. 以 V 表示正立方体的体积. 设 V 服从参数为 λ 的指数分布. 求其边长 l 的分布.

4. 分别举出连续型与离散型随机变量 ξ, η 的例子, 满足 $F_\xi = F_\eta$, 但 $P(\xi \neq \eta) = 1$.

5. 分别举出连续型与离散型随机变量 ξ, η, ζ 的例子, 满足 $F_\xi = F_\eta$, 但 $F_{\xi\zeta} \neq F_{\eta\zeta}$.

6. 设半径 r 的密度函数为 f , 求半径为 r 的圆的面积的分布和半径为 r 的球的体积的分布.

7. 设 $\xi \sim N(0, 1)$. 求 $e^\xi, \xi^n, |\xi|^n$ 的分布. 它们是连续型的吗?

8. 设 (ξ, η) 服从 $[0, 1]^2$ 上的均匀分布. 计算 $\xi + \eta$ 的分布.

9. 设 $\xi = (\xi_1, \dots, \xi_n) \sim N(\mu, \Sigma)$. 计算 $\sum_{i=1}^n a_i \xi_i$ 的分布, 其中 a_i 是常数.

10. 设 (ξ, η) 服从单位圆盘上的均匀分布. 计算 $\xi^2 + \eta^2$ 的分布.

11. 设 (ξ, η) 服从区域 $\{(x, y) : 0 \leq x \leq y \leq 1\}$ 上的均匀分布. 计算 ξ^n 的分布.

12. 设 X, Y, Z 为随机变量. 判断下面说法是否正确, 证明或给出反例.

- (a) X, Y 同分布时, $X + Z$ 与 $Y + Z$ 同分布;

(b) X, Y 不同分布时, $X + Z$ 与 $Y + Z$ 不同分布;

(c) X, Y 同分布时, XZ 与 YZ 同分布.

13. 设 (ξ, η) 服从 $[0, 1]^2$ 上的均匀分布. 令 $X := \xi \vee \eta, Y := \xi \wedge \eta$. 求 $P(X \leq x | Y \leq y)$.

14. 设 ξ 服从 $[0, 2\pi]$ 上的均匀分布. 求 $(\xi, \sin \xi), (\sin \xi, \cos \xi)$ 的联合分布.

15. 设 r 服从 $[0, 1]$ 上的均匀分布, θ 服从 $[0, 2\pi]$ 上的均匀分布, r 与 θ 独立. 令 $X := r \cos \theta, Y = r \sin \theta$. 求 (X, Y) 的联合分布.

16. 设 (X, Y) 服从单位圆盘上的均匀分布. 令 $r = \sqrt{X^2 + Y^2}, \theta := \text{Arg}(X, Y)$. 求 (r, θ) 的联合分布.

17. 设 ξ, η 独立同分布, 期望存在. 证明:

$$E[|\xi + \eta| - |\xi - \eta|] = \int_{-\infty}^{\infty} (1 - F(x) - F(-x))^2 dx,$$

其中 F 是 ξ 和 η 的公共分布函数.

18. 设 ξ 是非负随机变量, $r > 1$. 证明:

$$\int_0^{\infty} \frac{E[\xi \wedge x^r]}{x^r} dx = \frac{r}{r-1} E[\xi^{1/r}].$$

19. 设分布函数为

$$F(x) = \begin{cases} 0 & x \leq 0, \\ x & 0 < x \leq 1, \\ 1 & x \geq 1. \end{cases}$$

令

$$f(x) = \begin{cases} 0 & x \text{ 为有理数} \\ 1 & x \text{ 为无理数}. \end{cases}$$

证明 Riemann 积分

$$\int_0^1 f(x) dF(x)$$

不存在, 但 Lebesgue 积分

$$\int_0^1 f(x) d\mu$$

存在, 且等于 1.

5.10 方差与矩

随机变量的期望直观上是反映随机变量取值的平均大小的数字特征, 基于随机变量函数的期望可以定义出反映随机变量其他特征的量. 比如方差, 协方差, 相关系数和各种矩.

我们先看一个简单的例子.

n 人 $(\omega_1, \dots, \omega_n)$ 参加满分为100分的考试, ω_i 的成绩为 $\xi(\omega_i)$, 则平均成绩为

$$a = \frac{1}{n} \sum_{i=1}^n \xi(\omega_i)$$

每一 ω_i 的成绩偏离这个成绩的偏差为

$$\xi(\omega_i) - a$$

这引导我们考虑量

$$\frac{1}{n} \sum_{i=1}^n (\xi(\omega_i) - a)^2$$

又可写为

$$\sum_{k=0}^{100} (k - a)^2 \frac{n_k}{n},$$

其中 n_k 是得 k 分的人数. 因此上式等于

$$\sum_{k=0}^{100} (k - a)^2 P(\xi = k) = [E(\xi - E\xi)^2]$$

这就引导出一般的方差概念.

定义 5.10.1. 设随机变量 ξ 的期望有限, 方差定义为

$$D[\xi] := E[(\xi - E\xi)^2].$$

标准差定义为 $\sqrt{D[\xi]}$.

方差的另一个计算公式是:

$$D[\xi] = E[\xi^2] - (E[\xi])^2.$$

事实上

$$\begin{aligned} D[\xi] &= E[(\xi - E[\xi])^2] \\ &= E[\xi^2 - 2\xi E[\xi] + (E[\xi])^2] \\ &= E[\xi^2] - 2(E[\xi])^2 + (E[\xi])^2 \\ &= E[\xi^2] - (E[\xi])^2 \end{aligned}$$

我们常常用平均值来衡量一个随机变量的大小: 例如用平均身高衡量一个国家国民的高度, 用平均分数衡量一个班的学习成绩, 等等. 这样做的依据是什么呢? 依据就是下面的:

定理 5.10.2. 随机变量 ξ , 若期望和方差均有限, 则期望是常数中最接近 ξ 的, 在下面的意义下:

$$D[\xi] \leq E[(\xi - c)^2] \quad \forall c \in \mathbb{R}.$$

证明. 令

$$f(c) = E[(\xi - c)^2].$$

然后你立即得到 $E[\xi]$ 为 f 的最小值点(直接求导求极值也可以, 展开之后用一元二次函数的顶点也可以). □

方差还有如下基本性质:

命题 5.10.3. (i) 设 $\xi = c$ 是常数, 则

$$D[\xi] = 0.$$

(ii) 设 c 是常数, 则

$$D[\xi + c] = D[\xi]$$

(iii) 设 c 是常数, 则

$$D[c\xi] = c^2 D[\xi].$$

证明. (i) 若 $\xi = c$ 是常数, 则

$$E[\xi] = c.$$

所以

$$\xi - E[\xi] = 0.$$

于是

$$D[\xi] = 0.$$

(ii) 因为

$$E[\xi + c] = E[\xi] + c,$$

所以

$$\xi + c - (E[\xi] + c) = \xi - E[\xi].$$

从而

$$D[\xi + c] = E[(\xi + c - (E[\xi] + c))^2] = E[(\xi - E[\xi])^2] = D[\xi].$$

(iii)

$$\begin{aligned} D[c\xi] &= E[(c\xi - cE[\xi])^2] \\ &= c^2 E[(\xi - E[\xi])^2] \\ &= c^2 D[\xi]. \end{aligned}$$

□

注. 1. 设 ξ 的分布函数为 F , 那么根据上节的结果, 我们有:

$$D[\xi] = \int_{-\infty}^{\infty} (x - E[\xi])^2 dF(x).$$

2. 自然的问题是: 为什么不考虑用 $E[|\xi - E[\xi]|]$ 来衡量偏差? 回答是: 因为 $D[\xi]$ 更方便数学处理, 这跟 \mathbb{R}^n 中的常用距离是 $\sqrt{|x_1 - y_1|^2 + \cdots + |x_n - y_n|^2}$ 而不是 $|x_1 - y_1| + \cdots + |x_n - y_n|$ 是一个道理.

3. 方差中的“方”, 是平方的方.

对 $p \geq 1$, $E[|\xi|^p]$ 称为 ξ 的 p -阶矩. 由命题 5.9.1, 我们有

$$E[|\xi|^p] = \int_{-\infty}^{\infty} |x|^p dF(x).$$

我们看几个具体的例子.

例1. Bernoulli分布 $\xi \sim B(p)$.

$$\begin{aligned} E[\xi] &= p, \\ E[\xi^2] &= 1^2 \times p - 0^2 \times (1-p) = p, \\ D[\xi] &= p - p^2 = pq. \end{aligned}$$

例2. Poisson分布 $\xi \sim P(\lambda)$.

$$\begin{aligned} E[\xi] &= \lambda \\ E[\xi^2] &= \sum_{k=1}^{\infty} k^2 \times \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \sum_{k=1}^{\infty} k \times \frac{\lambda^k}{(k-1)!} e^{-\lambda} \\ &= \lambda \sum_{k=0}^{\infty} (k+1) \times \frac{\lambda^k}{k!} e^{-\lambda} \\ &= \lambda^2 + \lambda. \end{aligned}$$

$$D[\xi] = E[\xi^2] - (E[\xi])^2 = \lambda.$$

例3. 均匀分布 $\xi \sim U[a, b]$.

$$\begin{aligned} E[\xi] &= \frac{b+a}{2}, \\ E[\xi^2] &= \int_a^b x^2 \frac{1}{b-a} dx = \frac{b^2 + ab + a^2}{3}, \\ D[\xi] &= \frac{(b-a)^2}{12}. \end{aligned}$$

例4. 正态分布 $N(\mu, \sigma^2)$.

$$\begin{aligned} D[\xi] &= E[(\xi - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} dx \\ &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz \\ &= \sigma^2. \end{aligned}$$

二维随机变量 (ξ, η) 的协方差定义为.

$$\text{Cov}(\xi, \eta) := E[(\xi - E\xi)(\eta - E\eta)] = \int_{\mathbb{R}^2} (x - E[\xi])(y - E[\eta]) dF(x, y),$$

其中 F 是 (ξ, η) 的分布函数. 将 $(\xi - E\xi)(\eta - E\eta)$ 乘开后分别对每项求期望, 易见

$$\text{Cov}(\xi, \eta) = E[\xi\eta] - E[\xi]E[\eta].$$

若 $D[\xi] \cdot D[\eta] > 0$, 则 ξ 和 η 的相关系数定义为

$$\rho(\xi, \eta) := \frac{\text{Cov}(\xi, \eta)}{\sqrt{D\xi D\eta}}.$$

由Cauchy-Schwartz不等式, 我们有

$$|\rho| \leq 1.$$

若 $\text{Cov}(\xi, \eta) = 0$, 称 ξ, η 不相关. 易见 ξ, η 不相关的充要条件是

$$E[\xi\eta] = E[\xi]E[\eta].$$

例5. 设 $(\xi, \eta) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, 则 $\xi \sim N(\mu_1, \sigma_1^2)$ 和 $\eta \sim N(\mu_2, \sigma_2^2)$. 所以 $E[\xi] = \mu_1$, $D\xi = \sigma_1^2$, $E[\eta] = \mu_2$, $D\eta = \sigma_2^2$. 它们的协方差

$$\text{Cov}(\xi, \eta) = E[(\xi - E\xi)(\eta - E\eta)] = \int_{\mathbb{R}^2} (x - \mu_1)(y - \mu_2)p(x, y)dxdy,$$

其中 p 为密度函数

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \cdot \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right] \right\}.$$

对积分进行变量替换得

$$\begin{aligned} \text{Cov}(\xi, \eta) &= \sigma_1\sigma_2 \int_{\mathbb{R}^2} st \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} [s^2 - 2\rho st + t^2] \right\} dsdt \\ &= \sigma_1\sigma_2 \int_{\mathbb{R}} \frac{s}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds \int_{\mathbb{R}} \frac{t}{\sqrt{2\pi(1-\rho^2)}} \exp \left\{ -\frac{(t-\rho s)^2}{2(1-\rho^2)} \right\} dt \\ &= \rho\sigma_1\sigma_2 \int_{\mathbb{R}} \frac{s^2}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} ds \\ &= \rho\sigma_1\sigma_2. \end{aligned}$$

因此

$$\rho(\xi, \eta) = \rho.$$

现在我们终于知道二维正态分布的密度函数中的那个 ρ 的意义了: 它代表两个分量之间的相关关系!

也可以用下面的方式计算协方差. 回顾

$$\begin{aligned} p(x, y) &= \frac{1}{\sqrt{2\pi}\sigma_1} \exp \left\{ -\frac{(x-\mu_1)^2}{2\sigma_1^2} \right\} \\ &\quad \cdot \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_2} \exp \left\{ -\frac{\left[y - \left(\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1) \right) \right]^2}{2\sigma_2^2(1-\rho^2)} \right\} \\ &=: \phi(x)\psi(x, y). \end{aligned}$$

则

$$\begin{aligned} \text{Cov}(\xi, \eta) &= \int_{\mathbb{R}^2} (x - \mu_1)(y - \mu_2)\phi(x)\psi(x, y)dxdy \\ &= \int_{\mathbb{R}} (x - \mu_1)\phi(x)dx \int_{\mathbb{R}} (y - \mu_2)\psi(x, y)dy \\ &= \int_{\mathbb{R}} (x - \mu_1) \left(\rho\frac{\sigma_2}{\sigma_1}(x - \mu_1) \right) \phi(x)dx \\ &= \rho\sigma_1\sigma_2. \end{aligned}$$

协方差和相关系数的意义是什么呢? 注意与方差不同, 相关系数是可正可负的. 我们来大概想象一下, 设 ξ 代表语文成绩, η 代表数学成绩. 如果它们的相关系数是正的, 那么就是说 $\xi - E[\xi]$ 与 $\eta - E[\eta]$ 同号的情况比较多, 也就是说, 对于个人来说, 其语文与数学同时高于平均分或同时低于平均分的情况比较多, 所以是正向关联的, 因此叫做正相关; 反之, 如果它们的相关系数是负的, 就表明 $\xi - E[\xi]$ 与 $\eta - E[\eta]$ 异号的情况比较多, 因此叫做负相关. 正相关说明语文和数学或者都好或者都差的情况比较多, 负相关说明数学好语文就差, 或者语文好数学就差的同学比较多, 不是吗?

你能大约估计出吗, 你们班不同功课的成绩是正相关还是负相关呢? 你自己呢?

我们有下面的结果:

命题 5.10.4. 设 ξ_1, \dots, ξ_n 为随机变量. 那么

$$D \left[\sum_{i=1}^n \xi_i \right] = \sum_{i=1}^n D[\xi_i] + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(\xi_i, \xi_j).$$

特别地, 若 $i \neq j$ 时 ξ_i 与 ξ_j 不相关, 则

$$D \left[\sum_{i=1}^n \xi_i \right] = \sum_{i=1}^n D[\xi_i].$$

证明是直接展开了计算.

例6. 设 $\xi \sim B(n, p)$, 求 $D[\xi]$.

这个是可以直接计算的(你试试看). 但我们注意到

$$\xi = \xi_1 + \dots + \xi_n,$$

其中 $\xi_i \sim B(p)$ 且相互独立. 因为 $E[\xi_i] = p$, $E[\xi_i \xi_j] = p^2$, 因此 ξ_i 与 ξ_j 是不相关的, $\forall i \neq j$. 因此

$$D[\xi] = \sum_{i=1}^n D[\xi_i] = npq.$$

上述命题的一个直接推论是:

推论 5.10.5. 设 (ξ_1, \dots, ξ_n) 是 n 维随机变量, 且 ξ_i 与 ξ_j 不相关, $\forall i \neq j$. 则

$$D \left[\frac{1}{n} \sum_{i=1}^n \xi_i \right] = \frac{1}{n^2} \sum_{i=1}^n D[\xi_i].$$

我们可以把这个结果尝试地用到投资理财. 假设有 n 种股票, ξ_1, \dots, ξ_n , 其预期收益都一样, 即

$$E[\xi_i] = a, \quad \forall i = 1, \dots, n.$$

其波动也一样, 即

$$D[\xi_i] = \sigma^2, \quad \forall i = 1, \dots, n.$$

再假设这些股票是互不相关的(例如, ξ_1 是地产, ξ_2 是医疗, \dots). 如果你把资金平均分散到这 n 种股票上, 则预期收益是不变的, 即

$$E \left[\frac{1}{n} \sum_{i=1}^n \xi_i \right] = a,$$

而波动为

$$D\left[\frac{1}{n}\sum_{i=1}^n\xi_i\right]=\frac{\sigma^2}{n},$$

降低为原来的 $\frac{1}{n}$. 所以对厌恶风险的人来说, 他会选择把资金分散到各个行业进行投资, 叫做不把鸡蛋放在一个篮子里.

类似地, 还可以应用到测量的误差. 多次测量的算数平均的期望不变, 方差降低为原来的 $\frac{1}{n}$, 以此来减小误差.

习题

1. 证明: 若 $D[\xi] = 0$, 则 $\xi = E[\xi]$ a.s..
2. 证明: 当 $D[\xi]$ 与 $D[\eta]$ 均存在时, $|\rho| = 1$ 的充要条件是存在常数 a, b 使得 $P(\eta = a\xi + b) = 1$.
3. 设 ξ, η 为随机变量, $E[\xi] = E[\eta] = 0$, $D[\xi] = D[\eta] = 1$, ξ 与 η 的相关系数为 ρ . 证明:

$$E[\max(\xi^2, \eta^2)] \leq 1 + \sqrt{1 - \rho^2}.$$

4. 设 ξ, η 为随机变量, ξ 与 η 的相关系数为 ρ . 证明: $\forall \varepsilon > 0$,

$$P(\{|\xi - E[\xi]| \geq \varepsilon\sqrt{D[\xi]}\} \cup \{|\eta - E[\eta]| \geq \varepsilon\sqrt{D[\eta]}\}) \leq \frac{1}{\varepsilon^2}(1 + \sqrt{1 - \rho^2}).$$

5. 设 ξ 是取值于 $[0, 1]$ 的随机变量. ξ 的方差可能取得的最大值是多少? 什么情况下能达到这个值?
6. 设 ξ 是随机变量, a, b 是常数, 且 $P(\xi \in [a, b]) = 1$. 令 $m = E[\xi]$, $\sigma^2 = D[\xi]$. 证明: $\sigma^2 \leq (m - a)(b - m)$, 且等式成立的充要条件是 $P(\xi = a) + P(\xi = b) = 1$.
7. 在上题中, 设 $0 < a < b$. 证明:

$$E[\xi] \cdot E[\xi^{-1}] \leq \frac{(a + b)^2}{4ab}.$$

8. 分别构造离散型分布与连续型分布, 使得
 - (a) 期望不存在;
 - (b) 期望存在, 但 r 阶矩不存在, $r > 1$.

9. 设 $\xi = (\xi_1, \dots, \xi_d)$. 定义 ξ 的协方差矩阵为

$$\text{Cov}(\xi) := (\text{Cov}(\xi_i, \xi_j))_{1 \leq i, j \leq d}.$$

证明 $\text{Cov}(\xi)$ 为非负定矩阵.

5.11 几个等式与不等式

除了几个极少数的典型例子外, 一般来说具体算出期望是不可能的. 这时, 估计其值就是退而求其次的选择了. 因此, 不管理论上还是应用上, 我们都需要了解并娴熟地掌握估计期望值的基本工具.

下面是我们要建立的第一个等式. 它的意义在于, 直观上 $|\xi|$ 比较小时, $E[|\xi|^p]$ 应该会比较小. 而什么叫 $|\xi|$ 比较小呢? 似乎应该是当 x 比较大时, $P(|\xi| \geq x)$ 比较小. 因此, 如果直接用

$$E[|\xi|^p] = \int_{-\infty}^{\infty} |x|^p dF(x)$$

难以直接估计 p -阶矩的话(这个确实比较难, 因为不可能通过估计 $F(x)$ 即 $P(\xi \leq x)$ 来估计它), 那么可以尝试通过估计尾概率 $P(|\xi| \geq x)$ 来估计它. 之所以这是可能的, 是因为有下面的

命题 5.11.1. 设 ξ 是非负随机变量, $p \geq 1$. 则

$$E[\xi^p] = p \int_0^{\infty} r^{p-1} P(\xi \geq r) dr.$$

证明. 令

$$D := \{(x, r) : 0 \leq r \leq x\},$$

$$f(x, r) := |r|^{p-1} \mathbf{1}_D(x, r).$$

则由Fubini定理,

$$E \int_0^{\infty} r^{p-1} \mathbf{1}_D(\xi, r) dr = \int_0^{\infty} r^{p-1} E[\mathbf{1}_D(\xi, r)] dr.$$

即

$$E \int_0^{\infty} r^{p-1} \mathbf{1}(r \leq \xi) dr = \int_0^{\infty} r^{p-1} E[\mathbf{1}(r \leq \xi)] dr,$$

即

$$E \int_0^{\xi} r^{p-1} dr = \int_0^{\infty} r^{p-1} P(\xi \geq r) dr.$$

因此

$$E[\xi^p] = p E \int_0^{\xi} r^{p-1} dr = p \int_0^{\infty} r^{p-1} P(\xi \geq r) dr.$$

□

$P(|\xi| \geq x)$ 称为 ξ 的尾概率, 而估计尾概率常用的工具是前面证明了的Chebyshev不等式. 也即我们有

$$P(|\xi| \geq r) \leq \frac{E[f(|\xi|)]}{f(r)},$$

其中 f 是 \mathbb{R}_+ 上的正的单调上升函数. 由此得到

$$E[|\xi|^p] \leq p E[f(|\xi|)] \int_0^{\infty} r^{p-1} f^{-1}(r) dr.$$

为得到更多的不等式, 我们需要凸函数的概念.

定义 5.11.2. 设 I 是区间, $\varphi: I \rightarrow \mathbb{R}$ 为连续函数, 且

$$\varphi(\alpha x + \beta y) \leq \alpha \varphi(x) + \beta \varphi(y), \quad \forall x, y \in I, \alpha, \beta \geq 0, \alpha + \beta = 1,$$

则称 φ 为凸函数.

从定义看出, 凸函数的直观意思是 $\forall x, y \in I$, 曲线 φ 的 $[x, y]$ 这一段, 在过 $(x, \varphi(x))$ 和 $(y, \varphi(y))$ 两点的直线的下方. 即 $\forall x, y \in I$,

$$\varphi(z) \leq \frac{\varphi(y) - \varphi(x)}{y - x}(z - x) + \varphi(x), \quad \forall z \in [x, y].$$

利用归纳法不难证明, 若 φ 为凸函数, 则 $\forall n, \forall x_i \in I, \lambda_i \in (0, 1), i = 1, 2, \dots, n, \sum_{i=1}^n \lambda_i = 1$, 有

$$\varphi\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i \varphi(x_i).$$

这个不等式引起的自然猜测是, 既然它对任意凸组合都成立, 而期望也是一种凸组合(的极限), 所以它对期望也应该成立. 事实也的确如此, 这就是所谓的 Jensen⁸ 不等式, 并且它的确可以按照凸组合取极限这样的思路证明. 不过, 这个逼近过程太麻烦, 我们可以通过考察凸函数的性质找到一个简洁的证明.

由定义不难推得, 若 $x, y, z \in I, x < y < z$, 则

$$\frac{\varphi(x) - \varphi(y)}{x - y} \leq \frac{\varphi(x) - \varphi(z)}{x - z} \leq \frac{\varphi(y) - \varphi(z)}{y - z}.$$

因此对任意 $x \in I$, $\varphi'_-(x)$ 与 $\varphi'_+(x)$ 都存在(当 x 是 I 的端点时, 只有一种单边导数, 当然), 单调上升, 左导数左连续, 右导数右连续, 且对 $x < y$

$$\varphi'_-(x) \leq \varphi'_+(x) \leq \frac{\varphi(y) - \varphi(x)}{y - x} \leq \varphi'_-(y) \leq \varphi'_+(y). \quad (11.11)$$

反之我们有:

命题 5.11.3. 若 φ 是定义在 I 上的连续函数, φ 处处右可导且 φ'_+ 单调上升, 则 φ 为凸函数.

为证明这个结果, 我们需要下面的:

引理 5.11.4. 若 φ 是定义在 I 上的连续函数, $\varphi'_+(x) \geq 0, \forall x \in I$, 则 φ 单调上升.

证明. 首先证明若 $\varphi'_+(x) > 0, \forall x$, 则 φ 单调上升.

事实上, 设存在 $x_1 < x_2$ 使得 $\varphi(x_1) > \varphi(x_2)$. 令 x_3 是 φ 在 $[x_1, x_2]$ 上的极大值点, 则 $x_3 < x_2$. 于是显然有 $\varphi'_+(x_3) \leq 0$, 矛盾.

一般地, 对 $\delta > 0$, 令

$$\psi(x) := \varphi(x) + \delta x.$$

则 $\psi'_+ > 0$, 因此 ψ 单升. 令 $\delta \rightarrow 0$ 得到 φ 单调上升. □

现在我们可以给出命题 5.11.3 的证明了.

⁸Johan Ludwig William Valdemar Jensen, 1859-1925, 丹麦数学家.

证明. 设 $x_1, x_2 \in I$ 且 $x_1 < x_2$. 令 $\psi(x) := \varphi(x) + ax + b$, 其中 a, b 为满足 $\psi(x_1) = \psi(x_2) = 0$ 的常数. 则只需证

$$\psi(x) \leq 0, \quad \forall x \in [x_1, x_2].$$

设存在 $x_0 \in (x_1, x_2)$ 使得 $\psi(x_0) = \max_{x_1 < x < x_2} \psi(x) > 0$. 此时一定存在 $x_3 \in [x_1, x_0)$ 使得 $\psi(x_3) = 0$ 且

$$\psi(x) > 0, \quad \forall x \in (x_3, x_0].$$

于是 $\psi'_+(x_3) \geq 0$. 由于 φ'_+ 单升, 故在 $[x_3, x_2)$ 上 $\psi'_+ \geq 0$, 因此由刚刚证明的引理, ψ 在 $[x_3, x_2)$ 上单升, 又由于 ψ (左) 连续, 因此 $\psi(x_2) > 0$. 这与 $\psi(x_2) = 0$ 矛盾. \square

由命题 5.11.3, 下面的函数均为凸函数:

$$x^+, \quad |x|^p \quad (p \geq 1), \quad e^{cx} \quad (c > 0), \quad -\ln(x) \quad (x > 0),$$

等等.

我们都知道下面两个著名的结果. 第一个是几何平均小于算术平均, 即

$$\sqrt{st} \leq \frac{s+t}{2}, \quad \forall s, t \geq 0;$$

第二个是第一个的一个推论, 即

$$(s+t)^2 \leq 2(s^2+t^2).$$

我们可以利用凸函数推广这两个不等式, 得到:

命题 5.11.5. 设 $s, t \geq 0$, $\alpha, \beta \geq 0$ 且 $\alpha + \beta = 1$, $p \geq 1$. 则

$$s^\alpha t^\beta \leq \alpha s + \beta t; \tag{11.12}$$

$$(\alpha s + \beta t)^p \leq \alpha s^p + \beta t^p. \tag{11.13}$$

证明. (11.12)、(11.13) 分别是由于 $-\log(t)$ 、 t^p 是 \mathbb{R}_{++} 上的凸函数之故. \square

下面我们可以叙述前面预报了的 Jensen 不等式了.

定理 5.11.6 (Jensen 不等式). 设 φ 为 \mathbb{R} 上的凸函数. 则

$$\varphi(E[\xi]) \leq E[\varphi(\xi)].$$

证明. 注意 $\forall u, v \in \mathbb{R}$, 由 (11.11) 有

$$\varphi(u) - \varphi(v) \geq \varphi'_+(v)(u - v).$$

因此, 取 $u = \xi$, $v = E[\xi]$ 得

$$\varphi(\xi) - \varphi(E[\xi]) \geq \varphi'_+(E[\xi])(\xi - E[\xi]).$$

两边取期望即可. \square

借助于此引理, 我们可以证明两个重要的不等式. 为此引进记号: 对 $p \geq 1$,

$$\|\xi\|_p := (E[|\xi|^p])^{\frac{1}{p}}.$$

定理 5.11.7. 设 ξ, η 是随机变量.

(i) (Hölder⁹不等式) 设 $1 < p, q < \infty$ 满足 $p^{-1} + q^{-1} = 1$, 则

$$\|\xi\eta\|_1 \leq \|\xi\|_p \|\eta\|_q;$$

(ii) (Minkowski¹⁰不等式) $\forall p \geq 1$,

$$\|\xi + \eta\|_p \leq \|\xi\|_p + \|\eta\|_p.$$

正如我们早在命题5.2.8就证明了的, Hölder不等式在 $p = q = 2$ 的特殊情形又称为Cauchy-Schwarz不等式.

证明. 如果上面不等式右边中有一个为0或 ∞ , 则两个不等式都是显然的, 因此我们假定他们都为大于0的实数.

(i) 由(11.12)有

$$\left| \frac{\xi}{\|\xi\|_p} \cdot \frac{\eta}{\|\eta\|_q} \right| \leq \frac{1}{p} |\xi|^p \|\xi\|_p^{-p} + \frac{1}{q} |\eta|^q \|\eta\|_q^{-q},$$

两边取期望即得

$$\|\xi\|_p^{-1} \|\eta\|_q^{-1} \|\xi\eta\|_1 \leq \frac{1}{p} + \frac{1}{q} = 1.$$

(ii) 由(11.13)有

$$\begin{aligned} |\xi + \eta|^p &\leq (|\xi| + |\eta|)^p \\ &= \left| \frac{\|\xi\|_p}{\|\xi\|_p + \|\eta\|_p} (|\xi| \|\xi\|_p^{-1}) + \frac{\|\eta\|_p}{\|\xi\|_p + \|\eta\|_p} (|\eta| \|\eta\|_p^{-1}) \right|^p (\|\xi\|_p + \|\eta\|_p)^p \\ &\leq (\|\xi\|_p + \|\eta\|_p)^p \left(\frac{\|\xi\|_p}{\|\xi\|_p + \|\eta\|_p} (|\xi| \|\xi\|_p^{-1}) \right)^p + \frac{\|\eta\|_p}{\|\xi\|_p + \|\eta\|_p} (|\eta| \|\eta\|_p^{-1})^p \end{aligned}$$

两边取期望即得

$$\|\xi + \eta\|_p^p \leq (\|\xi\|_p + \|\eta\|_p)^p,$$

再两边开 p 次方即可. □

习题

1. 设 ξ, η 是非负随机变量, f 是 \mathbb{R} 上的非负连续函数. 令

$$F(x) := \int_0^x f(y) dy.$$

证明

$$\int_0^\infty E[\xi f(x) 1_{\eta \geq x}] dx = E[\xi F(\eta)].$$

⁹Otto Ludwig Hölder, 1859-1937, 德国数学家.

¹⁰Hermann Minkowski, 1864-1909, 德国数学家

2. 设 ξ, η 是非负随机变量, 且 $\forall \lambda > 0$ 有

$$P(\xi \geq \lambda) \leq \frac{1}{\lambda} E[\eta 1_{\xi \geq \lambda}].$$

证明

$$E[\xi^p] \leq q^p E[\eta^p], \quad \forall p > 1,$$

其中 $q := \frac{p}{p-1}$.

3. 设 $\xi \sim N(0, 1)$. 用Chebyshev不等式证明: 对任意 $a > 0$ 有

$$P(|\xi| > a) \leq 2e^{-\frac{a^2}{2}}.$$

4. 证明:

$$1 \leq p < q \implies \|\xi\|_p \leq \|\xi\|_q.$$

5. 设 ξ 是取值于 $\{x_1, \dots, x_n\}$ 的离散型随机变量. 证明:

$$\lim_{p \rightarrow \infty} (E[|\xi|^p])^{1/p} = \max\{|x_1|, \dots, |x_n|\}.$$

6. 设 ξ 是对称随机变量. 证明: $E[e^{\xi}] \geq 1$. 所谓 ξ 是对称随机变量是指

$$P(\xi \in A) = P(\xi \in -A), \quad \forall A \in \mathcal{B},$$

其中 $-A := \{-x : x \in A\}$.

7. 设 $\varphi: \mathbb{R}_+ \rightarrow [0, 1]$ 为增函数. 证明: $\forall x \geq 0$,

$$E[\varphi(|\xi|)] - \varphi(x) \leq P(|\xi| \geq x) \leq \frac{E[\varphi(|\xi|)]}{\varphi(x)}.$$

8. 证明: 若存在常数 $\alpha, \beta, \gamma > 0$ 使得

$$P(|\xi| > r) \leq \alpha e^{-\beta r^\gamma},$$

则 ξ 的任意阶矩存在.

9. 证明: 若存在 $\alpha > 1$ 使

$$\lim_{n \rightarrow \infty} \frac{P(|\xi| > \alpha n)}{P(|\xi| > n)} = 0,$$

则 ξ 的任意阶矩均存在.

10. 设 $P(\xi > 0) = 1$. 证明:

$$E[\xi^{-1}] \geq (E[\xi])^{-1},$$

$$E[\ln \xi] \leq \ln E[\xi],$$

$$E[\xi \ln \xi] \geq E[\xi] E[\ln \xi].$$

11. 设 ξ 是随机变量, $r \in \mathbb{R}$. 证明: 当 $rE[\xi] < 0$ 时, $E[\exp\{r\xi\}] < 1$.

12. 设 $a_i \in \mathbb{R}$, $i = 1, 2, \dots, n$, $r \geq 1$. 证明:

$$\left| \sum_{i=1}^n a_i \right|^r \leq n^{r-1} \sum_{i=1}^n |a_i|^r.$$

13. 证明, 对 $p \geq 1$, $E[|\xi|^p] < \infty$ 的充要条件是

$$\sum_{n=0}^{\infty} n^{p-1} P(|\xi| \geq n) < \infty.$$

14. 设 $p, q \geq 1$, $p^{-1} + q^{-1} = 1$, ξ 是非负随机变量, $\|\xi\|_p \leq \infty$.

(a) 证明: 对 $a \in (0, 1)$,

$$(1-a)E[\xi] \leq E[\xi 1_{[aE[\xi], \infty)}];$$

(b) 证明:

$$P(\xi \geq aE[\xi]) \geq (1-a)^q \frac{E[\xi]^q}{\|\xi\|_p^q}.$$

5.12 条件期望

设

$$\mathcal{P} = \{A_1, A_2, \dots\}$$

为 Ω 的一个分割. 任给 $A \in \mathcal{F}$, 有

$$P(A|\mathcal{P})(\omega) = \sum_{n=1}^{\infty} P(A|A_n)1_{A_n}(\omega).$$

由于对任意固定的 ω , $P(\cdot|\mathcal{P})(\omega)$ 是一个概率, 所以可以定义关于这个概率的期望. 具体地说, 设 ξ 为随机变量, 定义

$$E[\xi|\mathcal{P}](\omega) =: E[\xi(\omega')|\mathcal{P}](\omega) =: \int_{\Omega} \xi(\omega')P(d\omega'|\mathcal{P})(\omega).$$

当 ω 固定后, 它拥有普通期望的一切性质; 而当 ξ 固定后, 它是一个随机变量. 我们有:

命题 5.12.1.

$$E[\xi] = E[E[\xi|\mathcal{P}]].$$

证明. 任给 $A \in \mathcal{F}$, 有

$$E[1_A|\mathcal{P}](\omega) = \sum_{n=1}^{\infty} P(A|A_n)1_{A_n}(\omega).$$

两边取期望得

$$\begin{aligned} E[E[1_A|\mathcal{P}]] &= \sum_{n=1}^{\infty} P(A|A_n)P(A_n) = \sum_{n=1}^{\infty} P(AA_n) \\ &= P(A) = E[1_A]. \end{aligned}$$

因此当 $\xi = 1_A$ 时, 结论成立.

其次, 当 ξ 是简单随机变量, 即 ξ 具有形式 $\xi = \sum_{k=1}^m c_k 1_{A_k}$, 其中 c_k 是常数, $A_k \in \mathcal{F}$ 时, 由期望的线性性, 结论成立.

再次, 当 ξ 为非负随机变量时, 存在一列取有限值的简单随机变量 $\xi_m \uparrow \xi$, 因此由单调收敛定理, 结论依然成立.

最后, 对一般的 ξ , 利用分解 $\xi = \xi^+ - \xi^-$ 及期望的线性性, 结论照样成立. \square

为叙述下一个结果, 我们需要先引进两个概念.

定义 5.12.2. 设 $\mathcal{P} = \{A_i, i = 1, 2, \dots\}$ 为一个分割, ξ 为随机变量.

(i) 若

$$\xi = \sum_i c_i 1_{A_i},$$

其中 c_i 们为常数, 则称 ξ 为 \mathcal{P} -可测;

(ii) 若

$$P(\{\xi \in B\} \cap A_i) = P(\xi \in B)P(A_i) \quad \forall B \in \mathcal{B}, i,$$

则称 ξ 与 \mathcal{P} 独立.

然后我们有:

命题 5.12.3. 设 ξ, η 为随机变量, η 为 \mathcal{P} -可测. 则对 \mathbb{R}^2 上的任意非负 Borel 函数 f 有

$$E[f(\xi, \eta) | \mathcal{P}] = E[f(\xi, x) | \mathcal{P}]|_{x=\eta}.$$

特别地, 若还有 ξ 与 \mathcal{P} 独立, 则

$$E[f(\xi, \eta) | \mathcal{P}] = E[f(\xi, x)]|_{x=\eta}.$$

证明. 令

$$\mathcal{G} := \{A \in \mathcal{B}^2 : E[1_A(\xi, \eta) | \mathcal{P}] = E[1_A(\xi, x) | \mathcal{P}]|_{x=\eta}\}.$$

则

$$\mathcal{C} \subset \mathcal{G},$$

其中

$$\mathcal{C} := \{(a, b) \times (c, d) : a < b, c < d\}.$$

显然 \mathcal{C} 为 π -类, 又易证 \mathcal{G} 为 λ -类. 因此由 $\pi - \lambda$ 定理,

$$\mathcal{G} \supset \sigma(\mathcal{C}) = \mathcal{B}^2.$$

所以对任意 $A \in \mathcal{B}^2$ 有

$$E[1_A(\xi, \eta) | \mathcal{P}] = E[1_A(\xi, x)]|_{x=\eta}.$$

由此易见(利用期望的线性性), 对 \mathbb{R}^2 上的任意简单 Borel 函数 f , 都有

$$E[f(\xi, \eta) | \mathcal{P}] = E[f(\xi, x) | \mathcal{P}]|_{x=\eta}.$$

现设 $f \in \mathcal{B}^2$ 为任一非负 Borel 函数. 由命题 4.1.13, 存在一列简单 $f_n \in \mathcal{B}^2$ 使得 $f_n(x, y) \uparrow f(x, y), \forall x, y \in \mathbb{R}^2$. 因此单调收敛定理

$$\begin{aligned} E[f(\xi, \eta)|\mathcal{P}] &= \lim_{n \rightarrow \infty} E[f_n(\xi, \eta)|\mathcal{P}] \\ &= \lim_{n \rightarrow \infty} E[f_n(\xi, x)|\mathcal{P}]|_{x=\eta} \\ &= E[f(\xi, x)|\mathcal{P}]|_{x=\eta} \end{aligned}$$

如果 ξ 与 \mathcal{P} 独立, 则

$$E[f(\xi, x)|\mathcal{P}] = E[f(\xi, x)].$$

因此第二个结论成立. □

习题

1. 设 $\mathcal{P}_1, \mathcal{P}_2$ 是两个分割, 且 \mathcal{P}_1 是 \mathcal{P}_2 的加细. 证明:

$$E[E[\xi|\mathcal{P}_1]|\mathcal{P}_2] = E[E[\xi|\mathcal{P}_2]|\mathcal{P}_1] = E[\xi|\mathcal{P}_2].$$

2. 设 \mathcal{P} 为分割, ξ 为随机变量, η 为离散型随机变量, 且 η 为 \mathcal{P} -可测 (即 η 在 \mathcal{P} 的任何一个元素上均为常数). 证明: 对任意有界 Borel 可测函数 f , 有

$$E[f(\xi, \eta)|\mathcal{P}] = E[f(\xi, x)|\mathcal{P}]|_{x=\eta}.$$

3. 一个人乘飞机出行. 设当天天晴的概率为 p , 下雨的概率为 $q, p+q=1$. 在天晴的情况下, 飞机晚点的时间服从参数为 α 的指数分布; 在下雨的情况下, 晚点的时间服从参数为 β 的指数分布. 求飞机晚点的平均时间.
4. 一个养殖户养了 n_1 只鸡, n_2 只鸭, n_3 只鹅, 期望年末时每只鸡卖 y_1 元, 每只鸭卖 y_2 元, 每只鹅卖 y_3 元. 正常年份时, 每只家禽病死的概率分别为 p_1, p_2, p_3 ; 但如果有瘟疫, 相应的概率会增加到 q_1, q_2, q_3 . 设有瘟疫发生的可能性为 p . 问
- (a) 他的期望收益是多少?
- (b) 如果有个保险公司向他推销保险, 每只家禽保费分别为 r_1, r_2, r_3 , 死亡赔偿分别为 s_1, s_2, s_3 . 你是否建议他买保险?

