Pure Mathematics 理论数学, 2024, 14(2), 710-718
Published Online February 2024 in Hans. https://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2024.142070

渐近II型删失下广义半逻辑分布的统计推断

王海军

西北师范大学数学与统计学院，甘肃 兰州

收稿日期：2023年12月20日；录用日期：2024年1月10日；发布日期：2024年2月29日

摘 要

本文主要研究了广义半逻辑分布在渐近II型删失下的统计推断问题。首先，我们介绍了广义半逻辑分布的概念和性质，然后提出了渐近II型删失机制。接着，我们详细推导了在渐近II型删失下，使用最大似然估计和贝叶斯方法估计广义半逻辑分布参数的过程。最后，通过数值模拟和实际数据分析，验证了所提出方法的有效性和适用性。

关键词

广义半逻辑分布，渐近II型删失，参数估计，最大似然，贝叶斯方法

Statistical Inferences of Generalized Half Logistic Distribution under Progressive Type-II Censored

Haijun Wang

College of Mathematics and Statistics, Northwest Normal University, Lanzhou Gansu

Received: Dec. 20^{th}, 2023; accepted: Jan. 10^{th}, 2024; published: Feb. 29^{th}, 2024

文章引用: 王海军. 渐近II型删失下广义半逻辑分布的统计推断[J]. 理论数学, 2024, 14(2): 710-718.
DOI: 10.12677/pm.2024.142070

Abstract

In this paper, the parameter estimation problem of generalized semi-logical distribution under progressive type II censored is studied. Firstly, we introduce the concept and properties of generalized semi-logical distribution, and then propose the asymptotic type II deletion mechanism. Then, we derive in detail the procedure of estimating generalized semi-logical distribution parameters by using maximum likelihood estimation and Bayesian method under asymptotic type II deletion. Finally, the effectiveness and applicability of the proposed method are verified by numerical simulation and actual data analysis.

Keywords

Generalized Semi-Logical Distribution, Asymptotic Type II Deletion, Parameter Estimation, Maximum Likelihood, Bayesian Method

1. 引言

生存分析是一种在多种行业中使用的统计方法。例如，在制造业方面，它用于分析机器寿命；在金融方面，它被用来分析偿还房屋的时间抵押贷款；就保险业而言，它适用于分析保险和证券权利到期所需的时间。但是生存函数的数据是根据时间的流逝而产生的，一般来说很难获得完整的数据，所以在生存分析中，经常用到的数据类型是删失数据。广义半逻辑分布(GHL)是一种连续概率分布，在生存分析和可靠性研究中，估计渐进式II型删失样本下广义半逻辑分布的参数具有重要意义。最初，Balakrishnan [1]引入了半逻辑分布来对标准logistic随机变量建模。又在2000年 [2]提出了半logistic分布的广义版本，称为广义半逻辑分布。随机变量X 的概率密度函数和累积分布函数分别为:

$$f(x) = \alpha(\frac{2e^{-x}}{1+e^{-x}})^{\alpha}\frac{1}{1+e^{-x}}, x > 0, \alpha > 0,$$

$$F(x) = 1 - (\frac{2e^{-x}}{1+e^{-x}})^{\alpha}, x > 0, \alpha > 0,$$

生存函数为

$$S_j(x) = \bar{F}_j(x) = (\frac{2e^{-x}}{1+e^{-x}})^{\alpha_j}, x > 0.$$

本文研究了渐进式II型删失情景下的广义半逻辑分布的分析。首先，利用极大似然估计方法得到未知参数的估计，并证明其存在唯一性。此外，我们使用观察到的Fisher信息矩阵和delta方法构造近似置信区间。在此基础上，提出了基于平方误差损失函数和LINEX损失函数的贝叶斯估计方法。所有提出的估计方法通过仿真研究进行了评估。最后，我们使用真实数据集进行分析以验证我们评估的正确性。

2. 广义半逻辑分布的渐近II型删失模型

在众多研究领域，例如经济学、农业、医学、精算学、现代工业以及生物科学等领域，获取的数据常常存在观察不完整的情况。在传统寿命测试中，研究人员通常需要投入大量时间和资源等待所有单元失效。然而，为了降低实验成本，实验通常在所有单元失效之前就被终止，从而产生删失数据。在寿命测试和生存分析中，删失现象尤为常见，此时仅有观测值的部分信息可用。传统的I型、II型删失方案以及混合删失方案在灵活性上有所欠缺，因为它们无法在实验过程中移除单元。这一限制阻碍了相关研究的开展，特别是在单位丢失或需要在不同时间间隔进行观察的情况下。为了解决这个问题Cohen [3]引入了渐进式删失方案(PCS)。相较于传统方案，PCS在降低测试时间和成本方面具有一定的优势。

