DOI:10. 16356/j. 1005-2615. 2025. 03. 018

基于Yeo-Johnson变换和最大熵原理的结构可靠性分析方法

张雯静，李洪双
(南京航空航天大学航空学院,南京 210016)

摘要：针对航空航天、车辆等工程领域存在输入数据完全随机缺失的情况,提出了一种基于Yeo-Johnson变换和最大熵原理的结构可靠性分析方法。首先,通过Yeo-Johnson变换将原始数据变换为正态化数据。再次对正态化数据进行扩充抽样,推导Yeo-Johnson逆变换,并对扩充数据进行逆变换。最后利用最大熵原理对全部数据进行概率密度函数拟合及失效概率的计算。将本文方法应用在3个算例中,结果表明:该方法能够对缺失数据进行样本扩充,有效解决因为数据完全随机缺失无法正常进行可靠性分析的情况,所提方法具有较高的可行性和工程应用价值。

关键词：缺失数据;Yeo-Johnson变换;最大熵原理;结构可靠性分析;失效概率

中图分类号：V214.1　　**文献标志码**：A　　**文章编号**：1005-2615(2025)03-0562-10

A Structural Reliability Analysis Method with Yeo-Johnson Transformation and Maximum Entropy Principle

ZHANG Wenjing, *LI Hongshuang*
(College of Aerospace Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China)

Abstract: A structural reliability analysis method is proposed to solve the issue of completely random missing input data in aerospace, vehicle and other engineering fields, with Yeo-Johnson transformation and maximum entropy principle. Firstly, the original data is transformed into normalized one by Yeo-Johnson transformation. Secondly, the size of data is enlarged by adding new samples generated by normal distribution. Then the Yeo-Johnson inverse transformation is derived and the augmented data is converted by it. Finally, the maximum entropy principle is used to determine the probability density function which is further used to calculate failure probability. The proposed method is applied to three illustrative examples. The results show that the proposed method can enlarge the total number of samples in missing data and solve the issue of completely random missing input data in structural reliability analysis. The proposed method has a feasibility and engineering application value.

Key words: missing data; Yeo-Johnson transformation; maximum entropy principle; structural reliability analysis; failure probability

在工程领域存在多种不确定因素,包括材料性能参数、几何尺寸、测量误差以及制造误差等。这些不确定性因素会严重影响结构的安全性和可靠性,因此在对结构零部件进行分析、设计和生产的时候需要以结构可靠性理论作为参考。对于结构可靠性分析问题,其核心在于结构响应的功能函数

基金项目：国家科技重大专项(Y2019-I-0018-0017)。
收稿日期：2024-08-27;**修订日期**：2024-10-23
通信作者：李洪双,男,教授,博士生导师,E-mail:hongshuangli@nuaa.edu.cn。

引用格式：张雯静,李洪双. 基于Yeo-Johnson变换和最大熵原理的结构可靠性分析方法[J]. 南京航空航天大学学报(自然科学版),2025,57(3):562-571. ZHANG Wenjing, LI Hongshuang. A structural reliability analysis method with Yeo-Johnson transformation and maximum entropy principle[J]. Journal of Nanjing University of Aeronautics & Astronautics(Natural Science Edition),2025, 57(3):562-571.

是否达到许用值，并以此为基准结合可靠性方法来计算结构的失效概率[1]。

在实际工程应用中，输入数据完全随机缺失的情况对可靠性分析会造成诸多影响，例如分析结果的可信度等。为了解决这些问题，国内外许多学者针对性地提出了一些解决策略。比如 Rubin[2]提出了多重插补法来处理缺失数据问题，在保留单一插补易于理解和操作的优点的同时，通过含缺失值的数据集中生成多个完整数据集，有效处理和分析数据随机缺失的困难。多重插补法在处理缺失数据时与其他插补法相比更加稳定，对缺失比例较大的数据拟合效果较好，与真实数据误差较小[3]，但需要较多的计算资源来处理多个数据集，且在应用时需要谨慎考虑数据的缺失机制和模式。Breiman[4]建议采用的随机森林方法在处理数据缺失时具有一定的鲁棒性和灵活性，可通过节点分裂，从而减小缺失值对分析结果的影响。但当树的数量过多或者数据噪声较大时，仍然可能会发生过拟合。此外，随机森林模型包含了许多超参数，还易导致调参困难等问题。隐马尔可夫模型[5]在处理时间序列数据中的缺失值时，通过概率推断技术来估计缺失值，同时利用已观测到的数据来推断隐藏状态的序列。但其性能对参数设定和初始化非常敏感，选择不当的参数或初始值可能导致模型训练不稳定或结果不准确。

Box 等[6]在之前研究的基础上进行拓展和修正，进一步探讨统计数据分析中的变量转化问题。Taylor[7]讨论实现了分布对称的幂变换，对 Box-Cox 方法进行了稳健改进。Tibshirani[8]提出的 Lasso 回归方法以及 Zou 等[9]提出的 Elastic-net 回归方法都是在 Box-Cox 方法进行优化和改进，提高了鲁棒性和效率。Yang[10]提出一种修正的幂转化方法，可以根据特定的需求和数据特点对 Box-Cox 变换方法进行调整和改进，包括引入额外参数或变换形式，以更好地适应数据的分布形态。Abbasi 等[11]讨论了一种更简单有效的线性变换方法，降低数据的偏态。Peterson[12]论述了一种半参数变换方法，旨在解决在高通量数据分析中的归一化问题，并提供了一种可靠的归一化方法。

Yeo 等[13]在 Box-Cox 方法的基础上改善了数据的正态性或对称性，拓展了数据转化的范围。与传统的 Box-Cox 方法相较，Yeo-Johnson 方法具有更广泛的适用性。因为 Box-Cox 方法仅适用于处理正值的数据，而 Yeo-Johnson 方法可以处理包括负值和零值在内的数据。除此之外，异常值对 Yeo-Johnson 方法的影响相对较小并且因为引入的可调参数，可保持数据的分布形态，具有更好的鲁棒性和灵活性。Yeo 等[14]论述了一种基于经验特征函数的方法来选取最佳转化方式以实现正态化，该方法的优势在于：不受特定转化方式的限制，能够根据数据的特征动态选择转换方式。应用该方法可改善在正态性假设的前提下进行统计推断和模型建立的过程。