6 随机变量的独立性

什么叫独立? 在通常的理解中, 独立就是不依赖, 不受制. 国家的独立自由意味着摆脱了别的国家的欺凌, 个人的独立意味着不受他人奴役. 那么什么叫随机变量的独立? 简言之, 就是随机变量的取值不受制于其它随机变量取值的影响, 在概率的意义下. 这是一个在客观世界里常见的现象. 小到掷骰子的游戏, 大到悬浮在液体中的微粒所受到的液体分子运动在不同的时间区间内带来的冲击力. 另一方面, 独立性会给数学上的处理带来极大的方便, 是我们考虑概率问题时必须用好用尽的一个性质.

我们曾经对离散型随机变量定义了独立性. 现在把这个概念拓广到一般情况并对它进行系统的研究.

6.1 基本定义及性质

我们从最简单的情况即两个随机变量的情况开始.

定义 6.1.1. 设 ξ, η 是随机变量. 若

$$P(\xi \leq a, \eta \leq b) = P(\xi \leq a)P(\eta \leq b), \quad \forall a, b \in \mathbb{R},$$

则称 ξ, η 独立.

用分布函数表示, 独立性就是

$$F(x, y) = F_\xi(x)F_\eta(y), \quad \forall x, y \in \mathbb{R},$$

其中 F 是 (ξ, η) 的联合分布函数, F_ξ 与 F_η 分别是 ξ 与 η 的分布函数. 容易验证, 此条件等价于:
 $\forall x_1 < x_2, y_1 < y_2,$

$$F(x_2, y_2) - F(x_1, y_2) - F(x_2, y_1) + F(x_1, y_1) = [F_\xi(x_2) - F_\xi(x_1)][F_\eta(y_2) - F_\eta(y_1)].$$

或等价地, $\forall a < b, c < d,$ 有

$$P(a < \xi \leq b, c < \eta \leq d) = P(a < \xi \leq b)P(c < \eta \leq d).$$

我们需要说明, 对离散型随机变量, 这个定义和以前的定义是一致的.

设 (ξ, η) 的分布列为 $\{(x_i, y_j), p_{ij}\}$. 则两个边缘分布列分别为 $\{x_i, p_i^\xi\}$ 与 $\{y_j, p_j^\eta\}$, 其中 $p_i^\xi = \sum_j p_{ij}, p_j^\eta = \sum_i p_{ij}$. 按以前的定义, 独立性是指

$$p_{ij} = p_i^\xi \cdot p_j^\eta, \quad \forall i, j. \tag{1.1}$$

如果此式成立, 那么 $\forall x, y$,

$$\begin{aligned} F(x, y) &= \sum_{x_i \leq x, y_j \leq y} p_{ij} \\ &= \sum_{x_i \leq x, y_j \leq y} p_i^\xi p_j^\eta \\ &= \sum_{x_i \leq x} p_i^\xi \cdot \sum_{y_j \leq y} p_j^\eta \\ &= F_\xi(x) F_\eta(y). \end{aligned}$$

所以按现在的定义是独立的. 反之, 若按现在的定义是独立的, 那么 $\forall i, j$,

$$\begin{aligned} p_{ij} &= P(\xi = x_i, \eta = y_j) \\ &= \lim_{n \rightarrow \infty} P\left(x_i - \frac{1}{n} < \xi \leq x_i, y_j - \frac{1}{n} < \eta \leq y_j\right) \\ &= \lim_{n \rightarrow \infty} P\left(x_i - \frac{1}{n} < \xi \leq x_i\right) P\left(y_j - \frac{1}{n} < \eta \leq y_j\right) \\ &= P(\xi = x_i) P(\eta = y_j) \\ &= p_i^\xi p_j^\eta. \end{aligned}$$

所以按原来的定义也是独立的. 因此, 对离散型随机变量, 现在的定义和原来的定义是等价的.

另一种特殊情况是连续型随机变量. 设 (ξ, η) 的密度函数为 $p(x, y)$, 那么两个边沿密度分别为 $p_\xi(x)$ 与 $p_\eta(y)$, 其中

$$p_\xi(x) = \int_{\mathbb{R}} p(x, y) dy, \quad p_\eta(y) = \int_{\mathbb{R}} p(x, y) dx.$$

因此独立性意味着

$$\begin{aligned} F(x, y) &= F_\xi(x) F_\eta(y) \\ &= \int_{-\infty}^x p_\xi(u) du \int_{-\infty}^y p_\eta(v) dv \\ &= \int_{-\infty}^x \int_{-\infty}^y p_\xi(u) p_\eta(v) dudv. \end{aligned}$$

所以

$$p(x, y) = p_\xi(x) p_\eta(y),$$

只要 (x, y) 是 p 们的连续点. 反之, 若此式成立, 那么也容易证明 ξ 与 η 独立. 所以我们有

命题 6.1.2. 设 (ξ, η) 是离散型随机变量, 那么它们独立的充要条件是联合分布列等于两个边沿分布列的乘积.

设 (ξ, η) 是连续型随机变量, 那么它们独立的充要条件是联合密度函数在其连续点上等于两个边沿密度函数的乘积.

现在考虑下面的问题: 如果 ξ, η 是随机变量, $f \in \mathcal{B}^2$. 问: 在求期望 $E[f(\xi, \eta)]$ 时, 能否先固定一个, 即先求 $E[f(x, \eta)]$, 得到 x 的函数, 再将 ξ 代入到这个函数中去, 得到一个随机变量, 然后求这个随机变量的期望?

回答一般是否定的. 例如, 取 $f(x, y) = xy$. 此时, 对任意两个随机变量 ξ, η , 上面公式成立意味着 $E[\xi\eta] = E[\xi]E[\eta]$, 即 ξ, η 不相关. 这显然一般是不对的. 比如, $\xi = \eta$, ξ 是对称随机变量 (ξ 与 $-\xi$ 同分布), $E[\xi] = 0$ (例如 $\xi \sim N(0, 1)$). 则

$$\varphi(x) := E[x\eta] = 0, \quad \forall x,$$

于是

$$E[\varphi(\xi)] = 0.$$

但显然

$$E[f(\xi, \eta)] = E[\xi^2] \neq 0.$$

但在独立的情况, 这样做是可以的, 即我们有下面的定理:

定理 6.1.3. 设 $f: \mathbb{R}^2 \mapsto \mathbb{R}$ 为非负 Borel 可测, (ξ, η) 为二维随机变量. 令

$$\varphi(x) := E[f(x, \eta)].$$

则 $\varphi: \mathbb{R} \mapsto \mathbb{R}$ 为非负 Borel 可测, 且 ξ, η 为独立的充要条件对任意这样的 f ,

$$E[f(\xi, \eta)] = E[\varphi(\xi)].$$

证明. 先证 φ 的 Borel 可测性. 由于非负可测函数可由非负简单函数单调上升地逼近, 所以由单调收敛定理只需对 f 是非负简单函数证明. 而由期望的线性性, 这又归结为对示性函数证明. 因此我们设

$$f = 1_A, \quad A \in \mathcal{B}^2.$$

令

$$\mathcal{G} := \{A \in \mathcal{B}^2 : \varphi(x) := E[1_A(x, \eta)] \text{ 为 Borel 可测}\}.$$

则易证 \mathcal{G} 为 λ -类, 且包含了 π -类:

$$\Pi := \{(-\infty, a] \times (-\infty, b], \quad a, b \in \mathbb{R}\}.$$

因此由单调类定理有 $\mathcal{G} = \sigma(\Pi) = \mathcal{B}^2$.

下面证第二个结论. 充分性显然, 因为取

$$f(x, y) := 1_{(-\infty, a]}(x) \cdot 1_{(-\infty, b]}(y),$$

则

$$E[f(\xi, \eta)] = E[\varphi(\xi)]$$

就变为

$$P(\xi \leq a, \eta \leq b) = P(\xi \leq a)P(\eta \leq b).$$

往证必要性. 令

$$\mathcal{G} := \{A \in \mathcal{B}^2 : \text{定理对函数 } f := 1_A \text{ 成立}\}.$$

$$\mathcal{C} := \{(a, b] \times (c, d] : a < b, c < d\}.$$

则 \mathcal{C} 为 π -类, 且 $\mathcal{B}^2 = \sigma(\mathcal{C})$. 对 $A := (a, b] \times (c, d] \in \mathcal{C}$, 令

$$f := 1_A.$$

则

$$\varphi(x) = E[1_A(x, \eta)] = 1_{(a,b]}(x)E[1_{(c,d]}(\eta)] = 1_{(a,b]}(x)P(c < \eta \leq d).$$

所以

$$E[\varphi(\xi)] = E[1_{(a,b]}(\xi)]P(c < \eta \leq d) = P(a < \xi \leq b)P(c < \eta \leq d).$$

而由独立性也有

$$\begin{aligned} E[f(\xi, \eta)] &= E[1_{(a,b] \times (c,d]}(\xi, \eta)] \\ &= P(a < \xi \leq b, c < \eta \leq d) \\ &= P(a < \xi \leq b)P(c < \eta \leq d). \end{aligned}$$

所以, $\mathcal{C} \subset \mathcal{G}$.

往证 \mathcal{G} 为 λ -类. 显然 $\mathbb{R}^2 \in \mathcal{G}$. 设 $A, B \in \mathcal{G}$, $A \subset B$. 对任意 $C \subset \mathcal{B}^2$, 令

$$\varphi_C(x) := E[1_C(x, \eta)],$$

则

$$E[\varphi_B(\xi)] - E[\varphi_A(\xi)] = E[\varphi_{B \setminus A}(\xi)].$$

因此

$$\begin{aligned} E[1_{B \setminus A}(\xi, \eta)] &= E[1_B(\xi, \eta) - 1_A(\xi, \eta)] \\ &= E[1_B(\xi, \eta)] - E[1_A(\xi, \eta)] \\ &= E[\varphi_B(\xi)] - E[\varphi_A(\xi)] \\ &= E[\varphi_{B \setminus A}(\xi)]. \end{aligned}$$

所以 $B - A \in \mathcal{G}$. 再设 $A_n \in \mathcal{G}$, $A_n \uparrow A$. 则由单调收敛定理有

$$\begin{aligned} E[1_A(\xi, \eta)] &= \lim_{n \rightarrow \infty} E[1_{A_n}(\xi, \eta)] \\ &= \lim_{n \rightarrow \infty} E[\varphi_{A_n}(\xi)] \\ &= E[\lim_{n \rightarrow \infty} \varphi_{A_n}(\xi)] \\ &= E[\varphi_A(\xi)]. \end{aligned}$$

所以 $A \in \mathcal{G}$. 这样, 由 $\lambda - \pi$ 定理, \mathcal{G} 为 σ -代数. 于是 $\mathcal{G} = \mathcal{B}^2$.

现在, 若 f 是简单函数:

$$f := \sum_{k=1}^n a_k 1_{A_k}, \quad a_k \in \mathbb{R}, A_k \in \mathcal{B}^2,$$

则有

$$\begin{aligned} E[f(\xi, \eta)] &= \sum_{k=1}^n a_k E[1_{A_k}(\xi, \eta)] \\ &= \sum_{k=1}^n a_k E[\varphi_{A_k}(\xi)] \\ &= E\left[\sum_{k=1}^n a_k \varphi_{A_k}(\xi)\right] \\ &= E[\varphi(\xi)]. \end{aligned}$$

最后, 设 $f \geq 0$, $f \in \mathcal{B}^2$. 于是, 存在简单函数 $f_n \uparrow f$. 以 φ_n 表示 f_n 对应的 φ , 则由单调收敛定理, 有

$$\begin{aligned} E[f(\xi, \eta)] &= \lim_{n \rightarrow \infty} E[f_n(\xi, \eta)] \\ &= \lim_{n \rightarrow \infty} E[\varphi_n(\xi)] \\ &= E[\varphi(\xi)]. \end{aligned}$$

□

分别考虑正负部, 我们有

推论 6.1.4. 将上定理中的条件“ $f \geq 0$ ”换为“ f 使得 $f(\xi, \eta)$ 可积”, 结论依然成立.

我们还有:

推论 6.1.5. 记号与假设同推论 6.1.4. 令 $\zeta := f(\xi, \eta)$, $A \in \mathcal{B}(\mathbb{R})$. 则

$$P(\zeta \in A) = E[g(\eta)] = \int_{-\infty}^{\infty} g(x) dF_{\eta}(x),$$

其中

$$g(x) = P(f(\xi, x) \in A).$$

特别地, 若 η 有密度函数 ρ , 则

$$P(\zeta \in A) = \int_A g(x) \rho(x) dx.$$

证明. 用 $1_A \circ f$ 代替 f 用上一推论. □

下面是上面推论的一个重要特例.

推论 6.1.6. 设 ξ, η 是独立随机变量, $\zeta := \xi + \eta$. 则

$$F_{\zeta}(y) = \int_{-\infty}^{\infty} F_{\xi}(y-x) dF_{\eta}(x) = \int_{-\infty}^{\infty} F_{\eta}(y-x) dF_{\xi}(x).$$

这里当 F_{ξ} 或 F_{η} 不连续时, 中间和右边的积分理解为 Lebesgue-Stieltjes 积分; 当它们分段连续时, 按第四节习题 4 的意义理解.

特别地, 当 ξ 是连续型随机变量, 且有密度函数 ρ_{ξ} , 则 ζ 也是连续型的, 且密度函数为

$$\rho_{\zeta}(y) := \int_{-\infty}^{\infty} \rho_{\xi}(y-x) dF_{\eta}(x).$$

证明. 取 $A = (-\infty, y]$, $f(u, v) = u + v$. 则

$$g(x) = P(\xi + x \leq y) = P(\xi \leq y - x) = F_{\xi}(y - x).$$

所以

$$P(\zeta \leq y) = E[g(\eta)] = \int_{-\infty}^{\infty} F_{\xi}(y-x) dF_{\eta}(x).$$

若 ξ 有密度函数 ρ_ξ , 则上式右端成为

$$\begin{aligned} \int_{-\infty}^{\infty} dF_\eta(x) \int_{-\infty}^{y-x} \rho_\xi(u) du &= \int_{-\infty}^{\infty} dF_\eta(x) \int_{-\infty}^y \rho_\xi(u-x) du \\ &= \int_{-\infty}^y du \int_{-\infty}^{\infty} \rho_\xi(u-x) dF_\eta(x). \end{aligned}$$

因此 ζ 是连续型的, 且密度函数为

$$\rho_\zeta(y) = \int_{-\infty}^{\infty} \rho_\xi(y-x) dF_\eta(x).$$

□

我们还有一个推论:

推论 6.1.7. 设 ξ, η 独立, 且皆可积, 则 $\xi\eta$ 亦可积, 且

$$E[\xi\eta] = E[\xi]E[\eta].$$

证明. 取 $f(x, y) := |xy|$, 我们有

$$E[|\xi\eta|] = E[|\xi|]E[|\eta|] < \infty.$$

因此 $\xi\eta$ 可积. 再取 $f(x, y) = xy$ 就得到所要的等式.

□

设 ξ 为随机变量, 称

$$\sigma(\xi) := \xi^{-1}(\mathcal{B})$$

为 ξ 生成的 σ -代数. 即 $\sigma(\xi)$ 为使 ξ 可测的最小 σ -代数. 类似地, 设 I 是任一指标集, 定义 $\sigma(\xi_i, i \in I)$ 为使所有 $\xi_i, i \in I$, 可测的最小 σ -代数.

命题 6.1.8. 设 ξ, η 独立, 那么 $\sigma(\xi)$ 与 $\sigma(\eta)$ 独立. 即 $\forall A, B \in \mathcal{B}$,

$$P(\xi \in A, \eta \in B) = P(\xi \in A)P(\eta \in B).$$

证明. 记 $\mathcal{A} = \{(-\infty, x], x \in \mathbb{R}\}$, 则 $\mathcal{B} = \sigma(\mathcal{A})$. 因为 ξ, η 独立, 所以集类

$$\xi^{-1}(\mathcal{A}) = \{\{\xi \leq x\}, x \in \mathbb{R}\} \text{ 与 } \eta^{-1}(\mathcal{A}) = \{\{\eta \leq x\}, x \in \mathbb{R}\}$$

独立, 且它们都是 π -类. 由命题4.1.1有

$$\sigma(\xi) = \xi^{-1}(\mathcal{B}) = \xi^{-1}(\sigma(\mathcal{A})) = \sigma(\xi^{-1}(\mathcal{A})).$$

同理 $\sigma(\eta) = \sigma(\eta^{-1}(\mathcal{A}))$. 所以由命题3.4.3, $\sigma(\xi)$ 与 $\sigma(\eta)$ 独立.

□

由此我们可以得到:

推论 6.1.9. 设 ξ, η 独立, 则对任意Borel函数 f, g , $f(\xi)$ 与 $g(\eta)$ 独立.

证明. 对任意 $x, y \in \mathbb{R}$, $f^{-1}((-\infty, x]), g^{-1}((-\infty, y]) \in \mathcal{B}$. 所以

$$\{f(\xi) \leq x\} \in \sigma(\xi), \{g(\eta) \leq y\} \in \sigma(\eta).$$

再用上一命题即可.

□

因为上面的命题, 所以可以把

$$P(\xi \in A, \eta \in B) = P(\xi \in A)P(\eta \in B), \quad \forall A, B \in \mathcal{B}$$

作为独立性的定义.

下面是独立的一系列充分必要条件.

定理 6.1.10. 下列论断等价:

- (i) ξ, η 独立;
(ii) 对任意有界 Borel 函数 f, g 有

$$E[f(\xi)g(\eta)] = E[f(\xi)]E[g(\eta)].$$

- (iii) 对任意 $f, g \in C_b$ 有

$$E[f(\xi)g(\eta)] = E[f(\xi)]E[g(\eta)].$$

- (iv) 对任意 $f, g \in C_0$ 有

$$E[f(\xi)g(\eta)] = E[f(\xi)]E[g(\eta)].$$

- (v) 对任意 $f, g \in C_0^\infty$ 有

$$E[f(\xi)g(\eta)] = E[f(\xi)]E[g(\eta)].$$

证明. (i) \implies (ii): 由上一推论.

(ii) \implies (iii) \implies (iv) \implies (v) 显然.

以下反推.

(ii) \implies (i): $\forall x, y$, 取 $f(z) = 1_{(-\infty, x]}(z)$, $g(z) = 1_{(-\infty, y]}(z)$, 显然.

(iii) \implies (i): 设对任意 $f, g \in C_b$ 有

$$E[f(\xi)g(\eta)] = E[f(\xi)]E[g(\eta)]. \quad (1.2)$$

$\forall x, y$, 取

$$f_n(z) = 1_{(-\infty, x]}(z) + n \left(\frac{1}{n} + x - z \right) 1_{(x, x + \frac{1}{n}]},$$

$$g_n(z) = 1_{(-\infty, y]}(z) + n \left(\frac{1}{n} + y - z \right) 1_{(y, y + \frac{1}{n}]}$$

则

$$E[f_n(\xi)g_n(\eta)] = E[f_n(\xi)]E[g_n(\eta)].$$

令 $n \rightarrow \infty$ 得

$$P(\xi \leq x, \eta \leq y) = P(\xi \leq x)P(\eta \leq y).$$

(iv) \implies (iii): 现在, 设(1.2)对任意 $f, g \in C_0$ 成立. 对任意 $f, g \in C_b$, 取 $f_n, g_n \in C_0$ 使得 $f_n \rightarrow f$, $g_n \rightarrow g$ (可选取 $f_n(x) = f(x)1_{[-n, n]}$ 在 $x = \pm n$ 两点线性连续一下即可). 由于

$$E[f_n(\xi)g_n(\eta)] = E[f_n(\xi)]E[g_n(\eta)],$$

所以由有界收敛定理, 知

$$E[f(\xi)g(\eta)] = E[f(\xi)]E[g(\eta)],$$

即(1.2)对任意 $f, g \in C_b$ 成立.

(v) \implies (iv): 最后, 设(1.2)对任意 $f, g \in C_0^\infty$ 成立. 对任意 $f, g \in C_0$, 取 $f_n, g_n \in C_0^\infty$ 使得 $f_n \rightarrow f, g_n \rightarrow g$ (见附录11.4), 且

$$\sup_n \sup_x |f_n(x)| \leq \sup |f(x)|, \quad \sup_n \sup_x |g_n(x)| \leq \sup |g(x)|.$$

因此, 同样由有界收敛定理, 有

$$E[f(\xi)g(\eta)] = E[f(\xi)]E[g(\eta)],$$

即(1.2)对任意 $f, g \in C_0$ 成立. □

关于独立性的概念和结果可以很容易有以下推广.

第一, 推广到多个乃至无穷个随机变量的情形. 我们给出

定义 6.1.11. 设 $\xi_1, \xi_2, \dots, \xi_n$ 为随机变量. 若对任意 $A_i \in \mathcal{B}, i = 1, 2, \dots, n$,

$$P(\xi_1 \in A_1, \dots, \xi_n \in A_n) = \prod_{i=1}^n P(\xi_i \in A_i),$$

则称 ξ_1, \dots, ξ_n 独立.

设 ξ_1, ξ_2, \dots 为随机变量. 若对任意 $n, \xi_1, \xi_2, \dots, \xi_n$ 独立, 则称 ξ_1, ξ_2, \dots 独立.

这样推广之后, 除了涉及两个分布函数卷积的结论之外, 其它都有相应的推广, 结果的叙述和证明都和两个时相似, 恕不一一叙述. 因为没有多个分布函数的卷积这个概念, 所以涉及卷积的结论没有直接的对应物, 但理论上可从两个出发, 逐个推进, 求出多个独立随机变量和的分布函数.

我们还可以证明下面的结论:

命题 6.1.12. 设 $\{\xi_i, i \in \mathbb{N}_{++}\}$ 独立, $I, J \subset \mathbb{N}_{++}, I \cap J = \emptyset$, 则 $\sigma(\xi_i, i \in I)$ 与 $\sigma(\xi_i, i \in J)$ 独立.

证明. 令

$$\Pi_1 := \left\{ \bigcap_{i \in I_0} \xi_i^{-1}(A_i) : I_0 \subset I \text{ 为有限集}, A_i \in \mathcal{B} \right\},$$

$$\Pi_2 := \left\{ \bigcap_{j \in J_0} \xi_j^{-1}(A_j) : J_0 \subset J \text{ 为有限集}, A_j \in \mathcal{B} \right\}.$$

易见 Π_1 与 Π_2 独立且它们均为 π -类. 所以由命题3.4.3, 它们所分别生成的 σ -代数, 即 $\sigma(\xi_i, i \in I)$ 与 $\sigma(\xi_i, i \in J)$, 也独立. □

这个结果自然可推广到多个情形, 即我们有:

命题 6.1.13. 设 $\{\xi_i, i \in \mathbb{N}_{++}\}$ 独立, $I_k \subset \mathbb{N}_{++}, k = 1, 2, \dots$, 且两两不交, 则 $\{\mathcal{G}_k, k = 1, 2, \dots\}$ 独立, 其中 $\mathcal{G}_k := \sigma(\xi_i, i \in I_k)$.

由此立即得到:

命题 6.1.14. 设 $\{\xi_i, i \in \mathbb{N}_{++}\}$ 独立, $0 = n_0 < n_1 < n_2 < \dots$. 令 $r_i := n_i - n_{i-1}$. 则对 $f_i \in \mathcal{B}^{r_i}$, $f_1(\xi_1, \dots, \xi_{n_1}), f_2(\xi_{n_1+1}, \dots, \xi_{n_2}), f_k(\xi_{n_{k-1}+1}, \dots, \xi_{n_k}), \dots$ 独立.

我们看两个应用例子.

例1. 设 ξ_1, ξ_2, \dots 是独立随机变量序列. 令

$$S_n := \sum_{k=1}^n \xi_k.$$

设 $D \in \mathcal{B}$. 令 τ 表示 S_n 首次落入 D 的时刻:

$$\tau := \inf\{n : S_n \in D\},$$

即

$$\{\tau = n\} = \{S_k \notin D, k = 1, 2, \dots, n-1; S_n \in D\}.$$

则 $\{\tau = n\} \in \sigma(\xi_1, \dots, \xi_n)$, 因此它与 $\sigma(\xi_k, k \geq n+1)$ 独立.

例2. 设 τ_1, τ_2, \dots 为独立同分布随机变量, 公共的分布是参数为 λ 的指数分布. 令

$$\sigma_n := \sum_{i=1}^n \tau_i,$$

求 σ_n 的分布.

解. $\sigma_1 = \tau_1$, 所以 σ_1 的密度函数为

$$\rho_1(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

用上述推论和归纳法易证 σ_n 的密度函数为

$$\rho_n(x) = \begin{cases} \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

令

$$\xi_t := \sum_{n=1}^{\infty} 1_{[\sigma_n, \infty)}(t),$$

则对任意 t , ξ_t 为取非负整数值的随机变量, 表示的是有多少个 σ_n 小于 t . 其分布为

$$P(\xi_t = n) = P(\sigma_n \leq t) - P(\sigma_{n+1} > t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

所以 $\xi_t \sim P(\lambda t)$.

还记得那个找星星的小朋友吗? 设 σ_1 为捕捉到第一个星星的时刻, 那么

$$P(\sigma_1 > t) = e^{-\lambda t}.$$

记 σ_2 为捕捉到第二个星星的时刻. 由于是重新开始了捕捉, 所以 $\sigma_2 - \tau_1$ 与 σ_1 独立, 且与 σ_1 同分布. 依次下去, 可定义一系列 $\sigma_n - \sigma_{n-1}$ 与 σ_1 同分布, 且

$$\sigma_1, \sigma_2 - \sigma_1, \dots, \sigma_n - \sigma_{n-1}, \dots$$

相互独立. 这样就可以对 $\sigma_1, \sigma_2 - \sigma_1, \dots, \sigma_n - \sigma_{n-1}, \dots$ 用上述结果. 于是我们明白了:

哦, 却原来他在 $[0, t]$ 时段内找到的星星个数服从参数为 λt 的Poisson分布!

第二个方面的推广是可以推广到每个 ξ_i 都是多维随机变量的情形, 并且值空间的维数可以不同. 此时概念和结果的叙述与证明都可以通过对上面数值情形做明显的修改得到, 就不一一写出了.

习题

- 构造一个概率空间及其上的独立同分布随机变量列 ξ_1, ξ_2, \dots , 使得 $P(\xi_1 = 1) = P(\xi_1 = -1) = \frac{1}{2}$.
- 设 ξ, η 是独立随机变量, $E[|\xi + \eta|] < \infty$. 问是否有 $E[|\xi|] < \infty$? 证明或给出反例.
- (a) 设 ξ 与自己独立, 证明存在常数 c 使得 $P(\xi = c) = 1$.
(b) 设 ξ_1, ξ_2, \dots 独立, $\xi_n \xrightarrow{a.s.} \xi$. 证明存在常数 c 使得 $P(\xi = c) = 1$.
- 设 ξ, η, θ 是随机变量, ξ 与 η 同分布. 判断下列命题的真伪, 证明或给出反例.
 - $\xi + \theta$ 与 $\eta + \theta$ 同分布;
 - 在 ξ 与 η 独立的条件下, $\xi + \theta$ 与 $\eta + \theta$ 同分布;
 - 在三个随机变量独立的条件下, $\xi + \theta$ 与 $\eta + \theta$ 同分布;
- 设 (X, Y) 的密度函数 f 具有形式 $f(x, y) = f_1(x)f_2(y)$. 证明 X 与 Y 独立.
- 设 ξ_1, ξ_2, η 相互独立, $P(\eta = 1) = P(\eta = -1) = \frac{1}{2}$. 证明 $\cos(\xi_1 + \eta\xi_2)$ 与 $\cos(\xi_2 + \eta\xi_1)$ 同分布.
- 设 X, Y, Z, ξ, η 独立, $P(\xi = \pm 1) = P(\eta = \pm 1) = \frac{1}{2}$. 证明 $|X + \xi|Y + \eta Z|$ 与 $||X + \xi Y| + \eta Z|$ 同分布.
- 问: 是否存在一个二维随机向量 (X_1, X_2) 及两个随机变量 X_3, X_4 , 使得 $(X_1, X_2), X_3, X_4$ 不独立, 但 X_1, X_3, X_4 独立, X_2, X_3, X_4 也独立.
- 问: 是否存在两个独立的, 均取值于 $[0, 1]$ 的随机变量 ξ, η , 使得 $\xi + \eta$ 服从 $[0, 2]$ 上的均匀分布?
- 设 ξ 服从 $[0, 1]$ 上的均匀分布, η 是Bernoulli随机变量, $P(\eta = 1) = P(\eta = -1) = \frac{1}{2}$, ξ 与 η 独立. 求 $\xi^{-1/2}\eta$ 的分布.
- 设 $P(\xi = 1) = P(\xi = -1) = \frac{1}{2}$, ξ 与 η 独立. 证明 $\xi\eta$ 与 ξ 独立的充要条件是 η 是对称随机变量.
- 设 (Ω, \mathcal{F}, P) 为离散概率空间, ξ, η 是其上的随机变量, 且满足

$$\omega \neq \omega' \implies \xi(\omega) \neq \xi(\omega'), \eta(\omega) \neq \eta(\omega').$$

证明: 除非有一个元素 ω_0 使得 $P(\{\omega_0\}) = 1$, 否则 ξ 与 η 不可能独立.

13. 设 ξ 为连续型随机变量, 密度函数为 f ; η 与 ξ 独立. 证明 $\xi + \eta$ 也是连续型随机变量, 密度函数为

$$g(x) = E[f(x - \eta)].$$

14. 设 ξ, η 是有界随机变量, 且对任意非负整数 m, n 有

$$E[\xi^m \eta^n] = E[\xi^m]E[\eta^n].$$

证明 ξ, η 独立.

15. 设 ξ 与 η 独立, 且 $\xi + \eta = 0$. 证明 ξ 服从单点分布.

16. 设 (ξ, η) 为连续型随机变量, A 为 \mathbb{R}^2 中的一条简单曲线. 证明

$$P((\xi, \eta) \in A) = 0.$$

17. 设 ξ, η 是独立同分布随机变量, 且是连续型的. 证明:

$$P(\xi > \eta) = P(\xi < \eta) = \frac{1}{2}.$$

18. 设 $\xi_n, n = 1, 2, \dots$ 是独立同分布连续型随机变量序列. 令

$$A_n := \{\omega : \xi_m(\omega) < \xi_n(\omega), \forall m < n\}, n \geq 2.$$

证明 $A_n, n \geq 2$ 独立且 $P(A_n) = \frac{1}{n}$.

19. 设 ξ, η 独立, 且 η 是连续型的. 证明 $\xi + \eta$ 也是连续型的. 去掉独立性条件之后该结论仍然成立吗? 证明或举出反例.

20. 设对任意 $i = 1, \dots, n$,

$$\xi_i = \sum_{j=1}^{n_i} a_{ij} 1_{A_{ij}},$$

其中 $\mathcal{A}_i := \{A_{ij}, j = 1, \dots, n_i\}$ 为 Ω 的一个分割, a_{ij} 是常数. 证明: ξ_1, \dots, ξ_n 独立的充要条件是 $\mathcal{A}_1, \dots, \mathcal{A}_n$ 独立.

21. 设 ξ, η 为取值于 \mathbb{R} 的独立同分布随机变量, 密度函数为 f , 且方差存在. 令 $X := |\xi - \eta|$. 求 X 的分布, 期望与方差.

22. 设 ξ, η 是取值于 $[0, 1]$ 的独立同分布随机变量, 都服从均匀分布. 以 d 表示 (ξ, η) 离点 $(\frac{1}{2}, \frac{1}{2})$ 的距离, d' 表示它离四个顶点的距离的最小值. 求 $P(d \leq d')$.

23. 设 ξ_1, \dots, ξ_n 为独立同分布随机变量, 服从 $[0, 1]$ 上的均匀分布. 令

$$X := \min\{\xi_1, \dots, \xi_n\}, Y := \max\{\xi_1, \dots, \xi_n\}.$$

设 $0 \leq a \leq b \leq 1, x_1, \dots, x_n \in [0, 1]$. 计算 $P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n | a \leq X \leq Y \leq b)$

24. 设 ξ, η 是独立同分布随机变量, F 是其公共的分布函数. 证明

$$P(\xi \neq \eta) = \sum_{x \in \mathbb{R}} |F(x) - F(x-)|^2.$$

25. 设 (ξ, η) 服从单位圆盘上的均匀分布. 令 $X := \xi^2 + \eta^2$,

$$Y := \xi \sqrt{-\frac{2 \ln X}{X}}, \quad Z := \eta \sqrt{-\frac{2 \ln X}{X}}.$$

证明 Y 与 Z 独立同分布, 其共同的分布为 $N(0, 1)$.

26. 设 (ξ, η) 服从 $[0, 1]^2$ 上的均匀分布. 令

$$X := \sqrt{-2 \ln \xi} \cos(2\pi\eta), \quad Y := \sqrt{-2 \ln \xi} \sin(2\pi\eta).$$

证明 X 与 Y 独立同分布, 其共同的分布为 $N(0, 1)$.

27. 设 $\xi = (\xi_1, \dots, \xi_n)$ 是连续型随机向量, 密度函数具有形式 $f(x_1^2 + \dots + x_n^2)$.

(a) 设 $n \geq 2$. 以 $(r, \theta_1, \dots, \theta_{n-1})$ 表示 \mathbb{R}^n 上的某个球极坐标. 证明 $r(\xi), \theta_1(\xi), \dots, \theta_{n-1}(\xi)$ 相互独立;

(b) 设 A 是 n 阶正交阵, 证明 $A\xi$ 与 ξ 同分布.

28. 设 $\xi \sim P(\lambda), \eta \sim P(\mu)$, 且 ξ 与 η 独立. 计算 $\xi + \eta$ 与 $\xi - \eta$ 的分布列.

29. 设 ξ, η 是独立随机变量.

(a) 令

$$\mathcal{F} := \{A \in \mathcal{B}^2 : E[1_A(\xi, \eta)] = E[E[1_A(\xi, x)]|_{x=\eta}]\}.$$

证明 $\mathcal{F} = \mathcal{B}^2$.

(b) 证明对 \mathbb{R}^2 上的任意简单Borel函数 f ,

$$E[f(\xi, \eta)] = E[E[f(\xi, x)]|_{x=\eta}].$$

(c) 证明对 \mathbb{R}^2 上的任意非负Borel函数 f ,

$$E[f(\xi, \eta)] = E[E[f(\xi, x)]|_{x=\eta}].$$

6.2 不相关与独立的关系

我们曾经证明, 当 ξ, η 独立时, 只要 $E[\xi]$ 与 $E[\eta]$ 均存在, 那么就有 $E[\xi\eta] = E[\xi]E[\eta]$, 而这就是不相关. 所以在期望存在时, 独立意味着不相关 (但要注意独立性对任意随机变量, 而不只是对期望存在的随机变量, 均可定义; 而不相关性则必须对期望存在的随机变量定义).

自然的问题是, 不相关是否意味着独立?

即使简单地从定义来看, 这个问题的答案也是否定的, 因为不相关是个整体的概念(求期望是整体概念), 但独立是局部概念, 因为 ξ 和 η 独立就意味着对任何 $A, B \in \mathcal{B}$, 都有

$$P(\xi \in A, \eta \in B) = P(\xi \in A)P(\eta \in B).$$

顺着这个思路, 可以很容易找出反例.

例1. 设

$$\Omega := \{\omega_1, \omega_2, \omega_3\}, \quad P(\omega_i) = \frac{1}{3}, \quad i = 1, 2, 3;$$

$$\begin{aligned}\xi(\omega_1) &= 3, \xi(\omega_2) = -2, \xi(\omega_3) = -1; \\ \eta(\omega_1) &= \frac{1}{3}, \eta(\omega_2) = \frac{1}{6}, \eta(\omega_3) = \frac{2}{3}.\end{aligned}$$

则

$$E[\xi] = 0, E[\xi\eta] = 0.$$

所以 ξ 与 η 不相关. 但

$$P(\xi = 3, \eta = \frac{1}{3}) = P(\omega_1) = \frac{1}{3} \neq P(\xi = 3)P(\eta = \frac{1}{3}) = P(\omega_1)^2 = \frac{1}{9}.$$

例2. 设 η 是随机变量, 值域为 $\{0, \frac{\pi}{2}, \pi\}$, 且

$$P(\eta = 0) = P(\eta = \frac{\pi}{2}) = P(\eta = \pi) = \frac{1}{3}.$$

令

$$\xi_1 = \cos \eta, \quad \xi_2 = \sin \eta.$$

则 $\xi_1\xi_2 \equiv 0$. 故

$$E[\xi_1\xi_2] = 0.$$

又

$$E[\xi_1] = 1 \times \frac{1}{3} + 0 \times \frac{1}{3} + (-1) \times \frac{1}{3} = 0$$

所以 ξ_1 与 ξ_2 不相关. 但由于

$$\xi_1^2 + \xi_2^2 \equiv 1$$

所以 ξ_1 与 ξ_2 不可能独立. 例如我们有

$$P(\xi_1 = 0, \xi_2 = 0) = 0,$$

但

$$P(\xi_1 = 0) = \frac{1}{3}, \quad P(\xi_2 = 0) = \frac{2}{3},$$

所以

$$P(\xi_1 = 0, \xi_2 = 0) \neq P(\xi_1 = 0)P(\xi_2 = 0).$$

但在一个重要的特殊情形, 不相关和独立是等价的, 这就是 (ξ, η) 服从二维正态分布的情况.

设 (ξ, η) 服从二维正态分布, 即 $(\xi, \eta) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$, 其密度函数为

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \cdot \left[\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right] \right\}.$$

我们曾经证明

$$\xi \sim N(\mu_1, \sigma_1), \quad \eta \sim N(\mu_2, \sigma_2).$$

而它们的协方差为

$$\text{Cov}(\xi, \eta) = \rho\sigma_1\sigma_2.$$

所以 ρ 就是相关系数. ξ, η 不相关时 $\rho = 0$, 于是

$$p(x, y) = p_\xi(x)p_\eta(y).$$

从而 ξ 与 η 独立. 所以, 我们有

定理 6.2.1. 如果 (ξ, η) 服从二维正态分布, 那么 ξ 与 η 独立的充要条件是它们不相关.

注意这里的前提条件: (ξ, η) 服从二维正态分布. 即使 (ξ, η) 不服从二维正态分布, ξ, η 的边沿分布仍然有可能是正态分布, 此时定理不再成立.

从另一个角度看, 也可以这样说: 两个独立的正态分布的联合分布是二维正态分布, 两个不相关的正态分布的联合分布未必是二维正态分布.

例3. 令

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

设 g 为连续偶函数, 满足

$$\begin{aligned} |g(x)| &\leq \varphi(x), \\ \int_{-\infty}^{\infty} g(x) dx &= 0. \end{aligned}$$

定义

$$p(x, y) := \varphi(x)\varphi(y) + g(x)g(y),$$

因为 $p(x, y) \geq 0$, 且

$$\int_{-\infty}^{\infty} p(x, y) dy := \int_{-\infty}^{\infty} \varphi(x)\varphi(y) dy + \int_{-\infty}^{\infty} g(x)g(y) dy = \varphi(x),$$

$$\int_{-\infty}^{\infty} p(x, y) dx = \varphi(y),$$

则 $p(x, y)$ 是二维密度函数.

