

一种选择工程随机载荷统计分布的 累积剩余熵法*

杨 于^{1,2}, 梁英杰¹, 陈 文¹

(1. 河海大学 力学与材料学院 软物质力学研究所, 南京 211100;

2. 河海大学 理学院, 南京 210098)

(我刊编委陈文来稿)

摘要: 基于 Lévy 稳定分布,给出了一种选择工程随机载荷统计分布的累积剩余熵法.该方法采用候选分布估计载荷的累积分布和尾分布,并计算累积剩余熵,以及候选分布与真实分布的相对距离,相对距离越小表明候选分布越接近载荷的真实分布.结合永济第二节制闸上下游水位工程实例,对提出的方法进行了验证.计算结果表明:与正态分布、极值 I 型分布相比,Lévy 稳定分布的精度最高,能够较好地刻画水位分布的拖尾性;累积剩余熵的相对距离表明 Lévy 稳定分布与实际分布之间的误差最小,因此,基于 Lévy 稳定分布的累积剩余熵是一种有效工程随机载荷统计分布的方法,保证了工程结构可靠度计算的准确性.

关键词: Lévy 稳定分布; 累积熵; 累积剩余熵; 随机载荷; 可靠度

中图分类号: O232; O172; O373 **文献标志码:** A **DOI:** 10.21656/1000-0887.390157

引 言

在工程结构可靠性的研究中,根据功能函数判断和确定工程的可靠度和失效概率^[1-6].功能函数是工程随机载荷的函数,计算功能函数需要确定随机变量的统计分布^[7].工程随机载荷对应统计分布的尾部信息,在工程结构可靠度计算中起着关键作用^[8-9].目前,常用于描述工程随机载荷的统计分布统称为指数分布,正态分布常用于描述永久载荷^[3],极值 I 型分布常用于描述极大风速^[10].然而这些指数型分布,仅通过观测和试验资料的期望和方差确定工程随机载荷的统计分布,忽略了峰态、非对称、尾部等信息,且不能准确描述载荷概率密度的尖峰和尾部分布的特征^[7].虽然常规指数型统计分布可用于计算工程结构可靠度,但会导致可靠度的计算结果不同程度的偏低或者偏高,不能较全面和真实地反映服役结构的可靠性和剩余寿命等信息,甚至会得出与实际情况相悖的结论^[7].因此,需要新的统计分布描述工程随机载荷的统计特征.

* 收稿日期: 2018-05-31; 修订日期: 2018-10-16

基金项目: 111 引智计划(B12032);中央高校基本科研业务费(2017B01114)

作者简介: 杨于(1992—),女,硕士生(E-mail: 18705162653@163.com);

梁英杰(1988—),男,讲师,博士(E-mail: liangyj@hhu.edu.cn);

陈文(1967—2018),男,教授,博士,博士生导师(通讯作者. E-mail: chenwen@hhu.edu.cn).

Lévy 稳定分布^[11]是一类常用于描述非 Gauss(高斯)随机变量的统计分布,已用于 Brown(布朗)运动、金融领域、反常扩散、信号处理、股票市场模拟、楼面活荷载、水位波动等领域^[4,12-17]。该分布可以用于描述工程随机载荷的非指数型统计分布。描述该分布需要 4 个参数,分别是稳定指数 α 、倾斜指数 β 、尺度参数 γ 和位置参数 δ ^[18]。该分布的概率密度函数具有明显的尖峰、非对称和拖尾特征^[18]。Cauchy(柯西)分布($\alpha = 1$)和 Gauss 分布($\alpha = 2$)是 Lévy 稳定分布的两种特例。Lévy 稳定分布的特征函数具有解析表达式,但其概率密度函数和累积分布函数没有具体的解析表达式,利用现有的 MATLAB 工具箱^[19]可以准确数值计算该分布的统计量。采用 Lévy 稳定分布描述工程随机载荷的统计分布,面临的问题是如何从众多统计分布中选择工程随机载荷的最优统计分布。

香农熵^[20]常用于定量分析随机系统的不确定性。一般情况下,任给的随机系统,其随机变量对应统计分布的熵可以度量该随机系统的复杂程度。熵越大,表明该系统越混乱,反之则越有序。目前,累积熵^[21]是一种可以用于选择随机载荷统计分布的最优统计分布。其对应的相对距离函数的值越小,表明工程随机载荷的候选统计分布越接近实际分布。但是该方法不能直接利用统计分布的尾部信息,不能体现 Lévy 稳定分布用于描述工程随机载荷的优势。因此,需要提出一种直接结合工程随机载荷统计分布的尾部信息,选择工程随机载荷的最优统计分布。

