

文章编号: 1005-0523(2005)05-0158-04

logistic 模型在裂纹检测概率曲线拟和中的应用

肖 鹏, 白存儒

(西北工业大学 航空学院, 陕西 西安 710072)

摘要:提出了一种适合于描述无损检测概率曲线的经过修正的 logistic 回归模型, 并运用最小二乘原理, 利用非线性回归分析, 结合高斯-牛顿迭代法, 给出了求解该模型中未知参数的具体思路与方法. 同时结合具体算例, 对该模型的拟和优良性进行评估分析. 计算结果表明, 该模型的预测值与实验值吻合良好, 有较高的预测精度.

关键词:无损检测; 逻辑斯蒂模型; 非线性回归; 高斯-牛顿法; 预测精度

中图分类号: TP391.4

文献标识码: A

按飞机结构损伤容限设计要求, 在结构存在损伤时, 结构的剩余强度降低至规定值之前, 其损伤能被及时检测出来并加以维修, 以保证飞机结构的安全性. 因此可以说飞机结构的安全在一定程度上依赖于无损检测的能力. 无损检测可靠性是指无损检测方法对缺陷的检出能力, 是用该方法检出特定类型、特定尺寸缺陷有效性所做的一种度量^[1]. 为了直观地表征无损检可靠性的 高低, 引入裂纹检出概率(Probability Of Detection, short for POD)曲线, 它是无损检测可靠性研究的基础. 因而, 如何建立、以及建立怎样的适合于描述 POD 曲线的数学模型, 是检测问题的关键. 国际上 Packman、Lewis、Yang、Berens 和 Hovey 等学者曾对此进行了大量的研究^[2], 提出了十几种数学模型, 其中的几种如 Weibull 概率模型得到了广泛的应用.

置信度(Confidence Level, short for CL)是裂纹检出概率的可信度. 裂纹检出概率和置信度的选择是任意的, 通常选用 POD/CL 为 90/95^[1]. 本文也是按这种要求采用等裂纹尺寸间隔法对数据进行统计, 利用文献^[3]介绍的方法求解 P_L (长度为 a_i 的裂纹所对应的检出概率 P_i 的置信下限). 在对几种 POD 曲线模型进行分析研究的实践过程中, 提出了一种

拟合 POD 曲线的 logistic 模型, 通过非线性最小二乘法和高斯-牛顿迭代法确定该模型的未知参数, 并从回归方程预测的准确性与相关系数检验的角度, 对该模型进行分析评价. 实践表明, 该回归模型能够较好地描述检出裂纹长度的概率分布规律, 有一定的实用性.

1 裂纹数据统计方法^[4]

裂纹的每一次检测结果是一个二分类因变量: 裂纹被检出或未被检出. 同一裂纹区间的 N 次独立检测, 裂纹检出数 S 服从二项分布,

$$P(S = S_N) = C_N^{S_N} P^{S_N} (1 - P)^{N - S_N} \quad (1)$$

其中, P 未知, 若某次试验中裂纹检出数为 S_N , 则 P 的点估计值 \hat{P} 可用公式(2)表示,

$$\hat{P} = S_N / N \quad (2)$$

检出概率 P 的置信下限 P_L 应满足公式(3)给定的概率条件,

$$P(S = S_N) = \sum_{i=1}^N C_N^i P_L^i (1 - P_L)^{N - i} \quad (3)$$

置信下限 P_L 可按满足条件(3), 并经过简化得到的公式(4)求出,

收稿日期: 2005-03-15

作者简介: 肖鹏(1977-), 男, 陕西华县人, 2003 级硕士研究生, 研究方向为空中交通流量管理.

$$P_L = \frac{n_2}{n_1 + n_2 X} \quad (4)$$

其中,

$$n_1 = 2(N - S_N + 1) \quad F \text{ 分布上自由度};$$

$$n_2 = 2S_N \quad F \text{ 分布下自由度};$$

$$X = F_\alpha(n_1, n_2) \quad \text{对应于 } \alpha \times 100\% \text{ 的 } F \text{ 分布}$$

上百分位点.