最大熵原理是由 Jaynes[15]信息论中的熵理论和贝叶斯推理方法相结合而提出。信息熵是信息论的一个重要概念，表示随机变量的不确定性，熵越大表示分布的不确定性越大，同时也表示对未知条件的不确定性最少的情况。最大熵原理的核心思想是在一定的约束条件下，选择满足所有约束条件的概率分布中不确定性最大的分布。近年来，最大熵原理在实际工程中得到了广泛应用和发展，李昊燃等[16]为解决因为截断误差而导致的计算结果无法收敛的问题，提出一种改进的最大熵方法，并且结合单变量降维法进行可靠度分析。杜亦牧等[17]基于最大熵原理的复杂系统可靠性分析方法对病毒在宿主、易感染者和移除者等共存系统中的传播问题进行了系统的动力学理论分析。陈晓燕[18]提出了一种基于最大熵法的结构抗爆可靠性分析方法，研究了爆炸载荷作用下结构的可靠性分析问题。郑宏伟等[19]提出一种基于高阶矩最大熵法的结构混合可靠性分析方法，解决了随机-区间变量混合模型可靠性分析问题。周春晓[20]提出了对整数阶矩最大熵法的改进方案，基于非线性变换的失效概率预测精度的零熵变换准则，获得了最大熵改进方案中对非线性变换调制参数选区的判定依据。最大熵方法的优点是在缺乏先验信息的情况下，最大限度利用已知给定约束条件，建立反映数据真实情况的模型来拟合数据分布，且最大熵方法确定的分布类型是唯一的，故利用最大熵方法对补充后的原始数据进行可靠性分析可以最大程度地拟合原始分布。

综上所述，已有的研究在一定程度上解决了数据缺失问题，但计算都较为繁琐且对原始数据的要求较高，在解决实际工程问题时计算成本较高，同时无法直接进行可靠性分析。本文针对上述困难，提出了一种基于 Yeo-Johnson 变换和最大熵原理的结构可靠性分析方法，创新性地利用 Yeo-Johnson 逆变换结合正态分布的特征扩充原始数据，旨在简化计算过程，保留较高的计算精度，解决由于结构分析计算量过大造成的数据完全随机缺失问题。

1　Yeo-Johnson 变换及逆变换

本文通过 Yeo-Johnson 变换，推导了 Yeo-Johnson

逆变换,并提出利用Yeo-Johnson逆变换结合正态分布的特征来处理数据完全随机缺失问题。

1.1 Yeo-Johnson变换

Box-Cox变换是Yeo-Johnson变换的基础,一般用于处理数据中存在偏度的情况,将数据转换为更接近正态分布的形式,以便进行统计分析建模并且提高模型准确性。Box-Cox变换及其逆变换在回归分析中也有着重要的应用。Box-Cox变换函数表达式为

$$\psi^{\mathrm{BC}}(\lambda,x)=\begin{cases}(x^{\lambda}-1)/\lambda & \lambda\neq 0\\ \lg x & \lambda=0\end{cases} \tag{1}$$

式中:x为原始数据(正数),λ为实数。

为了使Box-Cox变换可以在整个实数域范围内都可以使用。Yeo-Johnson变换首先对式(1)进行修改,使其在针对正数和负数变换时的参数不同,扩大其适用范围[13]。

$$\psi(\lambda_+,\lambda_-,x)=\begin{cases}[(x+1)^{\lambda_+}-1]/\lambda_+ & x\geqslant 0,\lambda_+\neq 0\\ \lg(x+1) & x\geqslant 0,\lambda_+=0\\ -[(-x+1)^{\lambda_-}-1]/\lambda_- & x<0,\lambda_-\neq 0\\ -\lg(-x+1) & x<0,\lambda_-=0\end{cases} \tag{2}$$

式中:λ_+和λ_-分别表示变换在正数及零范围内、负数范围内的变换参数,取值范围为实数。

为了使式(2)所代表的变换是平滑且连续的,对上述变换增加约束条件:令$\psi(\lambda_+,\lambda_-,x)$关于$x$的二阶导数$\partial^2\psi(\lambda_+,\lambda_-,x)/\partial x^2$在$x=0$处为连续的,即满足[13]

$$\lim_{x\to 0^+}\partial^2\psi(\lambda_+,\lambda_-,x)/\partial x^2=\lim_{x\to 0^-}\partial^2\psi(\lambda_+,\lambda_-,x)/\partial x^2 \tag{3}$$

易知

$$\begin{cases}\lim\limits_{x\to 0^+}\partial^2\psi(\lambda_+,\lambda_-,x)/\partial x^2=\lambda_+-1\\ \lim\limits_{x\to 0^-}\partial^2\psi(\lambda_+,\lambda_-,x)/\partial x^2=-\lambda_-+1\end{cases} \tag{4}$$

由式(3,4)可知$\lambda_++\lambda_-=2$。结合式(2),令$\lambda_+=\lambda$,则$\lambda_-=2-\lambda$,λ的取值范围为实数,可得

$$\psi(\lambda,x)=\begin{cases}[(x+1)^{\lambda}-1]/\lambda & x\geqslant 0,\lambda\neq 0\\ \lg(x+1) & x\geqslant 0,\lambda=0\\ -[(-x+1)^{2-\lambda}-1]/(2-\lambda) & x<0,\lambda\neq 2\\ -\lg(-x+1) & x<0,\lambda=2\end{cases} \tag{5}$$

式(5)称为Yeo-Johnson变换。Yeo-Johnson变换具有以下性质:

(1)$x\geqslant 0,\psi(\lambda,x)\geqslant 0;x<0,\psi(\lambda,x)<0$。

(2)$\lambda>1$时,$\psi(\lambda,x)$是x的凸函数;$\lambda<1$时,$\psi(\lambda,x)$是x的凹函数。

(3)$\psi(\lambda,x)$是(λ,x)的连续函数。

(4)令$\psi^{(k)}=\partial^k\psi(\lambda,x)/\partial\lambda^k(k\geqslant 1)$,则有

$$\psi^{(k)}=\begin{cases}[(x+1)^{\lambda}\{\lg(x+1)\}^k-k\psi^{(k-1)}]/\lambda & \lambda\neq 0,x\geqslant 0\\ \{\lg(x+1)^{k+1}/(k+1)\} & \lambda=0,x\geqslant 0\\ -[(-x+1)^{2-\lambda}\{-\lg(-x+1)\}^k-k\psi^{(k-1)}]/(2-\lambda) & \lambda\neq 2,x<0\\ \{-\lg(-x+1)\}^{k+1}/(k+1) & \lambda=2,x<0\end{cases} \tag{6}$$

