

# 尼日利亚猴痘疫情不确定 Logistic 增长模型分析

陈丽丽,丁春晓\*

(聊城大学数学科学学院,山东聊城252000)

**摘要:**为了处理猴痘传播过程中的不确定性,利用不确定 Logistic 增长模型对尼日利亚猴痘疫情进行分析。首先,对尼日利亚疾控中心公布的猴痘累计确诊病例数据进行整理,并绘制出数据散点图。其次,建立 Logistic 增长模型,利用最小二乘法估计得到拟合 Logistic 增长模型,进而计算残差,通过 5 种检验方法验证残差频率是不稳定的,说明随机 logistic 增长模型研究尼日利亚猴痘疫情的不适用性,进而采用不确定 Logistic 增长模型进行建模。最后,给出尼日利亚猴痘累计确诊病例数的预测值和 95% 的置信区间,并且通过交叉验证方法和直观对比观测值、预测值和置信区间验证模型的可行性。本研究为尼日利亚猴痘疫情防控和政策制定提供一定的理论支撑。

**关键词:**不确定理论;不确定回归分析;不确定 Logistic 增长模型;尼日利亚猴痘疫情

**中图分类号:**O29 **文献标志码:**A

**引用格式:**陈丽丽,丁春晓. 尼日利亚猴痘疫情不确定 Logistic 增长模型分析[J]. 山东大学学报(理学版),2025,60(12):49-54,65.

## Analysis of Mpox outbreak in Nigeria based on uncertain Logistic growth model

CHEN Lili, DING Chunxiao\*

(School of Mathematical Sciences, Liaocheng University, Liaocheng 252000, Shandong, China)

**Abstract:** In order to deal with uncertainty in the transmission process of Mpox, an uncertain Logistic growth model was used to analyze the Mpox epidemic in Nigeria. Firstly, the cumulative confirmed cases of Mpox published by the Nigeria Center for Disease Control and Prevention were sorted out, and the scatter plot of the data was illustrated. Then, a Logistic growth model was developed and the fitted Logistic growth model was obtained using least squares estimation, which in turn calculated the residuals. Verification of the instability of residual frequencies through five tests illustrated the inapplicability of the stochastic Logistic growth model to study the Mpox outbreak in Nigeria. The uncertain Logistic growth model was further used for modeling. Finally, the forecast values of the cumulative number of confirmed Mpox cases in Nigeria were given, and 95% confidence intervals were obtained. And the feasibility of the model was verified by cross-validation method and the direct comparison among observed values, forecast values and confidence intervals. This study provided some theoretical support for the prevention and control of Mpox epidemic and policy formulation in Nigeria.

**Key words:** uncertainty theory; uncertain regression analysis; uncertainty Logistic growth model; Mpox outbreak in Nigeria

## 0 引言

猴痘是由猴痘病毒感染所导致的一种人畜共患病<sup>[1]</sup>,主要症状包括发热、皮疹、淋巴肿大等。猴痘病毒于 1970 年首次在一位刚果(金)的儿童体内发现。此后,非洲不断有猴痘病例出现,传播链比较短,具有地方流行性。而 2022 年爆发的猴痘疫情,规模空前盛大,涉及了很多从未有过猴痘病例的国家。严峻的猴痘疫情对全球旅游、进出口贸易以及文化交流等方面产生了严重影响,因此,基于实际数据的猴痘疫情研究分析刻不容缓。

收稿日期:2023-09-25;网络出版时间:2024-10-08 12:55:21

基金项目:国家自然科学基金资助项目(12026226);山东省自然科学基金资助项目(ZR2021MA089,ZR2022MF293)

第一作者:陈丽丽(2001—),女,硕士研究生,研究方向为不确定理论和生物数学. E-mail:1784957151@qq.com

\*通信作者:丁春晓(1987—),女,副教授,硕士生导师,博士,研究方向为不确定理论和生物数学. E-mail:dingchunxiao1987@163.com

Logistic 增长模型是描述种群增长或疾病传播的有效工具<sup>[3-5]</sup>,最初由比利时数学家 Verhulst 于 1838 年推导出来<sup>[2]</sup>。其特点表现为初期增长缓慢,随后逐渐加速,最终趋于平缓并接近饱和的“S”型增长趋势,因此 Logistic 增长曲线又被称为“S”型曲线。基于尼日利亚猴痘累计确诊病例的数据特点,本文采用 Logistic 增长模型建模分析。