PCS的删失流程如下：在寿命试验中，有n个元件。当第一次发生故障时，将n-1个工作元件和幸存元件中的R_1个随机退出该试验。在第二次失败时，$(n-R_1-2)$中的R_2个的存活元件随机退出本实验。我们据此进行实验，直到第m次失败。所有剩余的元件$R_m = n - m - R_1 - R_2 - R_3 - \cdots - R_{m-1}$从这个实验中撤出。显然，如果$R_1 = R_2 = \cdots = R_m - 1 = 0$且$R_m = n - m$，我们获得的样本是传统的II型删失样本。有关渐进式II型删失数据的进一步讨论，请参考Cramer [2]的研究。

3. 参数估计

3.1. 最大似然估计

在本节中，我们将推导最大似然估计量(MLE)的渐近分布，并构造未知参数α_1和α_2的近似置信区间。观察到的Fisher信息矩阵$I(\alpha_1\alpha_2)$由下式给出:

$$\begin{aligned} L(x \mid \alpha) &= C\prod_{i=1}^{m} f(x_{i:m:n};\theta)\left[1 - F(x_{i:m:n};\theta)\right]^{r_i} \\ &= C\prod_{i=1}^{m}(\frac{2e^{-x}}{1+e^{-x}})^{(\alpha(1+r_i))}(\frac{1}{1+e^{-x}}) \end{aligned} \tag{1}$$

此处

$$C = n(n-r_1-1)(n-r_1-2)\cdots(n-r_1-r_2-\cdots-r_{m-1}-m+1).$$

带入广义半逻辑分布的分布函数，密度函数，通过在方程两边计算相同的对数运算，则观测到的似然函数为

$$\log L(X|\alpha) \propto \sum_{i=1}^{m} \alpha(r_i+1)\log(\frac{2e^{-x_i}}{1+e^{-x_i}}) + m\log\alpha + \log(\frac{1}{1+e^{-x_i}}).$$

计算对α的偏导数，使其等于零，可以得到

$$\frac{\partial \log L(X|\alpha)}{\partial\alpha} = \sum_{i=1}^{m}(r_i+1)\log(\frac{2e^{-x_i}}{1+e^{-x_i}}) + \frac{m}{\alpha} = 0$$

α的最大似然估计可以很容易地得到

$$\hat{\alpha} = -\frac{m}{\sum_{i=1}^{m}(r_i+1)\log(\frac{2e^{-x_i}}{1+e^{-x_i}})}.$$

3.2. 近似置信区间

在这一节，我们将推导最大似然估计量(MLE)的渐近分布，并构造未知参数α的近似置信区间，观察到的Fisher信息矩阵$I(\alpha)$为：

$$I(\alpha) = \frac{\partial^2 \log L(\alpha)}{\partial\alpha^2}$$

故：

$$I(\alpha) = \frac{\partial^2 \log L(\alpha)}{\partial\alpha^2} = -\frac{m}{\alpha^2},$$

根据极大似然估计的渐近正态性，极大似然估计量的分布可以近似为二元正态分布。其均值为$(\hat{\alpha_1}, \hat{\alpha_2})$，协方差矩阵为$I$，表示为:

$$I(\alpha)^{-1} = I(\alpha)^{-1} = -\frac{\alpha^2}{m}$$

因此，未知参数的渐近$100(1-\varepsilon)$ 置信区间定义为

$$\left[\hat{\alpha} - Z_{\frac{\varepsilon}{2}}\sqrt{Var(\hat{\alpha})}, \hat{\alpha}_1 + Z_{\frac{\varepsilon}{2}}\sqrt{Var(\hat{\alpha})}\right],$$

这里，$Z_{\frac{\varepsilon}{2}}$表示标准正态分布的$\frac{\varepsilon}{2}$百分位数上。为了解决置信区间下限小于的可能性，可以采用