即$\psi^{(k)}$也是(λ,x)的连续函数。

(5)$\psi(\lambda,x)$是λ和x的单调递增函数。

(6)$x>0$时,$\psi(\lambda,x)$是λ的凸函数;$x<0$时,$\psi(\lambda,x)$是λ的凹函数。

通常,使用最大似然估计来计算Yeo-Johnson变换的关键参数λ的值。Asar等[21]提出通过拟合优度检验估计Box-Cox变换参数。本文利用最大似然估计法[13]计算得到λ的值,最大限度地提高数据的正态性或对称性。

Yeo-Johnson变换如图1所示。图中:$f(x)$表示存在完全随机缺失问题的原始数据的概率密度函数(Probability density function, PDF);y表示经过Yeo-Johnson变换后的正态化数据;$\varphi(y)$表示变换后正态化数据对应的概率密度函数。

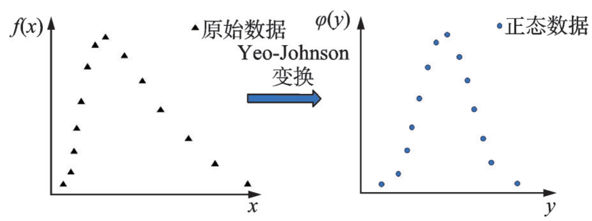

图1 Yeo-Johnson变换示意图

Fig.1 Diagram of Yeo-Johnson transformation

由于Yeo-Johnson变换可以灵活地处理不同类型的数据,使其逐渐成为数据分析领域的重要工具之一,目前对Yeo-Johnson变换的研究和应用较为成熟和广泛,由于调整了分布的形态,经过Yeo-Johnson变换后的数据呈正态化,与原始数据的分布不同,所以不宜直接用于结构可靠性分析。

1.2 Yeo-Johnson逆变换

利用Yeo-Johnson逆变换对Yeo-Johnson变换后的数据进行处理,反推出补充的原始数据,拟合

原始分布，并进行可靠性分析。

由于正态分布的特性，即均值和方差可以完全描述该正态分布，计算正态化数据的均值μ和标准差σ，便可唯一确定变换后数据的完整正态分布信息。因此，可以利用正态分布进行样本扩充，扩充过程如图 2 所示。

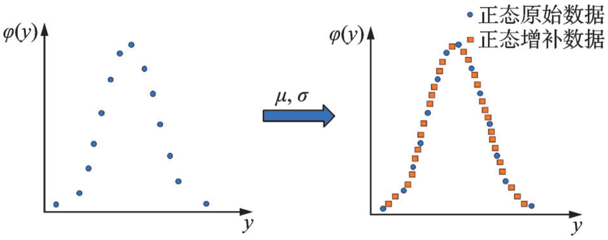

图 2　Yeo-Johnson 变换补充过程示意图

Fig.2　Diagram of Yeo-Johnson transformation replenishment process

在图 2 中，$\varphi(y)$表示 Yeo-Johnson 变换后的正态化数据y的概率密度函数。利用正态分布抽取足够多的扩充样本，记为y'。y'中样本量远远高于原始数据x的样本量。之后，对y'进行 Yeo-Johnson 逆变换，完成原始样本扩充。由于 Yeo-Johnson 变换为λ和x的单调递增函数，故其逆函数一定存在，且自变量与因变量一一对应，即在 Yeo-Johnson 逆变换过程中可保持数据的一一对应。接下来推导 Yeo-Johnson 逆变换过程。

将式(5)改写为

$$y'=\begin{cases}[(x'+1)^{\lambda}-1]/\lambda & x'\geqslant 0,\lambda\neq 0\\ \lg(x'+1) & x'\geqslant 0,\lambda=0\\ -[(-x'+1)^{2-\lambda}-1]/(2-\lambda) & x'<0,\lambda\neq 2\\ -\lg(-x'+1) & x'<0,\lambda=2\end{cases}\tag{7}$$

式中：y'为 Yeo-Johnson 变换后对正态分布抽样的扩充样本，x'为y'对应的原数据x的扩充样本，即经过 Yeo-Johnson 逆变换后得到的扩充样本。

由于 Yeo-Johnson 变换是一一对应的，且在变换过程中满足$x'\geqslant 0$时，$y'\geqslant 0$；$x'<0$时，$y'<0$。故对于 Yeo-Johnson 逆变换，其定义域与 Yeo-Johnson 变换的值域相同。Yeo-Johnson 逆变换的前提是：保持先前 Yeo-Johnson 变换的正态参数λ不变。对式(7)进行反演，可得 Yeo-Johnson 逆变换

$$x'=\begin{cases}(\lambda y'+1)^{1/\lambda}-1 & y'\geqslant 0,\lambda\neq 0\\ (y')^{10}-1 & y'\geqslant 0,\lambda=0\\ 1-[1-(2-\lambda)y']^{1/(2-\lambda)} & y'<0,\lambda\neq 2\\ 1-(-y')^{10} & y'<0,\lambda=2\end{cases}\tag{8}$$

需要强调的是：(1)本文研究的原始数据是完全随机缺失的，式(8)中的y'是原始转化数据及扩充样本的组合；(2)根据目前研究，本文方法仅针对单变量，无法处理多维数据问题，对于多维数据可以进行反复使用，从而进行可靠性分析和计算。Yeo-Johnson 逆变换的过程如图 3 所示。