在猴痘病毒的传播过程中,存在诸多不确定因素,这些因素可以通过引入扰动项来进行建模。如果将扰动项视为随机变量,则可运用随机 Logistic 增长模型进行建模分析,其建立在概率论基础之上;如果将扰动项视为不确定变量,则可运用不确定 Logistic 增长模型进行建模分析,其建立在不确定理论基础之上。不确定理论由清华大学刘宝碇教授于 2007 年建立<sup>[6]</sup>,2009 年进一步完善使其成为公理化体系<sup>[7]</sup>。不确定理论满足正则性、对偶性、次可加性及乘积公理<sup>[8]</sup>。不确定统计学是不确定理论的重要分支之一,由刘宝碇教授于 2010 年提出,是一套利用不确定理论来收集、分析和解释数据的数学技术<sup>[9]</sup>。不确定回归分析是一套利用不确定理论来探索解释变量与响应变量之间关系的统计技术<sup>[10]</sup>,它的切入点是把扰动项看作一个不确定变量。为了估计未知参数,专家学者提出了多种估计方法,如矩估计<sup>[11]</sup>、极大似然估计<sup>[12-13]</sup>和最小二乘估计<sup>[14]</sup>等多种方法。此外,利用不确定假设检验来评估不确定回归模型的合理性<sup>[15]</sup>,并对回归系数提出了一个不确定显著性检验<sup>[10]</sup>。迄今为止,不确定回归分析已成功应用于经济学<sup>[16]</sup>、疫情传播<sup>[17]</sup>和粮食产量<sup>[18]</sup>等领域,但对于猴痘疫情的研究仍是空白。本文首次将不确定 Logistic 增长模型应用于猴痘疫情的统计建模分析中。

## 1 尼日利亚猴痘疫情数据

本文采用尼日利亚猴痘累计确诊病例周数据,数据来源于尼日利亚疾病控制中心(<https://ncdc.gov.ng/>)。考虑到 2022 年猴痘疫情在全球范围内爆发,因此采用 2022 年以后的数据作为研究对象。

由于前期数据缺失,在保持数据的客观性和一致性的前提下,对现有有效数据进行重新排序,令  $t=1,2,\dots,37$  代表 2022 年第 4 周,2022 年第 8 周, $\dots$ ,2023 年第 3 周, $y_t$  表示第  $t$  周尼日利亚猴痘累计确诊病例人数。数据详情见表 1。设观测数据为

$$(t, y_t), \quad t=1, 2, \dots, 37. \quad (1)$$

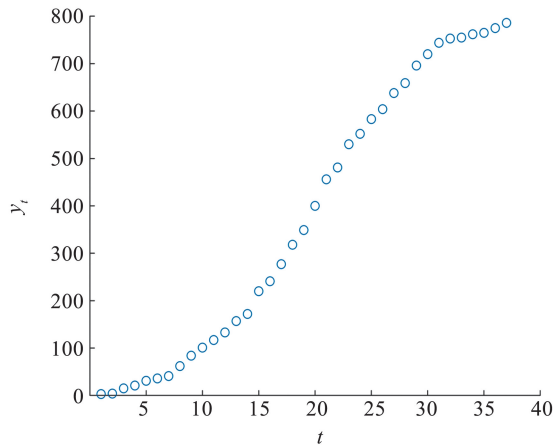
例如,  $t=5$  表示 2022 年第 22 周,  $y_5$  表示 2022 年第 22 周尼日利亚猴痘累计确诊病例数。

表 1 2022 年第 4 周至 2023 年第 3 周尼日利亚猴痘累计确诊病例周数据

Table 1 Weekly data of cumulative confirmed cases of Mpox in Nigeria from the 4th week of 2022 to the 3rd week of 2023