设 (ξ, η) 的联合密度函数为 $p(x, y)$, 则 ξ 的边沿密度函数为 $\varphi(x)$, η 的边沿密度函数为 $\varphi(y)$. 所以我们有:

(i) (ξ, η) 的两个边沿分布都是正态分布;

(ii) ξ 与 η 不相关, 因为

$$\begin{aligned} E[\xi\eta] &= \iint xy p(x, y) dx dy \\ &= \iint xy \varphi(x)\varphi(y) dx dy + \iint xy g(x)g(y) dx dy \\ &= 0 = E[\xi]E[\eta]; \end{aligned}$$

(iii) ξ 与 η 不独立, 因为存在 x, y , $p(x, y) \neq \varphi(x)\varphi(y)$;

(iv) (ξ, η) 不服从二维正态分布, 因为 $p(x, y)$ 不是正态分布的密度函数.

所以你不能贸然说：两个服从正态分布的随机变量如果不相关，则一定独立。在下这个结论的时候，一定要注意前提是否满足，即联合分布是否为正态分布。

我们可以把这个例子的方法一般化：设 $p(x, y)$ 为密度函数， $q(x, y)$ 是Riemann可积函数，满足：

$$|q(x, y)| \leq p(x, y),$$

$$\int_R q(x, y) dx = \int_R q(x, y) dy = 0.$$

则

$$p(x, y) + q(x, y)$$

也是密度函数，且与 $p(x, y)$ 有相同的边沿分布。

这又一次说明了，不同的联合分布可以对应同样的边沿分布。

习题

1. 设 $\xi \sim N(0, 1)$, $P(\eta = 1) = P(\eta = -1) = \frac{1}{2}$, η 与 ξ 独立. 令 $\zeta = \xi\eta$.
 - (a) 求 ζ 的分布;
 - (b) 求 (ξ, ζ) 的协方差矩阵;
 - (c) (ξ, ζ) 是否服从正态分布? 证明你的结论.
2. 举例说明两个正态随机变量之和未必是正态随机变量.
3. (a) 设 ξ 与自己不相关, 证明 ξ 为常数;
 (b) 设 ξ 与自己独立, 证明 ξ 为常数;
4. 设 ξ_i 与 η_j 不相关, c_i, d_j 为常数, $i = 1, \dots, n, j = 1, \dots, m$. 证明 $\sum_{i=1}^n c_i \xi_i$ 与 $\sum_{j=1}^m d_j \eta_j$ 不相关.
5. 设 ξ, η 是随机变量, 且 $\xi - \eta$ 既与 ξ 不相关也与 η 不相关. 证明 $\xi - \eta$ 为常数.
6. 设 ξ, η 为随机变量, $E[\xi] = E[\eta] = 0, D[\xi] = D[\eta] = 1$, ξ 与 η 的相关系数为 ρ . 证明:

$$E[\max(\xi^2, \eta^2)] \leq 1 + \sqrt{1 - \rho^2}.$$

7. 设 ξ, η 为随机变量, ξ 与 η 的相关系数为 ρ . 证明: $\forall \varepsilon > 0$,

$$P(\{|\xi - E[\xi]| \leq \varepsilon\} \cup \{|\eta - E[\eta]| \leq \varepsilon\}) \leq \frac{1}{\varepsilon^2}(1 + \sqrt{1 - \rho^2}).$$

8. 设 ξ 和 η 均为只取两个不同值的随机变量. 证明 ξ 与 η 独立的充要条件是它们不相关.

6.3 独立随机变量之和

本节我们研究独立随机变量之和的收敛性. 具体地说, 假设 $\{\xi_i\}$ 是独立随机变量序列. 我们要研究

$$S_n := \sum_{i=1}^n \xi_i$$

的收敛性问题.

有别于一般的随机级数, 这个由独立随机变量组成的级数有许多独特的性质, 其中第一个就是:

命题 6.3.1. S_n 几乎必然收敛和依概率收敛是等价的.

这个命题得以成立的道理是我们有下面的Ottaviani不等式:

引理 6.3.2 (Ottaviani不等式). 设有正常数 a, b 使得对任意 $1 \leq k \leq n$ 有

$$P(|S_n - S_k| \leq b) \geq a.$$

则对任意 $c > 0$,

$$P\left(\max_{1 \leq k \leq n} |S_k| > b + c\right) \leq \frac{1}{a} P(|S_n| > c).$$

证明. 令

$$\tau := \inf\{k : |S_k| > b + c\},$$

$$A_k := \{|S_n - S_k| \leq b\}.$$

则 $\{\tau = k\}, k = 1, \dots, n$ 互不相交, 且

$$\sum_{k=1}^n \{\tau = k\} = \left\{ \max_{1 \leq k \leq n} |S_k| > b + c \right\},$$

$$\sum_{k=1}^n \{\tau = k\} A_k \subset \{|S_n| > c\}.$$

由于 $\{\tau = k\} \in \sigma(\xi_1, \dots, \xi_k)$, $A_k \in \sigma(\xi_{k+1}, \dots, \xi_n)$, 所以 $\{\tau = k\}$ 与 A_k 独立, $\forall k$. 因此

$$\begin{aligned} P(|S_n| > c) &\geq P\left(\sum_{k=1}^n \{\tau = k\} A_k\right) \\ &= \sum_{k=1}^n P(\tau = k) P(A_k) \\ &\geq a \sum_{k=1}^n P(\tau = k) \\ &= a P\left(\max_{1 \leq k \leq n} |S_k| \geq b + c\right). \end{aligned}$$

□

下面我们证明:

定理 6.3.3. S_n 依概率收敛与几乎必然收敛等价.

证明. 只需证明前者蕴含后者. $\forall \varepsilon > 0, \delta > 0$, 取整数 N 使得 $m, n \geq N$ 时

$$P(|S_n - S_m| > \varepsilon) < \delta, \quad P(|S_n - S_m| \leq \varepsilon) > 1 - \delta.$$

于是由上一引理

$$P\left(\max_{1 \leq k \leq n} |S_{N+k} - S_N| > 2\varepsilon\right) \leq \frac{1}{1-\delta} P(|S_{N+n} - S_N| > \varepsilon) < \frac{\delta}{1-\delta}.$$

因为

$$\left\{\max_{m, n > N} |S_m - S_n| > 4\varepsilon\right\} \leq \left\{\max_{m > N} |S_m - S_N| > 2\varepsilon\right\} \cup \left\{\max_{n > N} |S_n - S_N| > 2\varepsilon\right\},$$

因此

$$P\left(\max_{m, n > N} |S_m - S_n| > 4\varepsilon\right) < \frac{2\delta}{1-\delta}.$$

由 ε, δ 的任意性和(4.6), 知 S_n 几乎必然收敛. □

为得到 S_n 依概率收敛的条件, 我们需要下面的不等式.

定理 6.3.4. 记号和条件同上. 设 $E[\xi_n] = 0, \forall n$, 则

(i) (Chatterji 不等式) 设 $1 \leq p < 2$, 则

$$\|S_n\|_p^p \leq 2^{2-p} \sum_{i=1}^n \|\xi_i\|_p^p.$$

(ii) (Rio 不等式) 设 $p \geq 2$. 则

$$\|S_n\|_p^2 \leq (p-1) \sum_{i=1}^n \|\xi_i\|_p^2.$$

我们先证明下述引理, 第二个结果的证明需要它.

引理 6.3.5. 设 $p \geq 2$, ξ, η 为独立随机变量, $E[|\xi|^p] < \infty, E[|\eta|^p] < \infty$, 且 $E[\eta] = 0$. 则

$$\|\xi + \eta\|_p^2 \leq \|\xi\|_p^2 + (p-1)\|\eta\|_p^2.$$

证明. 当 $p = 2$ 时, 只需对平方项展开后利用独立性和 $E[\eta] = 0$ 消除交叉项, 便可得到等式. 因此下面假设 $p > 2$. 令

$$\varphi_t := \|\xi + t\eta\|_p^p.$$

设 $x, y \in \mathbb{R}$. Taylor 展开到二阶项, 得

$$|x + ty|^p = |x|^p + pt|x|^{p-2}xy + p(p-1) \int_0^t ds \int_0^s |x + uy|^{p-2}y^2 du.$$

将 x, y 用 ξ, η 代入, 取期望, 用 Fubini 定理并注意

$$E[|\xi|^{p-2}\xi\eta] = E[\eta]E[|\xi|^{p-2}\xi] = 0,$$

及(由Hölder不等式)

$$E[|\xi + u\eta|^{p-2}\eta^2] \leq \|\xi + u\eta\|_p^{p-2}\|\eta\|_p^2,$$

我们有

$$\varphi_t \leq \|\xi\|_p^p + p(p-1)\|\eta\|_p^2 \int_0^t ds \int_0^s \varphi_u^{1-2/p} du.$$

于是由附录推论11.2.4有

$$\varphi_t \leq (\|\xi\|_p^2 + (p-1)t^2\|\eta\|_p^2)^{p/2}.$$

取 $t = 1$ 并在两边开 $p/2$ 次方即完成证明. □

有了这个引理, 定理6.3.4第二个结果的证明就呼之即出了. 实际上, 我们有:

$$\begin{aligned} \|S_n\|_p^2 &= \|S_{n-1} + \xi_n\|_p^2 \\ &\leq \|S_{n-1}\|_p^2 + (p-1)\|\xi_n\|_p^2 \\ &\leq \dots \\ &\leq (p-1) \sum_{i=1}^n \|\xi_i\|_p^2. \end{aligned}$$

现在我们证明第一个结论. 首先注意 $p = 1$ 时对任意 ξ_i 均成立, 且不等式后面的常数是1. 因此下面假设 $1 < p < 2$. 此时, 先证明下面的引理.

引理 6.3.6. 设 ξ, η 为两个独立的随机变量, $E[\eta] = 0$, $1 < p < 2$. 则

$$E[|\xi + \eta|^p] \leq E[|\xi|^p] + 2^{2-p}E[|\eta|^p].$$

证明. 令

$$C_p := \sup_{x \in \mathbb{R}} \frac{|1 + x|^p - 1 - px}{|x|^p} \leq 2^{2-p}.$$

则对任意 $x, y \in \mathbb{R}$ 有

$$|x + y|^p \leq |x|^p + p|x|^{p-1}\text{sgn}(x)y + C_p|y|^p.$$

以 $x = \xi$, $y = \eta$ 代入并取期望, 仍注意到

$$E[|\xi|^{p-1}\text{sgn}(\xi)\eta] = E[|\xi|^{p-1}\text{sgn}(\xi)]E[\eta] = 0,$$

即得

$$E[|\xi + \eta|^p] \leq E[|\xi|^p] + C_p E[|\eta|^p].$$

□

对 $\xi = S_{n-1}$, $\eta = \xi_n$ 用上面的引理有

$$E[|S_{n-1} + \xi_n|^p] \leq E[|S_{n-1}|^p] + 2^{2-p}E[|\xi_n|^p].$$

递推下去便得到第一个结论.

下面我们证明:

定理 6.3.7. 设 $1 \leq p < \infty$, 且

$$\sum_{n=1}^{\infty} \|\xi_i\|_p^{p \wedge 2} < \infty,$$

则 $S_n \xrightarrow{a.s.}$ 且 S_n 在 L^p 中收敛.

证明. 注意对 $n > m$, 由引理6.3.4有

$$\|S_n - S_m\|_p^{p \wedge 2} \leq C \sum_{i=m+1}^n \|\xi_i\|_p^{p \wedge 2} \rightarrow 0,$$

所以 S_n 在 L^p 是Cauchy列, 因此是依概率的Cauchy列, 于是 $S_n \xrightarrow{P}$. 再由上一定理知 $S_n \xrightarrow{a.s.}$. 设这个极限是 S , 则由Fatou引理

$$E[|S_n - S|^p] \leq \liminf_{m \rightarrow \infty} E[|S_n - S_m|^p] \leq C \left(\sum_{i=n+1}^{\infty} \|\xi_i\|_p^{p \wedge 2} \right)^{\frac{p}{p \wedge 2}}.$$

所以

$$\lim_{n \rightarrow \infty} E[|S_n - S|^p] = 0.$$

□

习题

1. 设 ξ, η_1, η_2 是随机变量, $\sigma(\xi)$ 与 $\sigma(\eta_1, \eta_2)$ 独立. 设 $f: \mathbb{R}^3 \rightarrow \mathbb{R}$ 为有界Borel函数. 证明

$$E[f(\xi, \eta_1, \eta_2)] = E[\varphi(\eta_1, \eta_2)],$$

其中

$$\varphi(x, y) := E[f(\xi, x, y)].$$

2. 设 φ 是凸函数, ξ, η 是独立随机变量, $E[\xi] = 0$. 设 $\zeta \in \sigma(\eta)$ 且下式中涉及的随机变量均可积. 证明

$$E[\varphi(\xi + \eta)\zeta] \geq E[\varphi(\eta)\zeta].$$

3. 设 $\xi_1, \xi_2, \dots, \xi_n$ 是独立随机变量, $E[\xi_k] = 0, \forall k$. 设 φ 是凸函数且下面涉及到的期望均有限. 令

$$S_n := \sum_{k=1}^n \xi_k,$$

$$\tau := \inf\{k : |S_k| \geq C\}.$$

证明:

$$E[\varphi(S_n)1_{\tau=k}] \geq E[\varphi(S_k)1_{\tau=k}].$$

4. 记号同上. 证明: $\forall p > 1$,

$$\sum_{k=1}^n E[\varphi(S_k)1_{\tau=k}] \leq E[\varphi(S_n)],$$

5. 设 $\xi_1, \xi_2, \dots, \xi_n$ 是相互独立随机变量, $E[\xi_i] = 0, \forall i$. 记 $S_k := \xi_1 + \dots + \xi_k$. 证明Kolmogorov不等式

$$P\left(\max_{1 \leq k \leq n} |S_k| \geq C\right) \leq \frac{E[S_n^2]}{C^2}.$$

6. 条件和符号与上一题相同. 证明:若 φ 还是非负偶函数, 且在正半轴上是递增的, 则

$$P\left(\max_{1 \leq k \leq n} \varphi(S_k) \geq \varepsilon\right) \leq \frac{1}{\varphi(\varepsilon)} E[\varphi(S_n)], \quad \forall \varepsilon > 0.$$

这就是广义Kolmogorov不等式 (取 $\varphi(x) := x^2$ 则得到本来的Kolmogorov不等式).

7. 从(广义)Kolmogorov不等式出发, 你可以在什么条件下得到 S_n 几乎一致收敛的结论?

8. 证明: 对 $p > 1$ 有

$$E\left[\left(\max_{1 \leq k \leq n} |S_k|\right)^p\right] \leq q^p E[|S_n|^p],$$

其中 $1/p + 1/q = 1$. 此不等式称为Doon不等式.

9. 设 ξ_1, \dots, ξ_n 独立, $p \geq 1, E[|\xi_i|^p] < \infty, E\xi_i = 0, \forall i$. 再设 η_1, \dots, η_n 是有界随机变量, η_1 是常数, $\eta_k \in \sigma(\xi_1, \dots, \xi_{k-1}), \forall k > 1$. 令

$$S_k := \eta_1 \xi_1 + \dots + \eta_k \xi_k.$$

证明 $\forall c > 0$,

$$P\left(\sup_{1 \leq k \leq n} |S_k| \geq c\right) \leq \frac{1}{c^p} E[|S_n|^p].$$

6.4 随机游动

上节我们研究了独立随机变量之和 S_n 的收敛问题. 从收敛的条件可以看出, 为保证 S_n 收敛, 基本上是需要 $\|\xi_n\|_p$ 随 n 的增加而变小. 现在考虑另一种情况, 即 ξ_n 们同分布的情况. 此时 $\|\xi_n\|$ 是常数, 因此除非它等于零(这种情况毫无必要研究), 否则不可能收敛. 但这时有别的问题需要研究.

这种情况称为随机游动, 因为直观上, 如果一个醉鬼在大街上失去控制地、踉跄地游动, 那基本上就是一个随机游动. 当然, 随机游动还可以作为其它许多实际问题的数学模型, 例如金融市场上股价的模型. 基于其立足的背景, 就需要研究各种各样的问题. 例如对于股价, 股民们关心什么? 无非是股价的波动, 预期的收益, 买卖的时机等等. 这些问题就构成了随机游动理论丰富的研究内容——当然这些远远地超出了本课程的范围. 本节将介绍一些最初等的概念和结果.

定义 6.4.1. 设 ξ_1, ξ_2, \dots 是独立同分布随机变量列, $E[\xi_1] = 0, E[\xi_1^2] = 1$. 则

$$S_0 := 0, \quad S_n := \sum_{k=1}^n \xi_k$$

称为随机游动.

最简单的随机游动的例子当然是

$$P(\xi_1 = 1) = P(\xi_1 = -1) = \frac{1}{2}.$$

我们称之为简单随机游动.

我们有

命题 6.4.2.

$$E[S_n] = 0, \quad E[S_n^2] = n, \quad \forall n.$$

证明. (i) 用期望的线性性有(注意这里不需要独立性)

$$E[S_n] = \sum_{k=1}^n E[\xi_k] = 0.$$

(ii) 利用独立性有(独立性派上用场了)

$$\begin{aligned} E[S_n^2] &= \sum_{k=1}^n E[\xi_k^2] + \sum_{i \neq j} E[\xi_i \xi_j] \\ &= n + \sum_{i \neq j} E[\xi_i] E[\xi_j] \\ &= n. \end{aligned}$$

□

下面我们把这个结果做一个重要推广, 即由确定性时刻 n 推广到一类随机时刻, 即所谓停时. 为说明什么是停时, 我们先引进记号

$$\mathcal{F}_n := \sigma(S_k, k \leq n) = \sigma(\xi_1, \dots, \xi_n).$$

\mathcal{F}_n 的直观含义是, 它代表了序列 $\{S_0, S_1, S_2, \dots\}$ 在时刻 n 之前(包含 n)所包含的全部信息.

现在我们可以给出:

定义 6.4.3. 设 $\tau: \Omega \mapsto \mathbb{N}_+ \cup \{\infty\}$. 若 $\forall n, \{\tau \leq n\} \in \mathcal{F}_n$, 那么 τ 就称为停时.

设 τ 为停时. 若存在常数 $C, \tau \leq C$ a.s., 称 τ 为有界停时; 若 $\tau < +\infty$ a.s., 称 τ 为有限停时.

可以这样来理解停时: 如果 τ 代表某件事发生的时刻, 那么这件事在 n 之前是否发生(即是否 $\tau \leq n$), 可以由到 n 时刻为止这个序列提供的全部信息(即 \mathcal{F}_n)来判断. 换一种说法, 这就是 $\{\tau \leq n\} \in \mathcal{F}_n$.

比如说, 你驾车从甲地去丙地, 中间要经过乙地. 公路上每隔10公里有一个标志, 说明此地是何地. 那么, 若以 τ 表示你第一次看到乙地地标的时刻, 那么 τ 就是一个停时; 相反, 如果 θ 表示你最后一次看到乙地地标的时刻, θ 就不是一个停时, 因为你必须看到下一个地标才能判断上一次看到的是不是最后一个乙地地标.

如果你初到一个国家, 得到一本日历, 且该日历上注明了所有的节假日, 说明今夕何夕. 那么, 你过第一个节日的时刻, 是一个停时, 但你过最后一个节日的时间, 就不是停时. 当然, 下这个结论的前提是, 你不会把日历往后翻, 也不会有人告诉你一年到底有几个节假日.

因为 $\{\tau \leq k\} \in \mathcal{F}_k$, 而 \mathcal{F}_k 与 $\xi_{k+1}, \xi_{k+2}, \dots$ 独立, 所以 $\{\tau \leq k\}$ 与 $\xi_{k+1}, \xi_{k+2}, \dots$ 独立. 这一事实将起到非常重要的作用. 例如我们利用它就可以证明:

定理 6.4.4. 设 S_n 是随机游动, τ 是有界停时, 则

$$E[S_\tau] = 0, \quad E[S_\tau^2] = E[\tau].$$

证明. 设 $\tau \leq n$, 此时 $\sum_{k=0}^n 1_{\tau=k} = 1$. 我们有

$$\begin{aligned} E[S_\tau] &= E\left[\sum_{k=0}^n S_k 1_{\tau=k}\right] \\ &= \sum_{k=0}^n (E[S_k; \tau=k] + E[S_n - S_k]E[1_{\tau=k}]) \\ &= \sum_{k=0}^n (E[S_k; \tau=k] + E[(S_n - S_k)1_{\tau=k}]) \\ &= \sum_{k=0}^n E[S_n 1_{\tau=k}] \\ &= E\left[S_n \sum_{k=0}^n 1_{\tau=k}\right] \\ &= E[S_n] = 0, \end{aligned}$$

以及

$$\begin{aligned} E[S_\tau^2] &= E[(S_\tau - S_n + S_n)^2] \\ &= E[(S_\tau - S_n)^2] + 2E[(S_\tau - S_n)S_n] + E[S_n^2] \\ &= \sum_{k=0}^n E[(S_k - S_n)^2; \tau=k] + E[S_n^2] \\ &\quad - 2 \sum_{k=0}^n E[(S_n - S_k)(S_k + S_n - S_k); \tau=k] \\ &= \sum_{k=0}^n (n-k)P(\tau=k) + n - 2 \sum_{k=0}^n (n-k)P(\tau=k) \\ &= n - \sum_{k=0}^n (n-k)P(\tau=k) \\ &= \sum_{k=0}^n kP(\tau=k) \\ &= E[\tau]. \end{aligned}$$

□

若 τ 仅为有限停时, 这个结论不一定成立, 但附加上一定的条件后是可以成立的. 即我们有下面的:

推论 6.4.5. 设 τ 是有限停时, 且有常数 C 使

$$\sup_n |S_{\tau \wedge n}| \leq C,$$

则上面结果依然成立, 且 $E[\tau] \leq C^2$.

证明. 令 $\tau_n := \tau \wedge n$, 易证 τ_n 是停时, 且有界. 因此

$$E[S_{\tau_n}] = 0, \quad E[S_{\tau_n}^2] = E[\tau_n].$$

由有界收敛定理,

$$\begin{aligned} E[S_\tau] &= \lim_{n \rightarrow \infty} E[S_{\tau_n}] = 0, \\ E[S_\tau^2] &= \lim_{n \rightarrow \infty} E[S_{\tau_n}^2] = \lim_{n \rightarrow \infty} E[\tau_n] \leq C^2. \end{aligned}$$

再由单调收敛定理,

$$\lim_{n \rightarrow \infty} E[\tau_n] = E[\lim_{n \rightarrow \infty} \tau_n] = E[\tau].$$

□

于是我们有:

推论 6.4.6. 设 S_n 是简单随机游动, a, b 是整数且 $a < 0 < b$. 令

$$\tau := \inf\{n : S_n \notin (a, b)\}.$$

则

$$E[S_\tau] = 0, \quad E[S_\tau^2] = E[\tau].$$

特别地,

$$\begin{aligned} E[\tau] &< \infty, \\ P(S_\tau = a) &= \frac{b}{b-a}, \quad P(S_\tau = b) = \frac{-a}{b-a}. \end{aligned}$$

证明. 令 $\tau_a := \inf\{n : S_n = a\}$, $\tau_b := \inf\{n : S_n = b\}$, 则 $\tau = \tau_a \wedge \tau_b$. 由习题1, τ_a, τ_b 为有限停时. 因为 $\{\tau > n\} = \{\tau_a > n\} \cap \{\tau_b > n\}$, 因此 τ 也为停时, 且 $\sup_n |S_{\tau \wedge n}| \leq (-a) \vee b$, 所以可以用上面的结果.

又

$$\begin{aligned} 0 = E[S_\tau] &= E[a1_{S_\tau=a}] + E[b1_{S_\tau=b}] \\ &= aP(S_\tau = a) + bP(S_\tau = b), \end{aligned}$$

$$P(S_\tau = a) + P(S_\tau = b) = 1,$$

解出得

$$P(S_\tau = a) = \frac{b}{b-a}, \quad P(S_\tau = b) = \frac{-a}{b-a}.$$

□

这个结果告诉我们, 从零点出发的随机游动先到达 a 的可能性与先到达 b 的可能性之比, 恰好是起点离它们的距离之比的倒数——这是符合直觉的.

习题

1. 设 $\{S_n\}$ 是简单随机游动.

(a) 证明: $\forall k \in \mathbb{N}_+$,

$$\lim_{n \rightarrow \infty} P(S_{2n} = 2k) = 0.$$

(提示: 利用Stirling公式)

(b) $\forall k \geq 0$, 令

$$A_k := \{\limsup_{n \rightarrow \infty} S_n \geq k\}.$$

证明 $P(A_k) \geq \frac{1}{2}$. (提示: 利用Fatou引理)

(c) 证明

$$P(\{\limsup_{n \rightarrow \infty} S_n = \infty\}) = 1,$$

$$P(\{\liminf_{n \rightarrow \infty} S_n = -\infty\}) = 1.$$

(提示: 利用Kolmogorov 0-1律.)

(d) 设 $a \in \mathbb{N}$. 令 $\tau_a := \inf\{n : S_n = a\}$. 证明 τ_a 为有限停时.

2. 举例说明, 在推论6.4.5中, 条件 $\sup_n |S_{\tau \wedge n}| \leq C$ 不能去掉.

6.5 条件独立性

正如事件之间的关系一样, 随机变量之间, 除了独立性, 还有条件独立性的概念.

定义 6.5.1. 设 ξ, η 分别是 m 维和 n 维随机变量.

(i) 设 C 是事件, $P(C) > 0$. 若

$$P(\xi \in A, \eta \in B | C) = P(\xi \in A | C)P(\eta \in B | C) \quad \forall A \in \mathcal{B}^m, B \in \mathcal{B}^n,$$

则称给定 C 时, ξ 与 η 条件独立.

(ii) 设 $\mathcal{P} := \{C_i, i = 1, 2, \dots\}$ 为分割. 若

$$P(\xi \in A, \eta \in B | C_i) = P(\xi \in A | C_i)P(\eta \in B | C_i) \quad \forall A \in \mathcal{B}^m, B \in \mathcal{B}^n, i = 1, 2, \dots,$$

则称给定 \mathcal{P} 时, ξ 与 η 条件独立.

(iii) 设 ζ 为离散随机变量, 值域为 $\{c_i, i = 1, 2, \dots\}$. 令 $C_i := \{\zeta = c_i\}$ $\mathcal{P} := \{C_i, i = 1, 2, \dots\}$. 若给定 \mathcal{P} 时, ξ 与 η 条件独立, 则称给定 ζ 时, ξ 与 η 条件独立. 此时, 条件 C_i 记为 $\zeta = c_i$.

回忆给定分割时的条件概率记号:

$$P(\cdot | \mathcal{P})(\omega) = \sum_{i=1}^{\infty} P(\cdot | C_i) 1_{C_i}(\omega).$$

利用此记号, 给定 \mathcal{P} 时, ξ 与 η 独立可写成

$$P(\xi \in A, \eta \in B | \mathcal{P}) = P(\xi \in A | \mathcal{P})P(\eta \in B | \mathcal{P}).$$

再引进记号:

$$P(\cdot | \zeta)(\omega) := \sum_{i=1}^{\infty} P(\cdot | \zeta = c_i) 1_{\zeta=c_i}(\omega).$$

则给定 ζ 时, ξ 与 η 条件独立可写成

$$P(\xi \in A, \eta \in B | \zeta) = P(\xi \in A | \zeta)P(\eta \in B | \zeta).$$

与无条件时的情形相似, 我们有

命题 6.5.2. 给定 \mathcal{P} 时, ξ 与 η 条件独立的充要条件是: 对任意 $f \in b\mathcal{B}^{m+n}$,

$$E[f(\xi, \eta) | \mathcal{P}] = E[E[f(\xi, x) | \mathcal{P}]_{x=\eta} | \mathcal{P}].$$

证明也是相似的, 留给读者.

习题

1. 设 ξ_1, ξ_2, \dots 是独立的离散随机变量序列, h_n 为Borel函数, $n = 1, 2, \dots$. 递归地定义

$$S_1 := \xi_1, S_{n+1} = S_n + h_n(S_n)\xi_{n+1}, n = 1, 2, \dots.$$

证明:

- (a) $S_n \in \sigma(\xi_1, \dots, \xi_n)$;
- (b) 给定 S_n 时, $\sigma(\xi_1, \xi_2, \dots, \xi_n)$ 与 S_{n+1} 条件独立;
- (c) 给定 S_n 时, $\sigma(\xi_1, \xi_2, \dots, \xi_n)$ 与 S_{n+1}, S_{n+2}, \dots 条件独立.
- (d) 设 $E[\xi_i] = 0, E[\xi_i^2] = 1$. 证明:

$$E[S_{n+1} | \sigma(\xi_1, \xi_2, \dots, \xi_n)] = S_n,$$

$$E \left[S_{n+1}^2 - \sum_{i=1}^n h_i^2(S_i) \middle| \sigma(\xi_1, \xi_2, \dots, \xi_n) \right] = S_n^2 - \sum_{i=1}^{n-1} h_i^2(S_i).$$

7 大数定律

扔一枚硬币, 我们说正面出现的概率为 p , 反面出现的概率为 $q := 1 - p$, 根据在哪里? 为什么合理? 对这个问题, 物理学家和数学家的思维是不同的.

物理学家的办法是, 做几次试验, 每次扔它成千上万次, 如果每次正反面出现的次数之比都接近于 $p : q$, 那么就代表这个假设是正确的, 合理的. 然后再给个误差范围, 基本上就可以了. 事实上, 实际操作中也只能是这个办法.

而数学家无法接受这种思维方式, 这简直是岂有此理嘛! 按这种方式, Goldbach¹ (哥德巴赫)猜想早就证明了嘛, 孪生素数猜想也早就证明了嘛, Bieberbach² 猜想至迟在1972年已经证明了嘛, 还需要等待De Branges³几乎搞疯了才在1983年证明?!

所以需要证明, 数学家认为.

第一个给出严格证明的是Jacob Bernoulli. 他的证明实际上是说明了用频率来定义概率的合理性, 即当重复试验的次数足够大时, 正反两面出现的次数之比会稳定在 $p : q$ 附近. 这就是所谓的大数定理, 而这里的“大数”指的就是试验的次数很大. Bernoulli的证明是纯分析的, 相当繁琐与困难, 并且是历史上首次意识到极限的概念([1]). 他的这个工作是概率论历史的名副其实的发令枪和起跑线.

我们现在就介绍大数定律, 但不必沿着Bernoulli的道路. 我们使用的是概率方法. 如果说Bernoulli当年走的是羊肠小道的話, 今天我们可是有高铁可以坐喽!

7.1 Markov大数定律

现在持续地抛上面的那枚硬币, 当第 n 次出现正面时记 ξ_n 为1, 反面时记为0. 则 $\xi_1, \dots, \xi_n, \dots$ 独立, 且 $\xi_n \sim B(p)$. 进行 n 次实验后, 正面出现的总次数为 $\mu_n = \xi_1 + \dots + \xi_n$, 正面出现的频率为 $\frac{\mu_n}{n} = \frac{1}{n} \sum_{k=1}^n \xi_k$. 上面的问题就是在问 $\frac{1}{n} \sum_{k=1}^n \xi_k$ 收敛吗? 以什么方式收敛? 如果收敛, 极限是 p 吗?

这种问题其实非常普遍.

再设想一下, 我们有 m 个球, 每个球上有一个数字 $l_k, k = 1, \dots, m$. 现在把球从袋中摸出来, 记录下球上的数字, 再放回袋中, 充分混合后在摸出来, 记录数字. 如此重复下去, 我们便得到一系列随机变量 $\xi_i, i = 1, 2, 3, \dots$. 这些随机变量是相互独立且同分布的. 考虑它们的部分和的平均值:

$$\eta_n := \frac{1}{n} \sum_{i=1}^n \xi_i.$$

¹Chin Goldbach, 1690-1764, 普鲁士数学家.

²Ludwig Bieberbach, 1886-1982, 德国数学家.

³Louis de Branges, 1932- , 美国数学家.

因为每个球每次都被等可能地取到, 所以当 n 趋于无穷大时, 该和直观上应该趋于其均值, 即

$$a := \frac{1}{m} \sum_{k=1}^m l_k.$$

那么, 这个猜想对吗?

诸如此类的问题构成了所谓大数定律的研究范畴. 由于随机变量序列收敛的方式有很多, 本节先从依概率收敛的大数定律讲起.

我们先明确什么是大数定律.

定义 7.1.1. 设 $\{\xi_n\}$ 是一列随机变量. 记 $S_n := \sum_{k=1}^n \xi_k$, 若

$$\frac{S_n - E[S_n]}{n} \xrightarrow{P} 0,$$

则说 $\{\xi_n\}$ 服从大数定律.

下面是Markov⁴大数定律. 这个定理的证明现在看起来是如此直接和简短, 以致于我们会怀疑它配得上Markov的大名吗? 不是阿猫阿狗都会证吗? 但实际上, 谁知道在他那个时代, 在许多概念都是模糊不清, 许多工具都是残缺不全的年代, Markov为得到这个结果受过怎样的折磨?

定理 7.1.2 (Markov大数定律). 若 $\{\xi_n, n \geq 1\}$ 是一列随机变量序列, $D[\xi_n] < \infty, \forall n \geq 1$ 且

$$\lim_{n \rightarrow \infty} \frac{D[S_n]}{n^2} = 0, \quad (1.1)$$

则 $\{\xi_n\}$ 服从大数定律.

证明. 由Chebyshev不等式有, $\forall \varepsilon > 0$

$$P\left(\left|\frac{S_n - E[S_n]}{n}\right| \geq \varepsilon\right) \leq \frac{D[S_n]}{n^2 \varepsilon^2} \rightarrow 0.$$

□

那么, 什么时候(1.1)满足呢?

首先, 设 $\{\xi_n\}$ 两两不相关(特别地, 若它们相互独立), 且有 $\alpha < 1$ 使

$$\limsup_{n \rightarrow \infty} D[\xi_n] n^{-\alpha} < \infty.$$

则有常数 C 使得

$$D[S_n] = \sum_{k=1}^n D[\xi_k] \leq n^{1+\alpha} C.$$

因此(1.1)满足. 特别地, 当 $\alpha = 0$ 时, 定理7.1.2称为 Chebyshev大数定律.

推论 7.1.3 (Chebyshev大数定律). 设 $\{\xi_n, n \geq 1\}$ 是一列两两不相关的随机变量序列, 且 $D[\xi_n] \leq C, \forall n \geq 1$, 则 $\{\xi_n, n \geq 1\}$ 服从大数定律.

⁴Andrei Andreyevich Markov, 1856-1922, 俄罗斯-苏联数学家.

在Chebyshev大数定律中, 取 $\xi_k \sim B(p_k)$, $\{\xi_k\}$ 相互独立, 则

$$D[\xi_k] = p_k(1 - p_k) \leq \frac{1}{4},$$

所以条件(1.1)满足. 特别地, 当 $p_k = p$ 时, 就退化为Bernoulli原来的结果, 因此称为Bernoulli大数定律.

推论 7.1.4 (Bernoulli大数定律). 设 μ_n 是事件 A 在 n 次独立试验中出现的次数, p 是 A 在每次试验中发生的概率, 则

$$\frac{\mu_n}{n} \xrightarrow{P} p.$$

证明. 只需定义

$$\xi_k := \begin{cases} 1, & \text{第}k\text{次实验中}A\text{出现,} \\ 0, & \text{否则,} \end{cases}$$

则 $\mu_n = \sum_{k=1}^n \xi_k$, $\xi_k \sim B(p)$ 且相互独立. 因此条件(1.1)满足, 由Chebyshev大数定律结论成立. \square

不过, 如果我们面对的是独立随机变量序列, 则上面有关方差的条件可以放宽.

定理 7.1.5. 设 $\{\xi_n\}$ 是独立随机变量序列, $E[\xi_n] = 0$, $\forall n \geq 1$, 且下列条件之一满足:

(i) $1 \leq p < 2$,

$$\frac{1}{n^p} \sum_{i=1}^n \|\xi_i\|_p^p \rightarrow 0;$$

(ii) $p \geq 2$ 且

$$\frac{1}{n^2} \sum_{i=1}^n \|\xi_i\|_p^2 \rightarrow 0.$$

则大数定律成立.

证明. 注意

$$P\left(\frac{|S_n|}{n} \geq \varepsilon\right) \leq \frac{E[|S_n|^p]}{n^p \varepsilon^p}.$$

再利用定理6.3.4即可. \square

最后, 我们来看看Bernoulli大数定律的一个有趣的应用, 即Weierstrass⁵定理的概率证明.

定理 7.1.6 (Weierstrass定理). 设 $f = f(x)$ 为 $[0, 1]$ 上的连续函数, 则Bernstein⁶多项式函数

$$B_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right) C_n^k x^k (1-x)^{n-k}$$

在 $[0, 1]$ 上一致收敛于 f .

⁵Karl Theodor Wilhelm Weierstrass, 1815-1897, 德国数学家.

⁶Sergei Natanovich Bernstein, 1880-1968, 俄罗斯-苏联数学家.

证明. 固定 $x \in [0, 1]$, 设一系列独立同分布随机变量序列 $\xi_n, n = 1, 2, \dots$, 其中 $P(\xi_1 = 1) = x, P(\xi_1 = 0) = 1 - x$. 记 $S_n = \sum_{k=1}^n \xi_k$, 则 $S_n \sim B(n, x)$. 因此

$$E \left[f \left(\frac{S_n}{n} \right) \right] = B_n(x).$$

由Bernoulli大数定律, $\frac{S_n}{n} \xrightarrow{P} x$. 因 f 连续, 因此 $f \left(\frac{S_n}{n} \right) \xrightarrow{P} f(x)$. 又因为 f 有界, 由有界收敛定理知, $B_n(x) \rightarrow f(x), n \rightarrow \infty, \forall x \in [0, 1]$.

下面证明该收敛也是一致的. 一方面, 因 f 有界, 存在常数 M , 使 $|f(x)| \leq M$. 另一方面, 因 f 在 $[0, 1]$ 上一致连续, 因此对任意的 $\varepsilon > 0$, 存在 $\delta > 0$, 当 $|x - y| < \delta$ 时, $|f(x) - f(y)| \leq \varepsilon$.

由Chebyshev不等式, 有

$$P \left(\left| \frac{S_n}{n} - x \right| > \delta \right) \leq \frac{1}{\delta^2 n^2} E \left[(S_n - nx)^2 \right] = \frac{nx(1-x)}{\delta^2 n^2} \leq \frac{1}{4n\delta^2}, \forall x.$$

对任意的 $\varepsilon > 0, \forall x$, 有

$$\begin{aligned} |B_n(x) - f(x)| &\leq E \left[\left| f \left(\frac{S_n}{n} \right) - f(x) \right| \right] \\ &= E \left[\left| f \left(\frac{S_n}{n} \right) - f(x) \right| 1_{\left\{ \left| \frac{S_n}{n} - x \right| > \delta \right\}} \right] \\ &\quad + E \left[\left| f \left(\frac{S_n}{n} \right) - f(x) \right| 1_{\left\{ \left| \frac{S_n}{n} - x \right| \leq \delta \right\}} \right] \\ &= 2MP \left(\left| \frac{S_n}{n} - x \right| > \delta \right) + \varepsilon \\ &\leq \frac{M}{2n\delta^2} + \varepsilon. \end{aligned}$$

因此

$$\lim_{n \rightarrow \infty} \sup_{x \in [0, 1]} |B_n(x) - f(x)| \leq \varepsilon.$$

由 ε 的任意性结论得证. □

习题

1. 设 $\{\xi_n\}$ 是随机变量序列. 令 $X_n = (\xi_1, \dots, \xi_n)$. 设有常数 C 及 $\alpha < 1$ 使得

$$\text{Cov}(X_n, X_n) \leq n^{1+\alpha} C I_n, \quad \forall n.$$

证明 $\{\xi_n\}$ 服从大数定律.

7.2 强大数定律

如果把大数定律的结论

$$\frac{S_n - E[S_n]}{n} \xrightarrow{P} 0$$

加强为

$$\frac{S_n - E[S_n]}{\varphi(n)} \xrightarrow{a.s.} 0,$$

其中 $\varphi(n) = O(n)$ 或最好是 $o(n)$, 则称为强大数定律.