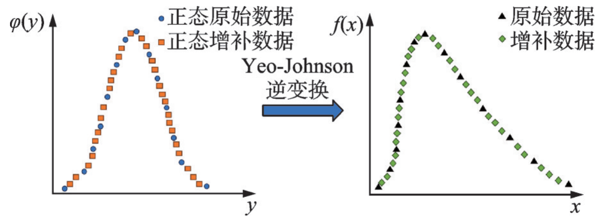

图 3　Yeo-Johnson 逆变换示意图

Fig.3　Diagram of Yeo-Johnson inverse transformation

在图 3 中，x'表示经过 Yeo-Johnson 逆变换后得到的数据，其样本数量远大于原始数据，便于之后的结构可靠性分析。

2　概率分布识别与失效概率计算

本文方法利用扩充数据集确定原始概率分布类型，并进一步完成失效概率计算。扩充数据集首先用于计算前四阶整数矩，并将前四阶整数矩信息作为最大熵方法的约束条件

$$\begin{cases}\mu=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i\\ s^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i-\mu)^2\\ \mathrm{ske}=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i-\mu)^3/s^3\\ \mathrm{kur}=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(x_i-\mu)^4/s^4-3\end{cases}\tag{9}$$

式中μ、s、ske 和 kur 分别为样本的均值、标准差、偏度和峰度。

具有概率密度函数$f(x)$的连续随机变量x的前四阶整数矩为

$$\mu_i=\int x^i f(x)\mathrm{d}x\quad i=1,2,3,4\tag{10}$$

概率密度函数$f(x)$具有非负性、归一性、可积性等性质，由其定义及归一性可知

$$\int f(x)\mathrm{d}x=1\tag{11}$$

对于随机变量X的信息熵定义为

$$H_X=-\int f(x)\ln f(x)\mathrm{d}x\tag{12}$$

式中$f(x)$为随机变量X的概率密度函数。最大熵方法具有唯一性、不变性、系统独立性和独立子集等性质，在无先验信息的条件下，利用给定约束条件，如统计矩信息，使式(12)数值最大的概率分布函数为最佳无偏估计分布[22]。

利用最大熵原理和整数矩,概率分布识别问题可以写成优化问题形式

$$\begin{cases}\text{Find:} & f(x)\\ \text{Maximize:} & H_X=-\int f(x)\ln f(x)\mathrm{d}x\\ \text{Subject to:} & \int f(x)\mathrm{d}x=1\\ & \int x^i f(x)\mathrm{d}x=\mu_i \quad i=1,2,3,4\end{cases}\tag{13}$$

再利用Lagrange乘子法求解式(13)的优化问题

$$L[f(x),\lambda]=-\int f(x)\lg f(x)\mathrm{d}x-(\lambda_0-1)\left[\int f(x)\mathrm{d}x-1\right]-\lambda_i\left[\int x^i f(x)\mathrm{d}x-\mu_i\right]\tag{14}$$

式中$L[f(x),\lambda]$为Lagrange函数。Lagrange函数的最优解条件为

$$\frac{\partial L[f(x),\lambda]}{\partial f(x)}=0\tag{15}$$

将式(14)代入式(15),可以得到如下最大熵分布

$$f(x)=\exp(-\lambda_0)\exp\left(-\sum_{i=1}^{4}\lambda_i x^i\right)\tag{16}$$

式中:$\lambda_i(i=0,1,2,3,4)$为Lagrange乘子,其中标准化因子λ_0为

$$\lambda_0=\ln\left[\int\exp\left(-\sum_{i=1}^{4}\lambda_i x^i\right)\mathrm{d}x\right]\tag{17}$$

本文将前四阶矩信息作为最大熵原理的约束条件,利用Nelder-Mead单纯形算法[23],计算得到相应的Lagrange乘子,进而确定最大熵概率密度函数,即结构响应的概率密度函数。给定结构响应阈值b,结构失效概率P_f为

$$P_f=\int_{-\infty}^{b}f(x)\mathrm{d}x\tag{18}$$

由于结构响应的概率密度函数是解析的,并且是一维函数,因此可采用高斯积分进行数值积分计算。

3 算　例

3.1 算例1:Weibull分布

Weibull分布是可靠性工程中最常用的概率分布之一,是可靠性分析和寿命检验的理论基础。由于该分布的拓展性很高且灵活性强,所以Weibull分布在机电类产品的疲劳可靠性分析中应用广泛。本节利用Weibull分布研究本文方法的适应性。

Weibull分布的概率密度函数可表示为

$$f_X(x)=\frac{\beta}{\eta}\left(\frac{x-r}{\eta}\right)^{\beta-1}\exp\left[-\left(\frac{x-r}{\eta}\right)^{\beta}\right]\tag{19}$$

式中:β为形状参数,控制Weibull分布概率密度函数的形状;η为尺度参数,表征分布的离散程度;r为位置参数,是Weibull分布概率密度曲线在x轴的起点与原点坐标的距离[24]。

Weibull分布的累积分布函数为

$$F_X(x)=1-\exp\left[-\left(\frac{x-r}{\eta}\right)^{\beta}\right]\tag{20}$$

指数分布和正态分布都可以看作是Weibull分布的特例。令$r=0$,$\eta=1.0$,当β从0.5变化至3.0时,分布曲线形状随β的变化从近似指数分布变化至近似正态分布,如图4所示。

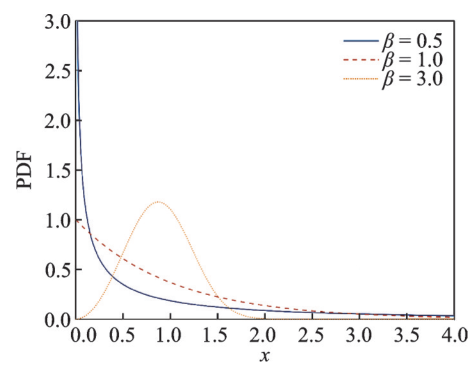

图4　形状参数对Weibull分布影响

Fig.4　Influence of shape parameters on Weibull distribution

针对Weibull分布,考虑当$r=0$、$\eta=1.0$时,对β进行取值,取值结果如表1所示。

表1　β取值范围

Table 1　Range of β

β_1	β_2	β_3	β_4	β_5
0.5	1.0	2.0	3.0	8.0

为了模拟数据完全随机缺失的情况,对不同β值的Weibull分布分别进行随机抽样,作为本例方法的的输入数据,样本量均为30个。将本文方法计算得出的Weibull分布概率密度函数与真实概率密度函数进行比较,结果如图5~9所示。

