

ISSN 1008-1399
CN61-1315/O1

- 国家自然科学基金委员会数学天元基金支持刊物
- 《中国期刊网》、《中国学术期刊》(光盘版)全文收录期刊
- 《中国学术期刊综合评价数据库》来源期刊
- 《万方数据-数字化期刊群》、《中文科技期刊数据库》全文收录期刊
- 《中国核心期刊(遴选)数据库》收录期刊
- 《CEPS思博网-中文电子期刊服务》收录期刊
- 《中国数学文摘》、《中国数学文献数据库》收录期刊
- 2000年获“陕西省优秀科技期刊一等奖”
- 国家新闻出版广电总局第一批认定学术期刊



高等数学研究

陳育身題

GAODENG SHUXUE YANJIU
STUDIES IN COLLEGE MATHEMATICS



2020 4

ISSN 1008-1399



第23卷 第4期 (总第198期)

2020年7月出版

西北工业大学 陕西省数学会 主办

高等数学研究

(双月刊)

第二十三卷第四期

(总第198期)

二〇二〇年七月

目 次

专题研究

- 先于极限的微积分中引入连续性 林 群, 童增祥, 张景中 (001)
- 从教育数学的角度探讨行列式教学 曾振柄, 黄 勇, 饶永生 (010)

探讨与研究

- 分形集上的 Iyengar 型不等式和 Ostrowski-Iyengar 型不等式 时统业, 曾志红 (022)
- 求自然数幂和的差分公式法 戴中林 (029)

推广与应用

- 微元法在一些积分问题中的应用 叶正麟, 林 伟 (032)
- 拉格朗日中值定理逆命题成立的一个充要条件 王银坤, 倪谷炎 (038)
- 均匀分布的一个性质及其应用 安 军 (041)
- 二项分布两种近似分布的收敛速度 李 斐, 陈晓娜 (044)
- 关于一元含参量变上限积分的等价无穷小 杨 芮 (047)
- 关于特征值反问题的唯一性 潘朝毅, 马玉雯 (050)
- 基于线性方程组同解证明矩阵秩相等的方法 张立卓 (052)
- 指数函数的七种等价定义 李翠香, 王梦娜 (054)
- 点到直线距离公式的 n 维推广 代丽芳, 梁茂林, 刘 薇 (057)

教学随议

- 期望定义中绝对收敛条件剖析 王红军 (059)
- 矩阵的几种关系浅析 尹小艳, 杨丹丹 (061)
- 关于离散数学教材中一个命题的注 谭尚旺, 宁文杰 (065)
- 条件数学期望及其应用的教学研究 林焱辉 (067)
- 案例教学法在概率论教学中应用 薛文娟, 沈 群, 许 冰 (070)
- 积分变换若干问题析疑 吴克坚, 刘 烁, 徐清华, 王瑞星, 赵清波 (073)
- 求解标准正态分布的特征函数的方法探讨和扩展 吴贤君 (077)
- 考研数学试题中的概率积分 王彭德, 戈 芳 (079)
- 思维导图——高效学习的得力助手 文生兰, 刘 倩, 韩艺兵 (082)
- 基于 AHP 法的高等数学课程教学方法的评价分析——以山西农业大学为例 张小英, 郝新生, 何云峰 (085)
- 矩阵的“秩”探源 王继强 (088)

方法与技巧

- 一道全国大学生数学竞赛决赛试题的另解 陈桂东 (090)
- 一题多解在线性代数教学中的作用 李为芹 (092)
- 矩阵秩的性质及秩不等式的五种证法 安 军 (096)
- 关于确定两个非独立随机变量和分布方法的探讨 张立卓, 张煦渤 (100)

教学改革

- 大数据背景下新建工科院校高等数学课程体系改革研究 周立新 (104)
- 依据学生学习现状 实施高等数学教学改革 王存荣 (112)
- 高等数学课程思政方法研究 齐新社, 李 国, 王 欣, 高翠翠 (118)
- 课程思政理念下数学史驱动的常数项级数教学设计 叶建兵 (120)
- 浅析高等数学中蕴含的思维观和方法论 滕吉红, 鲁志波, 黄晓英 (124)

doi:10.3969/j.issn.1008-1399.2020.04.008

二项分布两种近似分布的收敛速度

李 斐, 陈晓娜

(烟台大学 数学与信息科学学院, 山东 烟台 264005)