DOI: 10.12677/pm.2024.142070

对数变换和delta方法。以$\hat{\alpha}_k$为例，其渐近正态分布为$\hat{\alpha}_k \to N(\alpha_k, (I^{-1})_{kk})$。由此可以得到$\alpha_k$的双侧置信区间。

$$\left[\frac{\hat{\alpha}_1}{\exp(Z_{\frac{\varepsilon}{2}}\sqrt{Var(\hat{\alpha}_1)\frac{1}{\hat{\alpha}_1}})}, \hat{\alpha}_1\exp(Z_{\frac{\varepsilon}{1}}\sqrt{Var(\hat{\alpha}_1)\frac{1}{\hat{\alpha}_1}})\right],$$

$$\left[\frac{\hat{\alpha}_2}{\exp(Z_{\frac{\varepsilon}{2}}\sqrt{Var(\hat{\alpha}_2)\frac{1}{\hat{\alpha}_2}}}), \hat{\alpha}_2\exp(Z_{\frac{\varepsilon}{2}}\sqrt{Var(\hat{\alpha}_2)\frac{1}{\hat{\alpha}_2}})\right].$$

因此，我们可以得到关于$\hat{\alpha}$的$100(1-\varepsilon)\%$近似置信区间。

4. 贝叶斯估计

在这篇论文中，我们探讨了贝叶斯方法在广义半logistic分布参数估计中的应用，这一方法作为经典估计的有力补充。贝叶斯方法的一个重要特点是，它允许我们将待估计的未知参数视为随机变量，并将先验信息与样本数据有效结合。本节重点分析了广义半逻辑分布的参数点估计和区间估计的贝叶斯方法。依据Jun-Mei [4] 的研究建议，我们设α为随机变量，并假定它服从伽马先验分布。这一分布的密度函数为

$$\pi(\alpha; a, b) = \frac{b^a}{\Gamma(a)}\alpha^{a-1}e^{-b\alpha},$$

因此，贝叶斯估计的联合后验分布为：

$$\pi(\alpha|x) = \frac{L(x|\alpha)\pi(\alpha)}{\int_0^\infty L(x|\alpha)\pi(\alpha)d\alpha} \propto \alpha^{a-1}e^{-b\alpha}\prod_{i=1}^{m}(\frac{2e^{-x}}{1+e^{-x}})^{(\alpha(1+r_i))}(\frac{1}{1+e^{-x}})$$

这里，我们考虑两种不同类型的损失函数:对称平方误差损失函数和非对称LINEX损失函数。这两个损失函数的定义如下:

1) 平方误差损失函数($L(u, \hat{u}) = (u-\hat{u})^2$)。

2) 线性指数(LINEX)损失函数($L(u, \hat{u}) = e^{(s(\hat{u}-u))} - s(\hat{u}-u)$)。

假设$u(\alpha)$表示参数α的函数。函数$u(\alpha)$在平方误差损失函数和LINEX损失函数下的贝叶斯估计为:

$$\hat{u}_{se}(\alpha) = E\left[u(\alpha)\right] = \frac{\int_0^\infty u(\alpha)L(x|\alpha)\pi(\alpha)d\alpha}{\int_0^\infty L(x|\alpha)\pi(\alpha)d\alpha},$$

和

$$\hat{u}_{LI}(\alpha) = -\frac{1}{s}\log E\left[exp(-su)\right] = -\frac{1}{s}\log\frac{\int_0^\infty e^{-su(\alpha)}L(x|\alpha)\pi(\alpha)d\alpha}{\int_0^\infty L(x|\alpha)\pi(\alpha)d\alpha}.$$

根据上述推理，计算这两个积分的值存在一定挑战性。因此，我们采用蒙特卡罗马尔可夫链(MCMC)技术来解决这个问题。利用MCMC技术，我们可以得到函数$u(\alpha)$的贝叶斯估计来近似方程。

5. 模拟研究

利用R软件 [5] 进行仿真，在本节中，我们通过蒙特卡罗仿真研究，对不同估计方法得到的结果进行评价和比较。为了在偏差和均方误差(MSE)方面对点估计器的性能进行数值评估，对各种选择的滤波参数(即n、m和p)进行了模拟。此外，我们比较了区间估计在区间长度和覆盖概率方面的性能。首先，利用改进的广义算法 [6]生成II型渐进删失数据，。接下来，我们获得了α的最大似然估计(MLEs)和渐近置信区间(ACIs)。对于贝叶斯估计，我们考虑了α在平方误差损失函数和LINEX损失函数下的贝叶斯估计。在设定为d=1的LINEX损失函数下，贝叶斯估计的性能优于极大似然估计(MLEs)。获得了α的贝叶斯估计，以及它们相应的近似置信区间。