当$\beta\leqslant 1.0$时,此时偏度大于2.0,峰度大于9.0,Weibull分布近似指数分布,而指数分布的概率密度函数仅有一个延伸至正无穷的右尾,具有高度不对称性及较强的偏态。Yeo-Johnson变换及其逆变换难于消除原始数据较强的偏态对正态化的影响,变换后的数据仍存在一定程度偏离正态分布的情况,使得本文方法拟合的概率密度函数与Weibull分布概率密度函数之间存在较大偏差,拟合效果较差。

当 $\beta_3=2.0$ 时,偏度为 0.6,峰度为 3.4,Weibull 分布呈现右偏,且分布的不对称性减弱,通过 Yeo-Johnson 变换及其逆变换将原始数据转化为正态分布时,受到偏态影响,本文方法拟合的概率密度函数与 Weibull 分布概率密度函数之间存在一定偏差,但与 $\beta\leqslant 1.0$ 的情况相较误差减小。

当 $\beta\geqslant 3.0$ 时,偏度为 0.2,峰度为 2.8;正态分布的偏度为 0.0,峰度为 3.0。Weibull 分布近似正态分布,具有较高对称性和较小偏态,利用 Yeo-Johnson 变换及其逆变换时受偏态影响较小,所以本文方法拟合的概率密度函数与真实的概率密度函数之间误差相对较小,拟合效果较好,拟合精度也较高。当 $\beta_5=8.0$ 时,偏度为 -0.5,峰度为 3.3,本文方法拟合效果较好。

综上所述,本文方法对于偏态过大、不对称性程度较高的分布(例如指数分布),计算结果准确度

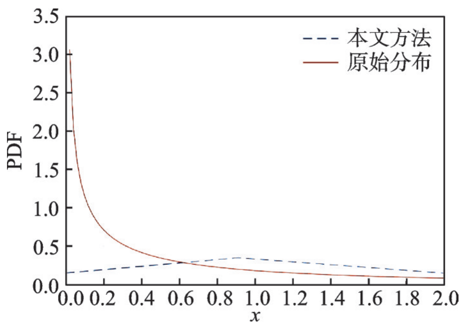

图 5　Weibull 分布 PDF 曲线($\beta_1=0.5,\eta=1.0$)

Fig.5　PDF of Weibull distribution($\beta_1=0.5,\eta=1.0$)

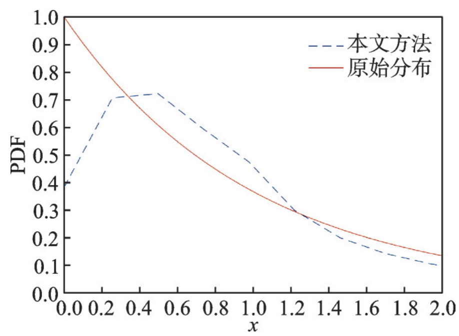

图 6　Weibull 分布 PDF 曲线($\beta_2=1.0,\eta=1.0$)

Fig.6　PDF of Weibull distribution($\beta_2=1.0,\eta=1.0$)

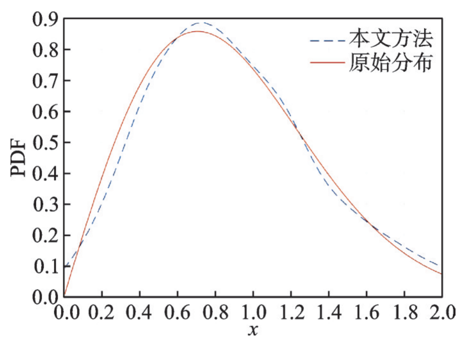

图 7　Weibull 分布 PDF 曲线($\beta_3=2.0,\eta=1.0$)

Fig.7　PDF of Weibull distribution($\beta_3=2.0,\eta=1.0$)

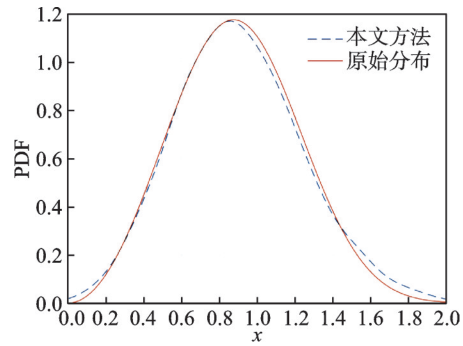

图 8　Weibull 分布 PDF 曲线($\beta_4=3.0,\eta=1.0$)

Fig.8　PDF of Weibull distribution($\beta_4=3.0,\eta=1.0$)

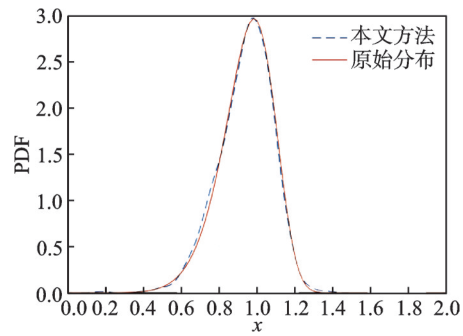

图 9　Weibull 分布 PDF 曲线($\beta_5=8.0,\eta=1.0$)

Fig.9　PDF of Weibull distribution($\beta_5=8.0,\eta=1.0$)

较低;对于偏态较小、对称性较高分布,偏度范围为 $[-0.5,0.5]$,峰度范围为 $[2.7,3.3]$ 的分布计算结果准确度很高。

3.2　算例 2:悬臂梁动力可靠性分析

考虑图 10 所示的悬臂梁在动载荷 $F(t)$ 作用下的结构可靠性分析问题[25]。设材料的密度为 $\rho=1$,横截面积为 A,弹性模量为 E,惯性矩为 I,两者乘积为弯曲刚度 EI,悬臂梁长度 $L=1$,各输入参数单位统一后不标注单位。