摘 要 以矩母函数为工具推导了二项分布 $B(n, p_n)$ 的泊松逼近的收敛速度, 指出其收敛速度与 np_n 收敛于 λ 的速度有关, 并与正态逼近的收敛速度做了对比. 最后做了模拟计算, 验证了定理的结果.

关键词 矩母函数; 泊松逼近; 二项分布; 收敛速度

中图分类号 O212.1 文献标识码 A 文章编号 1008-1399(2020)04-0044-03

Convergent Speeds of Two Approximations to the Binomial Distribution

LI Fei and CHEN Xiaona

(School of Mathematics and Information Science, Yantai University, Yantai 264005, China)

Abstract The convergent speed of the Poisson approximation to the Binomial distribution is studied using the tool of the moment generating function and is compared with the normal approximation. Simulations are used to verify the results.

Keywords moment generating function, Poisson approximation, Binomial distribution, convergent speed

1 引言

我们引入服从两点分布的随机变量 X_{ij} , 事件 A 发生时, $X_{ij} = 1$, 反之 $X_{ij} = 0$, 且 $P(X_{ij} = 1) = p_i$. 将 X_{ij} 写成如下形式, 注意到每一行的随机变量是独立同分布的.

$$\begin{aligned} & X_{11}, \\ & X_{21}, X_{22}, \\ & X_{31}, X_{32}, X_{33}, \\ & \dots, \\ & X_{n1}, X_{n2}, X_{n3}, \dots, X_{nm}. \end{aligned}$$

令 $X_n = \sum_{j=1}^n X_{nj}$, 则 X_n 服从二项分布 $B(n, p_n)$. 对于二项分布的计算, 我们知道可以用泊松分布和正态分布两种分布来近似计算. 当 $\lim_{n \rightarrow \infty} np_n = \lambda$ 时, 文献[1]指出此时泊松分布要好于正态分布. 很多文献(文献[2, 3]等)都介绍过二项分布的泊松逼近, 但鲜有文献涉及到泊松逼近的收敛速度. 文献[3]

利用矩母函数和累积生成函数推导出了 Edgeworth 展开式, 揭示了正态逼近的收敛速度. 受此启发, 本文借助矩母函数给出了二项分布泊松逼近的收敛速度, 并同正态逼近做了对比, 精确比较了在不同情况下两种近似分布的优劣.

2 泊松逼近的收敛速度

我们分两种情况来讨论泊松逼近的收敛速度, 分别为 $np_n = \lambda$, 以及 $\lim_{n \rightarrow \infty} np_n = \lambda$.

2.1 考虑 $np_n = \lambda$ 的情况

利用泊松分布、二项分布的矩母函数, 以及傅里叶变换等工具, 我们得到二项分布的泊松逼近的收敛速度如下.

定理 1 若 X_n 服从二项分布 $B(n, p_n)$, 且 $np_n = \lambda$. 令 $F_n(x)$ 表示 X_n 的分布函数, $G(x)$ 为 $P(\lambda)$ 的分布函数. 则有

$$F_n(x) = G(x) + O\left(\frac{1}{n}\right). \quad (1)$$

证明 X_{ni} 的矩母函数为

$$\psi_{X_{ni}}(t) = Ee^{tX_{ni}} = 1 + p_n(e^t - 1),$$

注意到 $\lim_{n \rightarrow \infty} p_n = 0$, 故对矩母函数取对数, 利用 $\log(1+x)$ 的幂级数展开得,

收稿日期: 2019-12-23 修改日期: 2020-02-10

基金项目: 国家自然科学基金(61503318).