周序号	周数	猴痘累计确诊病例/人	周序号	周数	猴痘累计确诊病例/人
1	2022 年第 4 周	3	20	2022 年第 37 周	400
2	2022 年第 8 周	4	21	2022 年第 38 周	456
3	2022 年第 17 周	15	22	2022 年第 39 周	481
4	2022 年第 21 周	21	23	2022 年第 40 周	530
5	2022 年第 22 周	31	24	2022 年第 41 周	552
6	2022 年第 23 周	36	25	2022 年第 42 周	583
7	2022 年第 24 周	41	26	2022 年第 43 周	604
8	2022 年第 25 周	62	27	2022 年第 45 周	638
9	2022 年第 26 周	84	28	2022 年第 46 周	659
10	2022 年第 27 周	101	29	2022 年第 47 周	696
11	2022 年第 28 周	117	30	2022 年第 48 周	720
12	2022 年第 29 周	133	31	2022 年第 49 周	744
13	2022 年第 30 周	157	32	2022 年第 50 周	753
14	2022 年第 31 周	172	33	2022 年第 51 周	755
15	2022 年第 32 周	220	34	2022 年第 52 周	762
16	2022 年第 33 周	241	35	2023 年第 1 周	765
17	2022 年第 34 周	277	36	2023 年第 2 周	775
18	2022 年第 35 周	318	37	2023 年第 3 周	786
19	2022 年第 36 周	349			

注:数据来源于尼日利亚疾病控制中心(<https://ncdc.gov.ng/>)。



注:图 1 为基于表 1 数据绘制。

图 1 2022 年第 4 周至 2023 年第 3 周尼日利亚猴痘累计确诊病例周数据

Fig.1 Weekly data of cumulative confirmed cases of Mpox in Nigeria from the 4th week of 2022 to the 3rd week of 2023

## 2 模型的建立与分析

对尼日利亚猴痘累计确诊病例数据进行如图 1 所示的可视化分析,发现猴痘疫情具有“S”型特征,因此采用 Logistic 增长模型

$$y = \beta_0 / (1 + \beta_1 \exp(-\beta_2 t)), \quad \beta_0, \beta_1, \beta_2 > 0 \tag{2}$$

对数据进行建模分析。其中,  $\beta_0$  表示模型的上限,即疫情在不受外部干预(如疫苗等)且环境条件保持不变的情况下,累计确诊病例可能达到的最大值。 $\beta_1$  表示初始条件参数,其值越大,初始病例数越小。 $\beta_2$  表示增长率参数, $\beta_2$  越大,增长曲线越陡峭,疾病传播速度越快。

基于观测数据(1),运用最小二乘估计求解模型(2)中未知参数  $\beta_0, \beta_1, \beta_2$ ,即

$$\min_{\beta_0, \beta_1, \beta_2 > 0} \sum_{t=1}^{37} (y_t - \beta_0 / (1 + \beta_1 \exp(-\beta_2 t)))^2. \tag{3}$$

利用 MATLAB 软件中“lsqnonlin”函数解得

$$\hat{\beta}_0 = 809.0984, \quad \hat{\beta}_1 = 59.9555, \quad \hat{\beta}_2 = 0.2032.$$

将  $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$  代入式(2),拟合 Logistic 增长模型为

$$y = 809.0984 / (1 + 59.9555 \exp(-0.2032t)). \tag{4}$$

将观测值  $y_t$  与拟合 Logistic 增长模型(4)所得的拟合值作差,即

$$\varepsilon_t = y_t - 809.0984 / (1 + 59.9555 \exp(-0.2032t)),$$

得到如表 2 所示的 37 个残差  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{37}$ 。

表 2 残差  
Table 2 Residual

-13.2044	-15.7662	-9.0871	-8.3176	-4.6322	-7.2315	-11.3419	-1.2137	7.8835	9.6695
7.8674	3.2226	3.5288	-8.3425	9.5912	-2.5461	-2.4660	0.2925	-8.6484	1.4620
16.4492	1.1487	11.3391	-3.3153	-6.3041	-16.2885	-10.0984	-13.7128	1.7689	7.1600
15.2170	10.6672	1.2299	-1.3689	-6.3857	-3.0542	2.4174			

