DOI:10.13718/j.cnki.xsxb.2020.05.004

修正的扩展林德利—威布尔分布①

张 泓, 李婷婷

西南大学 数学与统计学院, 重庆 400715

摘要: 介绍了一个新的分布——修正的扩展林德利—威布尔分布, 该分布表示为相互独立的林德利—威布尔分布和指数分布幂的商. 研究了该分布的概率密度函数、风险函数和矩, 并使用最大似然法估计出分布中的参数值, 最终利用实际数据验证了该分布对具有过度峰度的正数据建模的有效性.

关 键 词: 林德利—威布尔分布; 峰度; 最大似然估计; 扩展林德利—威布尔分布

中图分类号: O211.4 **文献标志码**: A **文章编号**: 1000-5471(2020)05-0017-07

在生存分析、环境科学分析、可靠性分析以及寿命测试分析中, 数据通常呈现尖峰厚尾的特征, 常见的伽马分布、对数正态分布、逆高斯分布和威布尔分布等对此类带有异常值的数据建模效果并不理想. 针对这类问题, 文献[1-3]通过添加新参数对一些常见的分布进行扩展得到一类新分布族, 这类新分布族在建模上具有更好的灵活性. 其中文献[4-6]使用这种扩展方法, 基于林德利分布提出林德利一般分布(LG). 本文对林德利—威布尔分布(LW)进行修正. LW 分布使用威布尔分布作为基线概率密度函数, 通过增加额外参数, 提高林德利分布的适用性和灵活性. 然而, LW 分布的厚尾特征并不明显, 不能对具有高峰度和异常观测值的数据集进行很好地拟合, 为了解决这个问题, 文献[7]引入含四参数的扩展林德利—威布尔分布(SLW), SLW 分布具有比 LW 分布更宽的峰度范围, 适用于具有非典型观测值的数据集. 在这个基础上, 本文提出了 SLW 分布的修正形式, 即修正的扩展林德利—威布尔分布(MSLW). 文献[8-10]指出, 类似修正扩展分布更容易修改一些常见的分布, 使其具有更高的峰度. 因此 MSLW 与 SLW 分布相比具有更重的尾部, 能更好地拟合带有异常观测值的数据, 可以作为 SLW 分布的替代模型.

1 修正的扩展林德利 — 威布尔分布

1.1 随机表达式

若随机变量 X 服从修正的扩展林德利 — 威布尔分布, 记为 $X \sim \mathrm{MSLW}(\lambda, \alpha, \theta, q)$, 具体表达式为

$$X = \frac{Z}{U^{\frac{1}{q}}},\ q > 0 \tag{1}$$

其中: $Z \sim \mathrm{LW}(\lambda, \alpha, \theta)$ 和 $U \sim \exp(2)$ 独立, $\lambda > 0$ 是尺度参数, $\alpha > 0$ 和 $\theta > 0$ 是形状参数, $q > 0$ 是峰度参数.

1.2 概率密度函数

命题 1 假设 $X \sim \mathrm{MSLW}(\lambda, \alpha, \theta, q)$, 则 X 的概率密度函数表示为

$$f_X(x; \lambda, \alpha, \theta, q) = \frac{2\alpha\theta^2 q\lambda^q}{(\theta+1)x^{q+1}} J_X(x; \lambda, \alpha, \theta, q),\ x > 0 \tag{2}$$

① 收稿日期: 2019-08-14
基金项目: 国家自然科学基金项目(11701469).
作者简介: 张 泓(1994-), 女, 硕士研究生, 主要从事极值统计分析的研究.
通信作者: 李婷婷, 博士, 副教授.

