

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

陳肖身題

GAODENG SHUXUE YANJIU
STUDIES IN COLLEGE MATHEMATICS



2021 1

第24卷 第1期 (总第201期)

2021年1月出版

西北工业大学 陕西省数学会 主办

ISSN 1008-1399



01>

9 771008 139214

高等数学研究

(双月刊)

第二十四卷第一期

(总第201期)

二〇二一年一月

目 次

方法与技巧

| | |
|------------------------------|----------------|
| 线性代数变换与仿射几何变换 | 叶正麟, 安晓虹 (001) |
| 计算期望的变形公式及其应用 | 王红军, 杨有龙 (005) |
| 应用正态分布的线性不变性研究二维正态分布性质 | 许忠好, 谭昕玥 (007) |
| 一类含三角函数级数求和的极限问题 | 赵 月, 沈荣鑫 (011) |
| 曲线曲面积分中利用对称性及曲线曲面方程化简 | 魏连鑫 (015) |
| 幂指数函数初探 | 郭艳鹏, 刘 妮 (018) |
| 利用特征矩阵求实对称矩阵的特征向量 | 孟宪萌, 牛 柯 (021) |

数学建模

| | |
|-----------------------------|-------------------------------|
| 高等院校数学课程教学质量的模糊综合评价研究 | 周 菲, 张 娜, 舒彩霞, 王 山, 李 硕 (024) |
|-----------------------------|-------------------------------|

推广与应用

| | |
|--------------------------------------|---------------------|
| 指数总体删失数据的极大似然估计 | 李 斐 (028) |
| 关于全期望和全概公式及其应用的教学研究——以新型冠状病毒为例 | 董迎辉 (031) |
| 抽屉原理构造方式的研究和演示 | 王丽丽, 王爱法 (033) |
| 矩阵若尔当标准形的一个新证明 | 刘淑霞, 麻常利, 曾丽伟 (036) |

学生园地

| | |
|-----------------------------|---------------------|
| 利用同时合同对角阵解决几类正定矩阵相关问题 | 刘潇奕 (040) |
| EP 矩阵的若干研究 | 王一涵, 陶 睿, 魏俊潮 (042) |

教学随笔

| | |
|-----------------------------|----------------|
| 无限维线性空间的注记 | 甘爱萍, 杨义川 (046) |
| 从不定积分概念的教学看数学课堂的思想性 | 刘 丹, 曹广福 (050) |
| 关于矩阵奇异值的若干性质 | 尹小艳, 杨丹丹 (056) |
| 从“分析”的角度探讨初、高等数学的衔接问题 | 张志平, 宁蒙蒙 (059) |
| 关于高等代数多项式理论的教学探讨 | 安 军 (063) |
| 一个不能被忽视的教学环节——结果分析 | 吴黎军, 鲁玉平 (068) |
| 实变函数中的集合论方法应用 | 方 益 (071) |
| 学习柯西积分定理及留数定理的体会 | 林志强 (074) |
| 函数的原函数存在与黎曼可积的关系 | 程 磊, 李 静 (077) |
| 关于微元法及其原理的探讨 | 李兴东, 张 睿 (080) |

教学改革

| | |
|---|------------------------------------|
| 概率论与数理统计慕课课件的构想与设计 | 张立卓 (084) |
| 关于本科几何课程教学方法的探讨 | 赵 唯, 陈云霞 (091) |
| “以学生为中心”基于多平台的线性代数在线教学探索与实践 | 杨 威, 李隐峰, 高淑萍 (095) |
| 河海大学 2020 版《高等数学》教学大纲修订的点滴体会 | 郑苏娟 (099) |
| 基于问题驱动—BOPPPS 教学模式在高等数学混合式教学的应用 | 张 丽, 武 燕 (104) |
| 微时代背景下师范生实践能力的培养研究——以本校《实变函数》课程翻转课堂为例 | 崔亚琼, 康淑瑰, 陈慧琴 (107) |
| 《概率论与数理统计》课程在管理类专业教学方法实践 | 于 淼, 吴素文, 王 倩, 宋 贲, 冯大光, 边振兴 (111) |
| 《初等数论》课题化教学方法研究——以中国剩余定理为例 | 路秀华, 张全雷 (116) |
| 基于经济类专业需求的高等数学教学改革探究 | 袁学帅 (120) |
| 公共数学全英文授课模式探析 | 文传军, 金 政, 任雪静 (124) |
| 数学竞赛试题选 | (119) |
| 2021 年科学突破奖 | (123) |
| 2020 年教育数学会 | (123) |

指数总体删失数据的极大似然估计

李 斐

(烟台大学 数学与信息科学学院, 山东 烟台 264005)

摘 要 对取自均值为 λ 的指数分布的缺失数据, 在完全随机缺失、右删失、非随机缺失三种缺失机制下, 分别推导了参数的极大似然估计, 并做了随机模拟.