作者简介: 李斐(1982-), 女, 山东烟台, 博士, 讲师, 研究方向为多元统计分析, Email: feili@ytu.edu.cn.

$$\begin{aligned} \log(\psi_{X_m}(t)) &= \log(1 + p_n(e^t - 1)) \\ &= p_n(e^t - 1) - \frac{1}{2}(p_n(e^t - 1))^2 \\ &\quad + \frac{1}{3}(p_n(e^t - 1))^3 + \dots, \end{aligned}$$

故有,

$$\begin{aligned} \psi_{X_m}(t) &= \exp\{p_n(e^t - 1) - \frac{1}{2}(p_n(e^t - 1))^2 \\ &\quad + \frac{1}{3}(p_n(e^t - 1))^3 + \dots\} \end{aligned}$$

因为 $X_n = \sum_{j=1}^n X_m$, 则 X_n 的矩母函数为

$$\begin{aligned} \psi_{X_n}(t) &= \exp\{np_n(e^t - 1) - \frac{1}{2}n(p_n(e^t - 1))^2 \\ &\quad + \frac{1}{3}n(p_n(e^t - 1))^3 + \dots\} \\ &= \exp\{\lambda(e^t - 1) - \frac{1}{2n}(\lambda(e^t - 1))^2 \\ &\quad + \frac{1}{3n^2}(\lambda(e^t - 1))^3 + \dots\} \\ &= \exp\{\lambda(e^t - 1) + \sum_{i=1}^{\infty} \sum_{j=2}^{\infty} a_{ij} \left(\frac{1}{n}\right)^i t^j\} \\ &= \exp\{\lambda(e^t - 1)\} \left[1 + \sum_{i=1}^{\infty} \sum_{j=2}^{\infty} b_{ij} \left(\frac{1}{n}\right)^i t^j\right] \end{aligned} \quad (2)$$

其中 a_{ij}, b_{ij} 分别为常数系数. 因为

$$\begin{aligned} \exp\{\lambda(e^t - 1)\} &= \exp\left\{\lambda t + \frac{\lambda t^2}{2} + \sum_{k=3}^{\infty} \frac{\lambda t^k}{k!}\right\} \\ &= \left(1 + \sum_{k=3}^{\infty} c_k t^k\right) \exp\left\{\lambda t + \frac{\lambda t^2}{2}\right\} \end{aligned} \quad (3)$$

将式(3)代入式(2), 得

$$\begin{aligned} \psi_{X_m}(t) &= e^{\lambda(e^t - 1)} + e^{\lambda t + \frac{\lambda t^2}{2}} \sum_{i=1}^{\infty} \sum_{j=2}^{\infty} r_{ij} \left(\frac{1}{n}\right)^i t^j \\ &= e^{\lambda(e^t - 1)} + e^{\lambda t + \frac{\lambda t^2}{2}} \sum_{i=1}^{\infty} \left(\frac{1}{n}\right)^i r_i(t). \end{aligned} \quad (4)$$

其中 $r_i(t) = \sum_{j=2}^{\infty} r_{ij} t^j$.

令 $G(x)$ 为 $P(\lambda)$ 的分布函数, $N(x)$ 表示正态分布 $N(\lambda, \lambda)$ 的分布函数. 不难验证,

$$\begin{aligned} \int_{-\infty}^{+\infty} e^{tx} dG(x) &= e^{\lambda(e^t - 1)}, \\ \int_{-\infty}^{+\infty} e^{tx} dN(x) &= e^{\lambda t + \frac{\lambda t^2}{2}}. \end{aligned}$$

以 $\frac{1}{n}$ 的一次项为例, 找到 $R_1(x)$, 使得下式成立,

$$\int_{-\infty}^{+\infty} e^{tx} dR_1(x) = \frac{1}{n} e^{\lambda t + \frac{\lambda t^2}{2}} r_1(t).$$

可验证,

$$R_1(x) = \frac{1}{n} r_1(\nabla)(N(x)) = \frac{1}{n} \left(\sum_{j=2}^{\infty} r_{1j} \nabla^j\right)(N(x)),$$

其中 ∇ 为求导算子. 同理可得到 $R_j(x), j = 1,$

2, ..., 满足

$$\int_{-\infty}^{+\infty} e^{tx} dR_j(x) = \frac{1}{n} e^{\lambda t + \frac{\lambda t^2}{2}} r_j(t).$$

对(4)式逆展开, 可得

$$F_n(x) = G(x) + R_1(x) + R_2(x) + \dots,$$

即得(1)式, 定理得证.