其中 $\lambda>0$ 是尺度参数，$\alpha>0$ 和 $\theta>0$ 是形状参数，$q>0$ 是峰度参数，且有

$$J_X(x;\lambda,\alpha,\theta,q)=\int_0^{\infty}u^{\alpha+q-1}(1+u^{\alpha})\exp\left\{-\theta u^{\alpha}-2\left(\frac{\lambda u}{x}\right)^{q}\right\}\mathrm{d}u \tag{3}$$

证 根据随机表达式(1)和雅可比行列式的方法，计算 X 的概率密度函数如下：首先做一个变量替换，令 $X=ZU^{-\frac{1}{q}}$ 以及 $T=U$，则有 $Z=XT^{\frac{1}{q}}$ 和 $U=T$，变换的雅可比行列式表示为

$$J=\begin{vmatrix}\dfrac{\partial Z}{\partial X} & \dfrac{\partial Z}{\partial T}\\ \dfrac{\partial U}{\partial X} & \dfrac{\partial U}{\partial T}\end{vmatrix}=t^{\frac{1}{q}}$$

因此，(X,T) 的联合概率密度函数为

$$f_{X,T}(x,t)=f_Z(x\,t^{\frac{1}{q}};\lambda,\alpha,\theta)\,f_U(t)\,t^{\frac{1}{q}},\ x>0,\ t>0$$

最后利用 $u=\dfrac{xt^{\frac{1}{q}}}{\lambda}$，就得到 X 最终的概率密度函数.

尺度参数 λ，形状参数 α 和 θ，峰度参数 q 对 MSLW 分布的概率密度函数的影响如图1所示，呈现单峰或非增形状.