关键词 极大似然估计; 指数分布; 缺失数据

中图分类号 O212.1 文献标识码 A 文章编号 1008-1399(2021)01-0028-03

Maximum Likelihood Estimation of Missing Data for Exponential Population

LI Fei

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

Abstract The maximum likelihood estimation of missing data for exponential population with mean λ is derived under three missing mechanisms, missing completely at random, right censoring, and missing not at random. The simulation results are also presented.

Keywords maximum likelihood estimation, exponential population, missing data

1 引言

Roderick J. A. 在文献[1]中将数据的缺失机制分为了三种, 简单来说, 就是缺失概率与所有变量无关的完全随机缺失(MCAR); 缺失概率与缺失变量无关的随机缺失机制(MAR); 缺失概率与缺失变量有关的非随机缺失(NMAR). 对缺失数据的处理主要有去掉缺失数据、插值方法和基于模型的处理方法, 例如极大似然方法. 文献[1]和[2]给出了正态总体缺失数据处理以及一些具体的实例.

本文给出了取自指数分布总体的缺失数据在完全随机缺失, 右删失和非随机缺失三种缺失机制下的参数的极大似然估计. 所得到的参数估计不仅包括指数分布均值的估计, 还有不同区间上的缺失概率的估计, 以及不同缺失区间的分段点的估计. 文

章最后给出了随机模拟, 验证上述估计的正确性, 随之样本容量的增大, 估计是越来越接近未知参数的真值.

2 不同删失情形下的极大似然估计

设总体 X 服从参数为 λ 的指数分布, 即 $X \sim e(\lambda)$. 取自该总体的容量为 n 的样本的一组删失数据的观测值, 不妨记为

$$x_1, x_2, \dots, x_{n_0}, y_{n_0+1}, y_{n_0+2}, \dots, y_{n_0+n_1},$$

其中 x_1, x_2, \dots, x_{n_0} 为观测到的 n_0 个数据, y_1, y_2, \dots, y_{n_1} 为缺失的 n_1 个数据, $n_0 + n_1 = n$. 随机变量 R 为缺失因子, $R=1$ 表示数据缺失, 反之 $R=0$ 为数据未缺失. 则对于上述观测值有,

$$R_i = 0, i = 1, 2, \dots, n_0,$$

$$R_i = 1, i = n_0 + 1, n_0 + 2, \dots, n.$$

令

$$P(R=1|x) = m(x)$$

表示缺失概率函数. 则上述观测结果发生的可能性的大小为

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.

$$\begin{aligned}
 & f(x_1, x_2, \dots, x_{n_0}, R_1 = 0, \dots, R_{n_0} = 0, \\
 & R_{n_0+1} = 1, \dots, R_{n_0+n_1} = 1) \\
 &= \prod_{j=n_0+1}^n P(R_j = 1) \prod_{i=1}^{n_0} f(x_i, R_i = 0) \\
 &= (P(R = 1))^{n_1} \prod_{i=1}^{n_0} (f(R_i = 0 | x_i) f(x_i)).
 \end{aligned}$$

注意到,

$P(R = 1) = E(R) = E(E(R | X)) = E(m(X))$, 结合指数分布的密度函数, 上述观测结果的似然函数为

$$L(\lambda) = (E(m(X)))^{n_1} \lambda^{n_0} e^{-\lambda \sum_{i=1}^{n_0} x_i} \prod_{i=1}^{n_0} (1 - m(x_i)). \quad (1)$$

下面, 我们对于几种常见的缺失形式分别给出参数的极大似然估计.

2.1 缺失方式为完全随机缺失

当缺失方式为完全随机缺失时, 即

$$P(R = 1 | x) = p, \quad (2)$$

此时缺失概率与变量值无关. 似然函数(1)可化为

$$L(\lambda, p) = p^{n_1} \lambda^{n_0} e^{-\lambda \sum_{i=1}^{n_0} x_i} (1 - p)^{n_0}.$$

其对数似然函数为

$$\ln(L(\lambda, p)) = n_1 \ln p + n_0 \ln \lambda - \lambda \sum_{i=1}^{n_0} x_i + \ln(1 - p),$$