Gibbs采样和Metropolis-Hastings算法是统计、数字通信、信号处理和机器学习等各个领域最常用的两种MCMC方法。通过将Metropolis-Hastings算法与Gibbs抽样相结合，我们可以从后验分布中生成样本。

表 1给出了最大似然估计和在不同损失函数下计算的α的MLEs和贝叶斯估计的偏差和均方误差(MSE)值。同样，表 2给出了α的最大似然估计和贝叶斯估计下的置信区间的长度和覆盖范围。综上所述，利用MCMC技术和MH算法，我们可以推导出函数$u(\alpha)$的贝叶斯估计来近似解。该方法由于其处理复杂积分计算的能力和在各个领域的适用性，在实际应用中得到了广泛的应用。

Table 1. Bias and MSE of MLE and the Bayes estimates of α

表 1. α的MLE和贝叶斯估计的偏差和均方误差

	MLE		SEL-BAYES		LL-BAYES	
(n,m)	BIAS	MSE	SBias	S MSE	LBias	LMSE
(40, 15)	0.0443	0.0035	0.0289	0.0008	0.0260	0.0007
(40, 25)	0.0330	0.0018	0.0209	0.0004	0.0227	0.0005
(40, 30)	0.0301	0.0015	0.0150	0.0002	0.0031	0.0001
(50, 15)	0.0443	0.0036	0.0404	0.0016	0.0424	0.0018
(50, 25)	0.0336	0.0019	0.0299	0.0008	0.0316	0.0010
(50, 30)	0.0308	0.0015	0.0369	0.0013	0.0385	0.0014
(60, 15)	0.0448	0.0039	0.0857	0.0073	0.0872	0.0076
(60, 25)	0.0328	0.0018	0.0532	0.0028	0.0548	0.0030
(60, 30)	0.0313	0.0016	0.0146	0.0002	0.0126	0.0001

为了评估点估计器的性能，我们测量了它们的偏差和均方误差。偏差反映了估计值与真实值之间的系统差异，而均方差结合了偏差和方差，提供了对估计方法的总体评估。进一步，我们通过考虑区间的长度和它们的覆盖概率来比较区间估计的性能。区间长度反映了估计的不确定程度，区间越短表明估计越精确。覆盖率表示真实值位于区间内的可能性，较高的覆盖率表示更可靠的区间估计。

Table 2. The average length and coverage probability of the ACI and BCI of α for distinct choices of n, m
表 2. 对于不同选择的n, m，αACI和BCI的平均长度和覆盖率

	MLE		SEL-BAYES	
(n,m)	CP	AL	BCP	BAL
$(40, 15)$	0.9510	0.2146	1	0.2275
$(40, 25)$	0.9510	0.1613	1	0.1385
$(40, 30)$	0.9570	0.1476	1	0.1451
$(50, 15)$	0.9430	0.2150	1	0.1589
$(50, 25)$	0.9500	0.1628	1	0.1317
$(50, 30)$	0.9470	0.1482	1	0.1153
$(60, 15)$	0.9350	0.2148	1	0.1531
$(60, 25)$	0.9570	0.1629	1	0.1137
$(60, 30)$	0.9590	0.1490	1	0.1326

通过表格，可以得出结论：1.随着n或$\frac{m}{n}$增加，最大似然估计和贝叶斯估计的偏差和均方误差都表现的更加优越。2.对于α，LINEX损失函数下的Bayes估计在偏置方面有更好的表现。3.随着n或$\frac{m}{n}$增加，覆盖率相应增大。

总体而言，贝叶斯估计优于最大似然估计，然而，在许多情况下，mle的性能接近其相应的贝叶斯估计，并且是可靠的。因此，当未知参数的先验知识不可用时，尤其是考虑到贝叶斯方法的计算需求，最大似然估计也是一种有效的方法。

通过对仿真结果的系统评价和比较，我们可以确定在给定参数组合下不同估计方法的相对优缺点，有助于选择最适合手头具体问题的估计方法。

6. 真实数据分析

考虑Nelson [7]给出的数据，这是一个绝缘流体测试实验表3。渐进II型删失样本是由Viverous和Balakrishnan [8]根据该数据生成的，审查方案如表4所示。在这个例子中，我们有$n = 16$,$m = 8$。根据第3节中的公式，我们可以计算λ估计值。