根据上述定理的结果, 泊松逼近的收敛速度为 $O\left(\frac{1}{n}\right)$. 由文献[3]的 Edgeworth 展开式(定理 1.

16) 可得, 正态逼近的收敛速度为 $O\left(\frac{1}{\sqrt{n}}\right)$. 所以当 $np_n = \lambda$ 时, 泊松逼近是好于正态逼近的.

2.2 考虑 $\lim_{n \rightarrow \infty} np_n = \lambda$ 的情况

当 $\lim_{n \rightarrow \infty} np_n = \lambda$ 时, 我们考虑泊松逼近的收敛速度应当与 np_n 收敛于 λ 的速度有关, 所以不妨设 $np_n = \lambda + O\left(\frac{1}{n^\alpha}\right)$, 则关于泊松逼近的收敛速度有如下定理.

定理 2 若 X_n 服从二项分布 $B(n, p_n)$, 且 $np_n = \lambda + O\left(\frac{1}{n^\alpha}\right), \alpha > 0$. 令 $F_n(x)$ 表示 X_n 的分布函数, $G(x)$ 为 $P(\lambda)$ 的分布函数. 有

$$F_n(x) = \begin{cases} G(x) + O\left(\frac{1}{n}\right), & \alpha \geq 1 \\ G(x) + O\left(\frac{1}{n^\alpha}\right), & \alpha < 1 \end{cases}. \quad (5)$$

证明 与式(2)类似的推导可得 X_n 的矩母函数为

$$\begin{aligned} \psi_{X_n}(t) &= \exp\{np_n(e^t - 1) - \frac{1}{2}n(p_n(e^t - 1))^2 \\ &\quad + \frac{1}{3}n(p_n(e^t - 1))^3 + \dots\} \end{aligned} \quad (6)$$

对(6)式作进一步化简, 并将 $np_n = \lambda + O\left(\frac{1}{n^\alpha}\right)$ 代入, 得,

$$\begin{aligned} \psi_{X_n}(t) &= \exp\{np_n(e^t - 1)\} \\ &\quad + \sum_{i=1}^{\infty} \sum_{j=2}^{\infty} \sum_{k=2}^{\infty} a_{ijk} \left(\frac{1}{n}\right)^i t^j (np_n)^k \\ &= \exp\{\lambda(e^t - 1)\} + O\left(\frac{1}{n^\alpha}\right)(e^t - 1) \\ &\quad + \sum_{i=1}^{\infty} \sum_{j=2}^{\infty} \sum_{k=2}^{\infty} a_{ijk} \left(\frac{1}{n}\right)^i t^j \left(\lambda + O\left(\frac{1}{n^\alpha}\right)\right)^k \\ &= \exp\{\lambda(e^t - 1)\} \left\{1 + \sum_{i=\alpha}^{\infty} \sum_{j=1}^{\infty} b_{ij} t^j O\left(\frac{1}{n^i}\right)\right\} \cdot \\ &\quad \left\{1 + \sum_{i=1}^{\infty} \sum_{j=2}^{\infty} c_{ij} O\left(\frac{1}{n^i}\right) t^j\right\} \\ &= \begin{cases} e^{\lambda(e^t - 1)} \left(1 + \sum_{i \geq \alpha}^{\infty} \sum_{j=1}^{\infty} r_{ij} O\left(\frac{1}{n^i}\right) t^j\right) & \alpha \leq 1 \\ e^{\lambda(e^t - 1)} \left(1 + \sum_{i=1}^{\infty} \sum_{j=2}^{\infty} r_{ij} O\left(\frac{1}{n^i}\right) t^j\right) & \alpha > 1 \end{cases}. \end{aligned}$$

采用定理 1 类似的方法,可证得(5)式.