### 2.1 随机 Logistic 增长模型建模分析

考虑到猴痘疫情传播过程中会受到各种各样不确定因素的影响,将扰动项  $\varepsilon$  视为随机变量,采用随机 Logistic 增长模型

$$y^{(1)} = 809.0984 / (1 + 59.9555 \exp(-0.2032t)) + \varepsilon \tag{5}$$

对尼日利亚猴痘累计确诊病例进行建模分析。此时,模型残差应构成一个满足独立同分布条件的随机序列,且其任意子序列应具有频率稳定性。

为了验证残差是否频率稳定,本文使用双样本柯尔莫可洛夫-斯米洛夫检验(two-sample Kolmogorov-

Smirnov test)、威尔科克森秩和检验(Wilcoxon rank sum test)、安萨里-布拉德利检验(Ansari-Bradley test)、双样本  $T$  检验(two-sample  $T$ -test)、双样本  $F$  检验(two-sample  $F$ -test) 5 种方法,验证残差不是来自同一分布。采用了 5 种方法进行检验。在利用 MATLAB 中函数命令对残差进行检验时, $p$  值是在原假设成立的前提下,得到当前样本统计量观测值或更极端值的概率。如果  $p$  值小于显著性水平(0.05),则拒绝原假设,认为观察到的数据集与参考分布不符(或两个数据集不是来自同一分布)。5 种检验方法的返回值  $p$  均小于 0.05。其中,two-sample  $T$ -test 以及 two-sample  $F$ -test 属于参数检验方法,均满足检验的前提条件,具体细节如下:

(1) two-sample Kolmogorov-Smirnov test

取 37 个残差中的两部分

$$(\varepsilon_1, \varepsilon_3, \dots, \varepsilon_7) \text{ 和 } (\varepsilon_{14}, \varepsilon_{15}, \dots, \varepsilon_{37}),$$

借助 MATLAB 软件中“kstest2”函数,通过 two-sample Kolmogorov-Smirnov test 得到  $p$  值为 0.003 9。这意味着,上述残差组具有不同的分布函数,因此不是来自同一个总体。

(2) Wilcoxon rank sum test

取 37 个残差中的两部分

$$(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_8) \text{ 和 } (\varepsilon_9, \varepsilon_{10}, \dots, \varepsilon_{37}),$$

借助 MATLAB 软件中“ranksum”函数,通过 Wilcoxon rank sum test 得到  $p$  值为 0.003 8。这意味着,上述残差组来自于具有不同中位数的分布,因此并不是来自于同一个总体。

(3) Ansari-Bradley test

取 37 个残差中的两部分

$$(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{15}) \text{ 和 } (\varepsilon_{33}, \varepsilon_{34}, \dots, \varepsilon_{37}),$$

借助 MATLAB 软件中“ansaribradley”函数,通过 Ansari-Bradley test 得到  $p$  值为 0.005 2。这意味着,上述残差组具有不同的分散性,因此并不来自同一个总体。

(4) two-sample  $T$ -test

取 37 个残差中的两部分

$$(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_8) \text{ 和 } (\varepsilon_9, \varepsilon_{10}, \dots, \varepsilon_{37}),$$

借助 MATLAB 软件中“ttest2”函数,通过 two sample  $T$ -test 得到  $p$  值为 0.003 0。这意味着,上述残差组在满足正态分布和方差齐性假设的前提下,具有不同的均值,因此并不来自同一个总体。

(5) two-sample  $F$ -test

取 37 个残差中的两部分

$$(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_8) \text{ 和 } (\varepsilon_{20}, \varepsilon_{21}, \dots, \varepsilon_{32}),$$

借助 MATLAB 软件中“vartest2”函数,通过 two-sample  $F$ -test 得到  $p$  值为 0.034 7。这意味着,上述残差组在满足正态分布假设的前提下,具有不同的方差,因此并不来自同一个总体。

由于 37 个残差不是来自同一总体,即残差不是来自同一分布,这意味着残差  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{37}$  不是白噪声,也不是概率论意义上的随机序列。如果在这种情况下仍然坚持应用概率回归分析对尼日利亚猴痘疫情进行建模是不合理的。

## 2.2 不确定 Logistic 增长模型建模分析

当残差频率不稳定时,将扰动项  $\varepsilon$  视为不确定变量,选择基于不确定理论的不确定 Logistic 模型

$$y^{(2)} = 809.0984 / (1 + 59.9555 \exp(-0.2032t)) + \varepsilon \quad (6)$$

对尼日利亚猴痘累计确诊病例进行建模分析。

假设残差服从正态不确定分布  $\mathcal{N}(e, \sigma)$ , 利用矩估计<sup>[11]</sup>得到不确定扰动项  $\varepsilon$  的期望值为

$$\hat{e} = \frac{1}{37} \sum_{i=1}^{37} \varepsilon_i = -1.1462, \quad (7)$$

方差是

$$\hat{\sigma}^2 = \frac{1}{37} \sum_{i=1}^{37} (\varepsilon_i - \hat{e})^2 = 8.5220^2. \quad (8)$$

因此,得到不确定 Logistic 增长模型

$$y^{(2)} = 809.0984 / (1 + 59.9555 \exp(-0.2032t)) + N(-1.1462, 8.5220). \quad (9)$$