本文将基于现有的工作^[4,21-22],提出了一种选择工程随机载荷最优统计分布的剩余累积熵法。该方法可以直接刻画随机载荷的尾部特征,并结合永济第二节制闸工程实例,全面考察上下游水位载荷的统计特征,并与累积熵法进行了比较。

1 Lévy 稳定分布

1.1 Lévy 稳定分布的定义

目前常用的两种参数化方法描述 Lévy 稳定分布,分别是由 Samorodnitsky、Taqqu^[22]和 Nolan^[23]提出的。对其 $X \sim S(\alpha, \gamma, \beta, \delta)$ 的具体表达式分别如下:

$$\varphi(t) = \begin{cases} \exp\left(-\gamma^\alpha |t|^\alpha \left[1 - i\beta \operatorname{sign}(t) \tan \frac{\pi\alpha}{2}\right] + i\delta t\right), & \alpha \neq 1, \\ \exp\left(-\gamma |t| \left[1 + i\beta \frac{2}{\pi} \operatorname{sign}(t) \ln |t|\right] + i\delta t\right), & \alpha = 1, \end{cases} \quad (1)$$

$$\varphi(t) = \begin{cases} \exp\left(-\gamma^\alpha |t|^\alpha \left[1 + i\beta \operatorname{sign}(t) \tan \frac{\pi\alpha}{2} [(\gamma |t|^{1-\alpha}) - 1]\right] + i\delta t\right), & \alpha \neq 1, \\ \exp\left(-\gamma |t| \left[1 + i\beta \frac{2}{\pi} \operatorname{sign}(t) \ln(\gamma |t|)\right] + i\delta t\right), & \alpha = 1. \end{cases} \quad (2)$$

式(2)克服了 $\alpha = 1$ 处不连续的缺点,其中参数的取值分别为 $0 < \alpha \leq 2$, $-1 \leq \beta \leq 1$, $\gamma > 0$, $\delta \in \mathbf{R}$ ^[18]。

1.2 Lévy 稳定分布的累积分布函数

目前直接积分法是常用于数值计算 Lévy 稳定分布累积分布函数的方法。直接积分法由 Zolotarev^[24]提出,在描述该方法时需要定义如下变量:

$$\zeta = \zeta(\alpha, \beta) = -\beta \tan \frac{\pi\alpha}{2}, \quad (3)$$

$$\mu = \mu(\alpha, \beta) = \frac{1}{\alpha} \arctan(-\zeta), \quad (4)$$

$$v(\theta; \alpha, \beta) = \begin{cases} (\cos(\alpha\mu))^{1/(\alpha-1)} \left(\frac{\cos \theta}{\sin(\alpha(\theta + \mu))} \right)^{\alpha/(\alpha-1)} \frac{\cos(\alpha\mu + (\alpha - 1)\theta)}{\cos \theta}, & \alpha \neq 1, \\ \frac{2}{\pi} \left(\frac{\pi/2 + \beta\theta}{\cos \theta} \right) \exp\left(\frac{1}{\beta}(\pi/2 + \beta\theta) \tan \theta\right), & \alpha = 1, \beta \neq 0. \end{cases} \quad (5)$$

令 $X \sim S(\alpha, \beta; 0)$, X 的表达式如下:

(a) 当 $\alpha \neq 1$ 且 $x > \zeta$ 时,

$$F(x; \alpha, \beta) = c_1(\alpha, \beta) + \frac{\text{sign}(1 - \alpha)}{\pi} \int_{-\mu}^{\pi/2} \exp(-(x - \zeta)^{\alpha/(\alpha-1)} V(\theta; \alpha, \beta)) d\theta, \quad (6)$$

其中

$$c_1(\alpha, \beta) = \begin{cases} \frac{1}{\pi} \left(\frac{\pi}{2} - \mu \right), & \alpha < 1, \\ 1, & \alpha > 1; \end{cases}$$

(b) 当 $\alpha \neq 1$ 且 $x = \zeta$ 时,

$$F(x; \alpha, \beta) = \frac{1}{\pi} \left(\frac{\pi}{2} - \mu \right); \quad (7)$$