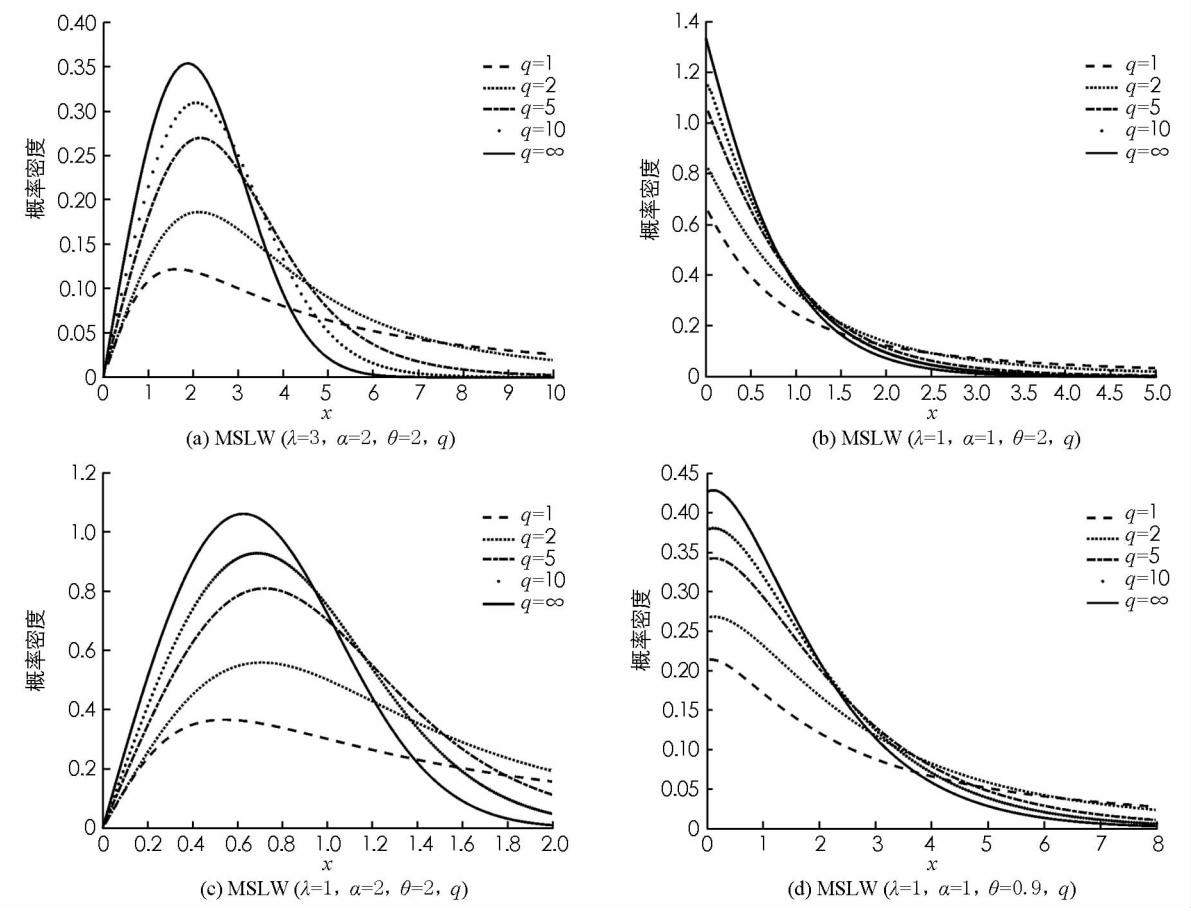

图1 MSLW$(\lambda,\alpha,\theta,q)$ 分布的概率密度函数

假设 $X\sim$ MSLW$(\lambda,\alpha,\theta,q)$，那么由 MSLW 分布的概率密度函数的具体表达式，可以推导以下性质：

1) $\lim\limits_{q\to\infty}f_X(x;\lambda,\alpha,\theta,q)=\dfrac{\alpha\theta^2}{\lambda(\theta+1)}\left(\dfrac{x}{\lambda}\right)^{\alpha-1}\left[1+\left(\dfrac{x}{\lambda}\right)^{\alpha}\right]\exp\left(-\theta\left(\dfrac{x}{\lambda}\right)^{\alpha}\right)$.

2) $\lim\limits_{q\to\infty}f_X(x;\lambda,1,\theta,q)=\dfrac{\theta^2}{\lambda(\theta+1)}\left(1+\dfrac{x}{\lambda}\right)\exp\left(\dfrac{-\theta x}{\lambda}\right)$.

3) $F_X(x;\lambda,\alpha,\theta,q)=\dfrac{2\alpha\theta^2q\lambda^q}{\theta+1}\displaystyle\int_0^x v^{-(q+1)}J_X(v;\lambda,\alpha,\theta,q)\mathrm{d}v$.

注 1　(i) 当 $q=1$ 时，称 X 服从典型的修正的扩展林德利 — 威布尔分布，表示为 $X=ZU^{-1}$，记为 $X\sim$CMSLW(λ，α，θ)，并且其概率密度函数为

$$f_X(x;\lambda,\alpha,\theta)=\frac{2\alpha\theta^2\lambda}{(\theta+1)x^2}J_X(x;\lambda,\alpha,\theta,1)\ ,\ x>0$$