由上述定理的结果,以及 Edgeworth 展开式,可以看出当 $\alpha > \frac{1}{2}$ 时,泊松逼近均好于正态逼近;而当 $\alpha = \frac{1}{2}$ 时,两者收敛速度差不多;当 $\alpha < \frac{1}{2}$ 时,正态逼近好于泊松逼近. 具体的比较结果见表 1.

表 1 两种近似分布的收敛速度比较

分布	$\alpha \geq 1$	$\frac{1}{2} < \alpha < 1$	$\alpha = \frac{1}{2}$	$\alpha < \frac{1}{2}$
泊松逼近	$O\left(\frac{1}{n}\right)$	$O\left(\frac{1}{n^\alpha}\right)$	$O\left(\frac{1}{\sqrt{n}}\right)$	$O\left(\frac{1}{n^\alpha}\right)$
正态逼近	$O\left(\frac{1}{\sqrt{n}}\right)$	$O\left(\frac{1}{\sqrt{n}}\right)$	$O\left(\frac{1}{\sqrt{n}}\right)$	$O\left(\frac{1}{\sqrt{n}}\right)$

3 模拟结果

针对上述两个定理的结果,随机模拟了几种情况下两种近似分布的收敛速度. 表 2 是当 $np_n = \lambda$ 时的情况,不妨取 X_n 服从 $B\left(n, \frac{8}{n}\right)$. 由表 2 可见,泊松分布在 $n=1000$ 时已经非常接近二项分布,而正态逼近在 $n=2000$ 时才接近二项分布.

表 2 $B\left(n, \frac{8}{n}\right)$ 两种近似分布的收敛速度比较

分布	$P(X_n \leq 1)$ $n=100$	$P(X_n \leq 5)$ $n=500$	$P(X_n \leq 10)$ $n=1\ 000$	$P(X_n \leq 15)$ $n=2\ 000$
二项分布	0.002319	0.189023	0.816685	0.991895
泊松分布	0.003019	0.191236	0.815886	0.991769
正态分布	0.004937	0.142479	0.761131	0.993428

当 $np_n = \lambda + O\left(\frac{1}{n^\alpha}\right)$ 时,我们模拟了 $\frac{1}{2} < \alpha < 1$ 时的情况,不妨取 $\alpha = \frac{6}{7}$,模拟结果见表 3. 可以看出,泊松逼近的效果要明显好于正态逼近,泊松逼近在 $n=500$ 时已非常接近二项分布.

表 3 $B\left(n, \frac{7}{n+n^{\frac{1}{7}}}\right)$ 两种近似分布的收敛速度比较

分布	$P(X_n \leq 1)$ $n=100$	$P(X_n \leq 5)$ $n=500$	$P(X_n \leq 10)$ $n=1\ 000$	$P(X_n \leq 15)$ $n=2\ 000$
二项分布	0.006809	0.303273	0.903548	0.997673
泊松分布	0.007295	0.300708	0.901479	0.997593
正态分布	0.010169	0.226573	0.874208	0.998798

表 4、表 5 分别是 $\alpha < \frac{1}{2}$ (取 $\alpha = \frac{1}{5}$), $\alpha = \frac{1}{2}$ 时的模拟情况,由表 4 可以看出此时泊松逼近的效果明显不如正态逼近(正态逼近在 $n=1000$ 时已非常接近二项分布). 由表 5 可以看出当 $\alpha = \frac{1}{2}$ 时,两者收敛速度差不多.

