

概率论与数理统计中的几个注记

李华灿¹, 李群芳²

¹赣南科技学院文法学院, 江西 赣州

²赣州师范高等专科学校数学系, 江西 赣州

收稿日期: 2025年10月4日; 录用日期: 2025年10月28日; 发布日期: 2025年11月6日

摘要

本文针对学生在学习概率论与数理统计中出现的問題, 对随机变量的数字特征、不相关与独立的关系、事件发生与概率的关系以及小概率事件的实际不可能性原理进行了整理, 并通过实例进行说明。

关键词

数字特征, 不相关, 必然事件, 小概率事件, 反证法

Several Notes on Probability Theory and Mathematical Statistics

Huacan Li¹, Qunfang Li²

¹School of Humanities and Law, Gannan University of Science and Technology, Ganzhou Jiangxi

²Department of Mathematics, Ganzhou Teachers College, Ganzhou Jiangxi

Received: October 4, 2025; accepted: October 28, 2025; published: November 6, 2025

Abstract

This paper addresses common issues encountered by students in learning probability theory and mathematical statistics. It systematically organizes key concepts such as the numerical characteristics of random variables, the relationship between uncorrelatedness and independence, the connection between events and probabilities, and the practical impossibility principle of small-probability events. These points are further illustrated with concrete examples.

Keywords

Numerical Characteristics, Uncorrelatedness, Sure Event, Small-Probability Event, Proof by Contradiction

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

概率论与数理统计[1]作为描述随机现象、进行统计推断的核心数学工具,其理论严谨性与应用广泛性已使其成为自然科学、社会科学及工程技术领域的基础课程。然而在教学实践中,学生常对“随机变量数字特征的存在性、不相关与独立的区别、概率为 0/1 事件的实际意义、小概率原理的逻辑本质”等核心概念产生认知偏差,这些偏差不仅源于概念本身的抽象性,也与不同教材对概念阐释的详略差异有关。因此,本文旨在系统梳理并澄清上述教学中的普遍难点,通过实例验证与理论补充,帮助学生建立对基础概念的准确认知。

从学科发展与教学实践来看,上述概念的重要性已被多部国际权威教科书明确强调。例如, Jaynes 在《Probability Theory: The Logic of Science》中指出,随机变量数字特征的“绝对收敛”要求是确保其“客观度量意义”的前提,若忽略这一条件,可能导致对随机变量“平均水平”的误判; Ross 的《Introduction to Probability and Statistics for Engineers and Scientists》则通过大量实例对比,说明“不相关 $\neq$ 独立”是概率统计中最易混淆的关系之一,尤其在非正态分布场景下,二者的差异直接影响统计模型的构建合理性;此外, Feller 的经典著作《An Introduction to Probability Theory and Its Applications》在阐述概率测度时,特别区分了“不可能事件”与“概率为 0 事件”,指出后者在测度论框架下的“几乎不可能”属性,而非绝对的“不可能发生”,而 Lehmann 的《Testing Statistical Hypotheses》则深入剖析了小概率原理与假设检验的逻辑关联,强调其作为“归纳推理工具”的核心地位。

本文立足上述学术背景,针对教学中反复出现的概念混淆问题,结合实例与理论透视展开分析,既呼应国际权威教材对核心概念的重视,也为教学实践提供更具针对性的概念澄清思路,助力学生打通“理论理解”与“应用实践”之间的壁垒。