其中 $\lambda>0$ 是尺度参数，$\alpha>0$ 和 $\theta>0$ 是形状参数，J_X 由(3)式给出.

(ii) 由性质 1) 可知，当 $q\to\infty$ 时，MSLW 分布收敛到一般的 LW 分布[4]；性质 2) 可知，当 $q\to\infty$ 且 $\alpha=1$ 时，MSLW 分布收敛到林德利指数分布[11].

1.3　可靠性分析

可靠性函数和风险率(失效率)函数是两项重要的可靠性指标. 其中可靠性函数 $R_T(t)$ 表示一个项目在某个时间 t 内未发生故障的概率，定义为 $R_T(t)=1-F_T(t)$. MSLW 分布的可靠性函数由下式给出

$$R_X(t\ ;\lambda,\alpha,\theta,q\ )=1-\frac{2\alpha\theta^2q\lambda^q}{\theta+1}\int_0^t v^{-(q+1)}J_X(v;\lambda,\alpha,\theta,q)\mathrm{d}v \tag{4}$$

假定某事件的存活时间达到时刻 t，那么该事件的风险率函数 $h_T(t)=\dfrac{f_T(t)}{1-F_T(t)}$ 可以粗略地解释为在超过时刻 t 瞬时死亡的条件概率. MSLW 分布的风险率函数如下

$$h_X(t;\lambda,\alpha,\theta,q)=\frac{J_X(t\ ;\lambda,\alpha,\theta,q)}{t^{q+1}}\left(\frac{\theta+1}{2\alpha\theta^2q\lambda^q}-\int_0^t\frac{J_X(v;\lambda,\alpha,\theta,q)}{v^{q+1}}\mathrm{d}v\right)^{-1} \tag{5}$$

其中 J_X 由(3)给出. 不同的参数 λ，α，θ，q 的取值下，MSLW 分布的可靠性函数和风险率函数如图 2 所示. 由图 2 可知，不同的参数取值下，MSLW 分布的两种可靠性指标函数表现出多种图形形状，这说明了新的分布 MSLW 的灵活性.