本节恒设 ξ_1, ξ_2, \dots 独立同分布, 且 $E[\xi_1] = 0$. 记 $S_n := \sum_{k=1}^n \xi_k$. 为得到强大数定律, 我们先准备两个分析引理. 第一个是:

引理 7.2.1 (Toeplitz引理). 设 $\{a_n\}$ 是非负数列, $a_1 > 0$, $b_n := \sum_{i=1}^n a_i \uparrow \infty$, $\{x_n\}$ 是数列且 $x_n \rightarrow x$. 则

$$\frac{1}{b_n} \sum_{i=1}^n a_i x_i \rightarrow x.$$

证明. $\forall \varepsilon > 0$, 取 N 使得

$$|x_n - x| < \frac{\varepsilon}{2}, \quad \forall n \geq N.$$

再取 $M \geq N$ 使得

$$\frac{1}{b_M} \sum_{i=1}^M a_i |x_i - x| < \frac{\varepsilon}{2}.$$

则当 $n \geq M$ 时

$$\begin{aligned} \left| \frac{1}{b_n} \sum_{i=1}^n a_i x_i - x \right| &\leq \frac{1}{b_n} \sum_{i=1}^n a_i |x_i - x| \\ &\leq \frac{1}{b_n} \sum_{i=1}^N a_i |x_i - x| + \frac{1}{b_n} \sum_{i=N+1}^n a_i |x_i - x| \\ &\leq \frac{1}{b_M} \sum_{i=1}^N a_i |x_i - x| + \frac{1}{b_n} \sum_{i=N+1}^n a_i |x_i - x| \\ &< \frac{\varepsilon}{2} + \frac{b_n - b_N}{b_n} \frac{\varepsilon}{2} \\ &< \varepsilon. \end{aligned}$$

□

第二个是:

引理 7.2.2 (Kronecker引理). 设 $\{x_n\}$ 是数列, $b_n > 0$ 且 $b_n \uparrow \infty$. 若

$$\sum_{n=1}^{\infty} \frac{x_n}{b_n}$$

收敛, 则

$$\lim_{n \rightarrow \infty} \frac{1}{b_n} \sum_{i=1}^n x_i = 0.$$

证明. 令 $b_0 = c_0 := 0$,

$$c_n := \sum_{i=1}^n \frac{x_i}{b_i}.$$

则 $x_n = b_n(c_n - c_{n-1})$. 于是由Abel分部求和公式有

$$\begin{aligned} \frac{1}{b_n} \sum_{i=1}^n x_i &= \frac{1}{b_n} \sum_{i=1}^n b_i (c_i - c_{i-1}) \\ &= c_n - \frac{1}{b_n} \sum_{i=1}^{n-1} c_i (b_{i+1} - b_i). \end{aligned}$$

于是由Toeplitz引理有

$$\lim_{n \rightarrow \infty} \frac{1}{b_n} \sum_{i=1}^n x_i = 0.$$

□

我们还需要下面的不等式:

引理 7.2.3. 设 ξ_1, ξ_2, η 为随机变量, $E[\eta] = 0$, 且 ξ_1, ξ_2 与 η 独立, $p \geq 1$. 则

$$E[|\xi_1 + \eta\xi_2|^p] \geq E[|\xi_1|^p].$$

证明. 由独立性, 定理6.1.3及Jensen不等式有

$$\begin{aligned} E[|\xi_1 + \eta\xi_2|^p] &= E[E[|x + y\eta|^p] |_{x=\xi_1, y=\xi_2}] \\ &\geq E[|E[x + y\eta]|^p |_{x=\xi_1, y=\xi_2}] \\ &= E[|\xi_1|^p]. \end{aligned}$$

□

下面这个不等式取自[8](在 $p = 2$ 时是由Hajek-Renyi证明的).

引理 7.2.4 (广义Hájek⁷-Rényi⁸不等式). 设 $\{a_n\}$ 是递增正数列, $m < n$, $\varepsilon > 0$. 令

$$A := \left\{ \max_{m \leq i \leq n} a_i^{-1} |S_i| \geq \varepsilon \right\}.$$

(i) 若 $1 \leq p \leq 2$, 则

$$\varepsilon^p P(A) \leq 2^{2-p} a_m^{-p} \sum_{k=1}^m E[|\xi_k|^p] + 2^{2-p} \sum_{k=m+1}^n a_k^{-p} E[|\xi_k|^p].$$

(ii) 若 $p > 2$, 且有常数 $0 < \alpha < 1, C > 0$ 使得

$$k^{\frac{1}{2}-\frac{1}{p}} \leq C a_k^\alpha, \quad \forall k.$$

则

$$\varepsilon^p P(A) \leq C_1 a_m^{\alpha p - p} \sum_{k=1}^m E[|\xi_k|^p] + C_1 \sum_{k=m+1}^{n-1} a_k^{\alpha p - p} E[|\xi_k|^p] + C_1 a_n^{\alpha p - p} \sum_{k=1}^n E[|\xi_k|^p],$$

其中 $C_1 = C^p(1 - \alpha)^{-1}$.

证明. 令

$$\tau := \inf \{ i \geq m : a_i^{-1} |S_i| \geq \varepsilon \}.$$

则对 $i \geq k \geq m$, 在引理7.2.3中取 $\xi_1 = S_k 1_{\tau=k}$, $\xi_2 = 1_{\tau=k}$, $\eta = S_i - S_k$, 有

$$E[|S_i|^p 1_{\tau=k}] \geq E[|S_k|^p 1_{\tau=k}].$$

⁷Jaroslav Hájek, 1926-1974, 捷克数学家.

⁸Alfréd Rényi, 1921-1970, 匈牙利数学家.

因此, 对 $k \geq m$ 有

$$\begin{aligned}
 \varepsilon^p P(\tau = k) &\leq a_k^{-p} E[|S_k|^p 1_{\tau=k}] \\
 &= \left(\sum_{i=k}^{n-1} (a_i^{-p} - a_{i+1}^{-p}) + a_n^{-p} \right) E[|S_k|^p 1_{\tau=k}] \\
 &= \left(\sum_{i=k}^{n-1} (a_i^{-p} - a_{i+1}^{-p}) E[|S_k|^p 1_{\tau=k}] + a_n^{-p} E[|S_k|^p 1_{\tau=k}] \right) \\
 &\leq \sum_{i=k}^{n-1} (a_i^{-p} - a_{i+1}^{-p}) E[|S_i|^p 1_{\tau=k}] + a_n^{-p} E[|S_n|^p 1_{\tau=k}].
 \end{aligned}$$

于是

$$\begin{aligned}
 \varepsilon^p P\left(\max_{m \leq i \leq n} a_i^{-1} |S_i| \geq \varepsilon\right) &= \varepsilon^p \sum_{k=m}^n P(\tau = k) \\
 &\leq \sum_{i=m}^{n-1} (a_i^{-p} - a_{i+1}^{-p}) E[|S_i|^p] + a_n^{-p} E[|S_n|^p] \\
 &= a_m^{-p} E[|S_m|^p] + \sum_{k=m+1}^n a_k^{-p} (E[|S_k|^p] - E[|S_{k-1}|^p]) \\
 &=: E[\eta].
 \end{aligned}$$

下面估计 $E[\eta]$.

(i) 当 $1 \leq p \leq 2$ 时, 由定理6.3.4中的Chatterji不等式有

$$E[|S_k|^p] \leq 2^{2-p} \sum_{i=1}^k E[|\xi_i|^p],$$

以及由引理6.3.6又有

$$E[|S_k|^p] - E[|S_{k-1}|^p] \leq 2^{2-p} E[|\xi_k|^p],$$

因此

$$\begin{aligned}
 E[\eta] &= a_m^{-p} E[|S_m|^p] + \sum_{k=m+1}^n a_k^{-p} (E[|S_k|^p] - E[|S_{k-1}|^p]) \\
 &\leq 2^{2-p} a_m^{-p} \sum_{k=1}^m E[|\xi_k|^p] + 2^{2-p} \sum_{k=m+1}^n a_k^{-p} E[|\xi_k|^p],
 \end{aligned}$$

完成证明.

(ii) 设 $p > 2$. 此时, 由定理6.3.4中的Rio不等式

$$\|S_i\|_p^2 \leq (p-1) \sum_{k=1}^i \|\xi_k\|_p^2,$$

和Hölder不等式有

$$\begin{aligned} E[|S_i|^p] &= (\|S_i\|_p^2)^{p/2} \\ &\leq \left((p-1) \sum_{k=1}^i \|\xi_k\|_p^2 \right)^{p/2} \\ &\leq (p-1)^{p/2} i^{\frac{p}{2}-1} \sum_{k=1}^i E[|\xi_k|^p]. \end{aligned}$$

所以

$$\begin{aligned} (p-1)^{-p/2} E[\eta] &\leq \sum_{i=m}^{n-1} \left(i^{\frac{p}{2}-1} (a_i^{-p} - a_{i+1}^{-p}) \sum_{k=1}^i E[|\xi_k|^p] \right) + a_n^{-p} n^{\frac{p}{2}-1} \sum_{k=1}^n E[|\xi_k|^p] \\ &= \sum_{k=1}^{n-1} \left(E[|\xi_k|^p] \sum_{i=m \vee k}^{n-1} i^{\frac{p}{2}-1} (a_i^{-p} - a_{i+1}^{-p}) \right) + a_n^{-p} n^{\frac{p}{2}-1} \sum_{k=1}^n E[|\xi_k|^p]. \end{aligned}$$

又

$$\begin{aligned} \sum_{i=m \vee k}^{n-1} i^{\frac{p}{2}-1} (a_i^{-p} - a_{i+1}^{-p}) &= \sum_{i=m \vee k}^{n-1} i^{\frac{p}{2}-1} \int_{a_i}^{a_{i+1}} \frac{p}{x^{p+1}} dx \\ &\leq C^p \sum_{i=m \vee k}^{n-1} \int_{a_i}^{a_{i+1}} \frac{p}{x^{p-\alpha p+1}} dx \\ &\leq C^p (1-\alpha)^{-1} a_{m \vee k}^{\alpha p-p}, \end{aligned}$$

所以

$$E[\eta] \leq C_1 a_m^{\alpha p-p} \sum_{k=1}^m E[|\xi_k|^p] + C_1 \sum_{k=m+1}^{n-1} a_k^{\alpha p-p} E[|\xi_k|^p] + C_1 a_n^{\alpha p-p} \sum_{k=1}^n E[|\xi_k|^p],$$

其中 $C_1 = C^p (1-\alpha)^{-1}$. 至此证明全部完成. □

下面是本节的主要结果:

定理 7.2.5 (强大数定律). 保持上一引理的记号及条件, 并设存在 $p \geq 1$ 满足

(i) $1 \leq p \leq 2$ 时

$$\sum_{n=1}^{\infty} a_n^{-p} E[|\xi_n|^p] < \infty;$$

(ii) $p > 2$ 时

$$\sum_{n=1}^{\infty} a_n^{\alpha p-p} E[|\xi_n|^p] < \infty.$$

则

$$\lim_{n \rightarrow \infty} \frac{S_n}{a_n} = 0 \quad a.s..$$

注. $p = 2$, $a_n = n$ 时称为 Kolmogorov 强大数定律.

证明. (i) 当 $1 \leq p \leq 2$ 时,

$$\varepsilon^p P\left(\max_{m \leq i \leq n} a_i^{-1} |S_i| \geq \varepsilon\right) \leq 2^{2-p} a_m^{-p} \sum_{i=1}^m E[|\xi_i|^p] + 2^{2-p} \sum_{i=m+1}^n a_i^{-p} E[|\xi_i|^p].$$

令 $n \rightarrow \infty$ 得

$$\varepsilon^p P\left(\max_{i \geq m} a_i^{-1} |S_i| \geq \varepsilon\right) \leq 2^{2-p} a_m^{-p} \sum_{i=1}^m E[|\xi_i|^p] + 2^{2-p} \sum_{i=m+1}^{\infty} a_i^{-p} E[|\xi_i|^p].$$

再令 $m \rightarrow \infty$, 并用 Kronecker 引理即得

$$\lim_{m \rightarrow \infty} P\left(\max_{i \geq m} a_i^{-1} |S_i| \geq \varepsilon\right) = 0.$$

(ii) $p > 2$ 时, 用对应的不等式, 证明类似, 请自行完成. \square

推论 7.2.6 (Borel 强大数定律). 设 μ_n 是事件 A 在 n 次独立试验中出现的次数, p 是 A 在每次试验中发生的概率, 则

$$\frac{\mu_n}{n} \xrightarrow{a.s.} p.$$

证明. 回顾 $\mu_n = \sum_{k=1}^n \xi_k$, 其中 $\xi_k \sim B(p)$ 且相互独立. 因为 $D[\xi_k] = p(1-p)$, 因此

$$\sum_{n=1}^{\infty} \frac{1}{n^2} E[|\xi_n - p|^2] = p(1-p) \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty.$$

由上一定理的(i), 结论成立. \square

习题

1. 设 $1 \leq p \leq 2$. 设 φ 为 \mathbb{R}_+ 上的正函数, 且存在 $N > 0$ 使得

$$\varphi'(x) \geq 1, \quad x \in [N, \infty),$$

$$\int_N^{\infty} \frac{dx}{\varphi(x)} < \infty.$$

令

$$a_n := \varphi\left(\sum_{k=1}^n E[|\xi_k|^p]\right)^{1/p}.$$

证明: 若 $a_n \rightarrow \infty$, 则

$$\lim_{n \rightarrow \infty} \frac{S_n}{a_n} = 0 \quad a.s..$$

2. 设 $p \geq 2$, 且

$$\sum_{n=1}^{\infty} \frac{E[|\xi_n|^p]}{n^{\frac{p}{2}+1}} < \infty.$$

证明

$$\lim_{n \rightarrow \infty} \frac{S_n}{n} = 0 \quad a.s..$$

7.3 Glivenko–Cantelli定理

设 ξ_1, ξ_2, \dots 是概率空间 (Ω, \mathcal{F}, P) 上的一列独立同分布的随机变量, $F(x)$ 为它们共同的分
布函数. 对每个 $n \geq 1$, 定义它们的经验分布函数为

$$F_n(x) := \frac{1}{n} \sum_{k=1}^n 1_{\{\xi_k \leq x\}}, \quad x \in \mathbb{R}.$$

则 $F_n(x)$ 单调递增, 右连续. 它表示随机变量 $\xi_1, \xi_2, \dots, \xi_n$ 中小于等于 x 的个数的频率, 且 $E[F_n(x)] = F(x)$. 由 Bernoulli 大数定律知, $\forall x \in \mathbb{R}$,

$$F_n(x) \xrightarrow{P} F(x), \quad n \rightarrow \infty.$$

由Borel强大数定律知, $\forall x \in \mathbb{R}$,

$$F_n(x) \xrightarrow{a.s.} F(x), \quad n \rightarrow \infty.$$

事实上, 我们能证明上面的几乎必然收敛也是关于 $x \in \mathbb{R}$ 一致收敛的. 这为用经验分布函数来近似分布函数 $F(x)$ 提供了理论依据.

定理 7.3.1 (Glivenko⁹–Cantelli¹⁰定理). 符号和条件同上, 则

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{a.s.} 0, \quad n \rightarrow \infty.$$

证明. 任给 $m \geq 2$, 定义 $x_{m,0} = -\infty$,

$$x_{m,k} = \inf \left\{ x \in \mathbb{R}; F(x) \geq \frac{k}{m} \right\}, \quad k = 1, 2, \dots, m-1$$

和 $x_{m,m} = +\infty$. 则对 $k = 0, 1, \dots, m-1$,

$$\begin{aligned} F(x_{m,k+1-}) - F(x_{m,k}) &= \lim_{\delta \downarrow 0} [F(x_{m,k+1} - \delta) - F(x_{m,k} + \delta)] \\ &\leq \frac{k+1}{m} - \frac{k}{m} \\ &\leq \frac{1}{m}. \end{aligned}$$

因此, 对任意的 $x \in \mathbb{R}$, 存在 $0 \leq k \leq m-1$, 使 $x \in [x_{m,k}, x_{m,k+1})$, 且由 $F_n(x)$ 和 $F(x)$ 关于 x 的单调性有

$$\begin{aligned} F_n(x) - F(x) &\leq F_n(x_{m,k+1-}) - F(x_{m,k}) \\ &= [F_n(x_{m,k+1-}) - F(x_{m,k+1-})] + [F(x_{m,k+1-}) - F(x_{m,k})] \\ &\leq [F_n(x_{m,k+1-}) - F(x_{m,k+1-})] + \frac{1}{m}. \end{aligned}$$

同理

$$\begin{aligned} F_n(x) - F(x) &\geq F_n(x_{m,k}) - F(x_{m,k+1-}) \\ &= [F_n(x_{m,k}) - F(x_{m,k})] - [F(x_{m,k+1-}) - F(x_{m,k})] \\ &\geq [F_n(x_{m,k}) - F(x_{m,k})] - \frac{1}{m}. \end{aligned}$$

⁹Valery Ivanovich Glivenko, 1896-1940, 俄罗斯-苏联数学家.

¹⁰Francesco Paolo Cantelli, 1875-1966, 意大利数学家.

结合上面两个不等式有

$$\begin{aligned} & \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \\ & \leq \max_{0 \leq k \leq m-1} \{|F_n(x_{m,k+1}-) - F(x_{m,k+1}-)|, |F_n(x_{m,k}) - F(x_{m,k})|\} + \frac{1}{m}. \end{aligned}$$

由Borel强大数定律知, $\forall k = 0, 1, \dots, m-1$,

$$F_n(x_{m,k}) \xrightarrow{a.s.} F(x_{m,k}), \quad n \rightarrow \infty.$$

同理, 因为

$$F_n(x-) := \frac{1}{n} \sum_{k=1}^n 1_{\{\xi_k < x\}}, \quad x \in \mathbb{R}$$

和 $E[F_n(x-)] = F(x-)$, 由Borel强大数定律也知, $\forall x \in \mathbb{R}$,

$$F_n(x-) \xrightarrow{a.s.} F(x-), \quad n \rightarrow \infty.$$

因此, $\forall k = 1, 2, \dots, m$,

$$F_n(x_{m,k}-) \xrightarrow{a.s.} F(x_{m,k}-), \quad n \rightarrow \infty.$$

综上, $\forall m \geq 2$, 有

$$\zeta := \limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \frac{1}{m}, \quad a.s..$$

最后, 因为 $F_n(x)$ 和 $F(x)$ 右连续, 则

$$\limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| = \limsup_{n \rightarrow \infty} \sup_{x \in \mathbb{Q}} |F_n(x) - F(x)|,$$

所以 ζ 为随机变量. 再由概率的连续性有

$$P(\zeta \neq 0) = P\left(\bigcup_m \left\{\zeta \geq \frac{1}{m}\right\}\right) = \lim_{m \rightarrow \infty} P\left(\zeta \geq \frac{1}{m}\right) = 0,$$

定理得证. □

8 特征函数

本章我们介绍一个研究随机变量的重要工具——特征函数.

当然, 这个工具并不是概率论所特有的. 从分析的角度看, 一个随机变量的特征函数其实就是其分布函数的Fourier变换. 不过, 和分析中Fourier变换的一般理论相比, 它具有强烈的概率特点. 这表现在作为Riemann-Stieltjes积分或Lebesgue-Stieltjes积分, 它既不可归类于光滑函数或 L^p 函数的Fourier变换这么一个狭小的范围内, 又不必放在广义函数的Fourier变换这么一个庞大的框架中, 表现在它在研究独立随机变量之和时发挥的巨大作用, 表现在有时候甚至连定义分布(例如退化的多维正态分布)这么一个“很概率”的事情也无法回避它……

8.1 复随机变量

研究特征函数会涉及到取复数值的随机变量, 所以我们需要在这方面做些准备工作. 我们迄今为止接触到的随机变量都是取实数值的, 所以就无修饰地用随机变量称呼之. 以后我们也保持这个习惯, 即单说随机变量时, 指的就是实值随机变量, 或至多是广义实值的, 即可以取 $\pm\infty$ 的随机变量, 虚随机变量则是随机变量乘以虚数单位 i , 而复随机变量就是由实随机变量和虚随机变量组成的东西, 即我们有:

定义 8.1.1. 设 ξ, η 是随机变量, 则

$$\zeta := \xi + i\eta$$

称为复随机变量.

若 ξ 和 η 的期望均存在, 那么 ζ 的期望定义为

$$E[\zeta] := E[\xi] + iE[\eta].$$

显然, 这样定义的期望保持了线性性, 即

$$E[z_1\zeta_1 + z_2\zeta_2] = z_1E[\zeta_1] + z_2E[\zeta_2], \quad \forall z_1, z_2 \in \mathbb{C}.$$

对复数 $z = x + iy$, 其共轭为

$$\bar{z} := x - iy,$$

其模为

$$|z| := \sqrt{x^2 + y^2} = (z\bar{z})^{\frac{1}{2}}.$$

因此, 随机变量 ζ 的共轭为

$$\bar{\zeta} := \xi - i\eta;$$

ζ 的模为

$$|\zeta| := \sqrt{\xi^2 + \eta^2}.$$

显然

$$\operatorname{Re}(E[\zeta]) = E[\operatorname{Re}(\zeta)], \quad \operatorname{Im}(E[\zeta]) = E[\operatorname{Im}(\zeta)], \quad \overline{E[\zeta]} = E[\bar{\zeta}].$$

我们有如下的Minkowski不等式:

命题 8.1.2.

$$|E[\zeta]| \leq E[|\zeta|].$$

证明. 注意

$$\operatorname{Re}(\overline{E[\zeta]}\zeta) \leq \left| \overline{E[\zeta]}\zeta \right| \leq |E[\zeta]||\zeta|,$$

两边取期望即得

$$|E[\zeta]|^2 = \overline{E[\zeta]}E[\zeta] = \operatorname{Re}(\overline{E[\zeta]}E[\zeta]) = \operatorname{Re}(E[\overline{E[\zeta]}\zeta]) = E[\operatorname{Re}(\overline{E[\zeta]}\zeta)] \leq |E[\zeta]|E[|\zeta|].$$

□

8.2 定义及例子

我们首先给出:

定义 8.2.1. 设 F 是分布函数. $\forall t \in \mathbb{R}$, 定义

$$f(t) := \int_{-\infty}^{\infty} e^{itx} dF(x), \quad (2.1)$$

称为 F 的特征函数. 若 ξ 是随机变量, 以 F 为分布函数, 则 f 也称为 ξ 的特征函数. 此时

$$f(t) := E[e^{it\xi}].$$

由于(2.1)右边的被积函数是有界连续的, 所以是可积的, 且此积分可理解为Riemman-Stieltjes积分. 当然, 它也能被理解为Lebesgue-Stieltjes积分.

当涉及多个随机变量的特征函数时, 为宣示主权, 会记 ξ 的特征函数为 $f_{\xi}(t)$. 显然, 当 $\eta = a\xi + b$ 时, 其中 $a, b \in \mathbb{R}$ 为常数, η 的特征函数为

$$f_{\eta}(t) = E[e^{it(a\xi+b)}] = e^{itb} f_{\xi}(at).$$

我们注意到, 若 F 是离散型的, 分布列为 $\{(x_n, p_n)\}$, 则

$$f(t) = \sum_{n=1}^{\infty} e^{itx_n} p_n;$$

若 F 是连续型的, 密度函数为 p , 则

$$f(t) = \int_{-\infty}^{\infty} e^{itx} p(x) dx,$$

且此积分可理解为Riemman积分.

下面我们计算一些常用随机变量的特征函数.

1. 单点分布: 设 ξ 服从单点分布: 即有 $a \in \mathbb{R}$ 使

$$P(\xi = a) = 1.$$

则

$$f(t) = e^{ita}.$$

2. Bernoulli分布: $\xi \sim B(p)$, $q = 1 - p$.

$$f(t) = q + pe^{it}.$$

3. 二项分布: $\xi \sim B(n, p)$.

此时 $\xi = \xi_1 + \cdots + \xi_n$, 其中 ξ_1, \cdots, ξ_n 独立且 $\xi_k \sim B(p)$. 所以

$$\begin{aligned} f_\xi(t) &= E \left[\exp \left\{ \sum_{k=1}^n it\xi_k \right\} \right] = \prod_{k=1}^n E[\exp\{it\xi_k\}] \\ &= \prod_{k=1}^n f_{\xi_k}(t) = (q + pe^{it})^n, \quad q = 1 - p. \end{aligned}$$

这个结果有个自然的推广: 若 $\xi = \xi_1 + \cdots + \xi_n$, 其中 ξ_1, \cdots, ξ_n 独立且 $\xi_k \sim B(p_k)$. 则

$$f_\xi(t) = \prod_{k=1}^n f_{\xi_k}(t) = \prod_{k=1}^n (q_k + p_k e^{it}), \quad q_k = 1 - p_k.$$

4. Poisson分布: $\xi \sim P(\lambda)$.

$$f(t) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k e^{itk}}{k!} = \exp(\lambda(e^{it} - 1)).$$

5. ξ 的分布列为: $P(\xi = 1) = P(\xi = -1) = \frac{1}{2}$.

$$f(t) = \frac{1}{2}(e^{it} + e^{-it}) = \cos t.$$

6. 标准正态分布: $\xi \sim N(0, 1)$.

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} dx.$$

这个分布的特征函数我们曾经算过. 它还有更多的计算方法, 例如可将 e^{itx} 展成Taylor级数后分别计算每一项, 然后对级数求和; 也可将被积函数用Euler公式写为实部和虚部, 因为虚部为奇函数所以在对称区间上的积分为零, 对实部用两次分部积分公式得到一个一阶齐次线性常微分方程, 然后解方程. 还可以利用复变函数的围道积分的方法如下: 首先注意

$$\begin{aligned} f(t) &= \lim_{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-n}^n e^{itx} e^{-\frac{x^2}{2}} dx \\ &= e^{-\frac{1}{2}t^2} \lim_{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-n}^n e^{-\frac{(x-it)^2}{2}} dx \\ &= e^{-\frac{1}{2}t^2} \lim_{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{C_n} e^{-\frac{z^2}{2}} dz, \end{aligned}$$

其中 C_n 为直线段 $\{z; x - it, |x| \leq n\}$. 设 C 是由 $\text{Im}(z) = 0, \text{Re}(z) = -n, \text{Re}(z) = n$ 及 $\text{Im}(z) = -t$ 组成的封闭曲线, 按顺时针方向取积分. 因为 $e^{-\frac{z^2}{2}}$ 解析, 所以

$$\int_C e^{-\frac{z^2}{2}} dz = \int_{C_n} e^{-\frac{z^2}{2}} dz + \varepsilon_n + \int_n^{-n} e^{-\frac{x^2}{2}} dx = 0.$$

其中 $|\varepsilon_n(t)| \leq ce^{-\frac{n^2}{2}}$. 令 $n \rightarrow \infty$ 得

$$f(t) = e^{-\frac{1}{2}t^2}.$$

若 $\xi \sim N(\mu, \sigma^2)$, 则 $\eta := \frac{\xi - \mu}{\sigma} \sim N(0, 1)$, 而

$$\xi = \sigma\eta + \mu.$$

所以

$$f_\xi(t) = E[\exp\{it(\sigma\eta + \mu)\}] = e^{it\mu - \frac{1}{2}\sigma^2 t^2}.$$

7. 均匀分布 $\xi \sim U(a, b)$.

$$f(t) = \frac{1}{b-a} \int_a^b e^{itx} dx = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$

特别, 当 $a = -1, b = 1$ 时,

$$f(t) = \frac{\sin t}{t}.$$

8. 指数分布 $E(\lambda)$.

$$\begin{aligned} f(t) &= \frac{1}{\lambda} \int_0^\infty e^{itx} e^{-\lambda x} dx \\ &= \frac{1}{\lambda} \int_0^\infty e^{(it-\lambda)x} dx \\ &= \frac{1}{\lambda} \cdot \frac{e^{(it-\lambda)x}}{it-\lambda} \Big|_0^\infty \\ &= \frac{1}{\lambda(\lambda - it)}. \end{aligned}$$

最后两个实变量复值的指数函数的积分计算参见附录11.5.

习题

1. 设 f 为特征函数, 证明 $g(t) := \overline{f(t)}$ 也是特征函数.
2. 设 φ 是 X 的特征函数. 证明:
 - (a) 若有 $t_n \rightarrow 0$ 使得 $|\varphi(t_n)| = 1$, 则存在 $b \in \mathbb{R}$ 使得 $X = b$ a.s..
 - (b) 若有 $t_n \rightarrow 0$ 使得 $\varphi(t_n) = 1$, 则存在 $b \in \mathbb{R}$ 使得 $X = b$ a.s..
3. 证明: 对 $\alpha > 2$, $\varphi(t) := \exp\{-|t|^\alpha\}$ 不是特征函数.
4. 设 φ 是特征函数. 证明

$$1 - |\varphi(2t)|^2 \leq 4(1 - |\varphi(t)|^2),$$

$$1 - \text{Re}\varphi(nt) \leq n(1 - (\text{Re}\varphi(t))^n) \leq n^2(1 - \text{Re}\varphi(t)), \quad n = 0, 1, 2, \dots,$$

$$\begin{aligned}
|\operatorname{Im}\varphi(t)|^2 &\leq \frac{1}{2}(1 - \operatorname{Re}\varphi(2t)); \\
1 + \operatorname{Re}\varphi(2t) &\geq 2(\operatorname{Re}\varphi(t))^2; \\
|\varphi(t) - \varphi(s)|^2 &\leq 4|1 - \varphi(t-s)|; \\
|\varphi(t) - \varphi(s)|^2 &\leq 2(1 - \operatorname{Re}\varphi(t-s)); \\
\frac{1}{2h} \left| \int_{t-h}^{t+h} \varphi(u) du \right| &\leq (1 + \operatorname{Re}\varphi(h))^{\frac{1}{2}}.
\end{aligned}$$

5. 设 ξ 的特征函数为 φ . 证明: 若有常数 c 使得 $\varphi \equiv c$, 则 $P(\xi = 0) = 1$.

8.3 基本性质

本节我们将证明特征函数的一些基本性质. 首先, 我们注意到特征函数有一个特别简单但特别有用的性质, 即

命题 8.3.1. 若 ξ, η 独立, 则

$$f_{\xi+\eta}(t) = f_{\xi}(t)f_{\eta}(t).$$

我们知道, 两个独立的随机变量之和的分布函数等于它们的分布函数的卷积——卷积是比较复杂的. 而本命题说明, 此和的特征函数等于它们的特征函数的乘积, 这就简单多了——这是特征函数的优点之一.

证明. 这是因为当 ξ, η 独立时, 容易验证

$$E[\exp\{it(\xi + \eta)\}] = E[\exp\{it\xi\} \exp\{it\eta\}] = E[\exp\{it\xi\}]E[\exp\{it\eta\}].$$

□

其次我们证明:

定理 8.3.2. 设 f 是随机变量 ξ 的特征函数, 则

- (i) $|f(t)| \leq f(0) = 1, \forall t$;
- (ii) $f(-t) = \overline{f(t)}, \forall t$;
- (iii) $|f(t) - f(s)| \leq E[|e^{i(t-s)\xi} - 1|], \forall t, s$;
- (iv) f 在 \mathbb{R} 上一致连续.

证明. (i) 由命题8.1.2,

$$|f(t)| = |E[e^{it\xi}]| \leq E[|e^{it\xi}|] = 1 = f(0).$$

(ii)

$$f(-t) = E[e^{-it\xi}] = E[\overline{e^{it\xi}}] = \overline{E[e^{it\xi}]} = \overline{f(t)}.$$

(iii)

$$\begin{aligned}
|f(t) - f(s)| &= |E[e^{it\xi}] - E[e^{is\xi}]| = |E[e^{is\xi}[e^{i(t-s)\xi} - 1]]| \\
&\leq E[|e^{is\xi}[e^{i(t-s)\xi} - 1]|] = E[|e^{i(t-s)\xi} - 1|].
\end{aligned}$$

(iv)

由(iii)及有界收敛定理有

$$\limsup_{|t-s| \rightarrow 0} |f(t) - f(s)| = \limsup_{|t-s| \rightarrow 0} E[|e^{i(t-s)\xi} - 1|] = 0.$$

□

例1. 判断函数

$$f(t) = \begin{cases} \frac{\sin t^2}{t^2}, & |t| < 1, (\frac{\sin 0}{0} := 1) \\ \sin t^2, & |t| \geq 1 \end{cases}$$

是否为特征函数.

解. 否, 因为它在 \mathbb{R} 上不一致连续. 比如取 $s_n^2 = 2\pi n$, $t_n^2 = 2\pi n + \frac{\pi}{2}$, 则

$$\sin t_n^2 - \sin s_n^2 = 1,$$

但当 n 趋于无穷时

$$t_n - s_n = \frac{\frac{\pi}{2}}{t_n + s_n} \rightarrow 0.$$

这说明了 $\sin t^2$ 在 \mathbb{R} 上不一致连续.如果 ξ 具有较好的可积性, 那么其特征函数也具有较好的光滑性. 更精确地说, 我们有:**定理 8.3.3.** 若 $E[|\xi|^n] < \infty$, 则 f 至少 n 次可导, 并有

$$f^{(k)}(t) = i^k E[\xi^k e^{it\xi}], \quad \forall k \leq n,$$

且 $f^{(k)}$ 一致连续, $\forall k \leq n$.**证明.** 由于

$$\frac{d}{dt} e^{it\xi} = i\xi e^{it\xi},$$

直接用积分号下求导的定理5.7.3, 可得 $f'(t)$ 存在, 且

$$f'(t) = iE[\xi e^{it\xi}].$$

而由于

$$|f'(t) - f'(s)| \leq E[|\xi| |1 - e^{i(t-s)\xi}|],$$

由控制收敛定理得 f' 一致连续性. 继续这一过程可得到 n 阶导数的结论. □

将上述结果应用到Taylor公式, 有

推论 8.3.4. 设 $E[|\xi|^n] < \infty$, 则有Taylor展式

$$\begin{aligned} f(t) &= \sum_{k=0}^{n-1} \frac{(it)^k}{k!} E[\xi^k] + i^n \int_0^t \int_0^{t_1} \cdots \int_0^{t_{n-1}} E[\xi^n e^{it_1\xi}] dt_1 dt_2 \cdots dt_n \\ &= \sum_{k=0}^n \frac{(it)^k}{k!} E[\xi^k] + \frac{1}{n!} \theta_n(t) (it)^n. \end{aligned}$$

其中 $|\theta_n(t)| \leq 2E[|\xi|^n]$, 且 $\lim_{t \rightarrow 0} \theta_n(t) = 0$.

证明. 因为 $E[|\xi|^n] < \infty$, 由定理8.3.3有

$$f^{(k)}(0) = i^k E[\xi^k], \quad \forall k \leq n.$$

所以 f 有 Taylor 展开

$$\begin{aligned} f(t) &= \sum_{k=0}^{n-1} \frac{1}{k!} f^{(k)}(0) t^k + \int_0^t \int_0^{t_n} \cdots \int_0^{t_2} f^{(n)}(t_1) dt_1 dt_2 \cdots dt_n \\ &= \sum_{k=0}^n \frac{1}{k!} f^{(k)}(0) t^k + \frac{1}{n!} \theta_n(t) (it)^n. \end{aligned}$$

其中

$$\begin{aligned} \theta_n(t) &= \frac{n!}{(it)^n} \int_0^t \int_0^{t_n} \cdots \int_0^{t_2} (f^{(n)}(t_1) - f^{(n)}(0)) dt_1 dt_2 \cdots dt_n \\ &= \frac{n!}{t^n} \int_0^t \int_0^{t_n} \cdots \int_0^{t_2} E[\xi^k (e^{it_1 \xi} - 1)] dt_1 dt_2 \cdots dt_n. \end{aligned}$$

注意到

$$\int_0^t \int_0^{t_n} \cdots \int_0^{t_2} dt_1 dt_2 \cdots dt_n = \frac{t^n}{n!},$$

因此

$$\theta_n(t) \leq \sup_{|s| \leq t} E[|\xi|^n |e^{is\xi} - 1|] \leq 2E[|\xi|^n].$$

而且, 由控制收敛定理有

$$\lim_{t \rightarrow 0} \theta_n(t) \leq \lim_{t \rightarrow 0} \sup_{|s| \leq t} E[|\xi|^n |e^{is\xi} - 1|] = E \left[|\xi|^n \lim_{t \rightarrow 0} \sup_{|s| \leq t} |e^{is\xi} - 1| \right] = 0.$$

□

若要将 f 展成幂级数, 则我们有下面的

推论 8.3.5. 设 f 是随机变量 ξ 的特征函数. 若 $E[|\xi|^n] < \infty, \forall n$, 且

$$\limsup_{n \rightarrow \infty} \frac{(E[|\xi|^n])^{1/n}}{n} = \frac{1}{eT} < \infty,$$

则

$$f(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} E[\xi^n], \quad \forall |t| < T.$$

证明. $\forall 0 < t_0 < T$, 有

$$\limsup_{n \rightarrow \infty} \frac{(E[|\xi|^n] t_0^n)^{1/n}}{n} = \frac{t_0}{eT}.$$

由 Stirling¹ 公式 (见附录11.1) 有

$$\lim_{n \rightarrow \infty} \left(\frac{n^n}{n!} \right)^{\frac{1}{n}} = e,$$

¹James Stirling, 1692-1770, 苏格兰数学家.

因此

$$\limsup_{n \rightarrow \infty} \left(\frac{E[|\xi|^n] t_0^n}{n!} \right)^{1/n} = \frac{t_0}{T} < 1.$$

由Cauchy检根判别法, 级数

$$\sum_{n=0}^{\infty} \frac{t_0^n}{n!} E[|\xi|^n] < \infty. \quad (3.2)$$

因为 $\left| \frac{(it)^n}{n!} E[\xi^n] \right| \leq \frac{t_0^n}{n!} E[|\xi|^n]$, $\forall |t| \leq t_0$, 由级数的M判别法, 级数

$$\sum_{n=0}^{\infty} \frac{(it)^n}{n!} E[\xi^n]$$

在 $|t| \leq t_0$ 上一致收敛.

由上一推论,

$$f(t) = \sum_{k=0}^n \frac{(it)^k}{k!} E[\xi^k] + R_n(t), \quad (3.3)$$

其中

$$|R_n(t)| \leq 2 \frac{t_0^n}{n!} E[|\xi|^n], \quad |t| \leq t_0.$$

右边为收敛级数(3.2)的通项, 所以

$$\limsup_{n \rightarrow \infty} |R_n(t)| \leq \lim_{n \rightarrow \infty} 2 \frac{t_0^n}{n!} E[|\xi|^n] = 0, \quad \forall |t| \leq t_0.$$

在等式(3.3)两边, 令 $n \rightarrow \infty$ 即得

$$f(t) = \sum_{k=0}^{\infty} \frac{(it)^k}{k!} E[\xi^k], \quad \forall |t| \leq t_0.$$

最后由 t_0 的任意性结论立得. □

以上都是由 ξ 的可积性导出 f 的光滑性; 若反过来, 要从 f 的光滑性判断 ξ 的可积性, 则有:

定理 8.3.6. 若 $f^{(2n)}(0)$ 存在且有限, 则 $E[\xi^{2n}] < \infty$.