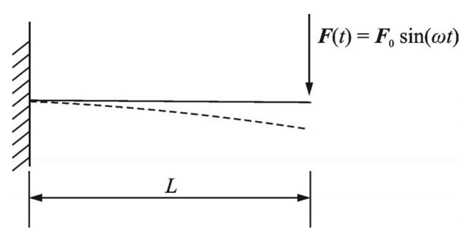

图 10　悬臂梁示意图

Fig.10　Diagram of cantilever

在横向载荷 $p(x,t)$ 的作用下,动力学控制方程为

$$\rho A\frac{\partial^2\omega}{\partial t^2}+EI\frac{\partial^4\omega}{\partial x^4}=p(x,t) \tag{21}$$

边界条件 $x=0$ 为固定端,$x=L$ 为自由端。为了避免悬臂梁发生共振失效,要求激励频率 ω 与

结构的一阶固有频率 ω_1 满足 $\frac{\omega}{\omega_1}<0.99$ 或 $\frac{\omega}{\omega_1}>1.01$。固有频率可由运动齐次方程确定,采用变量分离方法可得

$$\omega_1=\sqrt{\frac{12.362EI}{\rho A}}\times\frac{1}{L^2} \tag{22}$$

令功能函数为 $g=\frac{\omega}{\omega_1}$,当 $0.99\leqslant g\leqslant 1.01$ 时,悬臂梁发生失效。功能函数 g 的概率密度函数及失效概率如图11所示,阴影部分面积值为 A_f,该分布的偏度为0.17,峰度为3.11,在本文方法使用范围内。

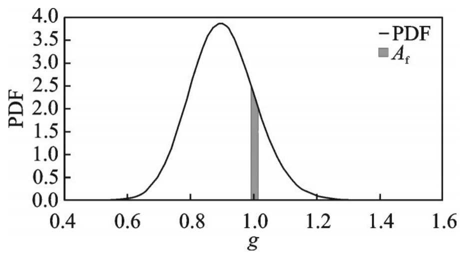

图11 悬臂梁功能函数 g 概率密度函数曲线

Fig.11 PDF of function g of cantilever beam

激励频率、弯曲刚度、横截面积为相互独立的正态分布输入变量,其分布参数如表2所示。

采用直接蒙特卡洛法(Monte Carlo sampling, MCS)和重要抽样法(Importance sampling, IS)估

表2 悬臂梁输入变量分布参数

Table 2 Distribution parameters of input variable of cantilever

变量	分布形式	均值	方差
ω	正态分布	1	0.1
EI	正态分布	0.1	0.01
A	正态分布	1	0.05

计悬臂梁的失效概率,样本数量取为 $N=1\times10^5$。IS计算时,取样本数量为 $N=1\times10^4$。同时采用本文方法、随机森林和多重插补法补充样本,然后计算该悬臂梁的共振失效概率。3种方法分别随机抽取样本量为10、20和30的原始样本,模拟完全随机缺失样本情况,补充后的数据总量为 1×10^3。随机森林法补充缺失数据时,缺失值位置随机分布在每组样本中,决策树数量设置为100,对补充后的数据使用最大熵方法进行可靠性分析。利用多重插补法对缺失数据进行补充时,缺失样本位置也随机分布在每组样本中。先利用初始化插值如均值插值、回归插值对初始样本进行预处理,构建预测模型,再通过多重插补链式方程进行迭代插补,设置迭代次数为10次,汇总分析结果得到插补数据,再对补充数据利用最大熵方法进行可靠性分析。以上计算结果均选择95%置信区间,结果如表3所示。

表3 悬臂梁可靠性分析结果

Table 3 Reliability analysis results of cantilever

方法	样本数量	失效概率95%置信区间	与MCS相对误差/%	与IS相对误差/%
MCS	1×10^5	[0.044 7,0.047 3]	—	—
IS	1×10^4	[0.044 9,0.047 4]	—	—
随机森林法	10	[0.059 8,0.073 0]	[25.25,35.21]	[24.92,35.07]
	20	[0.051 3,0.063 9]	[12.87,25.98]	[12.48,25.82]
	30	[0.047 6,0.053 7]	[6.09,11.92]	[5.67,11.73]
多重插补法	10	[0.055 9.0.071 4]	[20.04,33.75]	[19.68,33.61]
	20	[0.049 8,0.057 8]	[10.24,18.17]	[9.84,17.99]
	30	[0.047 0,0.052 4]	[4.89,9.73]	[4.47,9.54]
本文方法	10	[0.040 1,0.052 9]	[11.47,10.59]	[11.97,10.40]
	20	[0.041 0,0.051 3]	[9.02,7.80]	[9.51,7.60]
	30	[0.043 2,0.048 9]	[3.48,3.27]	[3.94,3.07]

当样本数量从10增加到30时,本文方法与MCS方法的相对误差逐渐减少。当样本数量较少时,由于抽样的不确定性以及样本信息过少,导致结果出现较大相对误差;当样本数量在30及以上时,可以保证较小相对误差。随机森林法与本文方法相比,在原始样本较少时,误差较大,且随机森林法计算量大,运算时间长,模型解释性差。多重插补法与本文方法相比,原始样本较少时,相对误差较大,同时多重插补法需要进行多次迭代运算,较为复杂,增加了计算量。

本算例中利用Yeo-Johnson变换将非正态数据转换为接近正态分布的数据,改善了数据的性能和稳定性,提高了计算的计算效率。同时,利用Yeo-Johnson逆变换扩充数据集,扩大了原始数据的尺度,提供更多的数据信息,便于后续的分析计算和验证。再将补充后的数据应用基于整数矩的最大熵方法进行可靠性分析,得到的失效概率与直接MCS方法计算悬臂梁的共振失效频率相比,相

对误差较小,说明本文方法保证了较高的计算精度,并且极大减少了样本数量,提高了计算效率;与随机森林法和多重插补法相比,本文方法简化了计算过程,减少了计算时间,对相同原始样本计算时精度更高,对本文方法的进一步研究可以探索其在航空航天领域中对飞机结构疲劳分析的应用。

3.3　算例 3:汽车前轴工字梁结构可靠性分析

汽车前轴是汽车的主要构造之一,利用转向节的摆转,实现汽车转向。它主要由工字梁、转向横拉杆、转向节、制动器、轮毂等部分构成。工字梁材料一般选用调质处理的碳钢或 Cr 钢,设计时要满足稳定性要求、抗震要求、刚度要求和强度要求,同时尽量减轻重量。前轴的危险截面常发生在其工字梁上,工字梁及其截面形状[26]如图 12、图 13 所示。