使用不确定假设检验判断不确定 Logistic 增长模型(9)是否很好地拟合尼日利亚猴痘累计确诊病例周数据,即检验 37 个残差  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{37}$  是否服从正态不确定分布:

$$\mathcal{N}(-1.1462, 8.5220)。$$

考虑下面两个假设

$$H_0: e = -1.1462 \text{ 且 } \sigma = 8.5220 \text{ 与 } H_1: e \neq -1.1462 \text{ 或 } \sigma \neq 8.5220。$$

取显著性水平  $\alpha=0.05$ , 根据不确定假设检验, 得出该检验为

$$W = \{(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{37}) : \text{至少存在 2 个下标 } t \text{ 满足 } 1 \leq t \leq 37 \text{ 使得 } \varepsilon_t < -18.3592 \text{ 或 } \varepsilon_t > 16.0667\}。$$

观察图 2 所示的检验结果, 只有  $\varepsilon_{21} \notin [-18.3592, 16.0667]$ , 可得  $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{37}) \notin W$ 。因此接受  $H_0$ , 即残差的期望为  $-1.1462$ , 方差为  $8.5220$ 。所以, 不确定 Logistic 增长模型很好地拟合了尼日利亚猴痘累计确诊病例周数据。

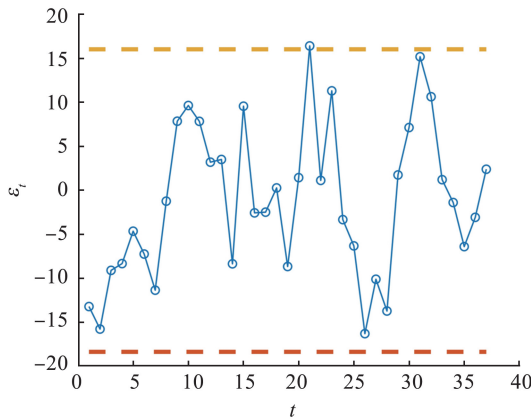


图 2 不确定假设检验的结果  
Fig.2 Results of uncertainty hypothesis test

利用不确定 Logistic 增长模型(9)预测 2023 年第 4 周尼日利亚猴痘累计确诊病例数。取  $t=38$ , 不确定变量  $\hat{y}_{38}^{(2)}$  为

$$\hat{y}_{38}^{(2)} = 809.0984 / (1 + 59.9555 \exp(-0.2032 \times 38)) + \mathcal{N}(-1.1462, 8.5220), \tag{10}$$

即  $\hat{y}_{38}^{(2)} \sim \mathcal{N}(787.0068, 8.5220)$ 。将  $\hat{y}_{38}^{(2)}$  的期望值作为预测值, 得到预测值为  $E[\hat{y}_{38}^{(2)}] = 787.0068$ 。因此预计 2023 年第 4 周尼日利亚猴痘累计确诊病例为 787 人。进一步得到 95% 的置信区间为

$$787.0068 \pm \frac{8.5220 \cdot \sqrt{3}}{\pi} \ln \frac{1+0.95}{1-0.95},$$

即  $787.0068 \pm 17.2130 = [769.7938, 804.2197] \approx [770, 804]$ , 2023 年第 4 周尼日利亚猴痘累计确诊病例人数 95% 的可能性介于 770~804 人之间。同理可以得到 2023 年第 5 周到第 9 周尼日利亚猴痘累计确诊病例人数的预测值和置信区间, 详见图 3 黄色星花部分和表 3 数值。

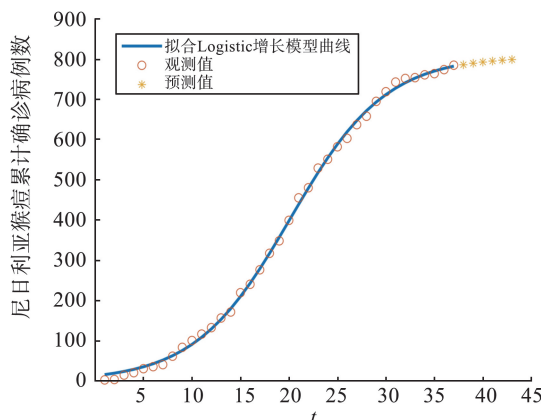


图 3 尼日利亚累计确诊病例数的观测值、拟合 Logistic 增长模型及预测值的对比图

Fig.3 Comparison of observed, fitted Logistic growth model and forecasts of the cumulative confirmed cases in Nigeria