证明. 先设 $n = 1$. 由L'Hopital法则有

$$\begin{aligned} f''(0) &= \lim_{t \downarrow 0} \frac{1}{2} \left[\frac{f'(2t) - f'(0)}{2t} + \frac{f'(0) - f'(-2t)}{2t} \right] \\ &= \lim_{t \downarrow 0} \frac{2f'(2t) - 2f'(-2t)}{8t} \\ &= \lim_{t \downarrow 0} \frac{f(2t) - 2f(0) + f(-2t)}{4t^2} \\ &= \lim_{t \downarrow 0} \int_{-\infty}^{\infty} \left(\frac{e^{itx} - e^{-itx}}{2t} \right)^2 dF(x) \\ &= - \lim_{t \downarrow 0} \int_{-\infty}^{\infty} x^2 \left(\frac{\sin tx}{tx} \right)^2 dF(x). \end{aligned}$$

于是由Fatou引理有

$$\begin{aligned}\int_{-\infty}^{\infty} x^2 dF(x) &= \int_{-\infty}^{\infty} x^2 \left(\lim_{t \downarrow 0} \frac{\sin tx}{tx} \right)^2 dF(x) \\ &\leq \lim_{t \downarrow 0} \int_{-\infty}^{\infty} x^2 \left(\frac{\sin tx}{tx} \right)^2 dF(x) \\ &= -f''(0).\end{aligned}$$

一般情形用归纳法. 设结论对 n 成立. 设 $f^{2(n+1)}(0) < \infty$. 则 $f^{(2n)}(0) < \infty$, 因此

$$\int_{-\infty}^{\infty} x^{2n} dF(x) < \infty.$$

若

$$E[\xi^{2n}] = \int_{-\infty}^{\infty} x^{2n} dF(x) = 0,$$

则 $\xi = 0$, a.s.. 因此

$$E[\xi^{2n+2}] = \int_{-\infty}^{\infty} x^{2n+2} dF(x) = 0,$$

结论已证. 再设

$$a := \int_{-\infty}^{\infty} x^{2n} dF(x) > 0.$$

由于

$$\begin{aligned}f^{(2n)}(t) &= \int_{-\infty}^{\infty} (ix)^{2n} e^{itx} dF(x) \\ &= (-1)^n \int_{-\infty}^{\infty} e^{itx} x^{2n} dF(x) \\ &= (-1)^n a \int_{-\infty}^{\infty} e^{itx} dG(x),\end{aligned}$$

其中

$$G(x) = a^{-1} \int_{-\infty}^x u^{2n} dF(u)$$

为一分布函数. 所以 $(-1)^n a^{-1} f^{(2n)}$ 是 G 的特征函数且在0点二次可导. 对此函数用 $n = 1$ 时的结果即得

$$\int_{-\infty}^{\infty} x^{2n+2} dF(x) = a \int_{-\infty}^{\infty} x^2 dG(x) < \infty.$$

□

自然会问, 若 $f'(0)$ 存在, 也会保证 $E[|\xi|] < \infty$ 吗? 答案是否定的. 比如, 令 ξ 的密度函数 $p(x)$ 为

$$p(x) = \begin{cases} \frac{e}{2} \left(\frac{1}{|x| \ln|x|} + \frac{1}{|x|(\ln|x|)^2} \right), & |x| \geq e, \\ 0, & 0 \leq x < e. \end{cases}$$

其分布函数为

$$F(x) = 1 - \frac{e}{2x \ln x}, \quad x \geq e.$$

则

$$\begin{aligned} \int_e^{+\infty} p(x) dx &= \frac{e}{2} \int_e^{+\infty} \left(\frac{1}{x \ln x} + \frac{1}{x(\ln x)^2} \right) dx \\ &= \frac{e}{2} \left(\frac{1}{\ln \ln x} - \frac{1}{\ln x} \right) \Big|_e^{+\infty} = +\infty, \end{aligned}$$

因此 $E[|\xi|] = +\infty$. 但因为密度函数为偶函数且

$$\lim_{x \rightarrow +\infty} x(1 - F(x) + F(-x)) = \lim_{x \rightarrow +\infty} 2x(1 - F(x)) = \lim_{x \rightarrow +\infty} \frac{e}{\ln x} = 0,$$

所以 $f'(0) = 0$. (证明比较麻烦, 参见文献[4]的P.656的定理)

由定理8.3.6和定理8.3.3 有

推论 8.3.7. 设 f 为特征函数. 若 $f^{(2n)}(0)$ 存在且有限, 则对任意 t , $f^{(2n)}(t)$ 存在且有限.

我们来看几个应用.

例1. 判断函数

$$f(t) := \begin{cases} \sqrt{1-t^2} & |t| < 1, \\ 0 & |t| \geq 1 \end{cases}$$

是否为特征函数.

解. 否. 因为 $f''(0)$ 有限, 若 f 为特征函数, 则 f'' 在任意点皆二次可导, 但显然在 $t = 1$ 时不可导.

例2. 判断函数

$$f(t) := |\cos(t)|$$

是否为特征函数.

解. 否. $f(t)$ 在 $t = 0$ 处无穷可导, 但在 $t = \frac{\pi}{2}$ 处不可导.

例3. 判断函数

$$f(t) := 1 - e^{-\frac{1}{|t|}}$$

是否为特征函数.

解. 否. 因为 $f^{(n)}(0) = 0, \forall n$, 所以若 f 为 ξ 的特征函数, 则 $E[\xi] = D[\xi] = 0$. 因此 $\xi = 0$ a.s.. 但此时 ξ 的特征函数为 $f(t) \equiv 1$.

例4. 判断函数

$$f(t) := e^{-|t|^\alpha}, \quad \alpha > 2$$

是否为特征函数.

解. 否. 此时 f 在 $t = 0$ 处二次可微, 且 $f'(0) = f''(0) = 0$. 再用上个例子的推理.

习题

1. 设 f_i 是特征函数, $i = 1, \dots, n$. 证明 $\prod_{i=1}^n f_i$ 也是特征函数.

2. 设 f 是特征函数, 证明 $|f(t)|^2$ 也是特征函数.
3. 设 φ, ψ 是分布函数 F, G 的特征函数, g 是母函数.

- (a) 证明 $\bar{\varphi}, \varphi^n$ ($n = 0, 1, 2, \dots$), $t \mapsto e^{ibt}\varphi(at)$ (a, b 是实数) 是特征函数并求出相应的分布函数;
- (b) 证明 $t \mapsto \psi(at)\varphi(bt)$ 是特征函数并求出相应的分布函数;
- (c) 证明 $g(\varphi)$ 是特征函数并求出相应的分布函数.
- (d) 问: $\operatorname{Re}\varphi$ 与 $\operatorname{Im}\varphi$ 是否依然为特征函数?
- (e) 证明 $e^{\lambda(\varphi-1)}$ 是特征函数并求出相应的分布函数.
- (f) 证明

$$t \mapsto \int_0^1 \varphi(ut) du, \quad t \mapsto \int_0^\infty e^{-u} \varphi(ut) du$$

是特征函数并求出相应的分布函数.

- (g) 证明对任意 $n \geq 1$,

$$t \mapsto n! \frac{e^{it} - \sum_{k=0}^{n-1} \frac{(it)^k}{k!}}{(it)^n}$$

是特征函数并求出相应的分布.

4. 设 φ 是一个特征函数, F 是一个分布函数. 证明

$$\psi(t) := \int_{-\infty}^{\infty} \varphi(tx) dF(x)$$

也是一个特征函数.

5. 设 $a \in \mathbb{R} \setminus \{0\}$, $\lambda > 0$. 若

$$P(\xi = ka) = \frac{1}{k!} e^{-\lambda} \lambda^k,$$

则称 $\xi \sim P(\lambda, a)$. 设 $\xi_k \sim P(\lambda_k, a_k)$, 且独立, 令 $\xi := \sum_{k=1}^n \xi_k$. 证明:

$$\varphi_\xi(t) = \exp \left\{ \sum_{k=1}^n \lambda_k (e^{ita_k} - 1) \right\}.$$

6. 设 ξ_1, \dots, ξ_n 独立, 都服从 $(-1, 1)$ 上的均匀分布. 证明 $n \geq 2$ 时, $\xi_1 + \dots + \xi_n$ 的密度函数为

$$f(x) = \frac{1}{\pi} \int_0^\infty \left(\frac{\sin t}{t} \right)^n \cos(tx) dt.$$

7. 利用特征函数证明: 若 ξ_1, ξ_2, \dots 相互独立, 则 $\xi_1, \xi_2 + \xi_3, \xi_4 + \xi_5 + \xi_6, \dots$ 相互独立.
8. 设 ξ 为整数值随机变量, φ 为其特征函数. 证明:

$$P(\xi = k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikt} \varphi(t) dt, \quad k = 0, \pm 1, \pm 2, \dots$$

9. 设 ξ, η 是独立随机变量, 且 $\xi + \eta$ 与 ξ 同分布. 证明 $\eta = 0$ a.s..
10. 设 φ 是 ξ 的特征函数, 且 $E[|\xi|^n] < \infty$. 证明对任意 t 及 $k \leq n$,

$$\varphi^{(k)}(t) = i^k E[\xi^k e^{it\xi}].$$

8.4 唯一性定理

特征函数之所以有用,植根于它和概率分布是一一对应的,即我们有下面的唯一性定理.

定理 8.4.1. 设 ξ 与 η 是随机变量. 若

$$E[\exp(it\xi)] = E[\exp(it\eta)],$$

则 ξ 与 η 同分布.

证明. 由定理5.8.2, 只需证明

$$E[f(\xi)] = E[f(\eta)], \quad \forall f \in C_0^\infty. \quad (4.4)$$

现设 $f \in C_0^\infty$. 由定理4.3.4及其后面的注, 有

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(t) e^{-itx} dt,$$

其中

$$\hat{f}(t) := \int_{-\infty}^{\infty} e^{ity} f(y) dy.$$

因此

$$\begin{aligned} E[f(\xi)] &= \frac{1}{2\pi} E \int_{-\infty}^{\infty} \hat{f}(t) e^{-it\xi} dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(t) E[e^{-it\xi}] dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(t) E[e^{-it\eta}] dt \\ &= \frac{1}{2\pi} E \int_{-\infty}^{\infty} \hat{f}(t) e^{-it\eta} dt \\ &= E[f(\eta)]. \end{aligned}$$

□

本定理说明, 一个随机变量的分布函数是由其特征函数唯一确定的. 我们来看一个应用.

例1. 设 ξ, η 独立同分布, 且 $E[\xi] = 0$, $E[\xi^2] = 1$. 证明: 若 ξ 与 $\frac{\sqrt{2}}{2}(\xi + \eta)$ 同分布, 则 ξ 服从标准正态分布.

证明. 以 f 表示 ξ 的特征函数. 因为 $E[\xi^2] < \infty$, 则 f 有二阶连续导数. 由带Peano余项的Taylor公式有

$$\begin{aligned} f(t) &= f(0) + f'(0)t + \frac{1}{2}f''(0)t^2 + o(t^2) \\ &= 1 - \frac{1}{2}t^2 + o(t^2). \end{aligned}$$

则

$$\begin{aligned} f(t) &= f\left(\frac{\sqrt{2}}{2}t\right)^2 \\ &= \dots \\ &= f\left(\left(\frac{\sqrt{2}}{2}\right)^n t\right)^{2^n} \\ &= \left[1 - \frac{1}{2^n} \cdot \frac{t^2}{2} + o\left(\frac{1}{2^n}t^2\right)\right]^{2^n}. \end{aligned}$$

由命题11.5.1, 令 $n \rightarrow \infty$, 得 $f(t) = e^{-\frac{1}{2}t^2}$. □

上面的定理说明了分布函数是由特征函数唯一确定的, 但没有说明具体是以怎样的方式确定的. 下面我们将证明一个由特征函数直接确定分布函数的公式, 即所谓的反演公式. 为此, 先证明一个分析引理.

引理 8.4.2 (Dirichlet²积分).

$$\lim_{T \rightarrow \infty} \int_0^T \frac{\sin t}{t} dt = \frac{\pi}{2}.$$

证明. 令

$$I_T := \int_0^T \frac{\sin t}{t} dt.$$

则

$$\begin{aligned} I_T &= \int_0^T \sin t dt \int_0^\infty e^{-ts} ds \\ &= \int_0^\infty ds \int_0^T \sin t e^{-ts} dt \\ &= \int_0^\infty \left(\frac{1}{1+s^2} - \frac{s \sin T + \cos T}{1+t^2} e^{-sT} \right) ds \\ &= \frac{\pi}{2} - \int_0^\infty \frac{s \sin T + T \cos T}{T^2 + s^2} e^{-s} ds. \end{aligned}$$

由于

$$\left| \int_0^\infty \frac{s \sin T + T \cos T}{T^2 + s^2} e^{-s} ds \right| \leq \frac{C}{T} \rightarrow 0, \quad T \rightarrow \infty,$$

引理得证. □

由于对 $a \in \mathbb{R}$ 有

$$\int_0^T \frac{\sin at}{t} dt = \int_0^{aT} \frac{\sin t}{t} dt,$$

所以我们有

²Johann Peter Gustav Lejeune Dirichlet, 1805-1859, 德国数学家

推论 8.4.3. 函数

$$I(T, a) := \int_0^T \frac{\sin at}{t} dt$$

$\mathbb{R}_+ \times \mathbb{R}$ 上有界且

$$\lim_{T \rightarrow \infty} I(T, a) = \frac{\pi}{2} \operatorname{sgn}(a), \quad \forall a \in \mathbb{R}.$$

下面证明反演公式. 我们先证明这个公式的初级版. 我们知道特征函数是有界连续函数, 因此在每个有限区间上都是可积的. 所谓初级版是在额外的条件

$$\int_{-\infty}^{\infty} |f(t)| dt < \infty$$

下考虑. 和最后的完整版相比, 这里的条件更强结论也更强, 公式也更加简洁. 但两者的思想是一样的, 只是有些在强条件下可以走得更远的地方, 在一般的条件下就到不了, 因此得不到简洁的公式而已.

定理 8.4.4 (反演公式初级版). 若

$$\int_{\mathbb{R}} |f(t)| dt < \infty,$$

则 F 是连续型分布函数, 且其密度函数为

$$p(x) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f(t) dt.$$

若进一步有

$$\int_{\mathbb{R}} |t|^n |f(t)| dt < \infty,$$

则 p 为 n 次连续可导, 且

$$p^{(n)}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} (-it)^n e^{-itx} f(t) dt.$$

证明. 令

$$p(x) := \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f(t) dt.$$

由控制收敛定理, p 是连续函数. 由Fubini定理有

$$\begin{aligned}
 \int_a^b p(x)dx &= \int_a^b dx \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f(t)dt \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(t)dt \int_a^b e^{-itx} dx \\
 &= \frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} f(t)dt \\
 &= \frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T dt \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} e^{itx} dF(x) \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dF(x) \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} e^{itx} dt \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dF(x) \lim_{T \rightarrow \infty} \int_{-T}^T \frac{\sin(x-a)t}{t} - \frac{\sin(x-b)t}{t} dt \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} (\pi(\operatorname{sgn}(x-a) - \operatorname{sgn}(x-b))) dF(x) \\
 &= \frac{1}{2} E[\operatorname{sgn}(\xi - a) - \operatorname{sgn}(\xi - b)],
 \end{aligned}$$

其中 ξ 的分布函数为 F . 因为

$$\operatorname{sgn}(x-a) - \operatorname{sgn}(x-b) = \begin{cases} 0 & x < a, x > b \\ 1, & x = a, b \\ 2, & a < x < b, \end{cases}$$

所以上式等于

$$\begin{aligned}
 &\frac{1}{2} [P(\xi = a) + 2P(a < \xi < b) + P(\xi = b)] \\
 &= \frac{1}{2} [F(a) - F(a-)] + F(b-) - F(a) + \frac{1}{2} [F(b) - F(b-)] \\
 &= \frac{F(b) + F(b-)}{2} - \frac{F(a) + F(a-)}{2}.
 \end{aligned}$$

所以当 a, b 均为 F 的连续点时, 有

$$F(b) - F(a) = \int_a^b p(x)dx.$$

但此式左边为 a, b 的右连续函数, 右边为 a, b 的连续函数, 所以对任意 $a < b$ 该等式均成立. 所以 p 是 F 的密度函数.

若

$$\int_{\mathbb{R}} |t| |f(t)| dt < \infty,$$

由于 f 是有界连续函数, 所以必有

$$\int_{\mathbb{R}} |f(t)| dt < \infty.$$

因此 F 有密度

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f(t) dt.$$

由定理5.7.3, 对 p 可在期望号下求导. 于是

$$p'(x) = \frac{-1}{2\pi} \int_{-\infty}^{\infty} ite^{-itx} f(t) dt.$$

递推下去可得到 n 阶导数的结果. □

我们注意到, 一旦 F 有密度函数 p , 那么 F 的特征函数 f 就是 p 的普通Fourier变换. 但有密度时却未必有

$$\int_{\mathbb{R}} |f(t)| dt < \infty.$$

而一旦此式成立, 则必有

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} f(t) dt. \quad (4.5)$$

这就是普通Fourier变换的反演公式.

另外我们注意到在上面的推理中, 条件

$$p \geq 0, \quad \int_{-\infty}^{\infty} p(x) dx = 1$$

其实不是本质的. 我们实际上是证明了下面的结果:

命题 8.4.5. 设 g 是 \mathbb{R} 上的Lebesgue可测函数且

$$\int_{-\infty}^{\infty} |g(x)| dx < \infty,$$

定义

$$\hat{g}(t) := \int_{-\infty}^{\infty} e^{itx} g(x) dx.$$

若存在 $n \in \mathbb{N}_+$ 使得

$$\int_{-\infty}^{\infty} |t|^n |\hat{g}(t)| dt < \infty, \quad (4.6)$$

则 g 有直到 n 阶的连续导数且 $\forall k \leq n$

$$g^{(k)}(x) = \frac{(-i)^k}{2\pi} \int_{-\infty}^{\infty} e^{-itx} t^k \hat{g}(t) dt.$$

那么什么时候(4.6)满足? 这是一个问题. 由它延申出去, 就会到达调和的一个重要领域——振荡积分.

现在我们转向一般情形. 现在需要注意不一定有

$$\int_{\mathbb{R}} |f(t)| dt < \infty,$$

所以不能定义 p . 除此之外, 也不能定义

$$\int_{-\infty}^{\infty} e^{-itx} f(t) dt.$$

对此, 我们用

$$\int_{-T}^T e^{-itx} f(t) dt$$

代替之, 并用引理8.4.2之本质结论, 即

$$\lim_{T \rightarrow \infty} \int_{-T}^T \frac{\sin ta}{t} dt = \frac{\pi}{2} \operatorname{sgn}(a).$$

经过这样的变通之后, 易见原证明除了前两个等式不复存在外, 其它步骤依然有效. 这样我们得到:

定理 8.4.6 (反演公式完整版). 对任意 $a < b$ 有

$$\frac{F(b) + F(b-)}{2} - \frac{F(a) + F(a-)}{2} = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} f(t) dt.$$

特别当 a, b 为 F 的连续点时, 有

$$F(b) - F(a) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} f(t) dt.$$

由这个公式我们也可以得到唯一性定理. 因为取 a, b 均为 F 的连续点, 再让 a 沿连续点趋于 $-\infty$, 就得到 $F(b)$ 由 f 唯一确定. 但 F 的连续点全体在 \mathbb{R} 中稠密, 且 F 是右连续函数, 所以对任意 x , $F(x)$ 由 f 唯一确定.

习题

1. 设 $\xi_k, k = 1, 2, \dots$ 是独立同分布随机变量, 特征函数为 $\exp(-|t|^\alpha)$, $\alpha \in (0, 2)$. 证明 $n^{-1/\alpha} \sum_{k=1}^n \xi_k$ 与 ξ_1 同分布.
2. 举例说明存在连续型分布, 其特征函数 φ 满足 $\int_{\mathbb{R}} |\varphi(t)| dt = \infty$.
3. 设 ξ 的特征函数为 φ . 证明:

(a) 若存在 $t_0 \neq 0$ 使得 $|\varphi(t_0)| = 1$, 则存在 a, h 使得

$$\sum_{n=-\infty}^{\infty} P(\xi = a + nh) = 1;$$

(b) 若存在 $t_0 \neq 0$ 使得 $\varphi(t_0) = 1$, 则存在 h 使得

$$\sum_{n=-\infty}^{\infty} P(\xi = nh) = 1;$$

4. 设 ξ 是整数值随机变量, φ 是其特征函数. 证明:

$$P(\xi = k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikt} \varphi(t) dt.$$

5. 设 F 是分布函数, φ 是其特征函数. 证明:

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T e^{-itx} \varphi(t) dt = F(x) - F(x-), \quad \forall x;$$

$$\lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T |\varphi(t)|^2 dt = \sum_{x \in \mathbb{R}} |F(x) - F(x-)|^2.$$

6. 设 ξ 是随机变量, φ 是其特征函数. 证明:

(a) 若存在奇函数 ψ 使得 $\varphi(t) = 1 + \psi(t) + o(t^2)$, $t \rightarrow 0$, 则 $\xi \equiv 0$.

(b) 若存在偶函数 ψ 使得 $\varphi(t) = \psi(t) + o(t)$, $t \rightarrow 0$, 则 $\xi \equiv 0$.

8.5 连续性定理

由上节的结果, 我们知道特征函数和分布函数相互唯一确定的. 本节我们要进一步研究的问题是, 这种一一对应的关系是否是连续的? 也即, 设 F_n 、 F 是分布函数, f_n 、 f 分别是其特征函数. 那么, 是否在某种意义上有

$$F_n \rightarrow F \iff f_n \rightarrow f?$$

我们从更一般性的视角来看这个问题. 设 F_n 是 ξ_n 的分布函数, F 是 ξ 的分布函数. 我们知道, 由控制收敛定理有

$$\xi_n \xrightarrow{a.s.} \xi \implies E[\varphi(\xi_n)] \rightarrow E[\varphi(\xi)], \quad \forall \varphi \in C_b(\mathbb{R}).$$

用分布函数表示就是

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \varphi(x) dF_n(x) = \int_{-\infty}^{\infty} \varphi(x) dF(x). \quad (5.7)$$

特别地, 取 $\varphi(x) := \exp\{itx\}$, 就得到 $f_n(t) \rightarrow f(t)$, $\forall t$.

但如何脱离随机变量, 直接通过 F_n 、 F 给出 (5.7) 成立的条件呢? 最容易想到的猜测当然是

$$F_n(x) \rightarrow F(x), \quad \forall x.$$

但有例子表明这一要求太强, 强到甚至很平凡的情况都无法满足.

例1. 设

$$\xi_n \equiv \frac{1}{n}, \quad \xi \equiv 0.$$

则

$$F_n(x) = \begin{cases} 1, & x \geq \frac{1}{n}, \\ 0, & x < \frac{1}{n}. \end{cases}$$

$$F(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

所以 $\xi_n \rightarrow \xi$, $\forall \omega$. 因此

$$E[\varphi(\xi_n)] \rightarrow E[\varphi(\xi)], \quad \forall \varphi \in C_b(\mathbb{R}).$$

但 $F_n(0) \not\rightarrow F(0)$. 并且这里面的原因也并不复杂, 无非是个 $y < x_n$ 时能否有 $y < \lim_{n \rightarrow \infty} x_n$ 的问题——这当然不行, 只能得到小于等于成立.

但也正是这个例子使我们也注意到, 除了 $x = 0$ 之外, 对其它任意 x , 均有 $F_n(x) \rightarrow F(x)$.

$x = 0$ 有什么特别之处? 它是 F 的间断点! 所以我们能期望的 $F_n \rightarrow F$ 的含义, 最多只能是 $F_n(x) \rightarrow F(x)$, 对 F 的任意连续点 x 成立.

经过以上的分析之后, 我们可以准备叙述我们的结果了. 为此我们先引进:

定义 8.5.1. 设 F_n, F 是分布函数. 若 $\forall \varphi \in C_b(\mathbb{R})$,

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \varphi(x) dF_n(x) = \int_{-\infty}^{\infty} \varphi(x) dF(x),$$

则称 F_n 弱收敛于 F , 记为 $F_n \xrightarrow{w} F$.

设 ξ_n, ξ 是随机变量(可以定义在不同的概率空间上), 分布函数分别为 F_n, F . 若 $F_n \xrightarrow{w} F$, 则称 ξ_n 弱收敛于 ξ , 或 ξ_n 依分布收敛于 ξ , 记为 $\xi_n \xrightarrow{w} \xi$.

易见, $\xi_n \xrightarrow{w} \xi$ 可以改写为

$$E[f(\xi_n)] \rightarrow E[f(\xi)], \quad \forall f \in C_b(\mathbb{R}).$$

由(5.7)前面的讨论可知, 我们有

命题 8.5.2.

$$\xi_n \xrightarrow{a.s.} \xi \implies F_n \xrightarrow{w} F \implies f_n(t) \rightarrow f(t) \quad \forall t.$$

对 \mathbb{R} 上的函数 F , 以 $C(F)$ 表示 F 的连续点全体. 若 F 为分布函数, 则 $C(F)$ 中至多包含可列个点.

下面我们证明:

定理 8.5.3. 设 $F_n \xrightarrow{w} F$, 则 $F_n(x) \rightarrow F(x), \forall x \in C(F)$.

证明. 设 $x \in C(F)$. 对 $\varepsilon > 0$, 令

$$\varphi_\varepsilon(y) = 1_{(-\infty, x]}(y) + \frac{1}{\varepsilon}(x + \varepsilon - y)1_{(x, x+\varepsilon]}(y).$$

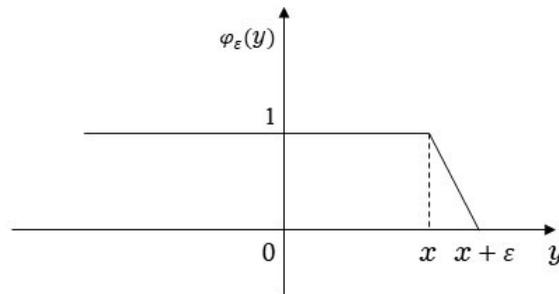


图 8.1: 函数 $\varphi_\varepsilon(y)$ 的图像

则 $\varphi_\varepsilon \in C_b(\mathbb{R})$, 且 $1_{(-\infty, x]}(y) \leq \varphi_\varepsilon(y) \leq 1_{(-\infty, x+\varepsilon]}(y)$. 因 $F_n \xrightarrow{w} F$, 因此

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \varphi_\varepsilon(y) dF_n(y) = \int_{-\infty}^{\infty} \varphi_\varepsilon(y) dF(y).$$

由此等式有

$$\begin{aligned} F(x - \varepsilon) &\leq \int_{-\infty}^{\infty} \varphi_\varepsilon(y + \varepsilon) dF(y) \\ &= \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \varphi_\varepsilon(y + \varepsilon) dF_n(y) \\ &\leq \liminf_{n \rightarrow \infty} F_n(x) \\ &\leq \limsup_{n \rightarrow \infty} F_n(x) \\ &\leq \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \varphi_\varepsilon(y) dF_n(y) \\ &= \int_{-\infty}^{\infty} \varphi_\varepsilon(y) dF(y) \\ &\leq F(x + \varepsilon). \end{aligned}$$

令 $\varepsilon \rightarrow 0$, 则

$$\lim_{n \rightarrow \infty} F_n(x) = F(x),$$

其中 $x \in C(F)$. □

下面证明这个定理的逆定理, 即

定理 8.5.4 (Helly³第二定理). 设 F_n, F 是分布函数, 且 $F_n(x) \rightarrow F(x), \forall x \in C(F)$. 则 $F_n \xrightarrow{w} F$.

现在我们看到了, $F_n \xrightarrow{w} F$ 要弱于 $F_n(x) \rightarrow F(x), \forall x$. 所以叫弱收敛啦!

证明. 设 $\varphi \in C_b(\mathbb{R})$. $\forall \varepsilon > 0$, 取 $T, -T \in C(F)$, 使得

$$F(-T) + 1 - F(T) < \varepsilon.$$

于是

$$\lim_{n \rightarrow \infty} (F_n(-T) + 1 - F_n(T)) < \varepsilon.$$

因为 φ 在 $[-T, T]$ 上一致连续, 所以可取 $-T = x_0 < x_1 < \cdots < x_m = T$ 使得 $x_i \in C(F) \cap_{n=1}^{\infty} C(F_n), \forall i = 1, \cdots, m$, 且

$$\sup_{x_{i-1} \leq x \leq x_i} |\varphi(x) - \varphi(x_{i-1})| < \varepsilon, \quad \forall i = 1, \cdots, m.$$

³Eduard Helly, 1884-1943, 奥地利-美国数学家.

于是 $\forall n$,

$$\begin{aligned}
 & \left| \int_{[-T, T]} \varphi(x) dF_n(x) - \sum_{i=1}^m \varphi(x_{i-1})(F_n(x_i) - F_n(x_{i-1})) \right| \\
 &= \left| \sum_{i=1}^m \int_{[x_{i-1}, x_i]} (\varphi(x) - \varphi(x_{i-1})) dF_n(x) \right| \\
 &\leq \sum_{i=1}^m \int_{[x_{i-1}, x_i]} |\varphi(x) - \varphi(x_{i-1})| dF_n(x) \\
 &\leq \varepsilon \sum_{i=1}^m \int_{[x_{i-1}, x_i]} dF_n(x) \\
 &\leq \varepsilon.
 \end{aligned}$$

同理

$$\left| \int_{[-T, T]} \varphi(x) dF(x) - \sum_{i=1}^m \varphi(x_{i-1})(F(x_i) - F(x_{i-1})) \right| \leq \varepsilon.$$

记

$$\begin{aligned}
 I_n &= \int_{\mathbb{R}} \varphi(x) dF_n(x) - \int_{\mathbb{R}} \varphi(x) dF(x) \\
 &= \left[\int_{-\infty}^{-T} \varphi(x) dF_n(x) - \int_{-\infty}^{-T} \varphi(x) dF(x) \right. \\
 &\quad \left. + \int_T^{\infty} \varphi(x) dF_n(x) - \int_T^{\infty} \varphi(x) dF(x) \right] \\
 &\quad + \left[\sum_{i=1}^m \varphi(x_{i-1})(F_n(x_i) - F_n(x_{i-1})) - \sum_{i=1}^m \varphi(x_{i-1})(F(x_i) - F(x_{i-1})) \right] \\
 &\quad + \left[\int_{[-T, T]} \varphi(x) dF_n(x) - \sum_{i=1}^m \varphi(x_{i-1})(F_n(x_i) - F_n(x_{i-1})) \right] \\
 &\quad + \left[\int_{[-T, T]} \varphi(x) dF(x) - \sum_{i=1}^m \varphi(x_{i-1})(F(x_i) - F(x_{i-1})) \right] \\
 &=: I_{1,n} + I_{2,n} + I_{3,n}.
 \end{aligned}$$

设 $|\varphi(x)| \leq C$, 则

$$\sup_n |I_{1,n}| \leq 2C\varepsilon.$$

对 $I_{3,n}$ 则有

$$\sup_n |I_{3,n}| \leq 2\varepsilon.$$

最后, 由于 $F_n(x_i) \rightarrow F(x_i)$, 所以

$$\lim_{n \rightarrow \infty} |I_{2,n}| = 0.$$

综合起来便得

$$\limsup_{n \rightarrow \infty} |I_n| \leq 2(C+1)\varepsilon.$$

由 ε 的任意性有

$$\lim_{n \rightarrow \infty} |I_n| = 0.$$

□

由以上两个定理, 我们证明了

$$F_n \xrightarrow{w} F$$

与

$$F_n(x) \rightarrow F(x), \forall x \in C(F)$$

是等价的. 所以这两个条件都可以作为弱收敛的定义. 实际上, 从历史的角度看, 第二个条件作为弱收敛的定义出现得更早一些. 但第一个条件则可用到更广泛的情况.

从定理8.5.4的证明可知, 我们实际上证明了下面更强的结果:

定理 8.5.5. 设 F_n, F 是 \mathbb{R} 上单调上升, 右连左极且一致有界的函数族, $F_n(x) \rightarrow F(x), \forall x \in C(F)$ 且 $F_n(-\infty) \rightarrow F(-\infty), F_n(\infty) \rightarrow F(\infty)$. 设 $\varphi_\theta \in C_b(\mathbb{R}), \forall \theta \in \Theta$, 且

- (i) $\sup_{\theta \in \Theta, x \in \mathbb{R}} |\varphi_\theta(x)| < \infty$;
- (ii) $\forall T > 0, \forall \varepsilon > 0, \exists \delta > 0$, 使得

$$|x - y| < \delta, x, y \in [-T, T] \implies \sup_{\theta \in \Theta} |\varphi_\theta(x) - \varphi_\theta(y)| < \varepsilon.$$

则

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta} \left| \int_{\mathbb{R}} \varphi_\theta(x) dF_n(x) - \int_{\mathbb{R}} \varphi_\theta(x) dF(x) \right| = 0.$$

注意到对任意 $M > 0$, 函数族 $\{e^{itx} : |t| \leq M\}$ 满足上面定理中的条件, 所以我们有

推论 8.5.6. 设 F_n, F 是分布函数, f_n, f 分别是其特征函数. 若 $F_n(x) \rightarrow F(x), \forall x \in C(F)$, 则对任意 $M > 0$,

$$\lim_{n \rightarrow \infty} \sup_{|t| \leq M} |f_n(t) - f(t)| = 0.$$

前面我们证明了Helly第二定理. 那么什么是Helly第一定理呢? 这就涉及到什么时候第二定理中的条件满足的问题, 即下面的

定理 8.5.7 (Helly第一定理). 设 F_n 是 \mathbb{R} 上的右连左极单调上升函数, 且

$$\sup_n |F_n(x)| < \infty, \forall x.$$

则存在 \mathbb{R} 上的右连左极单调上升函数 F 及子列 $\{n_k\}$, 使得 $F_{n_k}(x) \rightarrow F(x), \forall x \in C(F)$.

证明. 任取 \mathbb{R} 的可列稠密子集 $D = \{r_1, r_2, \dots\}$. 因为

$$\sup_n |F_n(r_1)| < \infty,$$

所以有子列 $\{n_{1,k}\}$, 使得

$$G(r_1) := \lim_{k \rightarrow \infty} F_{n_{1,k}}(r_1)$$

存在且有限. 同理, 由于

$$\sup_k |F_{n_{1,k}}(r_2)| < \infty,$$

所以有 $\{n_{1,k}\}$ 的子列 $\{n_{2,k}\}$, 使得

$$G(r_2) := \lim_{k \rightarrow \infty} F_{n_{2,k}}(r_2)$$

存在且有限. 继续这一进程, 知对任意 m , 存在 $\{n_{m-1,k}\}$ 的子列 $\{n_{m,k}\}$, 且

$$G(r_m) := \lim_{k \rightarrow \infty} F_{n_{m,k}}(r_m)$$

存在且有限. 显然, 对任意 $\alpha \leq m$, 都有

$$G(r_\alpha) = \lim_{k \rightarrow \infty} F_{n_{m,k}}(r_\alpha).$$

因此, 取子列 $n_m := n_{m,m}$, 就有

$$G(r_\alpha) = \lim_{m \rightarrow \infty} F_{n_m}(r_\alpha), \quad \forall \alpha.$$

这样 G 就在整个 D 上有定义, 且因 F_{n_m} 是单调上升的, 显然 G 是单调上升函数.

令

$$F(x) = \inf\{G(r) : r \in D, r > x\}.$$

则 F 为单调上升的右连续函数. $\forall x$, 取 $r', s' \in D, r < r' < x < s' < s$, 则有

$$F(r) \leq G(r') \leq F(x) \leq G(s') \leq F(s).$$

所以

$$\begin{aligned} F(r) \leq G(r') &= \lim_{m \rightarrow \infty} F_{n_m}(r') \\ &\leq \liminf_{m \rightarrow \infty} F_{n_m}(x) \\ &\leq \limsup_{m \rightarrow \infty} F_{n_m}(x) \\ &\leq \lim_{m \rightarrow \infty} F_{n_m}(s') \\ &= G(s') \leq F(s). \end{aligned}$$

若 x 为 F 的连续点, 则令 $r \uparrow x, s \downarrow x$, 得

$$F(x) \leq \liminf_{m \rightarrow \infty} F_{n_m}(x) \leq \limsup_{m \rightarrow \infty} F_{n_m}(x) \leq F(x).$$

所以

$$F(x) = \lim_{m \rightarrow \infty} F_{n_m}(x).$$

□

注1. 由定理的证明容易看出, 若存在 $m, M \in \mathbb{R}$ 使

$$m \leq F_n(x) \leq M, \quad \forall x, \forall n \geq 1,$$

则

$$m \leq F(x) \leq M, \quad \forall x.$$

2. 上面的证法通常称为“对角线法”. 这是一种很通用的方法. 例如常微分方程中的Ascoli-Arzelà定理也是使用这种方法证明的只不过在那里我们的出发点是一列一致有界同等连续的函数, 得到的极限是一个连续函数.

利用这个结果, 可以证明下面的结果——推论8.5.6的某种意义上的逆命题.

命题 8.5.8. 设 F_n 是分布函数, f_n 是其特征函数. 设 f 是定义在 \mathbb{R} 上的函数, 在 0 点连续, 且 $f_n(t) \rightarrow f(t), \forall t \in \mathbb{R}$. 则 f 是某一分布函数 F 的特征函数, 且 $F_n \xrightarrow{w} F$.

证明. 由上面的定理及注, 知存在单调上升且右连左极的函数 $F, \sup_x |F(x)| \leq 1$, 以及子列 $\{n_k\}$, 使得 $F_{n_k}(x) \rightarrow F(x), \forall x \in C(F)$. 为证明 F 为一分布函数, 只需 F 满足

$$F(\infty) - F(-\infty) = 1. \quad (5.8)$$

若否, 则

$$F(\infty) - F(-\infty) =: \delta < 1.$$

对任意 $x > 0$, 取 $z \geq x$ 使得 $z, -z \in C(F)$, 则

$$\begin{aligned} \alpha(x) &:= \limsup_{k \rightarrow \infty} (F_{n_k}(x) - F_{n_k}(-x)) \\ &\leq \limsup_{k \rightarrow \infty} (F_{n_k}(z) - F_{n_k}(-z)) \\ &= F(z) - F(-z) \leq \delta. \end{aligned}$$

于是, $\forall \varepsilon > 0$,

$$\begin{aligned} \left| \int_{-\varepsilon}^{\varepsilon} f_{n_k}(t) dt \right| &= \left| \int_{-\varepsilon}^{\varepsilon} dt \int_{-\infty}^{\infty} e^{ity} dF_{n_k}(y) \right| \\ &= \left| \int_{-\infty}^{\infty} dF_{n_k}(y) \int_{-\varepsilon}^{\varepsilon} e^{ity} dt \right| \\ &\leq \left| \int_{|y| < x} dF_{n_k}(y) \int_{-\varepsilon}^{\varepsilon} e^{ity} dt \right| + \left| \int_{|y| \geq x} dF_{n_k}(y) \int_{-\varepsilon}^{\varepsilon} e^{ity} dt \right| \\ &\leq 2\varepsilon(F_{n_k}(x) - F_{n_k}(-x)) + \frac{2}{x}. \end{aligned}$$

所以

$$\limsup_{k \rightarrow \infty} \left| \int_{-\varepsilon}^{\varepsilon} f_{n_k}(t) dt \right| \leq 2\varepsilon\delta + \frac{2}{x}.$$

但 x 是任意的, 所以

$$\limsup_{k \rightarrow \infty} \left| \int_{-\varepsilon}^{\varepsilon} f_{n_k}(t) dt \right| \leq 2\varepsilon\delta.$$

但另一方面, 由控制收敛定理有

$$\lim_{k \rightarrow \infty} \left| \int_{-\varepsilon}^{\varepsilon} f_{n_k}(t) dt \right| = \left| \int_{-\varepsilon}^{\varepsilon} f(t) dt \right|.$$

由于 $f(0) = 1$ 且 f 在 $t = 0$ 处连续, 故 ε 足够小时有

$$\left| \int_{-\varepsilon}^{\varepsilon} f(t) dt \right| > 2\varepsilon\delta.$$

因此

$$\lim_{k \rightarrow \infty} \left| \int_{-\varepsilon}^{\varepsilon} f_{n_k}(t) dt \right| > 2\varepsilon\delta.$$

这是一个矛盾. 因此必有

$$F(\infty) - F(-\infty) = 1,$$

即 F 是分布函数.