2. 不表是所有的随机变量的数字特征都存在

数字特征主要含数学期望、方差、协方差等,是描述随机变量分布特性的重要指标,这些数字特征的存在依赖于积分或级数的绝对收敛性。下面就数学期望不存在加以说明。

定义 1 若 X 是离散型随机变量,其概率函数为

$$P(X = x_i) = P(x_i) (i = 1, 2, \dots)$$

且级数 $\sum x_i p(x_i)$ 绝对收敛,则称

$$E(X) = \sum x_i p(x_i)$$

为离散型随机变量 X 的数学期望;若 X 是连续型随机变量,其概率密度函数为 $f(x)$,且积分

$$\int_{-\infty}^{\infty} x f(x) dx$$

绝对收敛,则称

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

为连续型随机变量 X 的数学期望。

由定义 1 可知,随机变量 X 的数学期望存在,必须要求对应的无穷级数或无穷限的广义积分绝对收

敛, 而不是条件收敛, 其根本目的是保证期望值不依赖求和或积分顺序(即期望值的唯一性)、避免因正负项抵消掩盖发散性(即期望值的稳定性)、确保期望作为随机变量“平均表现”的客观性与可操作性(即期望值的物理意义)。下面给出两个条件收敛但不绝对收敛的例子。

例 1 设离散型随机变量 X 的取值为

$$x_k = (-1)^k k \cdot \frac{\pi^2}{6}$$

且其概率分布为

$$P(X = x_k) = \frac{6}{\pi^2 k^2} \quad (k = 1, 2, 3, \dots),$$

则

$$\sum x_k p(x_k) = \sum_{k=1}^{+\infty} (-1)^k k \frac{\pi^2}{6} \frac{6}{\pi^2 k^2} = \sum_{k=1}^{+\infty} \frac{(-1)^k}{k}.$$

又因为级数 $\sum_{k=1}^{+\infty} \frac{(-1)^k}{k}$ 是交错调和级数, 由莱布尼茨判别法知, 该级数条件收敛, 但不绝对收敛, 故该随机变量的数学期望不存在。

注 1: 由例 1 知, 要构造离散型随机变量数学期望不存在的反例, 可选择一条件收敛但不绝对收敛的级数, 并以此级数作为 $\sum x_k p(x_k)$, 从而来寻找 x_k 与 $p(x_k)$ 。如在例 1 中选择 $\sum x_k p(x_k)$ 为调和交错级数 $\sum_{k=1}^{+\infty} \frac{(-1)^k}{k}$, 然后令

$$x_k = (-1)^k k \frac{\pi^2}{6}, \quad P(X = x_k) = \frac{6}{\pi^2 k^2};$$

或者令

$$x_k = \frac{(-1)^k}{k \cdot 2^k}, \quad P(X = x_k) = \frac{1}{2^k}$$

均可, 上式中 $\frac{6}{\pi^2}$ 是否概率密度的规范性得出的, 称为归一化常数。

例 2 (柯西分布) 设随机变量 X 的概率密度函数为

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}$$

则容易求得积分

$$\int_{-\infty}^{\infty} |x| f(x) dx = \int_{-\infty}^{\infty} \frac{|x|}{\pi(1+x^2)} dx = \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2} dx$$

不存在, 即积分

$$\int_{-\infty}^{\infty} x f(x) dx$$

不绝对收敛, 故其数学期望不存在。

理论透视: 黎曼级数定理的深层影响。黎曼级数定理明确指出, 条件收敛的无穷级数可通过改变项的求和顺序, 使级数和收敛到任意预先指定的实数, 甚至发散。这一特性直接决定了数学期望定义中“绝

对收敛”要求的必要性。若允许数学期望基于条件收敛的级数或积分定义, 同一随机变量的期望可能因计算顺序不同而出现多个结果, 完全违背“期望是随机变量平均水平客观度量”的本质属性。以柯西分布为例, 其概率密度关于原点对称, 若仅从条件收敛角度计算积分, 正负区域的积分会相互抵消并得到“0”的结果, 但这种结果缺乏实际意义: 柯西分布的取值会频繁出现极端值, 不存在稳定的平均水平, 而绝对收敛的要求恰好排除了这种“虚假平均”, 确保期望的定义与实际意义一致。