从形状参数为λ的广义半logistic分布中考虑一个$n = 9$,$m = 8$的渐进II型删失样本,删失方案为$r = (0, 0, 3, 4, 0, 4, 0, 0)$。然后，我们从$a = 0.2$,$b = 0.5$的伽马分布中生成$\Gamma$,使用生成的$\Gamma$，从中生成大小为$n = 19$,$m = 8$的渐进II型删失样本，删失方案$r = (0, 0, 3, 4, 0, 4, 0, 0)$。使用在第3节中的公式，计算$\Gamma$的估计。

Table 3. Failure log times to breakdown of an insulating fluid testing experiment
表 3. 渐进II型删失数据

i	1	2	3	4	5	6	7	8
x	0.270027	1.02245	1.15057	1.57898	2.11263	2.48989	3.60305	4.28895
r_i	0	0	2	3	0	3	0	0

Table 4. MLE and the Bayes estimates of α
表 4. 对α的估计偏差和均方误差

BIAS	MSE	SBIAS	SMSE	LBIAS	LMSE
0.0648	0.0100	0.0531	0.0028	0.0555	0.0030

7. 结论和展望

本文提出了几种统计推理方法来估计II型渐进删失模型下广义半逻辑分布的参数。方法包括极大似然估计和贝叶斯估计。通过蒙特卡罗仿真，验证了本文方法的实用性和可行性。

在此之前，许多学者对不同滤波方案下的广义半逻辑分布进行了参数估计或可靠性估计。然而，竞争风险模型的统计推断却很少被考虑。详细信息请参考 [9] 和 [10] 。在本文中，我们基于竞争风险模型来解决这个问题，并对这些问题进行了详细的讨论。

在实际应用中，从业者可以根据自身情况选择最合适的推理方法来准确估计参数并进行可靠性分析。值得注意的是，纳入估计结果，如偏倚或均方误差，并考虑预期的实验时间，可以极大地帮助设计最佳的评审方案，从而大大节省人力和财力。在未来的工作中，我们可以考虑删失过程中涉及竞争风险的模型。

基金项目

本文受到国家自然科学基金项目：不完全数据下相依竞争失效单调关联系统可靠性的统计推断(No. 12361060)的支持。

参考文献

[1] Balakrishnan, N. (1985) Order Statistics from the Half Logistic Distribution. *Journal of Statistical Computation and Simulation*, **20**, 287-309. https://doi.org/10.1080/00949658508810784

[2] Balakrishnan, N. and Aggarwala, R. (2000) Progressive Censoring: Theory, Methods, and Applications. Birkhaeuser, Basel. https://doi.org/10.1007/978-1-4612-1334-5

[3] Cohen, A. (1963) Progressively Censored Samples in Life Testing. *Technometrics*, **5**, 327-339. https://doi.org/10.1080/00401706.1963.10490102

[4] Jia, J.-M., Yan, Z.-Z. and Peng, X.-Y. (2018) Parameters Estimation of Burr-XII Distribution under First-Failure Progressively Unified Hybrid Censoring Schemes. *Statistical Analysis and Data Mining*, **11**, 271-281. https://doi.org/10.1002/sam.11391

[5] Lee, C., Grasso, C. and Sharlow, M.F. (2002) R: A Language and Environment for Statistical Computing.

[6] Balakrishnan, N. and Sandhu, R.A. (1995) A Simple Simulational Algorithm for Generating Progressive Type-II Censored Samples. *The American Statistician*, **49**, 229-230. https://doi.org/10.1080/00031305.1995.10476150

[7] Wayne, N. (1982) Applied Life Data Analysis. In: *Wiley Series in Probability and Statistics*, Wiley, Hoboken.

[8] Viveros, R. and Balakrishnan, N. (1994) Interval Estimation of Parameters of Life from Progressively Censored Data. *Technometrics*, **36**, 84-91. https://doi.org/10.1080/00401706.1994.10485403

[9] Barranco-Chamorro, I., Luque-Calvo, P.L., Jiménez-Gamero, M.-D. and Fernández, V.A. (2017) A Study of Risks of Bayes Estimators in the Generalized Half-Logistic Distribution for Progressively Type-II Censored Samples. *Mathematics and Computers in Simulation*, **137**, 130-147. https://doi.org/10.1016/j.matcom.2016.09.003

[10] Alghamdi, A.S. (2023) Statistical Inferences of Competing Risks Generalized Half-Logistic Lifetime Populations in Presence of Generalized Type-I Hybrid Censoring Scheme. *Alexandria Engineering Journal*, **65**, 699-708. https://doi.org/10.1016/j.aej.2022.10.018