于是, $F_{n_k} \xrightarrow{w} F$. 由命题8.5.2,

$$f(t) = \int_{-\infty}^{\infty} e^{itx} dF(x) = \lim_{k \rightarrow \infty} \int_{-\infty}^{\infty} e^{itx} dF_{n_k}(x).$$

下面证明 $F_n \xrightarrow{w} F$. 若否, 则存在 $x_0 \in C(F)$ 使得 $F_n(x_0) \not\rightarrow F(x_0)$. 因此存在子列 $\{n'\}$ 使得下面极限存在, 但

$$\lim_{n' \rightarrow \infty} F_{n'}(x_0) \neq F(x_0).$$

重复前面的推理, 可以在 $\{n'\}$ 中再抽取一个子列 $\{n''\}$ 及分布函数 G 使得 $F_{n''} \xrightarrow{w} G$, 且 f 为 G 的特征函数. 因此 F 与 G 有共同的特征函数. 但由于 $F(x_0) \neq G(x_0)$, 根据唯一性定理, 这是不可能的. \square

习题

1. 设 $\xi_n \xrightarrow{w} \xi$, $p \geq 1$. 证明: 若 $\liminf_{n \rightarrow \infty} E[|\xi_n|^p] < \infty$, 则 $E[|\xi|^p] < \infty$.
2. 设 F_n, F 是分布函数, 且 F 连续. 证明: 若 $F_n \xrightarrow{w} F$, 则在整个 \mathbb{R} 上, F_n 一致收敛于 F . 举例说明 F 的连续性假设不可去掉.
3. 设 ξ_1, ξ_2, \dots 是独立同分布随机变量, $E[|\xi_1|^p] < \infty$, $p \geq 1$. 证明:

$$n^{-\frac{1}{p}} \max_{1 \leq k \leq n} |\xi_k| \xrightarrow{w} 0.$$

4. 设 ξ_n 服从正态分布, $\xi_n \xrightarrow{P} \xi$. 证明 ξ 也服从正态分布.
5. 设 $\xi_n \xrightarrow{w} \xi$, f 是连续函数. 证明 $f(\xi_n) \xrightarrow{w} f(\xi)$.
6. 设 ξ_n, ξ, η_n, η 均定义在同一个概率空间上, $\xi_n \xrightarrow{w} \xi$, $\eta_n \xrightarrow{w} \eta$, 且 $\forall n$, ξ_n 与 η_n 独立. 设 $f \in C(\mathbb{R}^2, \mathbb{R})$. 证明: $f(\xi_n, \eta_n) \xrightarrow{w} f(\xi, \eta)$.
7. 设 ξ, ξ_1, ξ_2, \dots 是随机变量, 且

$$\lim_{n \rightarrow \infty} E[\xi_n^k] = E[\xi^k] \quad \forall k.$$

证明: 若级数

$$\sum_{n=1}^{\infty} \frac{E[\xi^k]}{k!}$$

的收敛半径大于0, 则 $\xi_n \xrightarrow{w} \xi$.

8. 设 F_n, F 为分布函数, 且有 \mathbb{R} 的稠子集 $\{x_n\}$ 使得

$$F_n(x_n) \rightarrow F(x) \quad \forall n.$$

证明 $F_n(x) \rightarrow F(x)$, $\forall x \in C(F)$.

9. 证明Helly第一定理(定理8.5.7)证明中定义的 $F(x)$ 是右连续函数.

8.6 多维情形

特征函数的概念可以推广到多维随机变量.

定义 8.6.1. 设 (ξ_1, \dots, ξ_n) 为 n -维随机变量, 则其特征函数定义为: $\forall t = (t_1, \dots, t_n)$,

$$f(t) = E[\exp\{it \cdot \xi\}] = E[\exp\{it\xi'\}] = E[\exp\{it_1\xi_1 + \dots + it_n\xi_n\}].$$

可以证明, 多维特征函数有与一维特征函数相似的性质.

我们来计算多维正态分布的特征函数.

例1. 设 $\xi = (\xi_1, \dots, \xi_n) \sim N(\mu, \Sigma)$. 这里 $\Sigma = (\sigma_{ij})$ 为 n 阶严格正定对称方阵, $\mu = (\mu_1, \dots, \mu_n)$ 为 n -维行向量. 则

$$f(t) = \exp\left\{i\mu t' - \frac{1}{2}t\Sigma t'\right\}.$$

证明. ξ 的密度函数为:

$$\begin{aligned} p(x_1, \dots, x_n) &= \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2} \sum_{j,k=1}^n \gamma_{jk}(x_j - \mu_j)(x_k - \mu_k)\right\} \\ &= \frac{1}{(2\pi)^{n/2}(\det \Sigma)^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x - \mu)\Sigma^{-1}(x - \mu)'\right\}, \end{aligned}$$

其中

$$\Sigma^{-1} = (\gamma_{ij}).$$

现在计算

$$f(t) = \int_{\mathbb{R}^n} \exp\{it \cdot x\} p(x) dx.$$

(i) 若 Σ 是对角阵:

$$\Sigma_1 := \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}.$$

则 ξ_1, \dots, ξ_n 独立, 且 $\xi_k \sim N(\mu_k, \sigma_k^2)$. 因此 ξ 的特征函数为

$$f(t) = \prod_{k=1}^n f_k(t_k),$$

其中 f_k 为 ξ_k 的特征函数, 即

$$f_k(t_k) = \exp\left\{i\mu_k t_k - \frac{1}{2}\sigma_k^2 t_k^2\right\}.$$

所以

$$f(t) = \prod_{k=1}^n \exp\left\{i\mu_k t_k - \frac{1}{2}\sigma_k^2 t_k^2\right\}.$$

(ii) 一般情况. 此时有正交阵 A 使得

$$\Sigma = A\Sigma_1A'.$$

令

$$\eta := (\xi - \mu)A.$$

则 $\xi = \eta A' + \mu$, $\eta \sim N(0, \Sigma_1)$. 因此

$$f_\eta(t) = \exp\left\{-\frac{1}{2}t\Sigma_1t'\right\}.$$

所以

$$\begin{aligned} f_\xi(t) &= E[\exp\{i\xi t'\}] \\ &= E[\exp\{i(\eta A' + \mu)t'\}] \\ &= \exp\{i\mu t'\} \exp\left\{-\frac{1}{2}tA\Sigma_1A't'\right\} \\ &= \exp\left\{i\mu t' - \frac{1}{2}t\Sigma t'\right\}. \end{aligned}$$

□

我们曾经证明, 若 ξ, η 独立, 那么

$$f_{\xi+\eta}(t) = f_\xi(t)f_\eta(t).$$

但反之不真. 不过, 独立性的确是可以通过特征函数刻画的, 但需要用到多维特征函数. 这就是下面的:

命题 8.6.2. 设 ξ_1, \dots, ξ_n 为随机变量, f 是 (ξ_1, \dots, ξ_n) 的特征函数, f_k 是 ξ_k 的特征函数. 则 ξ_1, \dots, ξ_n 独立的充要条件是

$$f(t_1, \dots, t_n) = f_1(t_1) \cdots f_n(t_n), \quad \forall (t_1, \dots, t_n) \in \mathbb{R}^n.$$

注意这里的充分性: 它要求对任意的 $(t_1, \dots, t_n) \in \mathbb{R}^n$ 等式成立. 若仅仅有

$$f(t, \dots, t) = f_1(t) \cdots f_n(t), \quad \forall t \in \mathbb{R},$$

即 ξ_1, \dots, ξ_n 和的特征函数等于特征函数的和是得不到 ξ_1, \dots, ξ_n 的独立性的. 反例见习题.

证明. 必要性显然, 往证充分性. 以 $n = 2$ 为例, 一般情形是一样的. 由定理6.1.10, 只需证对任意 $f_1, f_2 \in C_0^\infty$, 有

$$E[f_1(\xi_1)f_2(\xi_2)] = E[f_1(\xi_1)]E[f_2(\xi_2)]. \quad (6.9)$$

令

$$\hat{f}_k(t) := \int_{-\infty}^{\infty} f_k(x)e^{itx}dx, \quad k = 1, 2.$$

因为 $f_k \in C_0^\infty$, 所以分部积分两次后得到存在 $C > 0$ 使得

$$|\hat{f}_k(t)| \leq C \frac{1}{1+t^2}.$$

因此

$$\int_{-\infty}^{\infty} |\hat{f}_k(t)| dt < \infty, \quad k = 1, 2.$$

这样由命题8.4.5知

$$f_k(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}_k(t) e^{-itx} dt.$$

因此我们有

$$\begin{aligned} E[f_1(\xi_1)f_2(\xi_2)] &= \frac{1}{4\pi^2} E \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_1(t_1) e^{-it_1\xi_1} \hat{f}_2(t_2) e^{-it_2\xi_2} dt_1 dt_2 \right] \\ &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E \left[\hat{f}_1(t_1) e^{-it_1\xi_1} \hat{f}_2(t_2) e^{-it_2\xi_2} \right] dt_1 dt_2 \\ &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_1(t_1) \hat{f}_2(t_2) E \left[e^{-it_1\xi_1} e^{-it_2\xi_2} \right] dt_1 dt_2 \\ &= \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{f}_1(t_1) \hat{f}_2(t_2) E[e^{-it_1\xi_1}] E[e^{-it_2\xi_2}] dt_1 dt_2 \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}_1(t_1) E[e^{-it_1\xi_1}] dt_1 \cdot \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}_2(t_2) E[e^{-it_2\xi_2}] dt_2 \\ &= \frac{1}{2\pi} E \left[\int_{-\infty}^{\infty} \hat{f}_1(t_1) e^{-it_1\xi_1} dt_1 \right] \cdot \frac{1}{2\pi} E \left[\int_{-\infty}^{\infty} \hat{f}_2(t_2) e^{-it_2\xi_2} dt_2 \right] \\ &= E[f_1(\xi_1)] E[f_2(\xi_2)]. \end{aligned}$$

这正是我们要证明的. □

设 $F: \mathbb{R}^d \mapsto \mathbb{R}$, $a_i \leq b_i$. 定义

$$\begin{aligned} \Delta_{a_i, b_i}^i F(x_1, \dots, x_{i-1}, \hat{x}_i, x_{i+1}, \dots, x_d) \\ := F(x_1, \dots, x_{i-1}, b_i, x_{i+1}, \dots, x_d) - F(x_1, \dots, x_{i-1}, a_i, x_{i+1}, \dots, x_d), \end{aligned}$$

其中 \hat{x}_i 表示去掉这个变量. 对 $a = (a_1, \dots, a_d)$, $b = (b_1, \dots, b_d)$, $a_i \leq b_i, \forall i$, 定义

$$\Delta_{(a,b)} F := \Delta_{a_1, b_1}^1 \cdots \Delta_{a_n, b_n}^n F.$$

我们有:

定理 8.6.3 (d -维反演公式). 设 F 的特征函数为 f , $a \leq b$. 令 $A = \{x: x = (x_1, \dots, x_d), x_k \in \{a_k, b_k\}\}$. 若 A 中的点均是 F 的连续点, 则

$$\Delta_{(a,b)} F = \frac{1}{(2\pi)^d} \lim_{T_k \rightarrow \infty, k=1, \dots, d} \int_{-T_1}^{T_1} \cdots \int_{-T_d}^{T_d} \prod_{k=1}^d \frac{e^{-it_k a_k} - e^{-it_k b_k}}{it_k} f(t_1, \dots, t_d) dt_1 \cdots dt_d.$$

若有

$$\int_{\mathbb{R}^d} |f(t)| dt < \infty,$$

则 F 有密度函数

$$p(x) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-itx'} f(t) dt.$$

若进一步有

$$\int_{\mathbb{R}^d} |t|^n |f(t)| dt < \infty, \quad (6.10)$$

则对任意满足 $k := k_1 + \cdots + k_d \leq n$ 的 (k_1, \cdots, k_d) ,

$$\frac{\partial^k p}{\partial^{k_1} x_1 \cdots \partial^{k_d} x_d}$$

存在.

证明. 同一维的情形一样, 我们只在

$$\int_{\mathbb{R}^d} |f(t)| dt < \infty$$

的情况下给予详细证明. 不过我们特地选择了一个不一样的证明, 以便读者多一个视角看待同一个问题.

设 ξ 的分布函数和特征函数分别为 F 与 f , η 是与 ξ 独立的 d -维随机变量, 具有密度函数 g 和特征函数 φ , 且 φ 满足

$$\int_{\mathbb{R}^d} |\varphi(t)| dt < \infty \quad (6.11)$$

且

$$g(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \varphi(t) e^{-itx} dt, \quad \forall x. \quad (6.12)$$

由于 $|f(t)| \leq 1$, 所以

$$\int_{\mathbb{R}^d} |f(t)| |\varphi(t)| dt < \infty.$$

于是我们有

$$\begin{aligned} \int_{\mathbb{R}^d} e^{-itx} f(t) \varphi(t) dt &= \int_{\mathbb{R}^d} e^{-itx} E[\exp\{it\xi\}] E[\exp\{it\eta\}] dt \\ &= \int_{\mathbb{R}^d} e^{-itx} E[\exp\{it(\xi + \eta)\}] dt \\ &= E \left[\int_{\mathbb{R}^d} \exp\{it(-x + \xi + \eta)\} dt \right] \\ &= E \left[E \left[\int_{\mathbb{R}^d} \exp\{-it(x - y) + it\eta\} dt \right] \Big|_{y=\xi} \right] \\ &= E \left[\int_{\mathbb{R}^d} \exp\{-it(x - y)\} E[\exp\{it\eta\}] dt \Big|_{y=\xi} \right] \\ &= E \left[\int_{\mathbb{R}^d} \exp\{-it(x - \xi)\} \varphi(t) dt \right] \\ &= (2\pi)^d E[g(x - \xi)]. \end{aligned}$$

由于 $\forall a \leq b$ (即 $a_i \leq b_i, \forall i = 1, \dots, d$), 所以

$$\begin{aligned} \int_a^b E[g(x - \xi)]dx &= E \left[\int_a^b g(x - \xi)dx \right] \\ &= E \left[\int_a^b g(x - y)dx \Big|_{y=\xi} \right] \\ &= E \left[\int_{a-y}^{b-y} g(x)dx \Big|_{y=\xi} \right] \\ &= E [E[1_{(a,b)}(\eta + y)]|_{y=\xi}] \\ &= E[1_{(a,b)}(\eta + \xi)], \end{aligned}$$

所以

$$\int_a^b dx \int_{\mathbb{R}^d} e^{-itx} f(t) \varphi(t) dt = (2\pi)^d E[1_{(a,b)}(\eta + \xi)].$$

现在设 $\zeta \sim N(0, I)$. 取 $\eta = n^{-1}\zeta$. 显然 η 满足 (6.11) 与 (6.12). 因此

$$\int_a^b p_n(x) dx = E [1_{(a,b)}(n^{-1}\zeta + \xi)],$$

其中

$$p_n(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-itx} f(t) \varphi_n(t) dt,$$

而 φ_n 是 $n^{-1}\zeta$ 的特征函数. 由于 $n^{-1}\zeta \xrightarrow{a.s.} 0$, 所以 $\varphi_n(t) \rightarrow 1, \forall t$. 由控制收敛定理,

$$\lim_{n \rightarrow \infty} p_n(x) = p(x) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-itx} f(t) dt.$$

又由于

$$p_n(x) \leq \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |f(t)| dt,$$

所以再一次用控制收敛定理得

$$\begin{aligned} \int_a^b p(x) dx &= \int_a^b \lim_{n \rightarrow \infty} p_n(x) dx \\ &= \lim_{n \rightarrow \infty} E [1_{(a,b)}(n^{-1}\zeta + \xi)]. \end{aligned}$$

由Fatou引理,

$$\begin{aligned} E[1_{(a,b)}(\xi)] &\leq E \left[\liminf_{n \rightarrow \infty} 1_{(a,b)}(\xi + n^{-1}\zeta) \right] \\ &\leq \liminf_{n \rightarrow \infty} E [1_{(a,b)}(\xi + n^{-1}\zeta)] \\ &\leq \limsup_{n \rightarrow \infty} E [1_{(a,b)}(\xi + n^{-1}\zeta)] \\ &\leq E \left[\limsup_{n \rightarrow \infty} 1_{(a,b)}(\xi + n^{-1}\zeta) \right] \\ &\leq E[1_{[a,b]}(\xi)]. \end{aligned}$$

若 a, b 均为 F 的连续点, 则得到

$$F(b) - F(a) = \lim_{n \rightarrow \infty} E [1_{(a,b)}(\xi + n^{-1}\zeta)] = \int_a^b p(x)dx.$$

因此 F 是具有密度 p 的连续型分布. □

注. 如何证明(6.10)? 这是一个问题. 由于

$$f(t) = \int_{\mathbb{R}^d} \exp\{it \cdot x\} dF_\xi(x) = E[\exp\{it \cdot \xi\}],$$

故得到这样一个估计式基本上都要用到分部积分, 以把 t 从 $\exp\{it \cdot x\}$ 或 $\exp\{it \cdot \xi\}$ 上拉下来. 但利用第一个表达式进行分部积分是白日梦, 因为这需要 F_ξ 可微, 而我们要证明的就是此可微性. 那么利用第二个表达式呢? 这就要用到 ξ 的可微性. 但 ξ 是定义在概率空间上, 什么叫可微呢? 法国科学院前院士Paul Malliavin教授⁴对一类重要的概率空间定义了这种可微性, 建立了这种概率空间上的微分理论, 由此得到了分部积分公式, 并利用这个公式得到了一大类随机变量的密度函数的存在性. 由于这类概率空间是轨道函数空间, 所以Malliavin教授自己称之为随机变分学(Stochastic Calculus of Variations). 这是一套极其深奥的理论, 现在被学术界公称为Malliavin Calculus.

习题

1. 设 F 是 \mathbb{R}^d 上分布函数. 证明: 存在 $\{x_n\} \subset C(F)$, 使得 $x_n \rightarrow -\infty$.

解. $G(r) := F(r(1, \dots, 1))$ 是 \mathbb{R} 上的单调上升函数, 故至多只有可列个间断点.

2. 设 $\xi = (\xi_1, \dots, \xi_m)$ 的特征函数为 f , 设 $n \in \mathbb{N}_{++}$, $E[|\xi_i|^n] < \infty$, $\forall i = 1, 2, \dots, m$. 证明:

(a) $\forall \nu_1, \dots, \nu_m \in \mathbb{N}_+$, $\nu_1 + \dots + \nu_m \leq n$,

$$E[|\xi_1|^{\nu_1} \dots |\xi_m|^{\nu_m}] < \infty.$$

(b) 对满足上面条件的 ν_1, \dots, ν_m ,

$$\frac{\partial^{\nu_1 + \dots + \nu_m}}{\partial t_1^{\nu_1} \dots \partial t_m^{\nu_m}} f(t_1, \dots, t_m)$$

存在, 连续, 且在 $t = 0$ 处等于

$$i^{\nu_1 + \dots + \nu_m} E[\xi_1^{\nu_1} \dots \xi_m^{\nu_m}].$$

3. 设 (ξ, η) 的联合密度为

$$\varphi(x, y) = \frac{1}{4} (1 + xy(x^2 - y^2)) 1_{|x| \leq 1, |y| \leq 1}.$$

(a) 分别以 f, f_1, f_2 记 $(\xi, \eta), \xi, \eta$ 的特征函数. 证明:

$$f(t, t) = f_1(t)f_2(t), \quad \forall t \in \mathbb{R}.$$

(b) 证明 ξ 与 η 不独立.

⁴Paul Malliavin, 1925-2010, 法国数学家.

8.7 Skorokhod表现定理

本节我们证明弱收敛的Skorokhod⁵表现定理. 为方便计, 作为出发点即弱收敛的定义, 我们采用下面的:

定义 8.7.1. 设 F_n, F 是分布函数. 若 $\forall x \in C(F), F_n(x) \rightarrow F(x)$, 则称 F_n 弱收敛于 F .

由定理8.5.3与定理8.5.4, 此定义与原定义等价.

如果说在几乎必然收敛与依概率收敛中, 涉及到的随机变量都是定义在同一个概率空间上的话, 那么弱收敛则完全不同: 它只关心所涉及的随机变量的分布, 因此只要求它们取值于同一空间, 而对它们的定义域是否相同没有任何要求. 而即使所涉及的随机变量定义在同一概率空间上, 它关心的仍然只是它们取相同值的概率是否差不多, 而对在哪里取值则毫不在意.

由于这些原因, 下面的结果就十分引人注目了. (唉, Skorokhod得到这个定理的时候大概就是二十多岁. 难怪毛主席说: 我们“科学水平低……, 很多地方不如人家, 骄傲不起来.”)

定理 8.7.2 (Skorokhod表现定理). 设 $F_n \xrightarrow{w} F$, 则存在概率空间 (Ω, \mathcal{F}, P) 及定义在其上的随机变量 ξ_n, ξ , 使得它们分别以 F_n, F 为分布函数, 且 $\xi_n \xrightarrow{a.s.} \xi$.

在证明之前, 我们先看看这个问题的要点在哪里. 我们知道, 任意一个随机变量都可以在 $([0, 1], \mathcal{B}[0, 1], dx)$ 上实现. 好, 让我们看一个最简单的情形: 设 ξ_1, ξ_2 是两个Bernoulli随机变量(可以定义在不同的概率空间上), 那么当然存在 $([0, 1], \mathcal{B}[0, 1], dx)$ 的两个随机变量 η_1, η_2 , 使得它们都服从Bernoulli分布. 不过, η_1 的取值与 η_2 的取值可能处处都不相等, 例如可取

$$\eta_1(\omega) := 1_{[0, \frac{1}{2})}, \quad \eta_2 := 1_{[\frac{1}{2}, 1]}.$$

然而, η_1 和 η_2 当然也可以取为同一个. 虽然这是一种极端情况, 但依然可以给我们以启示. 也就是说, 当 ξ_1 与 ξ_2 同分布时, 可以在 $([0, 1], \mathcal{B}[0, 1], dx)$ 上定义两个恒等的随机变量 η_1 与 η_2 , 使得它们都与 ξ_1 和 ξ_2 的分布相同; 当 ξ_1 与 ξ_2 不同分布, 然而其分布比较接近时, 我们应该可以小心地分配 η_1 与 η_2 的值, 使得它们也比较接近. 这个事实对简单随机变量尤其清楚. 例如, 若

$$F_1(x) := \left(\frac{1}{2} + 10^{-2}\right) 1_{[0, 1)}(x) + 1_{[1, \infty)}(x)$$

$$F_2(x) = \frac{1}{2} 1_{[0, 1)}(x) + 1_{[1, \infty)}(x)$$

时, 便可取

$$\eta_1(\omega) = \begin{cases} 0, & \omega \in [0, \frac{1}{2} + 10^{-2}) \\ 1, & \omega \in [\frac{1}{2} + 10^{-2}, 1] \end{cases}$$

$$\eta_2(\omega) = \begin{cases} 0, & \omega \in [0, \frac{1}{2}) \\ 1, & \omega \in [\frac{1}{2}, 1] \end{cases}$$

易见仅对很少一部分的 ω , η_1 与 η_2 不相等.

下面就是这个定理的证明.

⁵Anatolii Volodymyrovych Skorokhod, 1930-2011, 苏联-乌克兰数学家.

证明. 我们用定理4.6.1之证明中的记号. 即概率空间取为 $([0, 1], \mathcal{B}, P)$, P 为 Lebesgue 测度. 设对 F , 构造出的随机变量为 ξ , 相应的逼近列为 $\{\eta_m\}$; 对 F_n , 构造出的随机变量为 ξ_n , 相应的逼近列为 $\{\eta_{n,m}\}$.

我们将定理4.6.1之证明中的所有分割的分点均取为 F 的连续点, 这样就可以在推进时避开地雷阵即不连续点集. 因此在所有分点上 F_n 均收敛于 F . 所以若 x, y 是分点, $\omega \in (F(x), F(y))$, 那么当 n 充分大时 $\omega \in (F_n(x), F_n(y))$, 于是 $\eta_{n,m}(\omega) = \eta_m(\omega)$.

所以对这样的 ω 有

$$\begin{aligned} & \limsup_{n \rightarrow \infty} |\xi(\omega) - \xi_n(\omega)| \\ & \leq |\xi(\omega) - \eta_m(\omega)| + \limsup_{n \rightarrow \infty} |\eta_m(\omega) - \eta_{n,m}(\omega)| + \limsup_{n \rightarrow \infty} |\eta_{n,m}(\omega) - \xi_n(\omega)| \\ & \leq 2^{-m+1}. \end{aligned}$$

令 $m \rightarrow \infty$ 即得结论. □

Skorokhod表现定理的威力极其强大, 极其有用. 下面我们看一些最简单的应用, 即用它给一些已知的结果以新的证明. 你比较一下新证明和原证明, 就能初步感受其威力了.

定理 8.7.3 (Helly-Bray定理). 设 $F_n \xrightarrow{w} F$, 则

(i) $\forall f \in C_b(\mathbb{R})$ 有

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f(x) dF_n(x) = \int_{-\infty}^{\infty} f(x) dF(x).$$

(ii) 设 $-\infty < a < b < \infty$, 且 $a, b \in C(F)$. 则 $\forall f \in C([a, b])$, 有

$$\lim_{n \rightarrow \infty} \int_a^b f(x) dF_n(x) = \int_a^b f(x) dF(x).$$

证明. (1) 用Skorokhod表现定理及控制收敛定理,

$$\begin{aligned} \lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} f(x) dF_n(x) &= \lim_{n \rightarrow \infty} E[f(\xi_n)] \\ &= E[f(\xi)] \\ &= \int_{-\infty}^{\infty} f(x) dF(x). \end{aligned}$$

(2) 设 $F(a) < F(b)$, 对充分大的 n , 令

$$\begin{aligned} \hat{F}_n(x) &:= (F_n(b) - F_n(a))^{-1} F_n((x \vee a) \wedge b), \\ \hat{F}(x) &:= (F(b) - F(a))^{-1} F((x \vee a) \wedge b). \end{aligned}$$

对 \hat{F}_n 及 \hat{F} 用上面结果即可. 请自证 $F(a) = F(b)$ 的情况. □

定理 8.7.4. 设 $F_n \xrightarrow{w} F$, f_n, f 分别为 F_n, F 的特征函数, 则在任何有限区间上, f_n 一致收敛于 f .

证明. 用Skorokhd表现定理及控制收敛定理, $\forall t$,

$$f_n(t) = E[e^{it\xi_n}] \rightarrow E[e^{it\xi}] = f(t).$$

为证有限区间上的一致性, 只需证 $\{f_n\}$ 是等度连续的. $\forall \varepsilon > 0$, 取 M 使得

$$F_n(-M) + 1 - F_n(M) < \varepsilon, \quad \forall n.$$

则

$$\sup_n P(|\xi_n| > M) < \varepsilon.$$

再取 $\delta = M^{-1}\varepsilon$. 则 $|t - s| < \delta$ 时

$$\begin{aligned} |f_n(t) - f_n(s)| &\leq E[|e^{it\xi_n} - e^{is\xi_n}|, |\xi_n| \leq M] + E[|e^{it\xi_n} - e^{is\xi_n}|, |\xi_n| > M] \\ &\leq |t - s|M + 2\varepsilon \\ &= 3\varepsilon. \end{aligned}$$

□

习题

1. 设 $\xi_n \xrightarrow{w} \xi$, $\{f_n\}$ 是连续函数列且在任意有限区间上一致收敛于函数 f . 证明 $f_n(\xi_n) \xrightarrow{w} f(\xi)$.
2. 设 $\xi_n \xrightarrow{w} \xi$, $\xi_n \geq 0, \forall n$. 证明: $\xi \geq 0$ 且

$$E[\xi] \leq \liminf_{n \rightarrow \infty} E[\xi_n].$$

9 特征函数的应用

特征函数之所以可能有用, 在于唯一性定理, 即特征函数和分布函数是相互唯一确定的(所以才是“特征”). 而之所以的确大有其用, 很大一部分原因在于, 独立随机变量和的分布函数等于其每个分布的卷积, 而独立随机变量和的特征函数等于其每个特征函数的乘积——后者容易处理多了.

在上一章我们已经穿插了一些特征函数的应用例子了, 本章我们再看一些更多的例子.

9.1 分布的计算

因为唯一性定理, 特征函数可用来帮助计算分布. 我们用几个具体例子来说明.

一. 设有 m 个桶, 一投手向它们投球. 设所投总数是随机的, 服从参数为 λ 的Poisson分布. 在投球次数确定之后, 每一球投入第 i 个桶的概率均为 p_i , $\sum_{i=1}^m p_i = 1$, 不随总投球次数的改变而改变, 且每次的投掷是相互独立的. 求投入到各桶的球数的分布, 并证明投入到各桶的球数是相互独立的.

从比例原则看, 投入每桶的球数也应该是Poisson分布, 且参数为 $p_i \lambda$. 但这需要严格证明. 我们来尝试建立这个问题的数学模型. 令

$$\Omega := \{\omega = (\omega_1, \dots, \omega_n) : n \in \mathbb{N}_+, \omega_1, \dots, \omega_n \in \{1, \dots, m\}\} \cup \{\Delta\}.$$

记 $|(\omega_1, \dots, \omega_n)| = n$, $|\Delta| = 0$. 以 Y 表示投掷次数, 即

$$Y(\omega) = |\omega|.$$

现在定义概率. 令

$$P(Y = n) = \frac{\lambda^n}{n!} e^{-\lambda}, \quad n = 0, 1, 2, \dots,$$

$$P((\omega_1, \dots, \omega_n) | Y = n) = p_{\omega_1} \cdots p_{\omega_n},$$

$$P(\Delta | Y = 0) = 1.$$

这样概率 P 就定义好了. 对 $n \in \mathbb{N}_{++}$, 令

$$\eta_n(\omega) = \begin{cases} \omega_n, & |\omega| \geq n, \\ 0, & |\omega| < n. \end{cases}$$

则在条件 $Y = n$ 下, $\eta_1, \eta_2, \dots, \eta_n$ 是独立同分布随机变量, 直观上表示 n 次投球中依次落入的桶号. 且

$$P(\eta_1 = \alpha | Y = n) = p_\alpha, \quad \forall \alpha \in \{1, \dots, m\}.$$

设 f 是定义在 $\{1, \dots, m\}$ 上的函数. 令

$$X := \sum_{k=1}^Y f(\eta_k) \quad \left(\sum_1^0 := 0 \right).$$

注意

$$E[\exp(if(\eta_1)) | Y = n] = \sum_{\alpha=1}^m \exp(if(\alpha)) p_\alpha,$$

再注意在条件 $Y = n$ 下, $\eta_1, \eta_2, \dots, \eta_n$ 独立同分布, 所以有

$$\begin{aligned} E \left[\exp \left(i \sum_{k=1}^Y f(\eta_k) \right) \middle| Y = n \right] &= E \left[\exp \left(\sum_{k=1}^n if(\eta_k) \right) \middle| Y = n \right] \\ &= \left(\sum_{\alpha=1}^m \exp(if(\alpha)) p_\alpha \right)^n. \end{aligned}$$

因此有

$$\begin{aligned} E[\exp(iX)] &= E \left[E \left[\exp \left(\sum_{k=1}^Y if(\eta_k) \right) \middle| Y = n \right] \middle|_{n=Y} \right] \\ &= E \left[\left(\sum_{\alpha=1}^m \exp(if(\alpha)) p_\alpha \right)^n \middle|_{n=Y} \right] \\ &= E \left[\left[\sum_{\alpha=1}^m \exp(if(\alpha)) p_\alpha \right]^Y \right] \\ &= \sum_{n=0}^{\infty} \left[\sum_{\alpha=1}^m \exp(if(\alpha)) p_\alpha \right]^n \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \exp \left(\lambda \left(\sum_{\alpha=1}^m \exp(if(\alpha)) p_\alpha - 1 \right) \right). \end{aligned} \quad (1.1)$$

若取 $f_j(\alpha) = \delta_{j\alpha}$, 则

$$\xi_j := \sum_{k=1}^Y f_j(\eta_k) \quad \left(\sum_1^0 := 0 \right).$$

为落在 j 号桶中的球数. 对 $g_j(\alpha) := t_j f_j(\alpha)$ 用(1.1)得

$$E[\exp(it\xi_j)] = \exp(\lambda p_j (e^{it_j} - 1)).$$

对 $g(\alpha) := \sum_{j=1}^m t_j f_j(\alpha)$ 用(1.1)得

$$E \left[\exp \left(i \sum_{j=1}^m t_j \xi_j \right) \right] = \exp \left(\lambda \left(\sum_{j=1}^m e^{it_j} p_j - 1 \right) \right)$$

故

$$E \left[\exp \left(i \sum_{j=1}^m t_j \xi_j \right) \right] = \prod_{j=1}^m E[\exp(it_j \xi_j)].$$

因此 ξ_1, \dots, ξ_m 独立且 $\xi_j \sim P(\lambda p_j)$.

本例可以进一步推广到以下情形:

二. 设 $\eta \sim P(\lambda)$, $\xi_i, i = 1, 2, \dots$ 同分布, 其公共的分布函数为 F . 设 $\eta, \xi_i, i = 1, 2, \dots$ 独立. 求 $X := \sum_{i=1}^{\eta} 1_A(\xi_i)$ 的分布 ($\sum_1^0 := 0$), 并证明当 A_1, \dots, A_m 两两不交时, X_1, \dots, X_m 是独立的, 其中 $X_k := \sum_{i=1}^{\eta} 1_{A_k}(\xi_i)$.

解. 我们来计算特征函数. 设 $f \in \mathcal{B}_b$. 由定理6.1.3有

$$\begin{aligned} E \left[\exp \left(i \sum_{\alpha=1}^{\eta} f(\xi_{\alpha}) \right) \right] &= E \left[E \left[\exp \left(i \sum_{\alpha=1}^n f(\xi_{\alpha}) \right) \right]_{n=\eta} \right] \\ &= \sum_{n=0}^{\infty} E \left[\exp \left(\sum_{\alpha=1}^n i f(\xi_{\alpha}) \right) \right] \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \sum_{n=0}^{\infty} \left(\int_{\mathbb{R}} \exp(i f(x)) dF(x) \right)^n \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \exp \left(-\lambda + \lambda \int_{\mathbb{R}} \exp(i f(x)) dF(x) \right). \end{aligned}$$

特别, 取 $f = t1_A$, 得

$$\begin{aligned} E[\exp(itX)] &= \exp \left(-\lambda + \lambda \int_{\mathbb{R}} \exp(it1_A(x)) dF(x) \right) \\ &= \exp \left(\lambda \int_A (\exp(it) - 1) dF(x) \right) \\ &= \exp(\lambda \nu(A)(e^{it} - 1)), \end{aligned}$$

其中 $\nu(A) = \int_A dF(x)$. 因此 $X \sim P(\lambda \nu(A))$. 取 $f = \sum_{k=1}^m t_k 1_{A_k}$, 则得

$$\begin{aligned} E[\exp(i \sum_{k=1}^m t_k X_k)] &= \exp \left(-\lambda + \lambda \int_{\mathbb{R}} \exp \left(i \sum_{k=1}^m t_k 1_{A_k}(x) \right) dF(x) \right) \\ &= \prod_{k=1}^m \exp(\lambda t_k \nu(A_k)(e^{it_k} - 1)) \\ &= \prod_{k=1}^m E[\exp(it_k X_k)]. \end{aligned}$$

所以 X_1, \dots, X_m 独立, 且 $X_k \sim P(\lambda \nu(A_k))$.

当 $\xi_1, \xi_2, \dots, \xi_n$ 服从同一类型的分布且独立时, 如果 $\xi_1 + \dots + \xi_n$ 也服从这一类型的分布, 那么就说这个分布具有再生性. 下面我们就考察几种分布的再生性.

命题 9.1.1. 设 $\xi_k \sim N(a_k, \sigma_k^2)$, $c_k \in \mathbb{R}$, $k = 1, \dots, n$, ξ_1, \dots, ξ_n 独立, 则

$$\xi := \sum_{k=1}^n c_k \xi_k \sim N \left(\sum_{k=1}^n c_k a_k, \sum_{k=1}^n c_k^2 \sigma_k^2 \right).$$

证明. 因为

$$\begin{aligned} E[\exp(it\xi)] &= \prod_{k=1}^n E[\exp(itc_k\xi_k)] \\ &= \prod_{k=1}^n \exp\left\{itc_k a_k - \frac{1}{2}c_k^2\sigma_k^2 t^2\right\} \\ &= \exp\left\{it \sum_{k=1}^n c_k a_k - \frac{1}{2}t^2 \sum_{k=1}^n c_k^2\sigma_k^2\right\}. \end{aligned}$$

□

命题 9.1.2. 设 ξ_1, \dots, ξ_n 独立, 且 $\xi_k \sim P(\lambda_k)$, $k = 1, \dots, n$. 令 $\xi := \sum_{k=1}^n \xi_k$. 则 $\xi \sim P(\sum_{k=1}^n \lambda_k)$.

证明. 我们来计算 ξ 的特征函数. 由独立性有

$$E[\exp\{it\xi\}] = \prod_{k=1}^n E[\exp\{it\xi_k\}] = \exp\left\{(e^{it} - 1) \sum_{k=1}^n \lambda_k\right\}.$$

□

命题 9.1.3. 设 ξ_1, \dots, ξ_n 独立, 且 $\xi_k \sim B(m_k, p)$. 令 $\xi := \sum_{k=1}^n \xi_k$. 则 $\xi \sim B(\sum_{k=1}^n m_k, p)$.

证明. 因为

$$E[e^{it\xi_k}] = (q + pe^{it})^{m_k},$$

所以

$$E[e^{it \sum_{k=1}^n \xi_k}] = (q + pe^{it})^{\sum_{k=1}^n m_k}.$$

□

习题

1. 证明: ξ 与 $-\xi$ 同分布的充要条件是 ξ 的特征函数是实值的.
2. 设 ξ_1, \dots, ξ_n 独立同分布, 且公共的分布为几何分布. 求 $\xi_1 + \dots + \xi_n$ 的分布.
3. 设 φ 是特征函数. 证明: 对任意 $\lambda \geq 0$, $e^{\lambda(\varphi-1)}$ 也是特征函数.
4. 设 ξ 是随机变量, F 是其分布函数, $m_n := E[\xi^n]$ 存在, $\forall n \in \mathbb{N}_+$. 证明: 若存在 $\delta > 0$ 使得

$$\sum_{n=1}^{\infty} \frac{|m_n|}{n!} \delta^n$$

收敛, 则 F 由 $\{m_n\}$ 唯一决定.

9.2 极限定理

迄今为止, 我们定义了几乎必然收敛, 依概率收敛及分布函数的弱收敛. 我们还证明了几乎必然收敛蕴含了依概率收敛和分布函数的弱收敛. 由于弱收敛只涉及到分布, 所以我们也常常将 $\xi_n \xrightarrow{w} \xi$ 中的 ξ 换为其分布. 例如, 若 $\xi \sim N(0, 1)$, 我们就记为 $\xi_n \xrightarrow{w} N(0, 1)$.

注意 $\xi_n \xrightarrow{w} \xi$ 并不要求 ξ_n 们和 ξ 定义在同一个概率空间上, 因此一般说来无法谈及弱收敛和依概率收敛谁强谁弱的问题. 但如果刚好它们都定义在同一个概率空间上, 则可以比较其强弱. 实际上, 我们在8.5节曾经证明了几乎必然收敛蕴含弱收敛. 现在我们进一步证明依概率收敛也蕴含弱收敛:

命题 9.2.1. 若 $\xi_n \xrightarrow{P} \xi$, 则 $\xi_n \xrightarrow{w} \xi$.

证明. 方法1. 首先我们注意, 对任意 $f \in C_b$ 有 $f(\xi_n) \xrightarrow{P} f(\xi)$. 这是因为, 假设结论不对, 则存在 $\delta_0 > 0$ 和 $\varepsilon_0 > 0$, 以及子列 ξ_{n_k} , 使

$$P(|f(\xi_{n_k}) - f(\xi)| > \delta_0) \geq \varepsilon_0.$$

但因为 $\xi_{n_k} \xrightarrow{P} \xi$, 因此存在子列, 为记号简单不妨仍记为它自己, $\xi_{n_k} \xrightarrow{a.s.} \xi$, 因此 $f(\xi_{n_k}) \xrightarrow{a.s.} f(\xi)$. 这与上面的不等式矛盾.

因此 $f(\xi_n) \xrightarrow{P} f(\xi)$. 由控制收敛定理, 有

$$\lim_{n \rightarrow \infty} E[f(\xi_n)] = E[f(\xi)].$$

此即 $\xi_n \xrightarrow{w} \xi$.