(c) 当 $\alpha \neq 1$ 且 $x < \zeta$ 时,

$$F(x; \alpha, \beta) = 1 - F(-x; \alpha, -\beta); \quad (8)$$

(d) 当 $\alpha = 1$ 且 $\beta = 0$ 时,

$$F(x; 1, \beta) = \frac{1}{2} + \frac{1}{\pi} \arctan x; \quad (9)$$

(e) 当 $\alpha = 1$ 且 $\beta \neq 0$ 时,

$$F(x; 1, \beta) = \begin{cases} \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} \exp\{-e^{-\pi x/(2\beta)} V(\theta; 1, \beta)\} d\theta, & \beta > 0, \\ 1 - F(x; 1, -\beta), & \beta < 0. \end{cases} \quad (10)$$

通过一些性质可以计算出一般形式下 Lévy 稳定分布的累积分布函数^[21].

2 累积熵和累积剩余熵

2.1 累积熵

离散随机变量 x_i 的香农熵为

$$H = - \sum_{i=1}^n f(x_i) \log_b f(x_i), \quad (11)$$

其中, $f(x_i)$ 为 x_i 出现的频率或者概率, 是根据实验数据的直方图得到的, n 为系统状态的数量, 即样本大小的长度, b 为对数的基底^[25].

累积熵根据式(11)推导得到, 已用于选择随机变量的最优统计分布^[21]:

$$E_1 = - \sum_{i=1}^n \frac{p(x_i)}{P(x)} \log_b \frac{p(x_i)}{P(x)}, \quad (12)$$

其中, $p(x_i)$ 为 x_i 的累积概率, n 为系统状态的数量, 即样本大小的长度, b 为对数的基底以及 $P(x) = \sum_{i=1}^n p(x_i)$. 通过验证可知累积熵是非负的和可扩展的. 累积熵和香农熵之间最明显的

差异在于前者与随机系统的累积分布函数相关,而后者取决于实验数据直方图的间隔长度即概率密度函数.与香农熵相比,累积熵的计算结果更准确^[21].

2.2 累积剩余熵

累积剩余熵的核心思想是将累积剩余概率替代累积熵中的累积概率,用以直接描述随机变量统计分布的尾部信息.累积剩余熵的定义为

$$E_2 = - \sum_{i=1}^n \left(\frac{1 - p(x_i)}{P(x)} \right) \log_b \left(\frac{1 - p(x_i)}{P(x)} \right), \quad (13)$$

式中, $p(x_i)$ 为 x_i 的累积概率, n 为系统状态的数量,即样本大小的长度, b 为对数的基底以及 $P(x) = n - \sum_{i=1}^n p(x_i)$.

2.3 相对距离

在文献[21]中,将候选分布对应的累积熵与随机变量的实际分布对应的标准累积熵之间的相对距离定义为 ε :

$$\varepsilon = |E_s - E_r|, \quad (14)$$

其中, E_s 为候选分布的累积熵, E_r 为实际分布的标准累积熵^[21].

结合累积剩余熵以及累积熵的相对距离可以定义累积剩余熵的相对距离为

$$\varepsilon = |E_t - E_r|, \quad (15)$$

其中, E_t 为候选分布的累积剩余熵, E_r 为实际分布的标准累积剩余熵.

根据式(14)和(15)可以估计候选分布对应的相对距离,由此得到距离集合,距离集合中最小值对应的分布模型即是候选分布中最优的统计分布^[21].

3 应用实例

本节以永济第二节制闸为例,采用1995年—2001年上游和下游水位载荷的实测数据^[3](由数据可知,上游水位的变化范围是50~300 cm,下游水位的变化范围是150~400 cm).上下游水位统计分布的规律直接影响水闸可靠度的计算^[4].根据上述方法,确定水位的最优统计分布.具体来讲:首先分别采用Lévy稳定分布、极值I型分布以及正态分布作为候选分布拟合水位的累积分布和尾部分布;然后采用累积熵和累积剩余熵分别计算相对距离;最后比较各候选分布对应的相对距离,确定上下水位的最优统计分布.

上游水位对应的拟合结果详见文献[4].文献[4]的结果表明,上游水位服从Lévy稳定分布.图1给出了正态分布、极值I型分布和Lévy稳定分布拟合下游水位的累积分布函数曲线.由图1可见,与极值I型分布相比,正态分布和Lévy稳定分布更接近下游水位的实际分布,并且Lévy稳定分布稳定指数的值为1.95,表明下游水位与正态分布较接近.

