《概率论》的教学中容易被忽略的三个问题

———一个习题的启示

张立新[1]

(浙江大学数学系, 杭州310027)

2010年10月5日

摘要: 本文从一个经典的习题出发, 分析在《概率论》的学习和教学乃至概率统计研究中容易被忽略的三个问题, 这些问题涉及随机向量的函数及其分布, 数学期望的存在性.

关键词: 概率论, 随机向量的函数, 分布, 数学期望.

1 一个习题及其推广变形版本

习题1 设$\xi_1,\xi_2,\ldots,\xi_n$ 为n个正的独立随机变量, 服从相同分布, 密度函数为$f(x)$. 试证明: 对任意$1\le k\le n$ 有

$$\mathsf{E}\left[\frac{\xi_1+\cdots+\xi_k}{\xi_1+\cdots+\xi_n}\right]=\frac{k}{n}. \tag{1}$$

这是一个《概率论》中的经典习题或例题, 最早出于何处已经难以考证. 在李贤平(1997), 林正炎和苏中根(2001), S. M. Ross (2009), 苏淳(2003), 魏宗舒等(1983), 周概容(1984) 等中都有相似的例题或习题. 习题结论的证明也相当简单, 下面的证明来自林正炎和苏中根(2001): 为了简便, 我们记$S_k=\xi_1+\cdots+\xi_n$, $k=1,\ldots,n$. 因为

$$\frac{\xi_1}{S_n},\frac{\xi_2}{S_n},\ldots,\frac{\xi_n}{S_n}\ \text{同分布}, \tag{2}$$

[1]作者简介: 张立新(1968-), 浙江大学数学系教授、系副主任, 教育部新世纪人才, 发表了学术论文110多篇, 研究领域为概率统计. Email: stazlx@zju.edu.cn

所以

$$\mathsf{E}\left[\frac{\xi_1}{S_n}\right] = \mathsf{E}\left[\frac{\xi_2}{S_n}\right] = \cdots = \mathsf{E}\left[\frac{\xi_n}{S_n}\right]. \tag{3}$$

从而对任意$1 \le k \le n$ 有

$$\begin{aligned} n\mathsf{E}\left[\frac{\xi_k}{S_n}\right] =& \mathsf{E}\left[\frac{\xi_1}{S_n}\right] + \cdots + \mathsf{E}\left[\frac{\xi_n}{S_n}\right] \\ =& \mathsf{E}\left[\frac{\xi_1 + \cdots + \xi_n}{\xi_1 + \cdots + \xi_n}\right] = 1. \end{aligned}$$

因此

$$\mathsf{E}\left[\frac{\xi_k}{S_n}\right] = \frac{1}{n}, \quad k = 1, \ldots, n.$$

故

$$\mathsf{E}\left[\frac{S_k}{S_n}\right] = \mathsf{E}\left[\frac{\xi_1}{S_n}\right] + \cdots + \mathsf{E}\left[\frac{\xi_k}{S_n}\right] = \frac{k}{n}.$$

结论得证.□

然而这个简单的证明中暗藏了不少玄机, 不少教师和学生没有觉察到, 以至于人们在试图推广这一习题时犯了一些错误. 在一些新版的教科书和教学课件中. 这个习题已有不少不同类型的改版. 下面列出的是典型的几种, 为了避免出现不必要的争论, 它们的出处不一一列出.

版本1 设$\xi_1, \xi_2, \ldots, \xi_n$ 为n个正的独立同分布随机变量. 试证明*:* 对任意$1 \le k \le n$ 有*(1).*

版本2 设$\xi_1, \xi_2, \ldots, \xi_n$ 为任意n个正的同分布随机变量. 试证明*:* 对任意$1 \le k \le n$ 有*(1).*

版本3 设$\xi_1, \xi_2, \ldots, \xi_n$ 为任意n个独立同分布随机变量, 有密度函数$f(x)$. 试证明*:* 对任意$1 \le k \le n$ 有*(1).*