方法2. 分别以 F_n 及 F 表示 ξ_n 与 ξ 的分布函数. $\forall x$ 及 $\varepsilon > 0$, 有

$$\begin{aligned} \{\xi \leq x - \varepsilon\} &= \{\xi \leq x - \varepsilon, \xi_n \leq x\} \cup \{\xi \leq x - \varepsilon, \xi_n > x\} \\ &\subset \{\xi_n \leq x\} \cup \{|\xi - \xi_n| > \varepsilon\}. \end{aligned}$$

所以,

$$P(\xi \leq x - \varepsilon) \leq P(\xi_n \leq x) + P(|\xi - \xi_n| > \varepsilon),$$

即

$$F(x - \varepsilon) \leq F_n(x) + P(|\xi - \xi_n| > \varepsilon).$$

类似地,

$$F_n(x) \leq F(x + \varepsilon) + P(|\xi - \xi_n| > \varepsilon).$$

令 $n \rightarrow \infty$ 得

$$F(x - \varepsilon) \leq \liminf_{n \rightarrow \infty} F_n(x) \leq \limsup_{n \rightarrow \infty} F_n(x) \leq F(x + \varepsilon).$$

若 $x \in C(F)$, 令 $\varepsilon \downarrow 0$ 就有

$$F(x) = \lim_{n \rightarrow \infty} F_n(x).$$

□

这个命题的逆命题是不成立的, 例如, 设 ξ 服从Bernoulli分布, 即 $P(\xi = 1) = P(\xi = -1) = \frac{1}{2}$. 令 $\xi_n := (-1)^n \xi$. 则 ξ_n 与 ξ 同分布, 因此 $\xi_n \xrightarrow{w} \xi$. 但显然没有 $\xi_n \xrightarrow{P} \xi$.

不过, 在一个特殊情况下, 一个重要的特殊情况下, 逆命题是成立的.

命题 9.2.2. 设 ξ_n 们和 ξ 定义在同一个概率空间上, 且 $\xi_n \xrightarrow{w} c$, 则 $\xi_n \xrightarrow{P} c$.

证明. 常数 c 的分布函数为

$$F(x) = \begin{cases} 0, & x < c, \\ 1, & x \geq c. \end{cases}$$

它唯一的间断点为 c . 因此

$$\lim_{n \rightarrow \infty} F_n(x) = \begin{cases} 0, & x < c, \\ 1, & x > c. \end{cases}$$

于是, $\forall \varepsilon > 0$, 我们有

$$\begin{aligned} P(|\xi_n - c| \geq \varepsilon) &= P(\xi_n \leq c - \varepsilon) + P(\xi_n \geq c + \varepsilon) \\ &\leq P(\xi_n \leq c - \varepsilon) + P(\xi_n > c + \varepsilon/2) \\ &= F_n(c - \varepsilon) + 1 - F_n(c + \varepsilon/2) \\ &\rightarrow F(c - \varepsilon) + 1 - F(c + \varepsilon/2) \\ &= 0 + 1 - 1 = 0. \end{aligned}$$

□

我们来看这个结果的一个有趣的应用.

推论 9.2.3. 设 $\{x_n\}$ 为实数列, 且对Lebesgue几乎所有的 $t \in [0, 1]$, $\exp\{itx_n\}$ 均收敛. 则 $\{x_n\}$ 收敛.

证明. 设 ξ 是服从 $[0, 1]$ 上均匀分布的随机变量. 令 $\xi_n := x_n \xi$. 任取两列 $m_k, n_k \rightarrow \infty$, 令 $\eta_k := (x_{m_k} - x_{n_k})\xi$. 则

$$\lim_{k \rightarrow \infty} E[\exp\{it\eta_k\}] = E[\lim_{k \rightarrow \infty} \exp\{it\eta_k\}] = 1.$$

因此 $\eta_k \xrightarrow{w} 0$. 于是由上一命题, $\eta_k \xrightarrow{P} 0$. 因此 $\lim_{k \rightarrow \infty} (x_{m_k} - x_{n_k}) = 0$. 由 $\{m_k\}, \{n_k\}$ 的任意性和Cauchy收敛准则, $\{x_n\}$ 收敛. □

当然这个结果可以直接证明, 不需要借助于概率. 事实上, 令

$$h(s) := \lim_{n \rightarrow \infty} e^{isx_n}.$$

则 $|h(s)| = 1$. 所以

$$\alpha(t) := \int_0^t h(s) ds$$

不可能恒为零. 选 t 使 $\alpha(t) \neq 0$. 则至少当 n 充分大时有 $\int_0^t e^{isx_n} ds \neq 0$, 且

$$x_n = \frac{e^{itx_n} - 1}{i \int_0^t e^{isx_n} ds} \rightarrow \frac{h(t) - 1}{i\alpha(t)}.$$

利用命题9.2.2可以证明:

命题 9.2.4 (Khinchin¹大数定律). 设 $\{\xi_n\}$ 是独立同分布随机变量列, 则 $\{\xi_n\}$ 服从大数定律的

¹Aleksandr Yakovlevich Khinchin, 1894-1959, 俄罗斯-苏联数学家.

充要条件是 $E\xi_1 < \infty$.

证明. 必要性显然. 下证充分性, 记 $E\xi_1 = \mu$. 要证

$$X_n := n^{-1} \sum_{k=1}^n \xi_k \xrightarrow{P} \mu.$$

设 ξ_i 们公共的特征函数为 f , 则 X_n 的特征函数为

$$f_n(t) = (f(n^{-1}t))^n.$$

因 ξ_i 期望存在和定理 8.3.3, $f(t)$ 连续可微. 因此 $f(t)$ 在零处的 Peano 型 Taylor 展开式为

$$f(t) = 1 + i\mu t + o(t), \quad t \rightarrow 0,$$

所以

$$\lim_{n \rightarrow \infty} f_n(t) = \lim_{n \rightarrow \infty} \left(1 + i\mu \frac{t}{n} + o\left(\frac{t}{n}\right) \right)^n = e^{i\mu t},$$

其中 $e^{i\mu t}$ 是单点分布 (取 μ 的概率为 1) 的特征函数. 这说明 X_n 特征函数收敛于常数 μ 的特征函数, 因此由命题 8.5.8, X_n 弱收敛于常数 μ . 再由命题 9.2.2, X_n 依概率收敛于 μ . \square

上述定理即是说,

$$\frac{S_n - E[S_n]}{n} \xrightarrow{P} 0.$$

其中 $S_n := \sum_{k=1}^n \xi_k$.

在数学分析里, 如果一个数列 a_n 满足

$$\frac{a_n}{n} \rightarrow 0,$$

那么就是说 a_n 是 n 的低阶无穷大. 这时人们往往被诱惑去考虑进一步的问题: $\alpha \in (0, 1)$ 为多少时, a_n 和 n^α 是等价无穷大?

所以现在我们也可以提出同样的问题: $\alpha \in (0, 1)$ 等于多少时, $S_n - E[S_n]$ 和 n^α 是等价无穷大? 也就是说,

$$\frac{S_n - E[S_n]}{n^\alpha} \xrightarrow{P} \xi?$$

其中 ξ 是随机变量, 且 $\xi \neq 0$ a.s..

但这个问题似乎太难, 现在也没有看到有什么结果, 且可能无法期待 α 不依赖于 ω . 不过, 如果在弱一点的意义下考虑这个问题, 即问是否有

$$\frac{S_n - E[S_n]}{n^\alpha} \xrightarrow{w} \xi?$$

就有大量的——也许我要说是海量的——结果. 关于这个问题、这类问题的研究在概率论的发展史上一直非常重要, 占有中心的地位. 因此, 这方面的结果就(被 Pólya² 带头)叫做中心极限定理. 今天研究这类问题的标准的也是极其强大有效的工具是特征函数, 其立命之基即特征函数与分布函数间的连续型定理. 发现这条路的是 Lyapunov³, 因此中心极限定理的代名词是 Lyapunov 定理.

²George Pólya, 1887-1985, 匈牙利-美国数学家

³Aleksandr Mikhailovich Lyapunov, 1857-1918, 俄罗斯数学家, 力学家

设 $\{\xi_n\}$ 是独立同分布随机变量列, $E[\xi_1] = 0$, $E[\xi_1^2] = 1$. 令

$$T_n := \frac{1}{\sqrt{n}} S_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n \xi_k.$$

则

$$E[T_n] = 0, E[T_n^2] = 1.$$

如果 ξ_1 的高阶矩存在, 如果你愿意, 你还可以多算几个 $E[T_n^k]$ (也许你至少可以算算 Bernoulli 分布的情形), 然后你会发现

$$\lim_{n \rightarrow \infty} E[T_n^k] = m_k,$$

其中 m_k 是标准正态分布的 k 阶矩.

这自然会——再次会——引起人们的联想:

是否 T_n 在某种意义下收敛到正态分布?

现在我们就介绍其中最早最经典的Laplace⁴定理.

定理 9.2.5 (Laplace中心极限定理). 设 ξ_1, ξ_2, \dots 是独立同分布随机变量, $E[\xi_1] = 0$, $E[\xi_1^2] = 1$. 则

$$T_n := \frac{S_n}{n^{1/2}} \xrightarrow{w} N(0, 1).$$

证明. 设 ξ_1 的特征函数为 f , 则 T_n 的特征函数为

$$f_n(t) = E[\exp\{itT_n\}] = \left(f\left(n^{-\frac{1}{2}}t\right)\right)^n.$$

由于 $E[\xi_1] = 0$, $E[\xi_1^2] = 1$, 所以

$$f\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{1}{2n}t^2 + o\left(\frac{1}{n}\right).$$

于是由命题11.5.1,

$$\lim_{n \rightarrow \infty} f_n\left(\frac{t}{\sqrt{n}}\right) = e^{-\frac{t^2}{2}}.$$

因此由逆连续定理, T_n 弱收敛于某一标准正态变量. □

引人注目的是, 不管诸 ξ_i 的分布如何, 只要满足一些矩条件, T_n 都弱收敛于同一个分布即标准正态分布. 也就是说, 当群体效应出现时, 个体的特征就消失了.

秦皇岛外打鱼船, 一片汪洋都不见, 都不见, 都不见.....

习题

1. 设 $\xi_n \sim P(n)$. 证明:

$$\frac{\xi_n - n}{n^{\frac{1}{2}}} \xrightarrow{w} N(0, 1).$$

⁴Pierre Simon de Laplace, 1749-1827, 法国数学家

2. 设 ξ_1, ξ_2, \dots 相互独立, $\xi_k \sim B(p_k)$, 且 $\sum_{k=1}^{\infty} p_k q_k = \infty$. 令

$$X_n := \frac{\sum_{k=1}^n \xi_k - \sum_{k=1}^n p_k}{\sqrt{\sum_{k=1}^n p_k q_k}}.$$

证明 $X_n \xrightarrow{w} N(0, 1)$.

3. 设 ξ_1, ξ_2, \dots 独立同分布, 且 $\xi_1 \sim N(0, 1)$. 令

$$\zeta_n := \sum_{k=1}^n \xi_k,$$

$$X_n := \sum_{k=1}^n \zeta_k.$$

证明:

$$\frac{X_n}{n^{\frac{3}{2}}} \xrightarrow{w} N(0, 1/3).$$

4. 设 ξ_n, η_n, ξ 均定义在同一个概率空间上, $\xi_n \xrightarrow{w} \xi$, $\eta_n \xrightarrow{w} c$, 其中 c 是常数. 设 $\varphi: \mathbb{R}^2 \mapsto \mathbb{R}$ 是连续函数, 且对任意 $M > 0$, 函数族

$$\{y \mapsto \varphi(x, y) : |x| \leq M\}$$

在 $y = c$ 处是等度连续的, 即 $\forall \varepsilon > 0, \exists \delta > 0$, 使得

$$|y - c| < \delta \iff \sup_{|x| \leq M} |\varphi(x, y) - \varphi(x, c)| < \varepsilon.$$

证明: $\varphi(\xi_n, \eta_n) \xrightarrow{w} \varphi(\xi, c)$. 特别地, 可取

$$\varphi(x, y) := xy, x + y.$$

并证明: 若 φ_1, φ_2 均满足题目的条件, 则 $\varphi_1 \varphi_2$ 也满足.

5. 设 $\{x_n\}$ 为实数列, ξ 为随机变量, 记 $\xi_n = x_n \xi$. 若 ξ_n 依概率收敛于零, 证明也 ξ_n 几乎必然收敛于零.

9.3 正态分布

我们曾经定义了 d -维正态分布 $N(\mu, \Sigma)$, 它是一种连续型分布. 其两个参数中, Σ 得是严格正定的. 在一维的时候, 这排除了单点分布的情形; 在多维的时候, 这排除了该分布集中在某个子空间上的情形, 这往往给正态分布的处理带来不少麻烦, 因为单点分布与集中在某个子空间上的正态分布都可以看作是退化的正态分布. 有可能以一种统一的方式定义非退化的与退化的正态分布吗?

退化到单点或者子空间上的正态分布首先不再是连续型分布, 因此不可能有密度函数. 所以如果能以统一的方式定义, 那么是 Σ 而不是 Σ^{-1} 就必须出现的定义中. 刚好, 正态分布的特征函数就有这样的特点. 所以特征函数是统一处理正态分布的不二工具选.

让我们更具体一点. 设 $\xi \sim N(\mu, \Sigma)$. 若 Σ 严格正定, 那么根据 7.5 节的结果, ξ 的特征函数是

$$f_\xi(t) = \exp \left\{ i\mu \cdot t - \frac{1}{2}t\Sigma t' \right\}$$

若 Σ 只是正定而非严格正定, 令

$$\Sigma_\varepsilon := \Sigma + \varepsilon I.$$

则 Σ_ε 严格正定, 因此

$$f_\varepsilon(t) := \exp \left\{ i\mu \cdot t - \frac{1}{2}t\Sigma_\varepsilon t' \right\}$$

是特征函数. 令 $\varepsilon \rightarrow 0$, 则

$$f_\varepsilon(t) \rightarrow f(t) := \exp \left\{ i\mu \cdot t - \frac{1}{2}t\Sigma t' \right\}.$$

由于 f 连续, 故由逆极限定理, f 是某个分布函数的(因此也是某个随机变量的)特征函数. 这样我们就有理由给出:

定义 9.3.1. 设 $\mu \in \mathbb{R}^d$, Σ 为 d 阶正定方阵. 则以

$$f(t) := \exp \left\{ i\mu \cdot t - \frac{1}{2}t\Sigma t' \right\}$$

为特征函数的随机变量 ξ 称为服从 d -维正态分布 $N(\mu, \Sigma)$.

正态分布有下面的性质:

命题 9.3.2. 若 ξ 服从正态分布, 则对任意 k , $E[|\xi|^k] < \infty$.

证明. 设 $\xi \sim N(\mu, \Sigma)$. 取 $\xi_n \sim N(\mu, \Sigma + n^{-1}I)$, 则 ξ_n 为非退化正态分布, 且 $\forall k$

$$\sup_n E[|\xi_n|^k] < \infty.$$

因此由 8.5 节习题 1, $E[|\xi|^k] < \infty$. □

命题 9.3.3. $\xi = (\xi_1, \dots, \xi_d)$ 服从 d -维正态分布的充要条件是 ξ_1, \dots, ξ_d 的任意线性组合均服从正态分布.

证明. 必要性: 设 $\xi \sim N(\mu, \Sigma)$. 则 $\forall (\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d$, $t \in \mathbb{R}$,

$$\begin{aligned} E \left[\exp \left\{ it \sum_{k=1}^d \lambda_k \xi_k \right\} \right] &= E \left[\exp \left\{ i \sum_{k=1}^d t \lambda_k \xi_k \right\} \right] \\ &= \exp \left\{ it \sum_{k=1}^d \lambda_k \mu_k - \frac{t^2}{2} \sum_{k,j=1}^n \sigma_{kj} \lambda_k \lambda_j \right\}. \end{aligned}$$

所以

$$\lambda \cdot \xi \sim N(\lambda \cdot \mu, \lambda \Sigma \lambda').$$

充分性: $\forall t \in \mathbb{R}^d$,

$$\begin{aligned} E[\exp\{it \cdot \xi\}] &= \exp\left\{iE[(t \cdot \xi)] - \frac{1}{2}D[(t \cdot \xi)]\right\} \\ &= \exp\left\{it \cdot E[\xi] - \frac{1}{2}t \text{Cov}(\xi)t'\right\}. \end{aligned}$$

所以 $\xi \sim N(E[\xi], \text{Cov}[\xi])$. □

习题

1. 证明: 若 $\xi \sim N(\mu, \Sigma)$, $\Sigma = (\sigma_{kj})$, 则 $\sigma_{kj} = \text{Cov}(\xi_k, \xi_j)$.
2. 证明: 若 $\xi = (\xi_1, \dots, \xi_n)$ 服从正态分布, 则 $\xi_1, \xi_2, \dots, \xi_n$ 独立的充要条件是它们不相关.
3. 设 ξ, η 独立, 且都服从标准正态分布.

(a) 证明 $\xi + \eta$ 与 $\xi - \eta$ 独立且均服从正态分布.

(b) 证明

$$\frac{\xi + \eta}{\xi - \eta}$$

为连续型随机变量且密度函数为 $\frac{1}{\pi(1+x^2)}$.

(c) 证明

$$\frac{2\xi\eta}{\sqrt{\xi^2 + \eta^2}}, \quad \frac{\xi^2 - \eta^2}{\sqrt{\xi^2 + \eta^2}}$$

为相互独立的正态随机变量.

10 中心极限定理中的余项估计

本章恒设 ξ_1, ξ_2, \dots 独立同分布, 且

$$E[\xi_1] = 0, \quad E[\xi_1^2] = 1.$$

令

$$S_n := \sum_{k=1}^n \xi_k.$$

则由中心极限定理, 对任意 $\varphi \in C_b$, 有

$$\lim_{n \rightarrow \infty} E[\varphi(n^{-\frac{1}{2}} S_n)] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \varphi(x) e^{-\frac{x^2}{2}} dx.$$

或者, 用分布函数的语言, 就是

$$\lim_{n \rightarrow \infty} F_n(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{x^2}{2}} dx \quad \forall y,$$

其中 F_n 是 $n^{-\frac{1}{2}} S_n$ 的分布函数. 但无论哪一种表述, 我们都没有得到收敛速度的定量估计. 本章我们就做出这样一些估计.

10.1 Lindeberg定理

我们曾经使用特征函数证明了中心极限定理. 本节我们将不使用特征函数而证明类似的结果. 为突出思想, 我们将只考虑独立同分布情形, 但实际上将定理的叙述和证明方法略加修改后, 本节的结果可推广到独立但不同分布的情形. 此外, 同样是为了突出思想, 同时也是为了得到简洁的收敛速度, 我们将假定三阶矩存在, 记 $\tau := E[|\xi_1|^3]$, 并假设 $\varphi \in C^3(\mathbb{R}, \mathbb{R})$ 且

$$\sup_{x \in \mathbb{R}} \frac{|\varphi(x)| + |\varphi'(x)| + |\varphi''(x)|}{1 + |x|^3} < \infty,$$

$$C := \sup_{x \in \mathbb{R}} |\varphi'''(x)| < \infty.$$

这些条件在更细致的分析下可以减弱, 但我们不展开讨论了.

我们将证明下面的:

定理 10.1.1 (Lindeberg¹定理). 设 $\tau := E[|\xi_1|^3] < \infty$, 则

$$\left| E[\varphi(n^{-\frac{1}{2}} S_n)] - \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \varphi(x) e^{-\frac{x^2}{2}} dx \right| \leq \frac{1}{3} \left(\sqrt{\frac{2}{\pi}} + \frac{1}{2} \tau \right) C n^{-\frac{1}{2}}.$$

¹Jarl Waldemar Lindeberg, 1876-1932, 芬兰数学家.

证明. 取 $\eta_1, \dots, \eta_n \sim N(0, 1)$, η_1, \dots, η_n 独立, 且 $\{\eta_1, \dots, \eta_n\}$ 与 $\{\xi_1, \dots, \xi_n\}$ 独立. 令

$$Z := n^{-\frac{1}{2}} \sum_{k=1}^n \eta_k.$$

则 $Z \sim N(0, 1)$. 因此

$$E[\varphi(n^{-\frac{1}{2}} S_n)] - E[\varphi(Z)] = E[\varphi(n^{-\frac{1}{2}} S_n)] - E[\varphi(n^{-\frac{1}{2}}(\eta_1 + \dots + \eta_n))].$$

为估计右边, 我们采用标准的逐项替换法. 这里的关键是用到了 Z 与 η_1 服从同一个分布, 因此正态分布是该分布的唯一选项, 见习题1.

现在我们就具体实施证明. 令

$$\zeta_j := n^{-\frac{1}{2}} \left(\sum_{k=1}^{n-j} \xi_k + \sum_{k=n-j+2}^n \eta_k \right),$$

其中规定

$$\sum_{k=1}^0 = \sum_{k=n+1}^n = 0.$$

则

$$\begin{aligned} I &:= E[\varphi(n^{-\frac{1}{2}} S_n)] - E[\varphi(Z)] \\ &= \sum_{j=1}^n \left\{ E[\varphi(\zeta_j + n^{-\frac{1}{2}} \xi_{n-j+1})] - E[\varphi(\zeta_j + n^{-\frac{1}{2}} \eta_{n-j+1})] \right\} \\ &= \sum_{j=1}^n \left\{ \left[E[\varphi(\zeta_j + n^{-\frac{1}{2}} \xi_{n-j+1})] - \varphi(\zeta_j) \right] - \left[E[\varphi(\zeta_j + n^{-\frac{1}{2}} \eta_{n-j+1})] - \varphi(\zeta_j) \right] \right\} \\ &=: \sum_{j=1}^n I_j. \end{aligned}$$

由Taylor公式,

$$\varphi(\zeta_j + y) - \varphi(\zeta_j) = \varphi'(\zeta_j)y + \frac{1}{2}\varphi''(\zeta_j)y^2 + r,$$

其中

$$|r| \leq \frac{1}{6}C|y|^3.$$

由独立性,

$$\begin{aligned} E[\xi_{n-j+1}\varphi'(\zeta_j)] &= E[\eta_{n-j+1}\varphi'(\zeta_j)] = 0, \\ E[\xi_{n-j+1}^2\varphi''(\zeta_j)] &= E[\eta_{n-j+1}^2\varphi''(\zeta_j)] = E[\varphi''(\zeta_j)]. \end{aligned}$$

所以

$$|I_j| \leq I_j^1 + I_j^2,$$

其中

$$I_j^1 \leq \frac{1}{6}CE[|\xi_1|^3]n^{-\frac{3}{2}},$$

$$I_j^2 \leq \frac{1}{6} CE[|\eta_1|^3] n^{-\frac{3}{2}} = \frac{1}{3} \sqrt{\frac{2}{\pi}} C n^{-\frac{3}{2}}.$$

由此立即得到要证的不等式. □

习题

1. 设 $E[\xi] = 0$, $E[\xi^2] = 1$, 且存在 $n > 1$, 使得 ξ_1, \dots, ξ_n 独立且均与 ξ 同分布, 而 $n^{-\frac{1}{2}}(\xi_1 + \dots + \xi_n)$ 也与 ξ 同分布. 证明 ξ 服从标准正态分布.
2. 设 ξ_1, ξ_2, \dots 独立同分布, $E[\xi_1] = 0$, $E[\xi_1^2] = 1$, $E[|\xi_1|^3] < \infty$. 设 $\varphi \in C(\mathbb{R}, \mathbb{R})$ 且存在常数 C 使得

$$|\varphi(x)| \leq C(1 + |x|^3), \quad \forall x.$$

(a) 证明存在 $\rho \in C^\infty(\mathbb{R}, \mathbb{R})$, $\rho \geq 0$, 且 $\rho(x) = 0, \forall |x| \geq 1$.

(b) 设 k 为正整数. 令

$$\varphi_k(x) = k \int_{-\infty}^{\infty} \rho(k(x-y)) \varphi(y) dy.$$

证明: 存在不依赖于 k 的常数 C 使得

$$|\varphi_k(x) - \varphi(x)| \leq C(1 + |x|^3).$$

(c) 令

$$\psi_R(x) := (1 + |x|^3) \alpha(xR^{-1}),$$

其中 $\alpha \in C_b^\infty(\mathbb{R}, [0, 1])$, $\alpha(x) = 0, \forall |x| \leq \frac{1}{2}$, $\alpha(x) = 1, \forall |x| \geq 1$. 证明:

$$\lim_{n \rightarrow \infty} E[\psi_R(n^{-\frac{1}{2}} S_n)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi_R(x) e^{-\frac{x^2}{2}} dx.$$

(d) 证明

$$\limsup_{n \rightarrow \infty} E[(1 + |n^{-\frac{1}{2}} S_n|^3)] = 0.$$

(e) 证明

$$E[(\varphi - \varphi_k)(n^{-\frac{1}{2}} S_n)]$$

(f) 证明

$$\lim_{k \rightarrow \infty} \int_{-\infty}^{\infty} |(\varphi - \varphi_k)(x)| e^{-\frac{x^2}{2}} dx \leq C \int_{|x| \geq R} (1 + x^2) e^{-\frac{x^2}{2}} dx.$$

10.2 几个分析引理

前节所介绍的Lindeberg定理给出了 $\varphi(n^{-\frac{1}{2}} S_n)$ 与 $\frac{1}{\sqrt{2\pi}} \int \varphi(x) e^{-\frac{x^2}{2}} dx$ 的差距的估计, 但没有给出 S_n 的分布函数和正态分布函数的差距的估计. 本章余下的篇幅就来估计这后一种估计. 本节做些准备工作, 证明几个分析方面的结果.

令

$$\gamma(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

是标准正态的密度函数.

引理 10.2.1. $\forall x > 0$,

$$\frac{x}{1+x^2}\gamma(x) \leq \int_x^\infty \gamma(y)dy \leq \frac{1}{x}\gamma(x).$$

证明. 由分部积分有

$$\int_x^\infty e^{-\frac{y^2}{2}} dy = \frac{1}{x}e^{-\frac{x^2}{2}} - \int_x^\infty \frac{1}{y^2}e^{-\frac{y^2}{2}} dy.$$

所以

$$\int_x^\infty e^{-\frac{y^2}{2}} dy \leq \frac{1}{x}e^{-\frac{x^2}{2}}.$$

由此得到第二个不等式.

同时又有

$$\begin{aligned} \int_x^\infty e^{-\frac{y^2}{2}} dy &= \frac{1}{x}e^{-\frac{x^2}{2}} - \int_x^\infty \frac{1}{y^2}e^{-\frac{y^2}{2}} dy \\ &\geq \frac{1}{x}e^{-\frac{x^2}{2}} - \frac{1}{x^2} \int_x^\infty e^{-\frac{y^2}{2}} dy, \end{aligned}$$

移项即得

$$\frac{1+x^2}{x^2} \int_x^\infty e^{-\frac{y^2}{2}} dy \geq \frac{1}{x}e^{-\frac{x^2}{2}},$$

这就是第一个不等式. □

推论 10.2.2. 设 $\xi \sim N(0, 1)$. 则 $\forall x > 0$, $E[\xi 1_{\xi > x}] = \gamma(x)$, 且 $P(\xi > x) \leq \sqrt{\frac{\pi}{2}}\gamma(x)$.

证明. 第一个等式直接计算可得, 即

$$E[\xi 1_{\xi > x}] = \int_x^{+\infty} y\gamma(y)dy = \gamma(x).$$

为证明第二个不等式, 令

$$f(x) := \gamma^{-1}(x) \int_x^{+\infty} \gamma(y)dy, x \geq 0,$$

则由引理10.2.1的第二个不等式, $\forall x \geq 0$,

$$f'(x) = x\gamma^{-1}(x) \int_x^{+\infty} \gamma(y)dy - 1 \leq 0.$$

因此 $f(x) \leq f(0) = \sqrt{\frac{\pi}{2}}$, $\forall x \geq 0$. □

下面的结果给了 γ 一个刻画.

引理 10.2.3. 设 f 二次连续可微, 且

$$\lim_{x \rightarrow \infty} f(x)\gamma(x) = 0, \int_{-\infty}^{\infty} |f(x)|\gamma(x)dx < \infty,$$

则

$$\int_{-\infty}^{\infty} xf(x)\gamma(x)dx = \int_{-\infty}^{\infty} f'(x)\gamma(x)dx.$$

反之, 若 g 为 \mathbb{R} 上的非负 Borel 函数, 满足

$$\int_{-\infty}^{\infty} g(x)dx = 1, \quad \int_{-\infty}^{\infty} |x|g(x)dx < \infty$$

且对任意 $f \in C_b^\infty(\mathbb{R})$ 有

$$\int_{-\infty}^{\infty} xf(x)g(x)dx = \int_{-\infty}^{\infty} f'(x)g(x)dx,$$

则 $g \equiv \gamma$.

证明. 第一部分是一个简单的分部积分. 至于第二部分, 设 ξ 是具有密度函数 g 的随机变量. 取 $f(x) = e^{itx}$, 则由定理5.7.3和假设, 有

$$\begin{aligned} \frac{d}{dt}E[e^{it\xi}] &= iE[\xi e^{it\xi}] \\ &= i \int_{-\infty}^{\infty} xf(x)g(x)dx \\ &= -t \int_{-\infty}^{\infty} f(x)g(x)dx. \end{aligned}$$

即 ξ 的特征函数 φ 满足微分方程

$$\varphi'(t) = -t\varphi(t), \quad \varphi(0) = 1.$$

因此

$$\varphi(t) = e^{-\frac{t^2}{2}}.$$

所以 $\xi \sim N(0, 1)$, 从而 $g \equiv \gamma$. □

引理 10.2.4. 设 h 是连续函数, h 几乎处处可导, 存在常数 A 使得 $|h'(x)| \leq A$ a.e., 且

$$h(x) - h(y) = \int_y^x h'(t)dt, \quad \forall x, y. \quad (2.1)$$

设 $\xi \sim N(0, 1)$. 令

$$c_0 := E[h(\xi)],$$

$$f(x) := \gamma^{-1}(x)E[(h(\xi) - c_0)\mathbf{1}_{\xi \leq x}], \quad (2.2)$$

则

$$f'(x) - xf(x) = h(x) - c_0, \quad (2.3)$$

$$|f(x)| \leq 2A, \quad |f'(x)| \leq 4A, \quad |f''(x)| \leq 2A. \quad (2.4)$$

注. 这里由于只假定 h 是连续函数, h' 几乎处处存在, 所以(2.2)中的积分及下面的许多积分应理解为Lebesgue积分. 同样是由于 h' 只是Borel可测, 并不一定连续, 所以 f 并不见得是处处可导的, 而只是几乎处处可导(在Lebesgue测度的意义下). 因此上面涉及到 f' 与 f'' 的不等式也只是几乎处处成立. 这里大家在现阶段可以把 h' 当成连续的, 就先这么混着, 等熟悉实变函数以后, 就会知道这里的一切和下面证明中所有的推理都是严格的, 都是正确的.

证明. 对(2.2)求导即得(2.3).

显然用 $\tilde{h}(t) := h(t) - h(0)$ 代替 h 而定义出来的 f 与用 h 本身定义出来的 f 是一样的, 所以必要时以 \tilde{h} 代替 h , 我们不妨假定 $h(0) = 0$. 这样就有

$$|h(x)| \leq \int_0^{|x|} |h'(t)| dt \leq A|x|.$$

因此

$$|c_0| = |E[h(\xi)]| \leq AE[|\xi|] = \sqrt{\frac{2}{\pi}}A.$$

又由于

$$E[h(\xi) - c_0] = 0,$$

所以

$$f(x) = -\gamma^{-1}(x)E[(h(\xi) - c_0)1_{\xi > x}]. \quad (2.5)$$

对 $x < 0$ 用(2.2), 对 $x \geq 0$ 用(2.5), 由推论10.2.2得

$$|f(x)| \leq \gamma^{-1}(x)E\left[A\left(|\xi| + \sqrt{\frac{2}{\pi}}\right)1_{\xi > |x|}\right] \leq 2A.$$

对(2.3)求导即得

$$f''(x) = xf'(x) + f(x) + h'(x), \quad (2.6)$$

因此

$$(\gamma(x)f'(x))' = \gamma(x)(f''(x) - xf'(x)) = \gamma(x)(f(x) + h'(x)). \quad (2.7)$$

又由(2.3)知

$$|f'(x)| \leq 3A|x| + c_0,$$

因此

$$\lim_{x \rightarrow \infty} \gamma(x)f'(x) = 0.$$

对(2.7)两边在区间 $(-\infty, x]$ 上积分得

$$f'(x) = \gamma^{-1}(x)E[(f(\xi) + h'(\xi))1_{\xi \leq x}],$$

又对(2.7)两边在区间 $(-\infty, +\infty)$ 上积分得

$$E[f(\xi) + h'(\xi)] = 0.$$

所以又有

$$f'(x) = -\gamma^{-1}(x)E[(f(\xi) + h'(\xi))1_{\xi > x}].$$

因此, 由推论10.2.2有

$$\begin{aligned} |f'(x)| &\leq \gamma^{-1}(x)E[(|f(\xi)| + |h'(\xi)|)1_{\xi > x}] \\ &\leq (2A + A)\gamma^{-1}(x)P(\xi > x) \\ &\leq 3A\sqrt{\frac{\pi}{2}} \leq 4A. \end{aligned}$$

下面估计 f'' . 将(2.3)代入(2.6), 我们有

$$\begin{aligned} |f''(x)| &= |f(x) + xf'(x) + h'(x)| \\ &= |(1 + x^2)f(x) + x(h(x) - c_0) + h'(x)|. \end{aligned}$$

然后注意到

$$\begin{aligned} h(x) - c_0 &= h(x) - \int_{-\infty}^{\infty} h(s)\gamma(s)ds \\ &= \int_{-\infty}^{\infty} (h(x) - h(s))\gamma(s)ds \\ &= \int_{-\infty}^x \gamma(s)ds \int_s^x h'(t)dt - \int_x^{\infty} \gamma(s)ds \int_x^s h'(t)dt \\ &= \int_{-\infty}^x h'(t) \int_{-\infty}^t \gamma(s)ds - \int_x^{\infty} h'(t) \int_t^{\infty} \gamma(s)ds \\ &= \int_{-\infty}^x h'(t)P(\xi \leq t)dt - \int_x^{\infty} h'(t)P(\xi > t)dt. \end{aligned}$$

类似地有

$$P(\xi \leq x)h(x) - E[h(\xi)1_{\xi \leq x}] = \int_{-\infty}^x h'(t)P(\xi \leq t)dt,$$

$$P(\xi > x)h(x) - E[h(\xi)1_{\xi > x}] = - \int_x^{\infty} h'(t)P(\xi > t)dt.$$

注意到 $E[h(\xi)1_{\xi \leq x}] = c_0 - E[h(\xi)1_{\xi > x}]$, 前一式乘以 $P(\xi > x)$, 后一式乘以 $P(\xi \leq x)$, 然后两式相减, 即得

$$E[(h(\xi) - c_0)1_{\xi > x}] = P(\xi > x) \int_{-\infty}^x h'(t)P(\xi \leq t)dt + P(\xi \leq x) \int_x^{\infty} h'(t)P(\xi > t)dt.$$

由(2.5), 所以

$$f(x) = -\gamma^{-1}(x) \left(P(\xi > x) \int_{-\infty}^x h'(t)P(\xi \leq t)dt + P(\xi \leq x) \int_x^{\infty} h'(t)P(\xi > t)dt \right).$$

此式代入到 f'' 的表达式, 并注意到引理10.2.1的第一个不等式, 有

$$\begin{aligned}
 |f''(x)| &\leq |h'(x)| + |(1+x^2)f(x) + x(h(x) - c_0)| \\
 &\leq A + A \left| x - (1+x^2)P(\xi > x)\gamma^{-1}(x) \right| \cdot \int_{-\infty}^x P(\xi \leq t) dt \\
 &\quad + A \left| (1+x^2)P(\xi \leq x)\gamma^{-1}(x) + x \right| \cdot \int_x^{\infty} P(\xi > t) dt \\
 &= A + A \left(-x + (1+x^2)P(\xi > x)\gamma^{-1}(x) \right) \left(xP(\xi \leq x) + \gamma(x) \right) \\
 &\quad + A \left(x + (1+x^2)P(\xi \leq x)\gamma^{-1}(x) \right) \left(-xP(\xi > x) + \gamma(x) \right) \\
 &= 2A,
 \end{aligned}$$

其中倒数第二个等式是因为由Fubini定理有

$$\begin{aligned}
 \int_{-\infty}^x P(\xi \leq t) dt &= xP(\xi \leq x) + \gamma(x), \\
 \int_x^{\infty} P(\xi > t) dt &= -xP(\xi > x) + \gamma(x).
 \end{aligned}$$

□

10.3 分布余项的积分估计

以 F 表 $N(0, 1)$ 的分布函数, 即

$$F(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy.$$

所谓分布余项的积分估计是指估计

$$\int_{-\infty}^{\infty} |F_n(x) - F(x)| dx.$$

为此我们先证明一个引理.

引理 10.3.1. 设 ξ_1, ξ_2 是可积随机变量, G_1, G_2 分别是其分布函数. 令

$$\varphi(x) := \int_0^x \operatorname{sgn}(G_1(y) - G_2(y)) dy.$$

则

$$\int_{-\infty}^{\infty} |G_1(x) - G_2(x)| dx = E[\varphi(\xi_2) - \varphi(\xi_1)]. \quad (3.8)$$

证明. (1) 先设 ξ_1 与 ξ_2 均有界, 比如 $|\xi_1| \vee |\xi_2| \leq n$, 其中 n 为常数. 此时, 由于当 $|x| > n$ 时 $G_1(x) - G_2(x) = 0$, 因此 $|\varphi(x)| \leq n$.