2 logistic 回归模型与二分类因变量

由于对应长度为 a_i 裂纹的每次检测结果 $P(D|a_i)$ 是一个二分类因变量, 并且裂纹检出概率 $P(D|a_i)$ 与裂纹长度 $a_i (i=1, 2, \dots, n)$ 之间的非线性关系通常是单调递增函数, 即随着裂纹长度 a_i 的增加其检出概率也应随之单调递增, 但这种增加是有限制的, 即要求 $P(D|a_i)$ 的值在区间 $(0, 1)$ 上. 一个自然的选择, 便是找到一个不依赖于对模型本身参数的约束而满足上述条件的、其分布曲线具有 S 形状的概率分布模型, 而 logistic 概率分布模型恰好具有这些特点. 同时, 大量实践发现, $a_i \sim P(D|a_i)$ 关系散点图和 logistic 分布模型曲线的趋势比较接近, 进一步研究表明, logistic 概率分布模型能够描述二分类因变量的分布特点, 比较适合于小裂纹段的裂纹检出概率曲线的拟合. 为了使用方便, 简化运算, 本文采用经过修正的 logistic 分布模型, 如公式 (5) 所示.

$$Y = f(X, \alpha, \beta, \gamma) = \frac{1}{\alpha + \exp(\beta - \gamma X)} \quad (5)$$

对无损检测数据经过统计分析处理后, 通过非线性回归分析, 估计出 logistic 回归模型的分布参数, 进而得到裂纹检出概率曲线.

3 logistic 回归模型参数估计

3.1 初始优良参数的获取

对公式 (5) 进行变换, 重新整理得到公式 (6), 显然, 公式 (6) 对 β, γ 是线性的, 对 α 则不然.

$$Z = \ln\left(\frac{1}{Y} - \alpha\right) = \beta - \gamma X \quad (6)$$

可以从 α 的合适的估计值 α_0 (比如 $\alpha_0 = 0.05$) 开始, 然后求解由公式 (6) 得到的线性最小二乘方程组 (7), 从而得到的 β, γ 的最小二乘估计 β_0, γ_0 . 由于 β_0, γ_0 是在 $\alpha = \alpha_0$ 条件下的得到的一组初始解, 因而, 必须使 α_0 按一定的步长增加, 重复求解方程组 (7), 得到使得公式 (8) 表示的残差平方和 RSS

达到最小的 $\hat{\alpha}_0, \hat{\gamma}_0$. 此时的 $\hat{\alpha}_0, \hat{\beta}_0, \hat{\gamma}_0$ 为 α, β, γ 初始参数优良估计值.

$$\left. \begin{aligned} n\beta - \gamma \sum_{i=1}^n X_i &= \sum_{i=1}^n \ln\left(\frac{1}{Y_i} - \alpha_0\right) \\ \beta \sum_{i=1}^n X_i - \gamma \sum_{i=1}^n X_i^2 &= \sum_{i=1}^n X_i \ln\left(\frac{1}{Y_i} - \alpha_0\right) \end{aligned} \right\} \quad (7)$$

$$RSS = \sum_{i=1}^n \left[Y_i - \frac{1}{\alpha_0 + \exp(\beta_0 - \gamma_0 X_i)} \right]^2 \quad (8)$$

随后这些初始值作为高斯-牛顿法^[5,6]初始值进行迭代, 仅需少数几次迭代就可求出公式 (5) 所表示的 logistic 概率分布模型的分布参数 α, β, γ 的最优估计值 $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$.

3.2 参数迭代求解

非线性回归分析中, 应用较多的参数分析估计方法是 Newton-Raphson 迭代法^[6,7]和 Gauss-Newton 迭代法^[5,8,9]. 这种方法是运用泰勒级数展开原非线性回归函数, 取其前两项, 以线性函数来逼近原非线性回归函数. 本文采用 Gauss-Newton 迭代法进行求解.

记 $\theta = (\alpha, \beta, \gamma)^T, \theta_0 = (\hat{\alpha}_0, \hat{\beta}_0, \hat{\gamma}_0)^T, X = (X_1, X_2, \dots, X_n)^T, Y = (Y_1, Y_2, \dots, Y_n)^T$. 其中, θ 表示公式 (5) 中的未知参数所组成的 3×1 阶矩阵, X 表示裂纹长度的观察值所组成的 $n \times 1$ 阶矩阵, Y 表示对应于裂纹长度 X 的检出概率观察值所组成的 $n \times 1$ 阶矩阵. 同时, 为表示方便, 记 $f(\theta) = f(X, \alpha, \beta, \gamma)$. 根据泰勒级数有公式 (9),

$$f(\theta) \approx f(\theta_0) + Z(\theta_0)(\theta - \theta_0) \quad (9)$$

如果迭代估计是收敛的, 经过简单变换, 得到未知参数最终近似迭代表达式, 如公式 (10) 所示,

$$\theta_{i+1} = \theta_i + [Z^T(\theta_i)Z(\theta_i)]^{-1} Z^T(\theta_i) [Y_i - f(\theta_i)] \quad (10)$$