为进一步考察下游水位统计分布的尾部特征,图2给出了正态分布、极值I型分布和Lévy稳定分布拟合下游水位尾部分布函数曲线.由图2可见,Lévy稳定分布可以准确地刻画尾部分布的特征.

根据累积分布和尾部分布的计算结果,结合式(12)~(15),分别计算各个候选分布对应累积熵和累积剩余熵的相对距离 ε , 见表1、2(表1中, ε_1 表示上游水位不同统计分布对应累积熵和标准累积熵的相对距离, ε_2 表示下游水位不同统计分布对应累积熵和标准累积熵的相对距离;表2中, ε_3 表示上游水位不同统计分布对应累积剩余熵和系统标准累积剩余熵的相对距离, ε_4 表示下游水位不同统计分布对应累积剩余熵和标准累积剩余熵的相对距离).由表1、表2可知,在4种情况下,Lévy稳定分布对应相对距离的值均最小,也表明Lévy稳定分布是

候选分布中最接近随机变量的实际分布,更进一步验证了 Lévy 稳定分布可有效描述上下游水位的统计分布.需要指出的是,在计算上游水位对应的累积剩余熵时,剔除了水位小于 100 cm 的数值,主要原因是小的水位数值对应的累积剩余概率较大,则对累积剩余熵的干扰比较大.

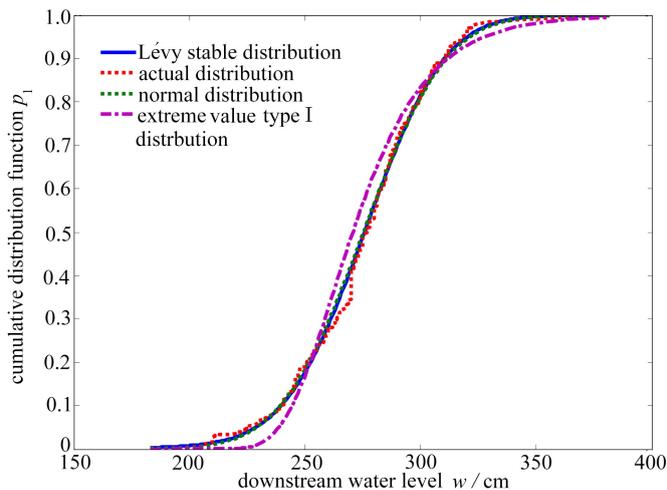


图 1 不同统计分布拟合下游水位累积分布函数的曲线

Fig. 1 Plots of the cumulative distribution function for the downstream water level fitted with 3 different statistical distributions

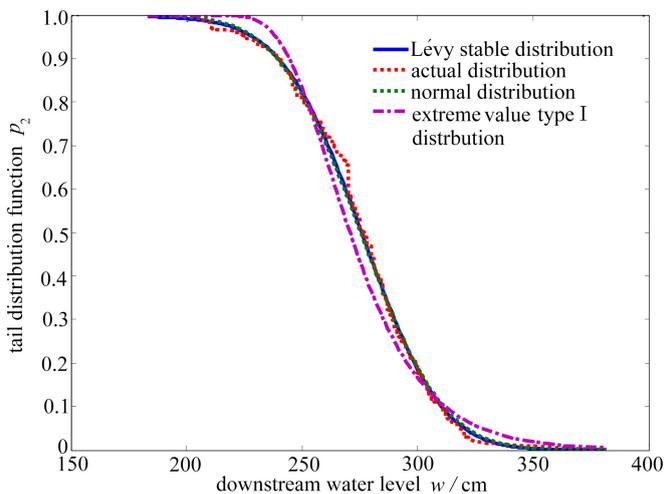


图 2 不同统计分布拟合下游水位尾部分布函数的曲线

Fig. 2 Plots of the tail distribution function for the downstream water level fitted with 3 different statistical distributions

表 1 上下游水位不同统计分布对应累积熵和标准累积熵的相对距离

Table 1 The relative distance between the cumulative entropy and the standard cumulative entropy of different statistical distributions for upstream and downstream water levels

distance function	ε_1	ε_2
extreme value type I distribution	0.014 1	0.013 3
normal distribution	0.028 8	0.005 8
Lévy stable distribution	0.002 6	0.002 8

表 2 上下游水位不同统计分布对应累积剩余熵和标准累积剩余熵的相对距离

Table 2 The relative distance between the cumulative residual entropy and the standard cumulative residual entropy of different statistical distributions for upstream and downstream water levels

distance function	ε_3	ε_4
extreme value type I distribution	0.021 4	0.017 1
normal distribution	0.011 5	0.009 8
Lévy stable distribution	0.006 0	0.006 4