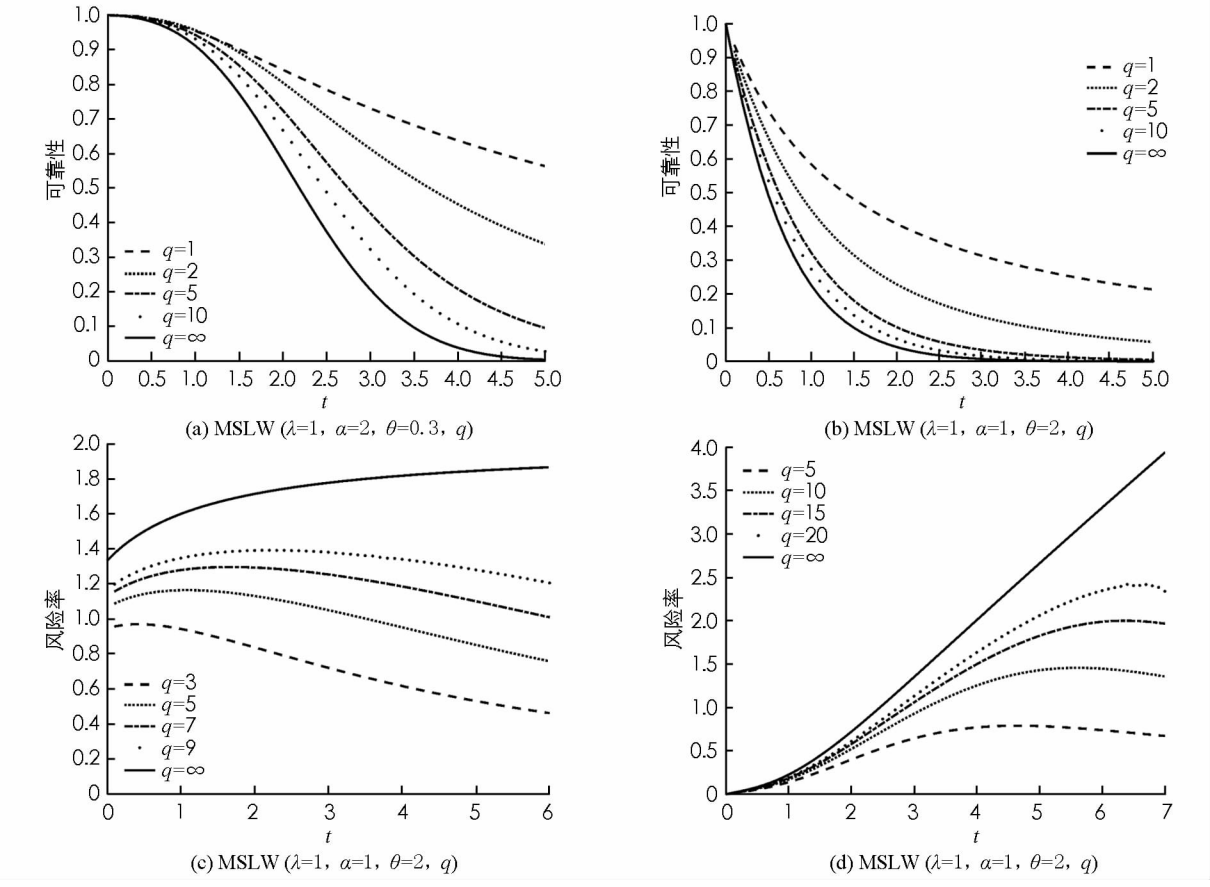

图 2　MSLW(λ，α，θ，q) 分布的可靠性(顶部)和风险率(底部)函数

1.4　分布的矩

引理 1　设 Z 服从 LW(λ，α，θ) 分布，则有

$$\mathbb{E}(Z^r)=\frac{r\lambda^r}{\alpha^2\theta^{\frac{r}{\alpha}}}\left(\alpha+\frac{r}{\theta+1}\right)\Gamma\left(\frac{r}{\alpha}\right),\ r=1,2,\cdots \tag{6}$$

其中 $\Gamma\left(\frac{r}{\alpha}\right)$，表示为伽马函数，见文献[7].

命题 2 设 X 服从 MSLW(λ，α，θ，q) 分布，那么，对于 $r=1,2,\cdots$，以及 $q>r$，X 的 r 阶矩如下

$$\mathbb{E}(X^r)=\frac{2^{\frac{r}{q}}r\lambda^r}{\alpha^2\theta^{\frac{r}{\alpha}}}\mathrm{d}(r) \tag{7}$$

其中 $d(r)=\left(\alpha+\frac{r}{\theta+1}\right)\Gamma\left(\frac{r}{\alpha}\right)\Gamma\left(\frac{q-r}{q}\right)$.

证 由(1)可知，Z 和 U 是相互独立的两个随机变量，所以就有

$$\mu_r=\mathbb{E}(X^r)=\mathbb{E}\left(\left(\frac{Z}{U^{\frac{1}{q}}}\right)^r\right)=\mathbb{E}(Z^r\,U^{\frac{r}{q}})=\mathbb{E}(Z^r)\mathbb{E}(U^{\frac{r}{q}}) \tag{8}$$

其中 $\mathbb{E}(U^{-\frac{r}{q}})$ 经计算为 $2^{\frac{r}{q}}\Gamma\left(\frac{q-r}{q}\right)$，$q>r$，$\mathbb{E}(Y^r)$ 由(6)式给出.

推论 1 若 $X\sim$ MSLW(λ，α，θ，q)，则有

1) $\mu_1=\mathbb{E}(X)=\frac{2^{\frac{1}{q}}\lambda}{\alpha^2\theta^{\frac{1}{\alpha}}}d(1)$

2) $\mu_2=\mathbb{E}(X^2)=\frac{2^{\frac{2}{q}}2\lambda^2}{\alpha^2\theta^{\frac{2}{\alpha}}}d(2)$，$q>2$；

3) $\mu_3=\mathbb{E}(X^3)=\frac{2^{\frac{3}{q}}3\lambda^3}{\alpha^2\theta^{\frac{3}{\alpha}}}d(3)$，$q>3$；

4) $\mu_4=\mathbb{E}(X^4)=\frac{2^{\frac{4}{q}}4\lambda^4}{\alpha^2\theta^{\frac{4}{\alpha}}}d(4)$，$q>4$；

注 2 偏度和峰度系数由下面两个式子给出

$$\beta_s=\frac{\mu_3-3\mu_1\mu_2+2\mu_1^3}{(\mu_2-\mu_1^2)^{\frac{3}{2}}},\ \beta_k=\frac{\mu_4-4\ \mu_1\mu_3+6\mu_2\mu_1^2-3\mu_1^4}{(\mu_2-\mu_1^2)^2}$$

由推论 1 可知，MSLW 分布的偏度和峰度系数值独立于尺度参数 λ.

图 3 表示形状参数 $\alpha=0.3$ 的 MSLW 分布和 SLW 分布的偏度和峰度系数，由图 3 可知 MSLW 分布的两个系数值均大于 SLW 分布，可用于拟合具有异常观测值数据的分布.