版本4 设$\xi_1, \xi_2, \ldots, \xi_n$ 为任意n个独立同分布随机变量, 它们的数学期望存在. 试证明*:* 对任意$1 \le k \le n$ 有*(1).*

版本5 设$\xi_1, \xi_2, \ldots, \xi_n$ 为任意n个独立同分布随机变量. 试证明*:* 对任意$1 \le k \le n$有*(1).*

版本6 设$\xi_1, \xi_2, \ldots, \xi_n$ 为任意n个非负的独立同分布随机变量. 试证明*:* 对任意$1 \le k \le n$ 有*(1).*

版本7 设$\xi_1, \xi_2, \ldots, \xi_n$ 为任意n个正的随机变量. 试证明*:* 对任意$1 \le k \le n$ 有*(1).*

其中版本2是最常见的一种.

2 容易被忽略的三个问题

在习题1的变形版本中肯定有不少是错误的, 有些错误是明显的. 例如:在版本7中,取$\xi_1 = \cdots = \xi_{n-1} \equiv 2$, 而$\xi_n \equiv 1$, 则结论(1) 显然不成立. 出现错误的原因是没有充分理解习题1的证明, 忽略了三个在学习和讲授《概率论》时容易忽略的问题. 下面我们从容易发现到难以发现的逐一进行分析. 通过这些分析, 就不难明白上述版本中哪些是正确的, 哪些是错误的.

问题1: $\frac{\xi_k}{S_n}$ 是随机变量吗? 或者说$\frac{\xi_k}{S_n}$有定义吗?

这是随机向量的实函数是否为随机变量的问题, 不难回答, 一般也不大容易在这个问题上犯错误. 但是仔细思考起来, 可能也会犯迷糊. 我们知道, 随机向量的函数是随机变量, 严格地说, 如果$f(x, y)$是Borel可测函数, (X, Y)是随机向量, 那么$f(X, Y)$是随机变量. 又知道, 连续函数,或者分段(分区域)连续的函数, 以及它们的极限都是Borel可测函数. 现在$f(x, y) = \frac{x}{y}$ 当$y > 0$ 或$y < 0$是都是连续的, $\frac{\xi_k}{S_n} = f(\xi_k, S_n)$不自然是随机变量吗? 但是我们忽略了$y = 0$的情形. 很明显, 若$S_n = 0$, 那么会出现$\frac{\cdot}{0}$型. 这时$\frac{\xi_k}{S_n}$没有定义. 当然, 如果$\xi_1, \xi_2, \ldots, \xi_n$ 为独立连续型随机变量, 那么$S_n = \xi_1 + \cdots + \xi_n$ 也是连续型随机变量, 这时$\mathsf{P}(S_n = 0) = 0$, 从而$\frac{\xi_k}{S_n}$几乎处处有定义. 而当$\xi_1, \xi_2, \ldots, \xi_n$为离散型随机变量, 就可能出现没有定义的情况. 例如,

若$\xi_1, \xi_2, \ldots, \xi_n$相互独立, 且均服从两点分布$b(1,p)$, 这时$S_n$服从二项分布$b(n,p)$, $\frac{\xi_k}{S_n}$没有定义的概率为p^n. 采用版本3的教师可能注意到了这一点, 他们不敢象采用版本4-7的教师那样做更大胆的推广.

在初等概率论中, 我们遇到的随机向量的函数变换常常是初等函数变换, 这样只要函数有定义, 随机向量的函数就是随机变量. 但是, 有时候验证样本空间上的一个函数是否为随机变量并不那么直接简单. 例如, 在数理统计中, 未知参数的估计量首先必须是随机变量, 不然就没有分布和以此为出发点的统计推断可言, 而要验证极大似然估计或者似然方程的解是随机变量就不那么直接简单, 很少有教科书涉及, 有兴趣的读者可以参看陈希孺(1999, Page 156-157)和R.J. Serfling (1980, Page 147-148).

问题2: $\frac{\xi_k}{S_n}, k = 1, \ldots, n$ 真的同分布吗?