设 ρ 为 \mathbb{R} 上的磨光函数(见附录11.4). 对 $\varepsilon > 0$, 令 $\rho(x) := \varepsilon^{-1} \rho(\varepsilon^{-1}x)$,

$$\varphi_\varepsilon(x) := \int_{\mathbb{R}} \varphi(y) \rho_\varepsilon(x-y) dy.$$

则

(i)

$$\varphi_\varepsilon \in C_b^1, \quad \sup_{x \in \mathbb{R}} |\varphi'_\varepsilon(x)| \leq 1;$$

(ii) 当 $\varepsilon \downarrow 0$ 时,

$$\varphi_\varepsilon(x) \rightarrow \varphi(x) \quad \forall x, \quad \varphi'_\varepsilon(x) \rightarrow \varphi'(x) \quad \text{a.e. Lebesgue 测度.}$$

由 5.8 节习题 8, 我们有

$$\int_{-n}^n \varphi'_\varepsilon(x)(G_1(x) - G_2(x)) dx = E[\varphi_\varepsilon(\xi_2) - \varphi_\varepsilon(\xi_1)].$$

令 $\varepsilon \downarrow 0$, 由有界收敛定理有

$$\int_{-n}^n \varphi'(x)(G_1(x) - G_2(x)) dx = E[\varphi(\xi_2) - \varphi(\xi_1)].$$

即

$$\int_{-n}^n |G_1(x) - G_2(x)| dx = E[\varphi(\xi_2) - \varphi(\xi_1)].$$

(2) 一般地, 令 $\xi_i^n := (-n) \vee \xi_i \wedge n, i = 1, 2$. 则由上一步有

$$\int_{-n}^n |G_1(x) - G_2(x)| dx = E[\varphi(\xi_2^n) - \varphi(\xi_1^n)].$$

令 $n \rightarrow \infty$, 左边用单调收敛定理, 右边用控制收敛定理(注意 $|\varphi(x)| \leq |x|$), 得

$$\int_{\mathbb{R}} |G_1(x) - G_2(x)| dx = E[\varphi(\xi_2) - \varphi(\xi_1)].$$

□

所以为估计(3.8)的左边, 只要估计其右边就行. 设 f 由(2.2)给出, 其中 h 代之以 φ , 此时 $A = 1$. 则由(2.3)有

$$\varphi(x) - \varphi(y) = f'(x) - xf(x) - [f'(y) - yf(y)].$$

因此

$$E[\varphi(\xi_2) - \varphi(\xi_1)] = E[f'(\xi_1) - \xi_1 f(\xi_1) - [f'(\xi_2) - \xi_2 f(\xi_2)]].$$

现在, 取 $\xi_1 \sim N(0, 1)$, $\xi_2 := n^{-\frac{1}{2}} S_n$. 由引理 2.2 知

$$E[f'(\xi_1) - \xi_1 f(\xi_1)] = 0,$$

所以

$$E[\varphi(n^{-\frac{1}{2}} S_n)] - E[\varphi(\xi_1)] = -E[f'(n^{-\frac{1}{2}} S_n) - n^{-\frac{1}{2}} S_n f(n^{-\frac{1}{2}} S_n)].$$

下面我们就来估计该式右边. 我们用 I 表示它. 于是

$$I = \sum_{m=1}^n \left(n^{-1} E[f'(n^{-\frac{1}{2}} S_n)] - n^{-\frac{1}{2}} E[\xi_m f(n^{-\frac{1}{2}} S_n)] \right).$$

令

$$\begin{aligned}\eta_m(t) &:= n^{-\frac{1}{2}}[S_n + (t-1)\xi_m], \\ g(t) &:= f(\eta_m(t)).\end{aligned}$$

用公式

$$g(1) = g(0) + g'(0) + \int_0^1 (g'(t) - g'(0))dt,$$

并注意 ξ_m 与 $S_n - \xi_m$ 即 $\eta_m(0)$ 独立, 从而

$$E[\xi_m g(0)] = E[\xi_m]E[g(0)] = 0,$$

$$E[\xi_m g'(0)] = n^{-\frac{1}{2}}E[\xi_m^2 f'(\eta_m(0))] = n^{-\frac{1}{2}}E[\xi_m^2]E[f'(\eta_m(0))] = n^{-\frac{1}{2}}E[f'(\eta_m(0))],$$

我们有

$$I = \frac{1}{n} \sum_{m=1}^n A_m - \frac{1}{n} \int_0^1 \sum_{m=1}^n B_m(t)dt,$$

其中

$$A_m := E[f'(\eta_m(1)) - f'(\eta_m(0))],$$

$$B_m(t) = E[\xi_m^2 [f'(\eta_m(t)) - f'(\eta_m(0))]].$$

由(2.4)和Hölder不等式有

$$\begin{aligned}|A_m| &\leq E[|f'(n^{-\frac{1}{2}}S_n) - f'(n^{-\frac{1}{2}}(S_n - \xi_m))|] \\ &\leq n^{-\frac{1}{2}}\|f''\|E[|\xi_m|] \leq 2n^{-\frac{1}{2}}.\end{aligned}$$

对 B_m 则有

$$\begin{aligned}|B_m(t)| &\leq E[\xi_m^2 |f'(\eta_m(t)) - f'(\eta_m(0))|] \\ &\leq n^{-\frac{1}{2}}t\|f''\|E[|\xi_m|^3] \leq n^{-\frac{1}{2}}2\tau t,\end{aligned}$$

其中 $\tau := E[|\xi_1|^3]$. 所以

$$\int_0^1 B_m(t)dt \leq n^{-\frac{1}{2}}\tau.$$

总结起来, 我们便证明了:

定理 10.3.2. 设 $\tau := E[|\xi_1|^3] < \infty$, 则

$$\int_{\mathbb{R}} |F_n(x) - F(x)|dx \leq n^{-\frac{1}{2}}(2 + \tau).$$

习题

1. 证明:

$$\int_{\mathbb{R}} \left| e^{-n} \sum_{k=0}^{[n\frac{1}{2}x+n]} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy \right| dx \leq 7n^{-\frac{1}{2}}.$$

2. 证明:

$$\int_{\mathbb{R}} \left| \frac{1}{\Gamma(\frac{n}{2})} \sqrt{\left(\frac{n}{2}\right)^n} \int_0^{1+x\sqrt{\frac{2}{n}}} y^{\frac{n}{2}-1} e^{-\frac{ny}{2}} dy - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy \right| dx \leq 17n^{-\frac{1}{4}}.$$

10.4 余项的一致估计

上节我们取特殊的 φ 得到了余项的积分估计. 本节我们说明取另一种 φ 可以得到余项的一致估计.

现在, 沿用上节的记号, 我们来证明:

定理 10.4.1.

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq 2\sqrt{2 + \tau} n^{-\frac{1}{4}}.$$

证明. $\forall \varepsilon > 0, x \in \mathbb{R}$, 取

$$\varphi(y) := \varphi_{\varepsilon, x}(y) := 1_{(-\infty, x]}(y) + \frac{1}{\varepsilon}(x + \varepsilon - y)1_{(x, x+\varepsilon]}(y).$$

因为

$$P(n^{-\frac{1}{2}}S_n \leq x) \leq E[\varphi(n^{-\frac{1}{2}}S_n)],$$

所以

$$\begin{aligned} & P(n^{-\frac{1}{2}}S_n \leq x) - P(\xi \leq x) \\ & \leq E[\varphi(n^{-\frac{1}{2}}S_n)] - P(\xi \leq x) \\ & = E[\varphi(n^{-\frac{1}{2}}S_n)] - E[\varphi(\xi)] + E[\varphi(\xi)] - P(\xi \leq x) \\ & \leq \left| E[\varphi(n^{-\frac{1}{2}}S_n)] - E[\varphi(\xi)] \right| + P(x \leq \xi \leq x + \varepsilon). \end{aligned}$$

因为

$$P(x \leq \xi \leq x + \varepsilon) = \int_x^{x+\varepsilon} \gamma(y) dy \leq \varepsilon,$$

由5.8节的习题8和定理10.3.2有

$$\begin{aligned} \left| E[\varphi(n^{-\frac{1}{2}}S_n)] - E[\varphi(\xi)] \right| &= \left| \int_{\mathbb{R}} \varphi'(y)(F_n(y) - F(y)) dy \right| \\ &\leq \varepsilon^{-1} \int_{\mathbb{R}} |F_n(y) - F(y)| dy \\ &\leq \varepsilon^{-1} n^{-\frac{1}{2}}(2 + \tau). \end{aligned}$$

所以

$$\sup_{x \in \mathbb{R}} \left(P(n^{-\frac{1}{2}}S_n \leq x) - P(\xi \leq x) \right) \leq \varepsilon^{-1} n^{-\frac{1}{2}}(2 + \tau) + \varepsilon.$$

同理

$$\begin{aligned} & P(\xi \leq x) - P(n^{-\frac{1}{2}}S_n \leq x) \\ & \leq P(\xi \leq x) - E[\varphi(\xi + \varepsilon)] + E[\varphi(\xi + \varepsilon)] - E[\varphi(n^{-\frac{1}{2}}S_n + \varepsilon)] \\ & = |E[\varphi(n^{-\frac{1}{2}}S_n + \varepsilon)] - E[\varphi(\xi + \varepsilon)]| + P(x \leq \xi \leq x + \varepsilon) \\ & \leq \varepsilon^{-1} n^{-\frac{1}{2}}(2 + \tau) + \varepsilon. \end{aligned}$$

于是

$$\sup_{x \in \mathbb{R}} \left| P(n^{-\frac{1}{2}} S_n \leq x) - P(\xi \leq x) \right| \leq \varepsilon^{-1} n^{-\frac{1}{2}} (2 + \tau) + \varepsilon.$$

因此,

$$\begin{aligned} \sup_{x \in \mathbb{R}} \left| P(n^{-\frac{1}{2}} S_n \leq x) - P(\xi \leq x) \right| &\leq \inf_{\varepsilon > 0} \left\{ \varepsilon^{-1} n^{-\frac{1}{2}} (2 + \tau) + \varepsilon \right\} \\ &= 2\sqrt{2 + \tau} n^{-\frac{1}{4}}. \end{aligned}$$

□

最后我们着重指出, 本章的各个结果只要求 ξ_1, ξ_2, \dots 独立同分布, 再加上一些矩条件, 而与具体是何种分布无关. 这种无关性通常被称为“不变性原理”, 表示不管 ξ_i 的分布如何变化, 最后的极限分布都是标准正态分布, 恒古不变.

习题

1. 证明:

$$\sup_{x \in \mathbb{R}} \left| e^{-n} \sum_{k=0}^{[n^{\frac{1}{2}} x + n]} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy \right| \leq 2\sqrt{7} n^{-\frac{1}{4}}.$$

2. 证明:

$$\sup_{x \in \mathbb{R}} \left| \frac{1}{\Gamma(\frac{n}{2})} \sqrt{\left(\frac{n}{2}\right)^n} \int_0^{1+x\sqrt{\frac{2}{n}}} y^{\frac{n}{2}-1} e^{-ny} dy - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy \right| \leq 2\sqrt{17} n^{-\frac{1}{4}}.$$

3. 设 ξ_1, \dots, ξ_{100} 独立, 且均服从 $B(p)$. 分别用 Poisson 定理和本节结果近似计算 $P(\xi_1 + \dots + \xi_n \leq k)$. 比较两种方法的优劣.

11 附录

11.1 Stirling公式

$$\lim_{n \rightarrow \infty} \frac{n!}{e^{-n} n^{n+1/2}} = \sqrt{2\pi}.$$

证明可见例如[5, p.382]. 下面给一个概率方法的证明.

证明. 设 ξ_1, ξ_2, \dots 独立同分布, 分布为 $E(1)$. 令 $\eta_k = \xi_k - 1$, 则 η_1, η_2, \dots 也独立同分布, 且

$$E[\eta_k] = 0, \quad E[\eta_k^2] = 1.$$

令

$$S_n := \sum_{k=1}^n \eta_k = \sum_{k=1}^n \xi_k - n,$$

由中心极限定理, $n^{-\frac{1}{2}} S_n \xrightarrow{w} N(0, 1)$.

由Skorokhod表现定理, 存在概率空间 (Ω, \mathcal{F}, P) 及定义在其上的随机变量 X_n, X , 使得 $n^{-\frac{1}{2}} S_n$ 与 X_n 同分布, $X \sim N(0, 1)$, 且 $X_n \xrightarrow{a.s.} X$. 又因为

$$E[X_n^2] = E[n^{-1} S_n^2] = 1,$$

所以

$$E[(X_n - X)^2] \leq 2(E[X_n^2] + E[X^2]) = 4,$$

因此,

$$\begin{aligned} E[|X_n - X|] &\leq E[1_{\{|X_n - X| > \varepsilon\}} |X_n - X|] + E[1_{\{|X_n - X| \leq \varepsilon\}} |X_n - X|] \\ &\leq 2\sqrt{P(|X_n - X| > \varepsilon)} + \varepsilon. \end{aligned}$$

令 $n \rightarrow \infty$, 并由 ε 的任意性, 有 $X_n \xrightarrow{L^1} X$. 特别地,

$$\lim_{n \rightarrow \infty} E[|X_n|] = E[|X|] = \sqrt{\frac{2}{\pi}}.$$

另外, 用归纳法易知, $\sigma_n := \sum_{k=1}^n \xi_k$ 的密度函数为 (见第6章例2)

$$\rho_n(x) = \begin{cases} \frac{x^{n-1}}{(n-1)!} e^{-x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

直接计算

$$\begin{aligned}
 E[|S_n|] &= E[|\sigma_n - n|] \\
 &= \int_0^\infty |x - n| \frac{x^{n-1}}{(n-1)!} e^{-x} dx \\
 &= \frac{n}{(n-1)!} \int_0^\infty \left| \frac{x}{n} - 1 \right| x^{n-1} e^{-x} dx \\
 &= \frac{n}{(n-1)!} n^n \int_0^\infty |u - 1| u^{n-1} e^{-nx} dx \\
 &= \frac{n}{(n-1)!} n^n \left(\int_0^1 (1-u) u^{n-1} e^{-nx} dx + \int_1^\infty (u-1) u^{n-1} e^{-nx} dx \right) \\
 &= \frac{1}{(n-1)!} n^n \left(u^n e^{-nx} \Big|_0^1 - u^n e^{-nx} \Big|_1^\infty \right) \\
 &= \frac{2}{(n-1)!} n^n e^{-n}.
 \end{aligned}$$

因此

$$\lim_{n \rightarrow \infty} E[|X_n|] = \lim_{n \rightarrow \infty} n^{-\frac{1}{2}} E[|S_n|] = \lim_{n \rightarrow \infty} \frac{2}{n!} n^{n+\frac{1}{2}} e^{-n} = \sqrt{\frac{2}{\pi}}.$$

□

11.2 Bihari-LaSalle不等式及其推论

定理 11.2.1. (Bihari¹-LaSalle²不等式) 设 φ 与 ψ 是 \mathbb{R}_+ 上的非负连续函数(ψ Lebesgue可积即可), h 是 \mathbb{R}_+ 上的递增函数, $h(0) = 0$, 且 $t > 0$ 时 $h(t) > 0$. 设 $x \geq 0$ 且

$$\varphi(t) \leq x + \int_0^t \psi(s) h(\varphi(s)) ds \quad \forall t \geq 0.$$

令

$$H(t) = \int_{t_0}^t h^{-1}(s) ds \quad t \geq 0.$$

这里 $t_0 \geq 0$, 且当

$$\int_{0+} h^{-1}(s) ds < \infty$$

时可取 $t_0 = 0$, 否则 $t_0 > 0$. 则

$$\varphi(s) \leq H^{-1}\left(H(x) + \int_0^t \psi(s) ds\right).$$

证明. 首先注意 $H: \mathbb{R}_+ \mapsto [-\infty, \infty)$ 是连续递增函数.

设 $x > 0$. 令

$$f(t) := x + \int_0^t \psi(s) h(\varphi(s)) ds.$$

¹Imre Bihari, 1915-1998, 匈牙利数学家.

²Joseph Pierre LaSalle, 1916-1983, 美国数学家.

由 h 的单调性有

$$h(\varphi(t)) \leq h(f(t)),$$

即

$$\frac{h(\varphi(t))\psi(t)}{h(f(t))} \leq \psi(t),$$

也即

$$H'(f(t)) \leq \psi(t).$$

从0到 t 积分得

$$H(f(t)) - H(f(0)) \leq \int_0^t \psi(s) ds.$$

但 $f(0) = x$, 故

$$H(f(t)) \leq x + \int_0^t \psi(s) ds.$$

所以

$$\varphi(t) \leq f(t) \leq H^{-1}\left(x + \int_0^t \psi(s) ds\right).$$

$x > 0$ 的情况获证. 若 $x = 0$, 则对任意 $x > 0$ 均有

$$\varphi(t) \leq x + \int_0^t \psi(s)h(\varphi(s)) ds \quad \forall t \geq 0.$$

故

$$\varphi(t) \leq f(t) \leq H^{-1}\left(H(x) + \int_0^t \psi(s) ds\right).$$

取 $x \downarrow 0$ 即得

$$\varphi(t) \leq f(t) \leq H^{-1}\left(H(0) + \int_0^t \psi(s) ds\right).$$

□

细心的读者可能已经注意到, 此定理中 t_0 的选取有无穷种可能. 但实际上所有的可能最终给出的界其实是一样的, 你可以自己验证一下.

很多熟知的结果都是Bihari不等式的推论. 例如我们有:

(1) 取 $h(t) = t$, $t_0 = 1$ (例如). 则得到Grönwall³不等式.

推论 11.2.2. (Grönwall不等式) 若

$$\varphi(t) \leq x + \int_0^t \psi(s)\varphi(s) ds \quad \forall t \geq 0,$$

则

$$\varphi(t) \leq x \exp\left\{\int_0^t \psi(s) ds\right\}.$$

(2) 取 $h(t) = t^{1-\alpha}$, $\alpha \in (0, 1)$, $t_0 = 0$, 则得到Zakai不等式.

³Thomas Hakon Grönwall, 1877-1932, 瑞典数学家.

定理 11.2.3 (Zakai⁴不等式). 设 φ_s 是 $[0, T]$ 上的非负连续函数, ψ_s 为非负Lebesgue可积函数, $\alpha \in (0, 1)$. 若

$$\varphi_t \leq \beta + \int_0^t \psi_s \varphi_s^{1-\alpha} ds,$$

则

$$\varphi_t \leq (\beta^\alpha + \alpha \int_0^t \psi_s ds)^{1/\alpha}.$$

下面是Zakai不等式的进一步推论:

推论 11.2.4. φ, ψ, α 及 β 如上. 若

$$\varphi_t \leq \beta + \int_0^t ds \int_0^s \psi_u \varphi_u^{1-\alpha} du,$$

则

$$\varphi_t \leq (\beta^\alpha + \alpha \int_0^t ds \int_0^s \psi_u du)^{1/\alpha}.$$

证明. 令

$$h_t := \max_{0 \leq s \leq t} \varphi_s.$$

则 h 是盖住了 φ 的最小增函数. 由于 $\int_0^t ds \int_0^s \psi_u \varphi_u^{1-\alpha} du$ 是增函数, 因此

$$\begin{aligned} h_t &\leq \beta + \int_0^t ds \int_0^s \psi_u \varphi_u^{1-\alpha} du \\ &\leq \beta + \int_0^t ds \int_0^s \psi_u h_u^{1-\alpha} du \\ &\leq \beta + \int_0^t h_s^{1-\alpha} ds \int_0^s \psi_u du. \end{aligned}$$

然后用上面推论即可. □

3.此外还有

推论 11.2.5. 若

$$\varphi(t) \leq \int_0^t \psi(s) h(\varphi(s)) ds \quad \forall t \geq 0,$$

且

$$\int_{0+} h^{-1}(s) ds = \infty,$$

则 $\varphi \equiv 0$.

事实上, 此时(随便取 $t_0 > 0$) $H(0) = -\infty$, 故

$$H(0) + \int_0^t \psi(s) ds = -\infty \quad \forall t \geq 0,$$

而 $H^{-1}(-\infty) = 0$.

⁴Moshe Zakai, 1926-2015, 爱尔兰数学家.

11.3 绝对连续函数

设 f 是定义在 \mathbb{R} 上的函数. 若 $\forall \varepsilon > 0, \exists \delta > 0$, 使得

$$x_1 < x_2 < \cdots, \sum_{n=1}^{\infty} |x_{n+1} - x_n| < \delta \implies \sum_{n=1}^{\infty} |f(x_{n+1}) - f(x_n)| < \varepsilon,$$

则称 f 为绝对连续函数.

我们有下面的

定理 11.3.1 (Newton⁵-Leibnitz⁶公式). 设 f 是绝对连续函数, 则在Lebesgue测度意义下, f' 几乎处处存在, 且

$$f(x) - f(y) = \int_y^x f'(t) dt.$$

这个结果的证明可见任何一本实变函数教材, 例如[18].

我们还可证明

定理 11.3.2. 设 f 是绝对连续函数, 且 $|f'| \leq C$ a.e., 则

$$|f(x) - f(y)| \leq C|x - y|.$$

若 f' 是绝对连续函数, 且 $|f''| \leq C$ a.e., 则

$$|f(y) - [f(x) + f'(x)(y - x)]| \leq \frac{C}{2}|y - x|^2.$$

证明. 第一式直接由Newton-Leibnitz公式得到. 由于

$$\begin{aligned} f(y) - f(x) &= \int_x^y f'(t) dt \\ &= \int_x^y \left(f'(x) + \int_x^t f''(s) ds \right) dt \\ &= f'(x)(y - x) + \int_x^y dt \int_x^t f''(s) ds, \end{aligned}$$

而

$$\left| \int_x^y dt \int_x^t f''(s) ds \right| \leq \int_x^y dt \int_x^t C ds = \frac{C}{2}|y - x|^2,$$

所以第二式成立. □

11.4 函数的磨光

在 \mathbb{R}^d 上定义函数:

$$\varphi(x) = \begin{cases} c^{-1} e^{-\frac{1}{1-|x|^2}}, & |x| < 1, \\ 0, & |x| \geq 1. \end{cases}$$

⁵Isaac Newton, 1643-1727, 英国物理学家, 数学家.

⁶Gottfried Wilhelm von Leibniz, 1646-1716, 德国数学家.

其中 $c := \int_{|x|<1} e^{-\frac{1}{1-|x|^2}} dx$.
 设 \mathbb{R}^d 上的函数 f 满足

$$\int_{\mathbb{R}^d} |f(x)| dx < \infty.$$

令

$$f_n(x) = n^d \int_{\mathbb{R}^d} f(y) \varphi(n(x-y)) dy.$$

易见当 $f(x)$ 在某个有界集外为零时, f_n 也在某个有界集(不是前面那个, 可能大一点)外也为零, f_n 无穷次可微, 且对 f 的任意连续点 x 有

$$\lim_{n \rightarrow \infty} f_n(x) = f(x).$$

11.5 指数函数的逼近

对于实数 a ,

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a}{n}\right)^n = e^a,$$

下面证明, 上面等式中的实数 a 换为复数 z 结论仍成立, 即 $\forall z \in \mathbb{C}$,

$$\lim_{n \rightarrow \infty} \left(1 + \frac{z}{n}\right)^n = e^z.$$

为此, 设 $z = a + ib$, 则 $e^z = e^a(\cos b + i \sin b)$, 以及

$$\begin{aligned} \left(1 + \frac{z}{n}\right)^n &= \left(1 + \frac{a+ib}{n}\right)^n \\ &= \left(\left(1 + \frac{a}{n}\right) + i\frac{b}{n}\right)^n \\ &= r_n(\cos \theta_n + i \sin \theta_n), \end{aligned}$$

其中

$$\begin{aligned} r_n &= \left(\left(1 + \frac{a}{n}\right)^2 + \left(\frac{b}{n}\right)^2\right)^{\frac{n}{2}}, \\ \theta_n &= n \arctan \frac{\frac{b}{n}}{1 + \frac{a}{n}}, \quad n > |a|. \end{aligned}$$

对 θ_n , 极限

$$\begin{aligned} \lim_{n \rightarrow \infty} \theta_n &= \lim_{n \rightarrow \infty} n \arctan \frac{\frac{b}{n}}{1 + \frac{a}{n}} \\ &= \lim_{n \rightarrow \infty} n \frac{\frac{b}{n}}{1 + \frac{a}{n}} = b. \end{aligned}$$

对 r_n , 取对数后的极限为

$$\begin{aligned} \lim_{n \rightarrow \infty} \ln r_n &= \lim_{n \rightarrow \infty} \frac{n}{2} \ln \left(1 + \frac{2a}{n} + \frac{a^2 + b^2}{n}\right) \\ &= \lim_{n \rightarrow \infty} \frac{n}{2} \left(\frac{2a}{n} + \frac{a^2 + b^2}{n}\right) = a. \end{aligned}$$

因此

$$\lim_{n \rightarrow \infty} \left(1 + \frac{a + ib}{n}\right)^n = e^a (\cos b + i \sin b).$$

命题 11.5.1. 设 z_n, z 是复数, 且 $|z_n| = o(n^{-1})$. 则

$$\lim_{n \rightarrow \infty} \left(1 + \frac{z}{n} + z_n\right)^n = e^z.$$

证明. 对任意复数 x, y 有

$$x^n - y^n = n(x - y) \int_0^1 (x + u(y - x))^{n-1} du.$$

而

$$|x + u(y - x)| \leq |x| + |y - x|.$$

所以当 n 充分大时

$$\begin{aligned} \left| \left(1 + \frac{t}{n} + z_n\right)^n - \left(1 + \frac{z}{n}\right)^n \right| &\leq n|z_n| \int_0^1 \left(1 + 2\frac{|z|}{n}\right)^{n-1} du \\ &\leq n|z_n| e^{2|z|} \rightarrow 0, \quad (n \rightarrow \infty). \end{aligned}$$

因此

$$\lim_{n \rightarrow \infty} \left(1 + \frac{z}{n} + z_n\right)^n = \lim_{n \rightarrow \infty} \left(1 + \frac{z}{n}\right)^n = e^z.$$

□

下面考虑实变量的复值函数

$$f(t) =: e^{zt} = e^{(a+ib)t}, \quad t \in \mathbb{R}.$$

则

$$\begin{aligned} f'(t) &= (e^{at} \cos(bt) + ie^{at} \sin(bt))' \\ &= e^{at}(a \cos(bt) - b \sin(bt)) + ie^{at}(a \sin(bt) + b \cos(bt)) \\ &= (a + ib)e^{(a+ib)t} \\ &= ze^{zt}. \end{aligned}$$

由此可得

$$\int_c^d e^{zt} dt = \frac{1}{z} e^{zt} \Big|_c^d, \quad \forall c, d \in \mathbb{R}..$$

这个公式是说, 对实变量复值的指数函数积分时, 你可以形式上将 z 看作实参数来积分就行, 与分成实部和虚部分开积分的结果是一致的.

11.6 常用分布

1. 离散均匀分布.

分布列:

$$P(\xi = k) = \frac{1}{N}, \quad k = 1, 2, \dots, N.$$

2. Bernoulli分布

$$P(\xi = 1) = p, \quad P(\xi = 0) = q, \quad 0 \leq p \leq 1, \quad q = 1 - p.$$

3. 二项分布

$$P(\xi = k) = C_n^k p^k q^{n-k}, \quad k = 1, 2, \dots, n, \quad 0 \leq p \leq 1, \quad q = 1 - p, \quad n = 1, 2, \dots$$

4. Poisson分布

$$P(\xi = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, 2, \dots, \quad \lambda > 0.$$

5. 几何分布

$$P(\xi = k) = pq^{k-1}, \quad k = 1, 2, \dots, \quad 0 \leq p \leq 1, \quad q = 1 - p.$$

6. 负二项分布

$$P(\xi = k) = C_{k-1}^{r-1} p^r q^{k-r}, \quad k = r, r+1, \dots, \quad 0 \leq p \leq 1, \quad q = 1 - p, \quad r = 1, 2, \dots$$

7. 均匀分布. 设 $a, b \in \mathbb{R}, a < b$. 令

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & \text{其它.} \end{cases}$$

8. 正态分布. 设 $\mu, \sigma \in \mathbb{R}, \sigma > 0$. 令

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}, \quad \mu \in \mathbb{R}, \quad \sigma > 0.$$

这个分布我们记为 $N(\mu, \sigma^2)$. $(\mu, \sigma) = (0, 1)$ 时称为标准正态分布.

为了方便, 往往将 $\sigma = 0$ 的情况也认为是正态分布, 是退化正态分布, 因为这时相当于 $\xi \equiv 0$.

9. 指数分布

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0.$$

10. Gamma分布

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}, \quad x \geq 0, \quad \alpha, \beta > 0.$$

11. Beta分布

$$f(x) = \frac{1}{B(r, s)} x^{r-1} (1-x)^{s-1}, \quad 0 \leq x \leq 1, \quad r, s > 0.$$

12. 双侧指数分布

$$f(x) = \frac{1}{2} \lambda e^{-\lambda|x-\alpha|}, \quad x \in \mathbb{R}, \quad \lambda > 0, \alpha \in \mathbb{R}.$$

13. χ^2 分布

$$f(x) = \frac{1}{\Gamma\left(\frac{n}{2}\right)} 2^{-\frac{n}{2}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, \quad x \geq 0, \quad n \geq 1.$$

14. Student分布

$$f(x) = \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}, \quad n \geq 1.$$

15. F分布

$$f(x) = \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}}}{B\left(\frac{m}{2}, \frac{n}{2}\right)} \cdot \frac{x^{\frac{m}{2}-1}}{\left(1 + \frac{mx}{n}\right)^{\frac{m+n}{2}}}, \quad x \geq 0, \quad m, n \geq 1.$$

16. Cauchy分布

$$f(x) = \frac{a}{\pi(x^2 + a^2)}, \quad x \in \mathbb{R}, \quad a > 0.$$

容易看出, 指数分布是 $\alpha = 1, \beta = \frac{1}{\lambda}$ 的Gamma分布, 自由度为 n 的 χ^2 分布是 $\alpha = \frac{n}{2}, \beta = 2$ 的Gamma分布.

表 11.1: 常见离散型分布表

分布名称	分布列	期望	方差	特征函数
离散均匀分布 $U(1, 2, \dots, n)$	$p_k = \frac{1}{n}, k = 1, 2, \dots, n$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$	$\frac{e^{it} - e^{i(n+1)t}}{n(1-e^{it})}$
二项分布 $B(n, p)$	$p_k = C_n^k p^k q^{n-k}, k = 1, 2, \dots, n,$ $0 < p < 1, q = 1 - p$	np	npq	$(pe^{it} + q)^n$
Poisson分布 $P(\lambda)$	$p_k = e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, \dots,$ $\lambda > 0$	λ	λ	$e^{\lambda(e^{it}-1)}$
几何分布 $Ge(p)$	$p_k = pq^{k-1}, k = 1, 2, \dots,$ $0 < p < 1, q = 1 - p$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^{it}}{1-qe^{it}}$
负二项分布 $NB(r, p)$	$p_k = C_{k+r-1}^{r-1} p^r q^k, k = 0, 1, \dots,$ $r = 1, 2, \dots, 0 < p < 1, q = 1 - p$	$\frac{rq}{p}$	$\frac{rq}{p^2}$	$\left(\frac{p}{1-qe^{it}}\right)^r$

表 11.2: 常见连续型分布表

分布名称	密度函数	期望	方差	特征函数
均匀分布 $U(a, b)$	$p(x) = \frac{1}{b-a}, x \in [a, b],$ $a, b \in \mathbb{R}, a < b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{itb} - e^{ita}}{it(b-a)}$
指数分布 $E(\lambda)$	$p(x) = \lambda e^{-\lambda x}, x \geq 0,$ $\lambda > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$(1 - \frac{it}{\lambda})^{-1}$
正态分布 $N(\mu, \sigma^2)$	$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R},$ $\mu \in \mathbb{R}, \sigma > 0$	μ	σ^2	$e^{i\mu t - \frac{\sigma^2 t^2}{2}}$
Gamma分布 $\Gamma(\alpha, \beta)$	$p(x) = \frac{\Gamma(\alpha)\beta^\alpha}{x^{\alpha-1}} e^{-\beta x}, x \geq 0,$ $\alpha, \beta > 0$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	$(1 - \frac{it}{\beta})^{-\alpha}$
Beta分布 $Be(r, s)$	$p(x) = \frac{1}{B(r, s)} x^{r-1} (1-x)^{s-1},$ $x \in [0, 1], r, s > 0$	$\frac{r}{r+s}$	$\frac{rs}{(r+s)^2(r+s+1)}$	$\varphi_{Be}(t)$
双侧指数分布 $Laplace(\alpha, \lambda)$	$p(x) = \frac{1}{2} \lambda e^{-\lambda x-\alpha }, x \in \mathbb{R}$ $\alpha \in \mathbb{R}, \lambda > 0$	α	$\frac{2}{\lambda^2}$	$\frac{e^{i\alpha t}}{1 + \lambda^{-2} t^2}$
χ^2 分布 $\chi^2(n)$	$p(x) = \frac{1}{\Gamma(\frac{n}{2})} 2^{-\frac{n}{2}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}},$ $x \geq 0, n \geq 1$	n	$2n$	$(1 - 2it)^{-\frac{n}{2}}$
Student分布 $t(n)$	$p(x) = \frac{\Gamma(\frac{1}{2}(n+1))}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}},$ $x \in \mathbb{R}, n \geq 1$	0	$\frac{n}{n-2}, n > 2$	$\varphi_S(t)$
F分布 $F(m, n)$	$p(x) = \frac{\left(\frac{m}{n}\right)^{\frac{m}{2}}}{B\left(\frac{m}{2}, \frac{n}{2}\right)} \cdot \frac{x^{\frac{m}{2}-1}}{\left(1 + \frac{mx}{n}\right)^{\frac{m+n}{2}}},$ $x \geq 0, n \geq 1$	$\frac{n}{n-2},$ $n > 2$	$\frac{2n^2(m+n-2)}{m(n-2)^2(n-4)},$ $n > 4$	$\varphi_F(t)$
Cauchy分布 $C(a)$	$p(x) = \frac{a}{\pi(x^2+a^2)}, x \in \mathbb{R},$ $a > 0$	不存在	不存在	$e^{a t }$

其中

$$\varphi_{Be}(t) = \frac{\Gamma(r+s)}{\Gamma(s)} \sum_{k=0}^{\infty} \frac{\Gamma(r+k)(it)^k}{\Gamma(r+s+k)\Gamma(k+1)},$$

$$\varphi_S(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} (1+t^2)^{-\frac{n+1}{2}},$$

$$\varphi_F(t) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} \left(\frac{n}{m}\right)^{\frac{n}{2}} \frac{\Gamma(\frac{m+n}{2} - it)}{\Gamma(\frac{m}{2} - it)\Gamma(\frac{n}{2} - it)}.$$

参考文献

- [1] J. Bernoulli: *Ars Conjectandi*, Impensis Thurnisiorum, Fratrum, 1733. (英译: On the law of Large Numbers, O. Sheynin译, Berlin 2005; 中译: 猜度术, 应坚刚译, 上海 2019.)
- [2] 崔尚斌: 偏微分方程现代理论引论, 科学出版社, 北京 2016.
- [3] R. Courant and D. Hilbert: *Methods of Mathematical Physics*. V. 1. Wiley Classics. 1989.
- [4] W. Feller: *An Introduction to Probability Theory and Its Applications*, Vol. 2, 3rd ed., New York: Wiley, 1971.
- [5] 华罗庚: 高等数学引论, 第一册. 高等教育出版社, 北京 2009.
- [6] 华罗庚: 高等数学引论, 第二册. 高等教育出版社, 北京 2009.
- [7] 伊藤清: 概率论, 闫理坦译, 人民邮电出版社, 北京 2021.
- [8] 柯嘉琪: Hajek-Renyi不等式的推广及其应用, 中山大学学士论文, 2024.
- [9] A.N.Kolmogorov: *Foundations of the Theory of Probability*. 2nd ed., translation edited by Nathan Morrison, with an added bibliography by A. T. Bharucha-Reid. Dover Publications, Inc., Mineola, New York 2020.
- [10] P. Lévy: *Calcul des Probabilités*, Réimpression autorisée de l'édition originale publiée par Gauthier-Villars en 1925. 2004, 2006(2^e tirage), Éditions Jaques Gabay.
- [11] P. Malliavin: *Integration and Probability*, Springer, 1995.
- [12] 李贤平: 概率论基础, 第三版, 高等教育出版社, 北京 2010.
- [13] Shafer, G. and Vovk, V.: The origins and legacy of Kolmogorov's Grundbegriffe, <https://arxiv.org/abs/1802.06071>
- [14] D.W.Stroock: *Probability Theory: An Analytic View*. Cambridge Univ. Press, 1993.
- [15] 苏淳: 概率论(第二版), 科学出版社, 北京 2010.
- [16] 王蒙: 暗杀3322, 人民文学出版社, 北京 2003.
- [17] 王梓坤: 概率论基础及其应用, 北京师范大学出版社, 北京 1996.
- [18] 杨力华: 实变函数论教程, 科学出版社, 北京 2017.

索引

- A
- a.s., 79
- B
- Bayes公式, 36, 41, 85
 - Bayes公式(可列概型), 57
 - Bernoulli分布, 32, 106
 - Bernoulli概型, 18, 27
 - Bernoulli概型(n 重), 18
 - Borel σ -代数, 94
 - Borel σ -代数(n -维), 98
 - Borel σ -代数(一维), 94
 - Borel σ -代数(二维), 97
 - Borel σ -代数, 77
 - Borel(可测)函数, 98
 - Borel集, 77
 - 不相关, 176
 - 不等式
 - Bihari-LaSalle, 281
 - Cauchy-Schwarz, 182
 - Chatterji, 203
 - Chebyshev, 138
 - Doob, 206
 - Gronwall, 282
 - Hölder, 182
 - Jensen, 181
 - Kolmogorov, 206
 - Minkowski, 182, 224
 - Ottaviani, 202
 - Rio, 203
 - Zakai, 283
 - 广义Hajek-Renyi, 217
 - 边沿分布, 117
 - 闭集, 94
- C
- $C(F)$, 241
 - C_0 , 166
 - C_0^∞ , 166
 - C_b , 166
 - 乘法公式, 36
 - 乘积概率空间, 82
 - 抽样
 - 无放回无顺序, 20
 - 无放回有顺序, 20
 - 有放回无顺序, 21
 - 有放回有顺序, 20
 - 次可列可加性, 56
 - 次可列可加性(可列概型), 57
- D
- De Morgan原理, 15
 - Dirichlet积分, 235
 - 代数, 49
 - 单调类, 74
 - 多项概型, 27
 - 大数定律, 213
 - Bernoulli, 214
 - Chebyshev, 213
 - Markov, 213
 - 定理
 - Fubini, 159, 160
 - Helly第一, 244
 - Helly第二, 242
 - Lindeberg, 268
 - Skorokhod表现定理, 254
 - Weierstrass, 214
 - 单调类1, 76
 - 单调类2, 76
 - 控制收敛, 156
 - 随机变量的存在性, 124

- 独立
 - 古典概型, 24
 - 有限概型, 45
- 独立性, 23
 - 事件, 86
 - 离散概型, 64
 - 随机变量, 187
 - 集类, 86
- E
 - 二项分布, 32, 106
 - 二项概型, 27
- F
 - 分割, 42, 62
 - 分布函数, 58, 103, 104
 - 分布函数(二维), 114
 - 分布函数(多维), 114
 - 分布列, 105
 - 分布列(可列概型), 57
 - 反演公式(d -维), 250
 - 反演公式初级版, 236
 - 反演公式完整版, 239
 - 复随机变量, 223
 - 方差, 61, 173
- G
 - 公式
 - 全概率, 39
 - 古典概型, 18
 - 概率, 79
 - 概率分布, 103
 - 概率空间, 79
 - 规范性, 79
- J
 - 几乎必然收敛, 148
 - 几乎必然相等, 104
 - 几何分布, 107
 - 几何概型, 81
 - 加法公式, 19, 27, 56, 57, 79
 - 加法原理, 19
 - 均匀分布, 287
 - 矩, 174
- 简单Borel函数, 100
- K
 - Kronecker引理, 216
 - 可列可加性, 56, 78, 79
 - 可列可加性(可列概型), 57
 - 可列概型, 56
 - 可略, 79
 - 可积(可列概型), 59
 - 可积(简单随机变量), 132
 - 开集, 94
- L
 - 0-1律
 - Kolmogorov, 92
 - λ -类, 75
 - λ_0 -类, 48
 - Lebesgue基本定理, 154
 - Lebesgue测度, 81
 - 离散型分布, 105
 - 离散型随机变量, 105
 - 离散型随机变量(多维), 116
 - 离散概型, 56
 - 联合分布, 117
 - 联合分布列, 58
 - 连续型分布, 107
 - 连续型分布(多维), 116
 - 连续型随机变量, 107
 - 连续型随机变量(多维), 116
- M
 - 密度函数, 107
 - 母函数, 67
- N
 - Newton-Leibnitz公式, 284
- P
 - $P(\lambda, a)$, 58
 - π -类, 47
 - $\pi - \lambda$ 定理, 76
 - $\pi - \lambda_0$ 定理, 51
 - p -阶矩, 174
 - Poisson分布, 58, 107

- Poisson分布(跳为 a), 58
- Q
- 全排列, 20
- 全期望公式, 43
- 全概率公式, 43, 85
- 全概率公式(可列概型), 57
- 强大数定律, 219
- Kolmogorov, 219
- 期望
- 一般情形, 136
- 关于参数的可微性, 161
- 关于参数的连续性, 160
- 简单随机变量, 132
- 非负随机变量, 135
- 期望(可列概型), 59
- 期望(有限概型), 29
- 期望(非负简单随机变量), 131
- R
- Riemann-Stieltjes积分, 162
- 弱收敛, 241
- S
- σ -代数, 74
- Stein-Chen方法, 68
- 上连续, 80
- 事件域, 79
- 实值随机变量, 100
- 生成的 λ_0 -类, 50
- 生成的代数, 50
- 示性函数, 14
- 随机变量, 100
- 随机变量(可列概型), 57
- 随机变量(多维), 101
- 随机变量(有限概型), 29
- T
- Toeplitz引理, 216
- 停时, 207
- 凸函数, 180
- 条件分布列(给定事件), 58
- 条件分布列(给定随机变量), 58
- 条件期望, 62, 184
- 条件期望(给定分割), 43
- 条件概率, 36, 85
- 条件概率(可列概型), 57
- 条件概率(给定分割), 43, 62
- 条件独立, 87
- 特征函数, 224
- 特征函数(多维), 248
- W
- 尾概率, 179
- X
- 下连续, 80
- 协方差, 175
- 相互独立
- 事件, 86
- 选排列, 20
- Y
- 一致可积, 158
- 依分布收敛, 241
- 依概率收敛, 148
- 原像, 93
- 引理
- Borel-Cantelli第一, 84
- Borel-Cantelli第二, 92
- Fatou, 155
- Pratt, 157
- 有限可加性, 27
- 有限概型, 26
- 样本点, 9
- 样本空间, 9
- Z
- 中心极限定理, 263
- 中心极限定理
- Laplace, 264
- 再生性(分布的), 259
- 指数分布, 108, 109
- 正态分布, 110, 287
- 多维, 117