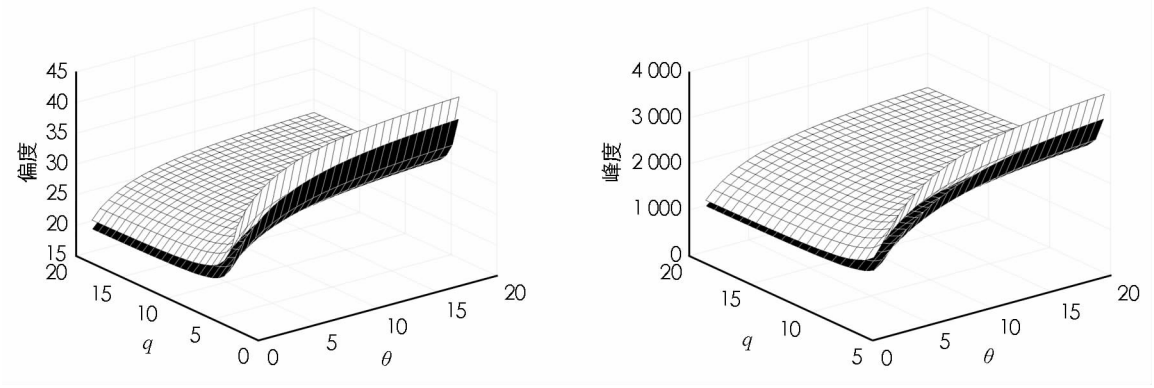

图 3 SLW 分布(下网格)和 MSLW 分布(上网格)的偏度和峰度系数

2 参数估计

本节阐述 MSLW 分布参数估计的最大似然(ML)方法. 若 $X_1,X_2,\cdots,X_n$ 是来自服从 MSLW(λ，α，θ，q) 分布，容量为 n 的随机样本. 假设参数未知，其中 $x_1,x_2,\cdots,x_n$ 表示观察值，则对数似然函数为

$$l(\lambda, \alpha, \theta, q; x) = c(\lambda, \alpha, \theta, q) - (q+1)\sum_{i=1}^{n}\log(x_i) + \sum_{i=1}^{n}\log J(x_i) \tag{9}$$

其中 $c(\lambda, \alpha, \theta, q) = n\log\left(\frac{2\alpha q}{\theta+1}\right) + 2n\log(\theta) + nq\log(\lambda)$，$J_X(x) = J_X(x; \lambda, \alpha, \theta, q)$ 由(3)式定义. 在实际中，为了得到参数的 ML 估计，本文选用数值方法求解优化问题

$$\begin{aligned} &\max \quad l(\lambda, \alpha, \theta, q) \\ &\text{s.t. } \lambda > 0, \alpha > 0, \theta > 0, q > 0 \end{aligned} \tag{10}$$

其中 $l(\lambda, \alpha, \theta, q)$ 由(9)式给出. 本文使用 R 软件中 optim 函数，应用 L-BFGS-B 算法来求解(10). 文献[7]指出，两个形状参数 α 和 θ 的存在导致参数可识别性的问题，但如果只有一个形状参数，这个问题就会消失. 因此，出于实际目的，在实证分析中，本文使用具有唯一形状参数的 MSLW 模型版本，即假定 $\alpha = \theta$.

3　实证分析

本节比较 4 种分布 MSLW，MSL，SLW 和 LW 对具有较高峰度的实际数据集建模的有效性. 数据来源于 http://lib.stat.cmu.edu/datasets/Plasma_Retinol，该数据表明人体中一些微量营养素的血浆浓度存在很大差异，可能会增加某些癌症发生的风险，其包含了 14 个变量，每个变量下有 314 个观测值，本文选择其中第 13 个变量 β-血浆进行分析.

表 1 是对 β-血浆数据进行描述性统计分析的一个总结，其中 n 表示样本容量，$\bar{x}$ 为样本均值，S^2 为样本方差，β_s 和 β_k 分别表示样本的偏度和峰度系数，揭示了数据具有较高的峰度，也可以更直观地在箱线图(图 4)中看出.