4 结 论

本文基于 Lévy 稳定分布,给出了一种选择工程随机载荷最优统计分布的累积剩余熵,并将其用于选择永济第二节制闸上下游水位载荷的统计分布.通过计算累积剩余熵的相对距离,确定了上下游水位的最优统计分布,并与累积熵法进行了比较,为准确计算水闸的可靠度提供了保障.所得结论如下:

- 1) 上下游水位的最优统计分布均是 Lévy 稳定分布.与正态分布、极值 I 型分布相比,Lévy 稳定分布具有更高的拟合精度,可以刻画上下游水位的拖尾特征.
- 2) 累积剩余熵的相对距离表明 Lévy 稳定分布与实际分布之间的误差最小,Lévy 稳定分布可以有效描述水位的统计分布.
- 3) 与累积熵法相比,累积剩余熵也是一种有效选择工程随机载荷最优统计分布的方法.

参考文献(References):

- [1] 赵国藩. 结构可靠度理论[M]. 北京: 中国建筑工业出版社, 2000. (ZHAO Guofan. *Structural Reliability Theory*[M]. Beijing: China Building Industry Press, 2000. (in Chinese))
- [2] 李全旺, 李春前, 孙健康. 基于结构可靠性理论的既有桥梁承载能力评估[J]. 工程力学, 2010, 27(S2): 142-151. (LI Quanwang, LI Chunqian, SUN Jiankang. Reliability-based capacity assessment of existing bridges[J]. *Engineering Mechanics*, 2010, 27(S2): 142-151 (in Chinese))
- [3] 邹春霞. 基于最大熵原理的水工建筑物可靠性分析[D]. 硕士学位论文. 呼和浩特: 内蒙古农业大学, 2004. (ZOU Chunxia. Reliability analysis of hydraulic construction based on the principle of maximum entropy[D]. Master Thesis. Hohhot: Inner Mongolia Agricultural University, 2004. (in Chinese))
- [4] LIANG Y, CHEN W. Reliability analysis for sluice gate anti-sliding stability using Lévy stable distributions[J]. *Signal Processing*, 2015, 107: 425-432.
- [5] PROPPE C. Estimation of failure probabilities by local approximation of the limit state function[J]. *Structural Safety*, 2008, 30(4): 277-290.
- [6] SU Y H, LI X, XIE Z Y. Probabilistic evaluation for the implicit limit-state function of stability of a highway tunnel in China[J]. *Tunnelling and Underground Space Technology*, 2011, 26(2): 422-434.
- [7] 梁英杰. 分数阶统计力学的研究及其在工程可靠性和反常扩散中的应用[D]. 博士学位论文. 南京: 河海大学, 2016. (LIANG Yingjie. A study on fractional statistical mechanics and its applications in engineering reliability and anomalous diffusion[D]. PhD Thesis. Nanjing: Hohai University. 2016. (in Chinese))
- [8] BLUDSZUWEIT H, DOMINGUEZ-NAVARRO J A, LLOMBART A. Statistical analysis of wind