$\frac{\xi_k}{S_n}, k = 1, \ldots, n$ 同分布是证明(1)的关键. 如果$\xi_1, \xi_2, \ldots, \xi_n$同分布, 那么$\frac{\xi_1}{S_n}, \frac{\xi_2}{S_n}, \ldots, \frac{\xi_n}{S_n}$同分布似乎应该不成问题. 正因为如此, 版本2广泛出现在新版和一些旧版教科书中. 这个问题涉及随机向量的函数的分布的问题. 不少教师和学生认为, 只要X_1与Y_1同分布, X_2与Y_2同分布, 就有$f(X_1, X_2)$与$f(Y_1, Y_2)$同分布. 事实上, 在一般情况下, 只有(X_1, X_2)和(Y_1, Y_2) 同分布, 即X_1, X_2的联合分布与Y_1, Y_2的联合分布相同, 才能保障$f(X_1, X_2)$与$f(Y_1, Y_2)$同分布. 联合分布问题常常在学习和讲授《概率论》时容易被忽略. 例如在某些考试或习题中出现了这样的题目: "设ξ服从参数为λ的指数分布, η服从二项分布$b(n,p)$, 求ξ与η的相关系数", "设$X_1, X_2, \ldots, X_n$ 均来自正态分布$N(\mu, \sigma^2)$总体, 证明样本均值与样本方差相互独立", 等等. 很明显, 在前者中没有ξ和η的联合分布是无法求得它们的相关系数的, 在后者中还应该假设$X_1, X_2, \ldots, X_n$相互独立.

在概率统计的研究中, 联合分布问题有时也容易被忽略, 从而导致错误. 例如在研究次序统计量、经验过程、分位数过程时, 类似于下面的结论常常被利用. 设$U_1, U_2, \ldots$为一列独立同分布的随机变量, 均服从$[0,1]$上的均匀分布,

记$U_{n,1},\ldots,U_{n,n}$ 为$U_1,\ldots,U_n$的次序统计量, 则对每个n,

$$(U_{n,1},\ldots,U_{n,n}) \text{ 与 } \left(\frac{T_1}{T_{n+1}},\ldots,\frac{T_n}{T_{n+1}}\right) \text{ 同分布}, \tag{4}$$

其中$T_k=E_1+\cdots+E_k$, $\{E_k;k\geq 1\}$为一列独立的指数随机变量, 均服从均值为1的指数分布. 这是一个将次序统计转化为独立随机变量之和的有效工具. 然而, 一些学者就直接用T_k/T_{n+1}代替$U_{n,k}$, 即"不妨设$U_{n,k}=T_k/T_{n+1}$". 这就发生了错误, 因为(4)不能保证

$$\text{组列 } \{U_{n,k}:k=1,\ldots,n;n\geq 1\} \text{ 与组列 } \left\{\frac{T_k}{T_{n+1}}:k=1,\ldots,n;n\geq 1\right\} \text{ 同分布}. \tag{5}$$

事实上, (5) 是不成立的. 因此, 次序统计量的这样一种转换手法在讨论弱收敛时十分有效而在研究强收敛时却常常失效.

在本文讨论的习题1中, $\xi_1,\xi_2,\ldots,\xi_n$同分布不能保证$\frac{\xi_1}{S_n},\frac{\xi_2}{S_n},\ldots,\frac{\xi_n}{S_n}$同分布. 例如, 若$\xi_1=\cdots=\xi_{n-1}$, 而$\xi_n$ 与ξ_1独立同分布, 且都为正的随机变量, 那么很明显$\xi_1,\xi_2,\ldots,\xi_n$同分布. 但是, 只要注意到ξ_1/ξ_n 与ξ_n/ξ_1 同分布, 且

$$\frac{\xi_1}{S_n}=\frac{1}{n+\xi_n/\xi_1},\quad \frac{\xi_n}{S_n}=\frac{1}{1+n\xi_1/\xi_n},$$

就知道$\frac{\xi_1}{S_n}$与$\frac{\xi_n}{S_n}$不同分布.