3. 不相关的两随机变量不一定独立

定义 2 设 X 、 Y 为任意两个随机变量, 若相关系数 $R(X, Y) = 0$ (或 $Cov(X, Y) = 0$), 则称随机变量 X 与 Y 不相关, 其中

$$R(X, Y) = \frac{Cov(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}.$$

定义 3 设 X 、 Y 为任意两个随机变量, A 、 B 分别是任一与随机变量 X 、 Y 有关的事件, 若

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

则称随机变量 X 与 Y 独立。

容易求得, 若随机变量 X 与 Y 独立, 则 X 与 Y 一定不相关; 但反过来不成立, 即: 若 X 与 Y 不相关, 则 X 与 Y 未必独立[2]。

例 3 设二维离散型随机变量 (X, Y) 的联合概率函数为

$X \backslash Y$	-1	0	1
-1	0.2	0	0.2
0	0	0.2	0
1	0.2	0	0.2

则关于 X 、 Y 的边缘概率分布为

X	-1	0	1
P	0.4	0.2	0.2

Y	-1	0	1
P	0.4	0.2	0.2

故

$$P(X=0, Y=0) \neq P(X=0)P(Y=0).$$

从而 X 与 Y 不独立。

但 $E(X) = E(Y) = E(XY) = 0$, 从而 X 与 Y 的协方差

$$Cov(X, Y) = E(XY) - E(X) \cdot E(Y) = 0.$$

即 X 与 Y 不相关。

例 4 设 $X \sim N(0,1)$, 且 $Y = X^{2k}$ (k 为整数), 则 X 与 Y 不独立且不相关。

事实上, 由 $Y = X^{2k}$ 知, X 与 Y 不独立。又

$$E(X) = 0, \quad E(XY) = E(X^3) = \int_{-\infty}^{+\infty} x^3 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0$$

故 $\text{Cov}(X, Y) = E(XY) - E(X) \cdot E(Y) = 0$, 即 X 与 Y 不相关。

例 5 设随机变量 (X, Y) 服从单位圆上的均匀分布, 即具有联合概率密度函数为

$$f(x, y) = \begin{cases} \frac{1}{\pi} & x^2 + y^2 \leq 1, \\ 0 & \text{其它.} \end{cases}$$

则 X 与 Y 不独立且不相关。

事实上, 由边缘概率密度的公式可得 X 与 Y 的概率密度分别为

$$f_X(x) = \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \frac{1}{\pi} dy = \frac{2\sqrt{1-x^2}}{\pi} \quad (-1 \leq x \leq 1)$$

$$f_Y(y) = \frac{2\sqrt{1-y^2}}{\pi} \quad (-1 \leq y \leq 1)$$

故 $f(x, y) \neq f_X(x) \cdot f_Y(y)$, 即 X 与 Y 不独立; 但由积分的“偶倍奇零”性质, 显然有

$$E(X) = \int_{-\infty}^{+\infty} x \cdot f_X(x) dx = \int_{-1}^1 x \frac{2\sqrt{1-x^2}}{\pi} dx = 0$$

$$E(Y) = \int_{-\infty}^{+\infty} y \cdot f_Y(y) dy = 0$$

$$E(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xy \cdot f(x, y) dx dy = 0$$

故 $\text{Cov}(X, Y) = E(XY) - E(X) \cdot E(Y) = 0$, 即 X 与 Y 不相关。

注 2: 由上可知, X 与 Y 不相关推不出 X 与 Y 独立, 但是若 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, r)$ (二维正态分布), 则此时 X 、 Y 不相关与独立等价。

延伸思考: 随机变量空间的“正交”与“因子分解”类比。可将随机变量视为某个函数空间中的向量, 此时“不相关”对应向量的“正交”关系。若两个随机变量不相关, 意味着它们在“中心化”(减去自身期望)后, 其“内积”(协方差)为 0, 即两个变量的线性关联程度为零。但“正交”仅反映线性层面的无关性, 无法排除变量间可能存在的非线性关联, 如同平面内垂直的向量仍可能通过非线性变换产生联系。而“独立”则对应更强的“因子分解”性质, 若两个随机变量独立, 其联合概率分布(或密度函数)可分解为各自边缘分布(或密度函数)的乘积, 这意味着一个变量的取值概率完全不受另一个变量取值的影响, 无论这种影响是线性还是非线性的。例如, 单位圆上均匀分布的随机变量 X 与 Y , 虽因对称性满足“正交”(不相关), 但联合密度无法分解为边缘密度的乘积, 故不独立。只有在特殊结构(如二维正态分布)中, “正交”才恰好等价于“因子分解”, 此时不相关与独立完全一致。