将初始优良估计值 $\theta_0 = (\hat{\alpha}_0, \hat{\beta}_0, \hat{\gamma}_0)^T$ 作为初值代入上式 (10) 即可进行计算. 其中,

$$Z(\theta_i) = \frac{\partial f(\theta_i)}{\partial \theta} = \begin{bmatrix} \partial f_i(\theta_i) / \partial \alpha & f_i(\theta_i) / \partial \beta & f_i(\theta_i) / \partial \gamma \\ \dots & \dots & \dots \\ f_i(\theta_i) / \partial \alpha & f_i(\theta_i) / \partial \beta & f_i(\theta_i) / \partial \gamma \\ \dots & \dots & \dots \\ f_n(\theta_i) / \partial \alpha & f_n(\theta_i) / \partial \beta & f_n(\theta_i) / \partial \gamma \end{bmatrix}_{n \times 3} \quad (11)$$

由裂纹检出概率观察值 Y_i 和 logistic 概率分布模型的预测值 $f(\theta_i)$ 之差 $(Y_i - f(\theta_i))$ 所组成的矩阵随着迭代次数的增加逐渐趋近于零矩阵, 因而保证了迭代的收敛.

3.3 模型优良性评估

模型未知分布参数求解完成后,需要评价模型是否有效地描述反映了变量及模型预测值匹配观察值的程度.如果模型预测值与对应观察值有较高一致性,就认为这一模型拟合数据.有许多方法可以对 logistic 回归模型的拟合优度进行评价^[10],本文仅从 logistic 概率分布模型预测的准确性和相关系数检验角度对该模型进行分析评价.

模型回归方程预测的准确性是评价模型拟合优良性的指标之一,它有效地表征了预测值与观察值的近似程度.设 y_i 与 \hat{y}_i ($i=1, 2, \dots, n$) 分别是对应长度为 a_i 的裂纹检出概率观察值和 logistic 概率分布模型预测值.定义下列三式,

$$Q = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (13)$$

$$R^2 = 1 - Q / \sum_{i=1}^n (y_i - \bar{y})^2 \quad (14)$$

$$\sigma = \sqrt{Q / (n-2)} \quad (15)$$

式中,

\bar{y} 表示裂纹检出概率观察值的平均值;

Q —残差平方和,其值越小,表明所配函数的相对误差越小;

σ —残差标准差,其值越小,表明所配函数的预测值精确度越高;

R —也称为相关系数,其值越大,表明所配曲线拟合效果越好;

关于回归方程的相关系数检验,常用的是相关系数检验法.值得注意的是,公式(14)所定义的非线性相关系数和线性回归的相关系数定义有所不同.在显著水平为 α , 自由度为 $n-2$, 且当 $R > \gamma_{\alpha, n-2}$ 时认为该回归方程显著 ($\gamma_{\alpha, n-2}$ 为相关系数临界值,可查表获取).

4 算例

该算例的数据来源于 45# 钢的孔边角裂纹某次磁粉探伤的检测结果(数据项 a , N , S_N 为原始数据),如下表 1 所示,表中有关数据已按文中介绍的裂纹数据统计方法经过处理.

表 1 无损检测试验数据

α	N	S_N	\bar{p}	n_1	n_2	X	P_L
2.0—3.0	243	81	0.3333	326	162	1.3215	0.2733
3.0—4.0	126	61	0.4841	132	122	1.3487	0.4066
4.0—5.0	135	84	0.6222	104	168	1.3646	0.5421
5.0—6.0	117	92	0.7863	52	184	1.4505	0.7093
6.0—7.0	108	104	0.9630	10	208	1.9027	0.8455
7.0—8.0	90	87	0.9667	8	174	2.0155	0.9152
8.0—9.0	63	62	0.9841	4	124	2.4497	0.9268
9.0—10.0	63	63	1.0000	2	126	3.0695	0.9535

注:表 1 中 α 表示检测到的裂纹长度(单位:mm), X 为对应于 $\alpha \times 100\%$ 的 F 分布上百分位点,其它请参考本文第一小节.

4.1 参数估计

为了减少计算,该算例仅求出对应置信水平为 95% 的 logistic 概率模型分布参数(取表中裂纹区间的上限值和 P_L 进行分析求解),平均检出概率模型的分布参数的求解与此类似.按照本文介绍的方法,经过计算,logistic 概率模型的初始优良参数分别为 $\hat{\alpha}_0 = 1.04$, $\hat{\beta}_0 = 3.5584$, $\hat{\gamma}_0 = 0.7977$, 残差平方和 $RSS = 0.0053$. 然后利用 Gauss—Newton 迭代法进行迭代,仅经过 5 次迭代即可得到最终分布参数 $\alpha = 1.010599$, $\beta = 3.024607$, $\gamma = 0.660799$, 残差平方和 $RSS = 0.001780$. 因此,长度为 x 的裂纹概率分布函数为