为了对模型可行性进行评估,借助 MATLAB 软件的“cvpartition”函数,采用 5 折交叉验证的方法,得到模型的平均绝对误差为 7.031 9,均方根误差为 8.261 9。尼日利亚猴痘累计确诊病例数据的范围为 3~789,在这个数据范围下,上述误差较小,说明模型性能较好。此外,比较表 3 的观测值和预测值、置信区间,可以发现观测值均位于置信区间内,从而进一步说明了模型的可行性。

表 3 2023 年第 4 周到第 9 周的预测值和置信区间  
Table 3 Forecast values and confidence intervals for the 4th to 9th week of 2023

周数	观测值	预测值	置信区间
2023 年第 4 周	789	787	[770,804]
2023 年第 5 周	796	791	[774,808]
2023 年第 6 周	800	794	[777,811]
2023 年第 7 周	805	796	[779,814]
2023 年第 8 周	807	799	[781,816]
2023 年第 9 周	812	800	[783,817]

### 3 结论

本文采用不确定回归分析,针对尼日利亚猴痘累计确诊病例建立不确定 Logistic 增长模型,并进行参数估计、残差分析、不确定假设检验和预测。通过交叉验证以及直接对比观测值和预测值的方法,验证了利用不确定 Logistic 增长模型对尼日利亚猴痘疫情分析预测的适用性和准确性。

本研究所建立的 Logistic 增长模型主要用于猴痘疫情的短期趋势预测。由于疫情发展易受到公共卫生干预、传播途径变化及病毒变异等多种外部因素影响,因此该模型在长期预测中的适用性存在局限。未来,还将结合最新数据,对猴痘传播风险、峰值、达峰时间、传染规模等疫情技术指标进行预测分析,并对模型进行进一步优化。

#### 参考文献:

- [1] KUMAR N, ACHARYA A, GENDELMAN H E, et al. The 2022 outbreak and the pathobiology of the monkeypox virus[J]. *Journal of Autoimmunity*, 2022, 131:102855.
- [2] 李春喜,姜丽娜,邵云,等. 生物统计学[M]. 5 版. 北京:科学出版社,2013:163-166.  
LI Chunxi, JIANG Lina, SHAO Yun, et al. *Biostatistics*[M]. Fifth ed. Beijing: Science Press, 2013:163-166.
- [3] SHEN C Y. Logistic growth modelling of COVID-19 proliferation in China and its international implications[J]. *International Journal of Infectious Diseases*, 2020, 96:582-589.
- [4] MULLIGAN G F. Logistic population growth in the world's largest cities[J]. *Geographical Analysis*, 2006, 38(4):344-370.
- [5] MARTELLONI G, MARTELLONI G. Analysis of the evolution of the Sars-Cov-2 in Italy, the role of the asymptomatics and the success of Logistic model[J]. *Chaos, Solitons & Fractals*, 2020, 140:110150.
- [6] LIU B D. *Uncertainty Theory*[M]. 2nd ed. Berlin: Springer-Verlag, 2007.
- [7] LIU B D. Some research problems in uncertainty theory[J]. *Journal of Uncertain Systems*, 2009, 3(1):3-10.
- [8] 温艳清,刘宝亮,师海燕,等. 基于不确定理论的多部件竞争失效系统的可靠性评估[J]. *山东大学学报(理学版)*, 2023, 58(5):76-83.  
WEN Yanqing, LIU Baoliang, SHI Haiyan, et al. Reliability evaluation for multi-component competing failure system based on uncertainty theory[J]. *Journal of Shandong University (Natural Science)*, 2023, 58(5):76-83.
- [9] LIU B D. *Uncertainty theory: a branch of mathematics for modelling human uncertainty*[M]. Berlin: Springer-Verlag, 2010.
- [10] YAO K, LIU B D. Uncertain regression analysis: an approach for imprecise observations[J]. *Soft Computing*, 2018, 22(17):5579-5582.
- [11] LIO W, LIU B D. Residual and confidence interval for uncertain regression model with imprecise observations[J]. *Journal of Intelligent & Fuzzy Systems*, 2018, 35(2):2573-2583.
- [12] LIO W, LIU B D. Uncertain maximum likelihood estimation with application to uncertain regression analysis[J]. *Soft Computing*, 2020, 24(13):9351-9360.