4. 概率为 0 的事件未必不发生, 概率为 1 的事件未必一定发生

众所周知, 不可能事件 Φ 概率为零, 在每次试验中一定不发生; 必然事件概率为 1, 在每次试验中一定发生。但是, 反过来, 不成立, 即概率为 0 的事件未必不发生, 概率为 1 的事件未必一定发生。

例 6 设 X 服从区间 $[a, b]$ 上的均匀分布, 即 $X \sim U[a, b]$, 其概率密度函数为

$$f(x) = \begin{cases} 1 & a \leq x \leq b, \\ 0 & \text{其它} \end{cases}$$

则 $P\left(X = \frac{a+b}{2}\right) = \int_{\frac{a+b}{2}}^{\frac{a+b}{2}} f(x) dx = 0$ 。但 $\frac{a+b}{2}$ 是区间 $[a, b]$ 上的一个可能取值, 这意味着 “ $X = \frac{a+b}{2}$ ” 是一个概率为 0 但可能发生的事件。

注 3: 若称概率为 0 的事件称为零测集, 则概率为 0 的事件可视为可能发生, 但发生的可能性在测度论中可被忽略的事件, 相当于几乎处处不发生的事件; 在例 6 中, $P\left(X = \frac{a+b}{2}\right) = 0$, 故 $P\left(X \neq \frac{a+b}{2}\right) = 1$, 显然事件 “ $X \neq \frac{a+b}{2}$ ” 也不是必然事件。

理论透视: 测度论框架下的概率本质。从测度论角度看, 概率本质上是定义在 “事件 σ 代数” 上的规范化测度(总测度为 1), 而 “长度、面积” 等几何测度是概率测度的特殊情形。在区间 $[0, 1]$ 上的均匀分布中, 单点的 “长度测度” 为 0, 对应概率为 0, 但单点本身是样本空间中的合法元素, 仍有可能被随机选中(如随机投点恰好落在该点)。这表明 “概率为 0” 仅表示事件在测度意义上 “体积为零”, 而非 “不存在于样本空间”。类似地, 概率为 1 的事件(几乎必然事件)对应测度为 1 的集合, 但其补集(概率为 0 的事件)仍可能存在。

5. 小概率事件的实际不可能性原理是一种概率反证法, 具有一定的容错率

定义 4 在统计学中, 当某一事件的概率极低(如低于预设的显著性水平 α , 通常为 0.05 或 0.01)时, 认为该事件在实际中不可能发生, 从而拒绝原假设, 此原理称为小概率事件的实际不可能性原理。

小概率事件的实际不可能性原理是一种基于概率的经验推断, 从而有一定的容错率(即小概率), 而非绝对逻辑证明(即不保证百分之百正确), 常应用于科学实验、质量控制、医学研究等需要统计推断的领域。例如, 在假设检验中, 若某现象的 P 值远小于 α , 则拒绝原假设(如 “药物无效”), 接受备择假设(如 “药物有效”)。

例 7 某制药公司研发了一款新型降压药 “MediLower”, 声称其降压效果显著优于安慰剂。为验证这一说法, 研究人员将 200 名高血压患者随机分为两组: 实验组(100 人): 每日服用 MediLower, 治疗周期为 8 周, 治疗后测量得到每位患者的收缩压的平均值为 $\bar{X}_1 = 12.5$ mmHg, 标准差 $s_1 = 3.2$; 对照组(100 人): 每日服用外观相同的安慰剂, 治疗周期为 8 周, 治疗后测量每位患者的收缩压的平均值为 $\bar{X}_2 = 5.0$ mmHg, 标准差 $s_2 = 2.8$, 问 “MediLower 是否比安慰剂更有效降低收缩压(单位: mmHg)”。