- power forecast error[J]. *IEEE Transactions on Power Systems*, 2008, **23**(3): 983-991.
- [9] MAHANI A, KAVIAN Y S, NADERI M, et al. Heavy-tail and voice over internet protocol traffic: queueing analysis for performance evaluation[J]. *Iet Communications*, 2011, **5**(18): 2736-2743.
- [10] KWON K, FRANGOPOL D M. Bridge fatigue reliability assessment using probability density functions of equivalent stress range based on field monitoring data[J]. *International Journal of Fatigue*, 2010, **32**(8): 1221-1232.
- [11] WERON R J. *Computationally Intensive Value at Risk Calculations*[M]//GENTLE J E, HÄRDLE W, MORI Y, ed. *Handbook of Computational Statistics*. Berlin: Springer, 2004.
- [12] DAVID A. Lévy processes-from probability to finance and quantum groups[J]. *Notices of the AMS*, 2004, **51**(11): 1336-1347.
- [13] MARIANI M C, BIANCHINI A, BANDINI P. Normalized truncated Levy walk applied to flexible pavement performance[J]. *Transportation Research Part C*, 2012, **24**: 1-8.
- [14] GONZALEZ D S, KURUOGLU E E, RULZ D P. Modelling with mixture of symmetric stable distributions using Gibbs sampling[J]. *Signal Processing*, 2010, **90**(3): 774-783.
- [15] 徐绪松, 马莉莉, 陈彦斌. Lévy 分布在中国股票市场中的实证检验[J]. 科技进步与决策, 2005, **22**(8): 103-105.(XU Xusong, MA Lili, CHEN Yanbin. The empirical tests of Lévy distribution on China's stock market[J]. *Science & Technology Progress and Policy*, 2005, **22**(8): 103-105.(in Chinese))
- [16] 梁英杰, 陈文. Lévy 稳定分布对住宅楼面活荷载的统计分析[J]. 工程力学, 2014, **31**(6): 166-172.(LIANG Yingjie, CHEN Wen. Statistical analysis of live load on residence floor using Lévy stable distributions[J]. *Engineering Mechanics*, 2014, **31**(6): 166-172.(in Chinese))
- [17] 陈文, 孙洪广, 李西成. 力学与工程问题的分数阶导数建模[M]. 北京: 科学出版社, 2010.(CHEN Wen, SUN Hongguang, LI Xicheng. *Fractional Derivative Modeling in Mechanical and Engineering Problems*[M]. Beijing: Science Press, 2010.(in Chinese))
- [18] NOLAN J P. *Stable Distribution: Models for Heavy-Tailed Data*[M]. Boston: Birkhauser, 2010.
- [19] LIANG Y, CHEN W. A survey on computing Lévy stable distributions and a new MATLAB toolbox[J]. *Signal Processing*, 2013, **93**(1): 242-251.
- [20] MURALI R. More on a new concept of entropy and information[J]. *Journal of Theoretical Probability*, 2005, **18**(14): 967-981.
- [21] LIANG Y, CHEN W. A cumulative entropy method for distribution recognition of model error[J]. *Physica A: Statistical Mechanics and Its Applications*, 2015, **419**: 729-735.
- [22] SAMORODNITSKY G, TAQQU M S. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*[M]. Boca Raton: CRC Press, 1994.
- [23] NOLAN J P. Parameterizations and modes of stable distributions[J]. *Statistics & Probability Letters*, 1998, **38**(2): 187-195.
- [24] ZOLOTAREV V M. *One-Dimensional Stable Distributions*[M]. Providence: American Mathematical Society, 1986.
- [25] DIONISIO A, MENEZES R, MENDES D A. An econophysics approach to analyse uncertainty in financial markets: an application to the Portuguese stock market[J]. *The European Physical Journal B: Condensed Matter and Complex Systems*, 2006, **50**(1/2): 161-164.

A Cumulative Residual Entropy Method in Selection of Random Load Distributions

YANG Yu^{1,2}, LIANG Yingjie¹, CHEN Wen¹

- (1. *Institute of Soft Matter Mechanics, College of Mechanics and Materials, Hohai University, Nanjing 211100, P.R.China;*
2. *College of Science, Hohai University, Nanjing 210098, P.R.China*)
(Contributed by CHEN Wen, M. AMM Editorial Board)

Abstract: A cumulative residual entropy method was proposed to select the most suitable statistical distribution for the random load based on the Lévy stable distribution. In this method, the cumulative distribution and the tail distribution of the water level were fitted with the candidate distributions, and the cumulative residual entropy and the corresponding relative distances between the estimated distribution and the real distribution were respectively calculated, where the smaller the relative distance is, the more accurate the candidate distribution will be. The proposed method was validated by the upstream and downstream water levels of the Yongji 2nd sluice gate. The results show that, the Lévy stable distribution has the highest accuracy compared with the normal and the extreme value type I distributions in describing the heavy tail of the water level. The relative distances between the cumulative entropy and the cumulative residual entropy reveal that the fitting errors of the Lévy stable distribution are the smallest. Thus, based on the Lévy stable distribution, the cumulative residual entropy method is effective in selection of the distribution of the random load, which provides new ideas for the selection of random load distributions, and helps accurately calculate the reliability of engineering structures.

Key words: Lévy stable distribution; cumulative entropy; cumulative residual entropy; random load; reliability

引用本文/Cite this paper:

杨于, 梁英杰, 陈文. 一种选择工程随机载荷统计分布的累积剩余熵法[J]. 应用数学和力学, 2018, **39**(12): 1343-1350.

YANG Yu, LIANG Yingjie, CHEN Wen. A cumulative residual entropy method in selection of random load distributions[J]. *Applied Mathematics and Mechanics*, 2018, **39**(12): 1343-1350.