表 1　描述性统计分析

n	$\bar{x}$	s^2	β_s	β_k
314	190.50	33 481	3.55	19.94

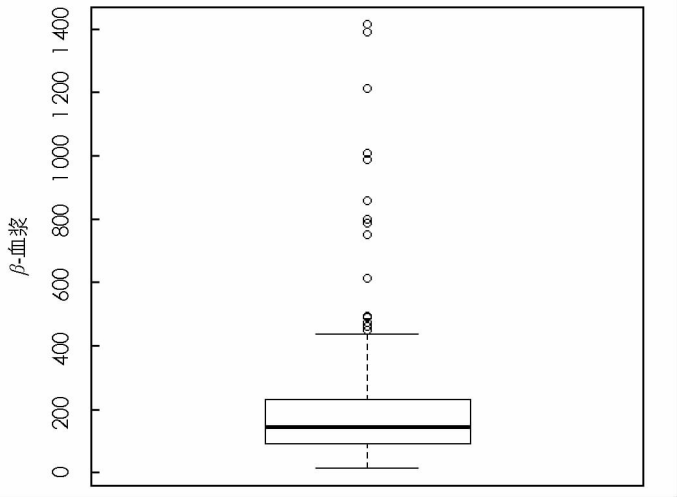

图 4　β-血浆的箱线图

表 2 分别给出了 4 个模型 MSLW，MSL，SLW 和 LW 的参数的最大似然估计以及相应的最大对数似然函数值，表明 MSLW 模型对应的对数似然函数值最大.

表 2　4 个模型的最大似然估计

估计值	MSLW	MSL	SLW	LW
$\hat{\lambda}$	133.306 1	—	134.191 1	1.162 9
$\hat{\alpha}$	2.495 3	0.015	2.250 8	0.942 4
$\hat{\theta}$	—	—	—	0.016 7
$\hat{q}$	2.306 8	4.053	1.855 9	—
对数似然函数值	−1 907.13	−1 919.40	−1 908.71	−1 930.75

为了比较分布的拟合效果，本文考虑 Akaike 信息准则(AIC)和 Bayesian 信息准则(BIC)．它们分别表示为

$$\mathrm{AIC}=2k-2\log lik,\ \mathrm{BIC}=k\log n-2\log lik$$

其中：k 为分布参数个数，n 为样本容量，$\log lik$ 为对数似然函数的最大值．表 3 显示了每个模型对应的 AIC 和 BIC 值，以及应用拟合优度 Kolmogorov-Smirnov 进行检验得到的 Kolmogorov-Smirnov 统计量(K-S 统计量)和检验的 P 值．可以看出，基于 AIC,BIC,K-S 统计量和 P 值，MSLW 模型比另外 3 个模型的拟合效果更好．除此之外，可以更直观地在图 5(a)发现，对于真实数据，MSLW 拟合程度更优．图 5(b)直方图尾部的放大图揭示了 MSLW 分布对具有较高峰度的数据集适用性更强．图 6 分别描述了 MSLW,MSL,SLW 和 LW 同原始数据的 Q-Q 图，也同样表明 MSLW 模型在该数据集上具有更好的拟合效果．