解: 利用小概率事件的实际不可能性原理检验 “药物的有效性检验”。

1) 提出假设: 令原假设(H_0): 两组患者的平均收缩压下降值无差异($\mu_1 - \mu_2 = 0$); 备择假设(H_1): 实验组的平均收缩压下降值显著高于对照组($\mu_1 - \mu_2 > 0$)。

2) 计算检验统计量: 使用独立样本 t 检验, 合并方差后得:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{12.5 - 5.0}{\sqrt{\frac{3.2^2}{100} + \frac{2.8^2}{100}}} \approx 16.7.$$

3) 确定 P 值: 经查表可得: 在单侧检验下自由度为 198, $t \approx 16.7$ 对应的 P 值远小于 0.005。

4) 给定显著性水平: 预设 $\alpha = 0.05$ 。

5) 结论与解释: 由于 $P < \alpha$, 则拒绝原假设 H_0 , 接受备择假设 H_1 。

6) 实际推断: 根据 “小概率事件的实际不可能性原理”, 若药物无效(H_0 为真), 则观察到如此大的

降压差异(12.5 vs 5.0 mmHg)的概率极低($P < \alpha$), 因此, 认为 MediLower 的降压效果显著优于安慰剂。

例 8 证明“素数无穷多”。

证明: (利用严格反证法) [3]。假设素数有限, 设为 $\{q_1, q_2, \dots, q_n\}$, 令 $Q = q_1 q_2 \cdots q_n + 1$, 则 $Q = q_1 q_2 \cdots q_n + 1$ 不被任何已知素数整除, 故存在新素数, 与假设矛盾, 即素数有无限多个。

延伸思考: 归纳推理与演绎推理的哲学差异。小概率事件的实际不可能性原理[4]本质上是一种归纳推理, 其核心逻辑是“基于大量经验观察, 小概率事件在单次试验中几乎不会发生”, 但这种推理无法保证结论的绝对正确性, 即使事件概率极低, 仍存在发生的可能性(如 $\alpha = 0.05$ 对应的“第一类错误”)。归纳推理的优势在于适应不确定的现实世界, 能基于有限数据做出可接受的推断, 广泛应用于医学、社会学等无法进行完全枚举的领域。而严格反证法(如证明素数无穷多)属于演绎推理, 其逻辑基础是“若假设成立则会导致矛盾, 故假设必然不成立”, 推理过程遵循严格的逻辑规则, 结论具有绝对的必然性和普遍性。演绎推理的前提是必须有明确的公理或定义, 适用于数学、逻辑等抽象领域, 但难以直接应用于存在随机扰动的现实问题。两者的差异反映了“理论确定性”与“现实不确定性”的对立统一, 在概率论中, 既需要演绎推理构建严格的理论框架(如概率公理、期望定义), 也需要归纳推理将理论应用于实际决策(如假设检验、风险评估), 而小概率原理正是连接两者的关键桥梁[5]。

参考文献

- [1] 缪铨生, 等. 概率与数理统计[M]. 上海: 华东师范大学出版社, 2000.
- [2] 刘宣, 马海强. 关于两随机变量不相关概念的思考[J]. 大学数学, 2024, 40(4): 73-76.
- [3] 张建华. 数学分析中证明函数极限存在性的若干方法[J]. 景德镇学院学报, 2021, 36(3): 104-108.
- [4] 秦秉杰. 小概率原理及其应用[J]. 太原学院学报(自然科学版), 2019, 37(3): 18-21.
- [5] 吴宏锬, 梁璜. 小概率事件原理的一点补充[J]. 高等数学研究, 2009, 12(1): 84-85.