在习题1中, 保障$\frac{\xi_1}{S_n},\frac{\xi_2}{S_n},\ldots,\frac{\xi_n}{S_n}$同分布的关键是$\xi_1,\xi_2,\ldots,\xi_n$具有可交换性, 即如果$(i_1,\ldots,i_n)$是$(1,\ldots,n)$的一个置换, 那么向量$(\xi_{i_1},\ldots,\xi_{i_n})$与向量$(\xi_1,\ldots,\xi_n)$同分布, 这是因为它们的联合分布函数都是$F(x_1)\cdots F(x_n)$, 其中$F(x)$是$\xi_k(k=1,\ldots,n)$的分布函数. 由可交换性知

$$(\xi_{i_1},\xi_{i_1}+\cdots+\xi_{i_n}) \text{ 与 } (\xi_1,\xi_1+\cdots+\xi_n) \text{ 同分布}$$

和

$$\frac{\xi_{i_1}}{\xi_{i_1}+\cdots+\xi_{i_n}} \text{ 与 } \frac{\xi_1}{\xi_1+\cdots+\xi_n} \text{ 同分布}.$$

从而,$\frac{\xi_1}{S_n}, \frac{\xi_2}{S_n}, \ldots, \frac{\xi_n}{S_n}$同分布. 值得注意的是, 在推导$\xi_1, \xi_2, \ldots, \xi_n$具有可交换性时, 既利用了$\xi_1, \xi_2, \ldots, \xi_n$的相互独立性, 也利用了它们的同分布性.

由上面的阐述可见, " $\xi_1, \xi_2, \ldots, \xi_n$独立同分布" 这一个条件可以换为更一般的条件"$\xi_1, \xi_2, \ldots, \xi_n$可交换". 是不是只要$\xi_1, \xi_2, \ldots, \xi_n$可交换并且$\mathsf{P}(S_n = 0) = 0$, 就有结论(1)呢? 这就涉及最后一个容易被忽略的问题.

问题3: $\frac{\xi_k}{S_n}$, $k = 1, \ldots, n$ 的数学期望存在吗?

在引入数学期望的定义时, 我们会强调要验证数学期望的存在性, 即绝对可积性. 而在应用甚至研究中常常会忘记这一点. 例如在一些统计文献中, 会出现这样的推导

$$\text{由于} Y_n = o_P(a_n), \quad \text{所以} \mathsf{E}Y_n = o(a_n).$$

且不说依概率收敛(或者几乎处处收敛)不能保证数学期望的收敛, 连数学期望的存在性都常常缺乏验证.

在习题1中,由于$\xi_1, \ldots, \xi_n$是正的随机变量, 我们有$0 < \xi_k \le S_n$. 所以$0 \le \frac{\xi_k}{S_n} \le 1$, 这就保证了$\frac{\xi_k}{S_n}$的数学期望是存在的. 如果去掉"随机变量为正"这一条件, 期望的存在性就会出问题. 以$n = 2$为例. 设ξ_1, ξ_2为独立的标准正态随机变量, 则ξ_2/ξ_1服从Cauchy分布, 密度函数为

$$\frac{1}{\pi(1 + x^2)}.$$

从而$\frac{\xi_1+\xi_2}{\xi_1} = 1 + \frac{\xi_2}{\xi_1}$的密度函数为

$$\frac{1}{\pi(1 + (x-1)^2)}.$$

因此$\frac{\xi_1}{\xi_1+\xi_2}$的密度函数为

$$\frac{1}{\pi(1 + (1/x - 1)^2)} \frac{1}{x^2} = \frac{1}{\pi(x^2 + (x-1)^2)}, \quad x \ne 0.$$

显然, $\frac{\xi_1}{\xi_1+\xi_2}$的数学期望不存在. 值得注意的是, 上面的例子也说明, 即使$\xi_1, \ldots, \xi_n$的数学期望都存在, $\frac{\xi_k}{S_n}$ 的数学期望也不一定存在, $k = 1, \ldots, n$.