表 4 $B\left(n, \frac{7}{n+n^{\frac{4}{5}}}\right)$ 两种近似分布的收敛速度比较

分布	$P(X_n \leq 1)$ $n=100$	$P(X_n \leq 5)$ $n=500$	$P(X_n \leq 10)$ $n=1\ 000$	$P(X_n \leq 15)$ $n=2\ 000$
二项分布	0.036862	0.540115	0.972321	0.999688
泊松分布	0.007295	0.300708	0.901479	0.997593
正态分布	0.033086	0.425995	0.969098	0.999945

表 5 $B\left(n, \frac{8}{n+\sqrt{n}}\right)$ 两种近似分布的收敛速度比较

分布	$P(X_n \leq 1)$ $n=100$	$P(X_n \leq 5)$ $n=500$	$P(X_n \leq 10)$ $n=1\ 000$	$P(X_n \leq 15)$ $n=2\ 000$
二项分布	0.004649	0.222531	0.840272	0.993340
泊松分布	0.003019	0.191236	0.815886	0.991769
正态分布	0.007857	0.166571	0.790860	0.994915

最后,我们还模拟了 $\alpha > 1$ 时的情况,取 $\alpha = 2$,即 $X_n \sim B\left(n, \frac{7}{n+n^{-1}}\right)$,结果见表 6. 由表 6 可知,当 $n=200$ 时,泊松分布的值就已经很接近二项分布了,而在 $n=2000$ 时,正态分布才很接近二项分布,因此当 $\alpha > 1$ 时,二项分布的泊松逼近收敛速度快于正态逼近的收敛速度.

表 6 $B\left(n, \frac{7}{n+n^{-1}}\right)$ 两种近似分布的收敛速度比较

分布	$P(X_n \leq 1)$ $n=100$	$P(X_n \leq 5)$ $n=500$	$P(X_n \leq 10)$ $n=1\ 000$	$P(X_n \leq 15)$ $n=2\ 000$
二项分布	0.006017	0.296177	0.902228	0.997640
泊松分布	0.007295	0.300708	0.901479	0.997593
正态分布	0.009351	0.220812	0.872416	0.998773

参考文献

- [1] E. L. Lehmann, Elements of Large - Sample Theory [M], New York: Springer, 1999.
- [2] 李贤平. 概率论基础[M]. 北京: 高等教育出版社, 1997.
- [3] Jun Shao. Mathematical Statistics [M]. New York: Springer, 2003.

热忱欢迎订阅2020年《高等数学研究》

高校 科研院所图书馆、资料室必备参考资料, 数学
教师科研、教学的得力助手, 大学生学习数学的良师益友

《高等数学研究》是西北工业大学和陕西省数学会联合主办的以高校教师、大学生和科技工作者为主要的服务对象、国内外公开发行的数学期刊, 是1950年初期我国创办的几份数学杂志之一, 2014年国家新闻出版广电总局第一批认定的学术期刊。多年来, 刊物深受读者喜爱, 得到了关心数学教育的众多数学家、数学教育家、广大数学教育工作者的关心和支持, 在此谨向热情关爱支持本刊的专家、学者和广大作者、读者表示衷心的感谢和敬意!

本刊为双月刊, A4开本, 单月月底出版。2020年第2, 3, 5, 6期各64页, 定价9.0元; 第1, 4期各128页, 定价20元。全年6期总定价76元。一、二年级大学生读者, 可以只订阅其中与高等数学、线性代数、概率统计课程内容密切相关的4期, 即2020年第2, 3, 5, 6期, 共价36元。

订阅方式:

- (1) 可以在全国各地邮局订阅, 邮发代号: 52-192。
- (2) 直接汇款至本刊编辑部订阅。

近若干年的部分期刊尚有少量存量, 如有需要可直接与编辑部联系购买(含邮购)。

Studies in College Mathematics

(Bimonthly, Started in 1954)

Vol. 23, No. 4 (Serial No. 198), Jul. 2020

Sponsored by Northwestern Polytechnical University and Shaanxi Mathematical Society

Edited and published by Editorial Board of Studies in College Mathematics

(Northwestern Polytechnical University, Xi'an 710072, PRC)

Editor-in-Chief CUI Junzhi

Printed by Xi'an Kunming Printing House

Distributed by China International Book Trading Corporation

(P. O. Box 399, Beijing 100044, PRC)

CN 61-1315/O1/ISSN1008-1399

投稿邮箱: gdsxyj@yeah.net

在线投稿: <http://gdsxyj.cbpt.cnki.net>

邮发代号: 52-192

国内定价: 20.00元