分别关于 λ, p 求偏导, 建立似然方程组如下,

$$\begin{cases} \frac{\partial \ln(L(\lambda, p))}{\partial \lambda} = \frac{n_0}{\lambda} - \sum_{i=1}^{n_0} x_i = 0 \\ \frac{\partial \ln(L(\lambda, p))}{\partial p} = \frac{n_1}{p} - \frac{1}{1-p} = 0 \end{cases},$$

求解得 λ, p 的极大似然估计为,

$$\begin{cases} \hat{\lambda} = \frac{1}{n} \\ \hat{p} = \frac{n_1}{n} \end{cases}. \quad (3)$$

2.2 缺失方式为右删失

当缺失方式为右删失时, 即

$$P(R = 1 | x) = \begin{cases} 1 & x > c \\ 0 & x \leq c \end{cases}, \quad (4)$$

此时数据为右删失数据. 似然函数(1)可化为

$$L(\lambda, p) = P(X > c)^{n_1} \lambda^{n_0} e^{-\lambda \sum_{i=1}^{n_0} x_i} = e^{-\lambda c n_1} \lambda^{n_0} e^{-\lambda \sum_{i=1}^{n_0} x_i},$$

$$x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n_0)} \leq c,$$

其中 $x_{(i)}, i = 1, \dots, n_0$ 为将观测值按从小到大的顺序排列第 i 位的观测值. 其对数似然函数为

$$\ln(L(\lambda, p)) = -\lambda c n_1 + n_0 \ln \lambda - \lambda \sum_{i=1}^{n_0} x_i, \quad (5)$$

由(5)式易知, 对数似然函数为 c 的单调递减函数, 故 c 的极大似然估计为

$$\hat{c} = x_{(n_0)}.$$

(5)式关于 λ 求导, 建立似然方程如下,

$$\frac{d \ln(L(\lambda, p))}{d \lambda} = -c n_1 + \frac{n_0}{\lambda} - \sum_{i=1}^{n_0} x_i = 0,$$

求解得 λ, c 的极大似然估计为,

$$\begin{cases} \hat{\lambda} = \frac{n_0}{x_{(n_0)} n_1 + \sum_{i=1}^{n_0} x_i} \\ \hat{c} = x_{(n_0)} \end{cases}. \quad (6)$$

2.3 非随机缺失方式

考虑更为一般的缺失方式,

$$P(R = 1 | x) = \begin{cases} p & x > c \\ q & x \leq c \end{cases}, \quad (7)$$

不妨设 $p > q$, 将观测值按从小到大的顺序重新排序为 $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n_0)}$. 此时似然函数为

$$L(\lambda, p, q, c) = (p e^{-\lambda c} + q(1 - e^{-\lambda c}))^{n_1} \lambda^{n_0} e^{-\lambda \sum_{i=1}^{n_0} x_i} (1 - p)^{n_0 - \sum_{i=1}^{n_0} I_{\{x_i \leq c\}}} (1 - q)^{\sum_{i=1}^{n_0} I_{\{x_i \leq c\}}}. \quad (8)$$

求(8)式的最大值, 可令 c 分别取 $x_{(i)}, i = 1, \dots, n_0$,

此时 $\sum_{i=1}^{n_0} I_{\{x_i \leq c\}} = i$, 求 $L(\lambda, p, q, x_{(i)})$ 的最大值, 即

$$\begin{cases} \frac{\partial}{\partial p} L(\lambda, p, q, x_{(i)}) = 0 \\ \frac{\partial}{\partial q} L(\lambda, p, q, x_{(i)}) = 0, \\ \frac{\partial}{\partial \lambda} L(\lambda, p, q, x_{(i)}) = 0 \end{cases}$$

经过化简得,

$$\begin{cases} p = \frac{n_1 q + q i - n_1}{n q - n + i} \\ e^{\lambda x_{(i)}} = \frac{n(1-q)}{n - n q - i} \\ -n_1 x_{(i)} + x_{(i)} n q + \frac{n_0}{\lambda} - \sum_{i=1}^{n_0} x_i = 0 \end{cases},$$

该方程组含超越方程, 可通过 matlab 求出数值解, 记为 $\hat{\lambda}_i, \hat{p}_i, \hat{q}_i$. 此时似然函数的最大值为 $L(\hat{\lambda}_i, \hat{p}_i,$

$\hat{q}_i, x_{(i)}$). 易找到 k , 使得

$$L(\hat{\lambda}_k, \hat{p}_k, \hat{q}_k, x_{(k)}) = \max_{i=1, \dots, n_0} L(\hat{\lambda}_i, \hat{p}_i, \hat{q}_i, x_{(i)}),$$

则 λ, p, q, c 的极大似然估计分别为

$$\hat{\lambda} = \lambda_k, \hat{p} = p_k, \hat{q} = q_k, \hat{c} = x_{(k)}.$$

3 模拟结果

取均值为 $2(\lambda = 0.5)$ 的指数分布, 分别按照下述三种缺失方式产生相应的随机数, 并求参数的极大似然估计.