3 习题1的解答与启示

根据上一节的讨论, 在习题的推广版本1-7中, 只有版本1是正确的. 版本1所涉及的随机变量不局限于连续型, 可以是一般的随机变量. 这样, 似乎版本1是这一习题的最佳表述.

但是, 为了证明(1), 我们除了要证明$\frac{\xi_k}{S_n}$的期望存在性外, 还要证明$\frac{\xi_k}{S_n}$, $k=1,\ldots,n$, 同分布. 为了证明同分布, 如前面所述, 我们需要一个性质, 那就是, 如果随机向量$\boldsymbol{X}$与$\boldsymbol{Y}$同分布, $g(\boldsymbol{x})$为Borel可测函数(或者连续函数), 那么$g(\boldsymbol{X})$与$g(\boldsymbol{Y})$同分布. 但是这一性质的严格证明需要测度论的知识, 对于只有数学分析或微积分背景知识的低年级学生来说是不易被理解的, 在不少初等概率论中不提及这一个性质. 当然, 证明也可以对连续型随机变量和离散型随机变量分别进行, 但这不是版本1的严格证明, 因为随机变量不止连续型和离散型两种. 因此, 从严格意义上来说, 版本1在初等概率论的范畴是没有解答的.

然而, 如果象在习题1中一样, 把随机变量局限为连续型的随机变量(或者离散型随机变量), 就可以回避一些复杂的知识背景. 下面我们给出习题1的解答.

因为

$$\begin{aligned}\mathsf{E}\left|\frac{\xi_k}{S_n}\right| &= \int_0^\infty\cdots\int_0^\infty\left|\frac{x_k}{x_1+\cdots x_n}\right| f(x_1)\cdots f(x_n)\,dx_1\cdots dx_n\\ &\le \int_0^\infty\cdots\int_0^\infty 1\cdot f(x_1)\cdots f(x_n)\,dx_1\cdots dx_n = 1.\end{aligned}$$

所以$\frac{\xi_k}{S_n}$的数学期望存在. 又

$$\begin{aligned}\mathsf{E}\left[\frac{\xi_k}{S_n}\right] &= \int_0^\infty\cdots\int_0^\infty \frac{x_k}{x_1+\cdots x_n} f(x_1)\cdots f(x_n)\,dx_1\cdots dx_n\\ &= \int_0^\infty\cdots\int_0^\infty \frac{y_1}{y_1+\cdots y_n} f(y_1)\cdots f(y_n)\,dy_1\cdots dy_n = \mathsf{E}\left[\frac{\xi_1}{S_n}\right].\end{aligned}$$

这就证明了(3). 由此易得(1). □

由此可见, 习题1是经过精心设计的, 它巧妙地考察了学生运用数学期望的线性性质的能力, 又考虑到了学生的知识背景. 我们知道, 概率论中的许多概念只有在测度论的框架下才有严格的定义, 不少性质和定理也只有在测度论的框架下才有严格的证明. 在初等概率论的教学中, 既要让学生掌握概率论的本质内容又兼顾到数学的严谨性和学生的知识背景不是一个十分容易的工作, 任课教师既要有深厚的高等概率论基础, 又要掌握将高等概率论思想融贯在初等概率论教学之中的技巧.

参考文献:

[**1**] 概率论. 第一册, 概率论基础, 复旦大学编, 复旦大学出版社, 1979

[**2**] 高等数理统计, 陈希孺编, 中国科学技术大学出版社, 1999.

[**3**] 概率论, 第二版, 林正炎和苏中根编著, 浙江大学出版社, 2001.

[**4**] *A First Course in Probability, Seventh Edition*, S. M. Ross Eds., Pearson Education, Inc, 2009.

[**5**] *Approximation Theorem of Mathematical Statistics*, R.J. Serfling, John Wiley & Sons, INC, 1980.

[**6**] 概率论, 苏淳编著, 科学出版社, 2003.

[**7**] 概率论与数理统计教程, 魏宗舒等编著, 高等教育出版社, 1983.

[**8**] 概率论与数理统计教程, 周概容编著, 高等教育出版,1984.