$$P(D | \alpha_i = x) = f(x) =$$

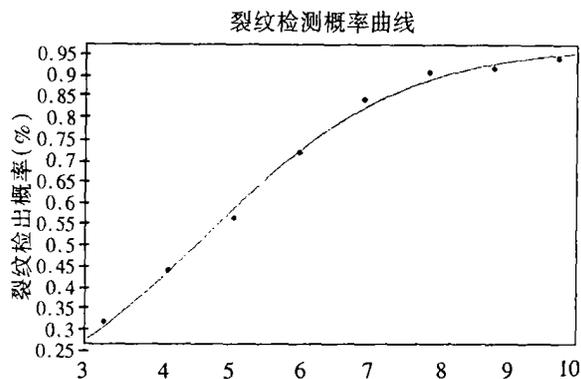
$$\frac{1}{1.010599 + \exp(3.024607 - 0.660799x)}$$

裂纹检出概率曲线如图 1 所示(图中的圆点表示裂纹检出概率观察值的置信下限).

4.2 模型评估

经过上述计算分析得到残差平方和 $Q = 0.00178$, 相关系数 $R = 0.999815$, 残差标准差 $\sigma = 0.0172$, 这三个参数均表明所选 logistic 回归模型能够较好地描述检出裂纹的概率分布.同时,对于在 $\alpha = 0.05$ 显著水平下通过查表得到相关系数临界值 $\gamma_{0.05, 6} = 0.707$, 显然相关系数满足关系 $R > \gamma_{\alpha, n-2}$, 因此该 logistic 回归方程在置信水平 $\alpha = 0.05$ 时显

著.

图1 裂纹检测概率曲线(置信水平 $1 - \alpha = 95\%$)

5 小结

本文提出一种经过修正的 logistic 回归模型,并给出求解这种模型未知参数的一般方法和思路,结合算例,对 logistic 回归模型的拟合优良性进行评估,结果表明该模型的预测值与实验值吻合良好,有较高的预测精度.实践表明,该回归模型相对 Yang-Trapp 提出的幂函数模型对未知参数没有约束,使用方便.同时需要指出的是,文中所用到的相关系数是在非线性方程化为线性方程的过程中对 $f(x)$ 进行变换的基础上得到的,已不具有线性回归

方程的相关系数的意义,但一般来说,它们可以作为曲线拟和好坏的参考,而不能确切地表明 x 与 $f(x)$ 间的拟和情况^[11].值得注意的是,利用柯尔莫哥洛夫检验与卡方拟和优度检验均不能对检出裂纹长度是否服从 logistic 分布进行检验,这是由于本文对裂纹采用与上述两种检验不同的数据统计方法所致.

参考文献:

- [1] 李家伟,陈积懋.无损检测手册[M].北京:机械工业出版社,2002.
- [2] 傅惠民,钱若力.无损检测可靠性与寿命控制方法[M].航空动力学报,1994,9(3):227~232.
- [3] 林富甲.结构可靠性[M].西安:西北工业大学出版社,1991.
- [4] 林富甲,黄玉珊.裂纹检测概率曲线的统计测定[M].航空学报,1982,9(3):227~232.
- [5] (澳)拉特科斯基著,洪再吉,韦博成等译.非线性回归模型统一的实用方法[M].南京:南京大学出版社,1986.
- [6] 范金城,梅长林.数据分析[M].北京:科学出版社,2002.
- [7] 钟万勰,何夯,刘正.数值计算方法[M].北京:中国建筑工业出版社,1991.
- [8] 韦博成.近代非线性回归[M].南京:东南大学出版社,1989.
- [9] 朱军.线性模型分析原理[M].北京:科学出版社,2000.
- [10] 王济川,郭志刚.Logistic 回归模型—方法与应用[M].北京:高等教育出版社,2001.
- [11] 丁振良.误差理论与数据处理(第2版)[M].哈尔滨:哈尔滨工业大学出版社,2002.

The Application of Modified Logistic Model in Fitting of the Flaw Testing Probability Curve

XIAO Peng, BAI Cun-ru

(School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: The research in this paper presented a new modified logistic model which could satisfy the demand of data processing of nondestructive testing. Used least squares, adopted nonlinear regression, and applied Gauss-Newton iteration, by which introduced the detail estimation method of distribution parameter. The goodness of fit of the model was also evaluated by integrating a group of concrete reliability test data. Computational result indicates that prediction values of the proposed modified logistic model are in good agreement with the experimental results and the model is suitable for matching testing probability curve and has the higher prediction accuracy and better flexibility.

Key words: nondestructive testing; logistic model; nonlinear regression; gauss-Newton; predictive accuracy