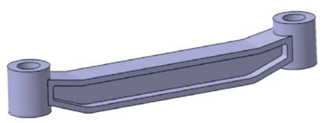

图 12　汽车前轴工字梁示意图

Fig.12　Diagram of I-beam of car front axle

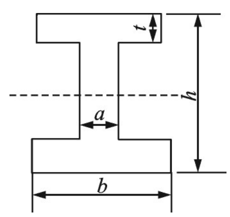

图 13　工字梁截面示意图

Fig.13　Cross-section of I-beam

已知危险截面的最大正应力的值为 $\sigma = M/W_x$ 和 $\tau = T/W_\rho$,其中 M 和 T 分别为前轴所受的弯矩和转矩的值,单位为 N•mm,W_x 和 W_ρ 分别为结构的截面系数和极截面系数,且有

$$W_x = \frac{a(h-2t)^3}{6h} + \frac{b}{6h}\left[h^3 - (h-2t)^3\right] \tag{23}$$

$$W_\rho = 0.8bt^2 + 0.4\left[a^3 \frac{h-2t}{t}\right] \tag{24}$$

式中:工字梁的几何参数 a、b、t、h 的单位为 mm,由前轴的材料特性可知,该前轴的静强度极限为 $\sigma_s = 460$ MPa,故由静强度分析并构建模型的功能函数为

$$g = \sigma_s - \sqrt{\sigma^2 + 3\tau^2} \tag{25}$$

功能函数 g 的概率密度函数如图 14 所示,阴影部分面积值为 A_f,该分布的偏度为 −0.31,峰度为 3.18,在本文方法适用范围内。

前轴所受的弯矩和转矩以及工字梁的几何参

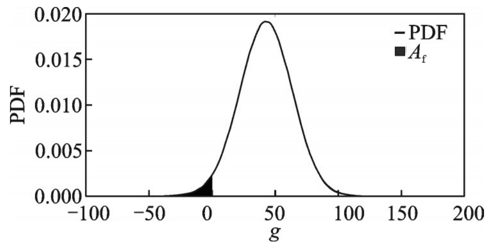

图 14　工字梁功能函数 g 概率密度函数曲线

Fig.14　PDF of the functional function g of the I-beam

数 a、b、t 和 h 为相互独立的正态分布输入变量,其分布参数[27]如表 4 所示。

表 4　工字梁输入变量分布参数

Table 4　Distribution parameters of input variable of I-beam

变量	分布形式	均值	方差
a/mm	正态分布	12	0.060
b/mm	正态分布	65	0.325
t/mm	正态分布	14	0.070
h/mm	正态分布	85	0.425
M/(N•mm)	正态分布	3.5×10^6	1.75×10^5
T/(N•mm)	正态分布	3.1×10^6	1.55×10^5

本节算例与 3.2 节算例的抽样过程相同,计算结果如表 5 所示。由表 5 结果可得,当样本数量 N=10 时,本文方法与 MCS、IS 的相对误差超过

表 5　工字梁可靠性分析结果

Table 5　Reliability analysis results of I-beam

方法	样本数量	失效概率 95% 置信区间	与 MCS 相对误差/%	与 IS 相对误差/%
MCS	1×10^6	[0.018 9, 0.020 6]	—	—
IS	1×10^4	[0.019 1, 0.020 5]	—	—
随机森林法	10	[0.026 9, 0.034 1]	[29.74, 39.59]	[29.00, 39.88]
	20	[0.024 5, 0.028 9]	[22.86, 28.72]	[22.04, 29.07]
	30	[0.022 9, 0.026 5]	[17.47, 22.26]	[16.59, 22.64]
多重插补法	10	[0.025 4, 0.033 5]	[26.17, 38.51]	[25.39, 38.81]
	20	[0.022 8, 0.027 7]	[17.11, 25.63]	[16.23, 25.99]
	30	[0.021 5, 0.025 9]	[12.09, 20.46]	[11.16, 20.85]
本文方法	10	[0.015 3, 0.026 7]	[24.34, 22.85]	[25.66, 22.93]
	20	[0.016 9, 0.023 8]	[11.83, 12.34]	[13.02, 12.77]
	30	[0.017 5, 0.021 9]	[8.00, 5.94]	[9.14, 6.39]

20%,难以保证计算结果的精度,当样本数量$N=30$时,与MCS、IS的相对误差可以控制在10%以内。算例2和算例3说明:应用本文方法时,样本数量至少要在20以上,若要保证较高的精度,应使用30个以上样本。但MCS和IS都需足够的样本量来保证结果收敛,计算量大且耗时长;随机森林法和多重插补法对于缺失数据拟合效果较好,对于样本量较少的数据无法保证分布的计算精度且计算复杂、时间较长。而本文方法的原始输入样本数量仅为30时可补充较多原始数据信息,保证可靠性分析结果的置信性,说明了本文方法对非正态分布在数据完全随机缺失下的可靠性分析计算具有较高的准确性,可为类似的数据缺失可靠性分析提供参考。上述算例本文方法均应用于独立输入变量的可靠性分析问题,对于相关或不独立的输入变量情况,可通过将本文方法与Copula模型相结合的方式进行分析和计算。

4 结 论

针对工程领域中常见的输入数据完全随机缺失问题,本文提出了一种基于Yeo-Johnson变换和最大熵原理的结构可靠性分析方法。通过Yeo-Johnson变换及逆变换实现原始数据的扩充,利用最大熵原理拟合数据分布,进而完成失效概率评估。该方法提供了一种简单而有效的可靠性分析工具,为解决数据完全随机缺失问题提供新的选择。利用3个算例验证了本方法的可行性,并得出以下结论:

(1)本文方法适合偏度在[−0.5,0.5],峰度在[2.7,3.3]的中等偏度及近似对称的数据集,偏度过大或者对称性不高可能导致分析结果较差。

(2)原始数据量大于20时,可以使用本文方法;数据量大于30时,本文方法表现较好,可以保证在95%置信水平下,失效概率评估结果具有较高精度。

参考文献:

[1] 吕震宙,宋述芳,李洪双,等.结构机构可靠性及可靠性灵敏度分析[M].北京:科学出版社,2009.
LYU Zhenzhou, SONG Shufang, LI Hongshuang, et al. Structural system reliability and sensitivity analysis of reliability[M]. Beijing: Science Press, 2009.