(1) 当缺失方式为完全随机缺失时, 不妨取式(2)中 $p=0.3$. 先产生服从均值为 2 的指数分布的 n 个随机数, 再按此种缺失方式产生一组缺失数据, 代入到(3)式中计算估计值, 具体计算结果见表 1. 可见当 $n=100$ 时估计值已经比较接近参数的真值.

表 1 $p=0.3, \lambda=0.5$ 模拟结果

| 参数 | $n=50$ | $n=100$ | $n=1000$ |
|-----------|--------|---------|----------|
| p | 0.28 | 0.3000 | 0.3020 |
| λ | 0.5683 | 0.5183 | 0.5072 |

(2) 当缺失方式为右删失时, 不妨取式(4)中 $c=2.8$. 先产生服从均值为 2 的指数分布的 n 个随机数, 再按此种缺失方式产生一组缺失数据, 代入到(6)式中计算估计值, 具体计算结果见表 2. 可见当 $n=1000$ 时估计值比较接近参数的真值.

表 2 $c=2.8, \lambda=0.5$ 模拟结果

| 参数 | $n=50$ | $n=100$ | $n=1000$ |
|-----------|--------|---------|----------|
| c | 2.7964 | 2.7964 | 2.7989 |
| λ | 0.5511 | 0.5534 | 0.5083 |

(3) 非随机缺失方式, 不妨取(7)式中 $c=2.8, p=0.8, q=0.2, \lambda=0.5$. 先产生服从均值为 2 的指数分布的 n 个随机数, 再按此种缺失方式产生一组缺失数据, 具体计算结果见表 3. 可见由于未知参数较多, 当 $n=1000$ 时估计值才比较接近未知参数的真值.

表 3 $c=2.8, p=0.8, q=0.2, \lambda=0.5$ 模拟结果

| 参数 | $n=50$ | $n=100$ | $n=1000$ |
|-----------|--------|---------|----------|
| c | 2.7964 | 2.8232 | 2.8017 |
| λ | 0.3670 | 0.4881 | 0.5145 |
| p | 0.7767 | 0.8056 | 0.7962 |
| q | 0.0806 | 0.1817 | 0.2078 |

参考文献

- [1] Roderick J. A., Donald B. R., Statistical Analysis with Missing Data(Second Edition) [M]. New Jersey: John Wiley & Sons, 2002.
- [2] Craig K. Enders. Applied Missing Data Analysis [M]. New York: Springer, 2010.

(上接第 27 页)

学生是否会和老师课后及时交流(0.105)所占比重较大, 它是评估的重点, 也是加强和改进的重点.

5 结论

本研究采用模糊数学原理对教师课堂教学质量进行综合评价, 合理考虑各种因素对教学质量的影响, 它可以与定性和定量因素相结合, 扩大信息量, 具有明确的理论和可操作性, 评价结果优良、公正, 为教学质量等级确定提供了科学依据. 然而, 在模糊综合评判中, 评价矩阵和权重, 评价矩阵和权重向量被确定为某些主观因素. 由于调查的样本量不够大, 仅限于我校的一些教师, 缺少高水平和经验丰富的专家进行打分统计权重, 课堂教学的评价因素较少, 这些都是值得今后进一步调查研究, 并得到更加科学合理的课堂教学评价.

参考文献

- [1] 李小梅. 基于模糊数学的课堂教学质量评价[J]. 浙江科技学院学报, 2010, 22(4): 247-252.
- [2] 谢季坚, 刘承平. 模糊数学方法及其应用(第 3 版) [M]. 武汉: 华中科技大学, 2006.
- [3] 教育部. 2008 年教育事业统计公报 [EB/OL]. 2009.
- [4] 谢季坚, 刘承平编著. 模糊数学方法及其应用 [M]. 武汉: 华中科技大学出版社, 2000-5-1.
- [5] 白思俊. 系统工程 [M]. 北京: 电子工业出版社, 2009: 204-205.
- [6] 李功越. 层次分析法在企业生产质量控制中的应用研究[J]. 现代制造工程, 2007, (1): 84-85.
- [7] 徐晓敏. 层次分析法的运用[J]. 统计与决策, 2008, (1): 157-158.

热忱欢迎订阅2021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刊物, 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, 单月月月底出版。2021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2021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. 24, No. 1 (Serial No. 201), Jan. 2021

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