[2] RUBIN D B. Multiple imputation for nonresponse in surveys[M]. [S.l.]: Wiley, 1987.

[3] 席梦瑶,赖俊峰,张改梅.基于数据不同缺失率的插补方法比较[J].内蒙古工业大学学报(自然科学版), 2023, 42(5): 391-395.
XI Mengyao, LAI Junfeng, ZHANG Gaimei. Comparison of interpolation methods based on different missing rates of data[J]. Journal of Inner Mongolia University of Technology (Natural Science Edition), 2023, 42(5): 391-395.

[4] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45: 5-32.

[5] TRUQUET L. Local stationarity and time-inhomogeneous Markov chains[J]. Institute of Mathematical Statistics, 2019, 47(4): 2023-2050.

[6] BOX G E P, COX D R. An analysis of transformations revisited, rebuffed[J]. Journal of the American Statistical Association, 1982, 77(377): 209.

[7] TAYLOR J M G. The retransformed mean after a fitted power transformation[J]. Journal of the American Statistical Association, 1986, 81(393): 114-118.

[8] TIBSHIRANI R. Regression shrinkage and selection via the lasso: A retrospective[J]. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2011, 73(3): 273-282.

[9] ZOU H, HASTIE T. Regularization and variable selection via the elastic net[J]. Journal of the Royal Statistical Society Series B: Statistical Methodology, 2005, 67(2): 301-320.

[10] YANG Z L. Estimating a transformation and its effect on Box-CoxT-ratio[J]. Test, 1999, 8(1): 167-190.

[11] ABBASI B, AKHAVAN NIAKI S T, SEYEDAN S E. A simple transformation method in skewness reduction[J]. International Journal of Engineering, Transactions A: Basics, 2011, 24(2): 169-175.

[12] PETERSON R A. Finding optimal normalizing transformations via best normalize[J]. The R Journal, 2021, 13(1): 310-329.

[13] YEO I K, JOHNSON R A. A new family of power transformations to improve normality or symmetry[J]. Biometrika, 2000, 87(4): 954-959.

[14] YEO I K, JOHNSON R A, DENG X W. An empirical characteristic function approach to selecting a transformation to normality[J]. Communications for Statistical Applications and Methods, 2014, 21(3): 213-224.

[15] JAYNES E T. Information theory and statistical mechanics[J]. Physical Review, 1957, 106(4): 620-630.

[16] 李昊燃,李刚,曾岩.基于改进最大熵方法的可靠性分析[C]//2014工程结构可靠性设计理论、方法与应用学术研讨会.大连,中国:[s.n.], 2014: 156-159, 189.
LI Haoran, LI Gang, ZENG Yan. Reliability analysis based on improved maximum entropy method[C]// Proceedings of the 2014 Symposium on Theory, Method and Application of Reliability Design for Engineering Structures. Dalian, China: [s.n.], 2014: 156-159, 189.

[17] 杜亦牧,孙昌璞.基于最大熵原理可靠性分析的病毒

传播模型[J]. 科学通报，2020，65(22)：2356-2362.
DU Yimu，SUN Changpu. A virus propagation model based on maximum entropy principle for reliability analysis[J]. Science Bulletin，2020，65(22)：2356-2362.

[18] 陈晓燕. 基于最大熵法的结构抗爆可靠性分析[D]. 哈尔滨：哈尔滨工程大学，2013.
CHEN Xiaoyan. Reliability analysis of structural blast resistance based on maximum entropy method[D]. Harbin：Harbin Engineering University，2013.

[19] 郑宏伟，孟广伟，李锋，等. 基于高阶矩最大熵方法的结构混合可靠性分析[J]. 机械工程学报，2021，57(14)：282-290，303.
ZHENG Hongwei，MENG Guangwei，LI Feng，et al. Structural hybrid reliability analysis based on high-order moment maximum entropy method[J]. Journal of Mechanical Engineering，2021，57(14)：282-290，303.

[20] 周春晓. 改进最大熵方法及其在水下航行体可靠性分析中的应用[D]. 大连：大连理工大学，2019.
ZHOU Chunxiao. Improved maximum entropy method and its application in reliability analysis of underwater vehicles[D]. Dalian：Dalian University of Technology，2019.

[21] ASAR Ö，ILK O，DAG O. Estimating Box-Cox power transformation parameter via goodness-of-fit tests[J]. Communications in Statistics-Simulation and Computation，2017，46(1)：91-105.

[22] 王宇，李洪双. 最大熵准则识别材料疲劳寿命分布[J]. 航空工程进展，2015，6(3)：297-305.
WANG Yu，LI Hongshuang. Maximum entropy principle for identifying the fatigue life distribution of material[J]. Advances in Aeronautical Science and Engineering，2015，6(3)：297-305.

[23] LAGARIAS J C，REEDS J A，WRIGHT M H，et al. Convergence properties of the nelder-mead simplex method in low dimensions[J]. SIAM Journal on Optimization，2006，9(1)：112-147.

[24] 李洪双，马远卓. 结构可靠性分析与随机优化设计的统一方法[M]. 北京：国防工业出版社，2015.
LI Hongshuang，MA Yuanzhuo. Unified methods for structural reliability analysis and stochastic optimization design[M]. Beijing：National Defense Industry Press，2015.

[25] AU S K，WANG Y. Engineering risk assessment with subset simulation[M]. Singapore：Wiley，2014.

[26] 张磊刚，吕震宙，陈军. 基于失效概率的矩独立重要性测度的高效算法[J]. 航空学报，2014，35(8)：2199-2206.
ZHANG Leigang，LYU Zhenzhou，CHEN Jun. Efficient algorithm for moment independent importance measurement based on failure probability[J]. Acta Aeronautica et Astronautica Sinica，2014，35(8)：2199-2206.

[27] WANG P，LU Z Z，TANG Z C. An application of the Kriging method in global sensitivity analysis with parameter uncertainty[J]. Applied Mathematical Modelling，2013，37(9)：6543-6555.

（编辑：陈珺，王婕）