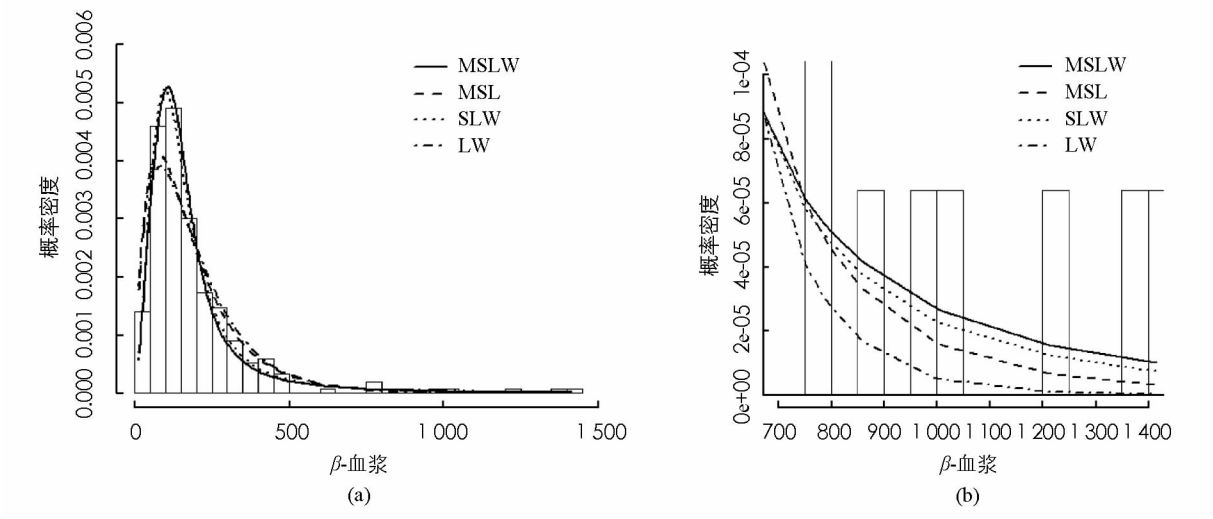

图 5　4 个模型的直方密度图

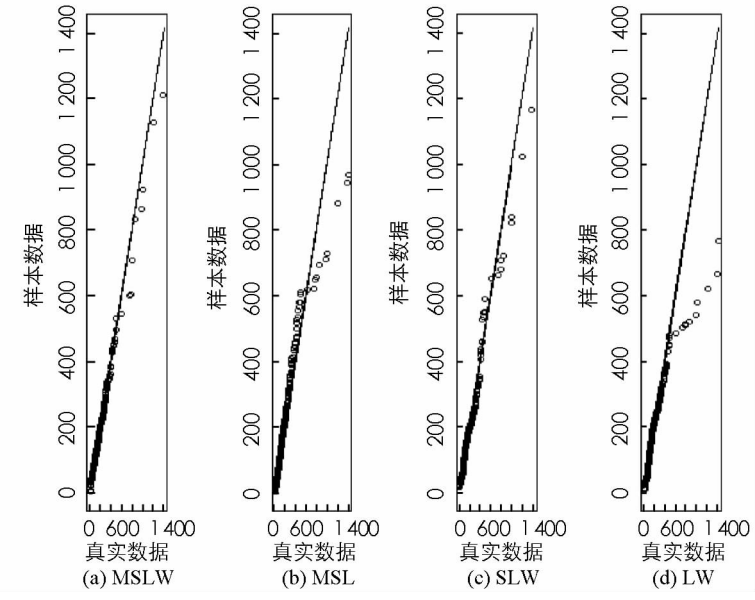

图 6　4 个模型的 Q-Q 图

表 3　4 个模型的 AIC,BIC,K-S 统计量及 P 值

	MSLW	MSL	SLW	LW
AIC	3 820.26	3 842.80	3 823.42	3 867.50
BIC	3 831.51	3 850.30	3 834.67	3 878.75
K-S 统计量	0.054	0.105	0.057	0.080
P 值	0.747	0.062	0.681	0.272

4 结　论

引入了 LW 分布的一个新的扩展形式，新分布表示为 LW 分布和指数分布的幂这两个独立随机变量的比值，称为修正的扩展林德利—威布尔分布(MSLW). 该分布扩大了峰度范围，可用于模拟具有过度峰度和异常观测值的正数据集，实证分析验证了该分布的可行性.

参考文献：

[1] MARSHALL A W, OLKIN I. A New Method for Adding a Parameter to a Family of Distributions with Application to the Exponential and Weibull Families [J]. Biometrika, 1997, 84(3): 641-652.

[2] SHAW W T, BUCKLEY I R C. The Alchemy of Probability Distributions: Beyond Gram-Charlier Expansions, and a Skew-Kurtotic-Normal Distribution from a Rank Transmutation Map [EB/OL]. (2009-01-05)[2019-07-05]. https://arxiv.org/abs/0901.0434.

[3] ZOGRAFOS K, BALAKRISHNAN N. On Families of Beta- and Generalized Gamma-Generated Distributions and Associated Inference [J]. Statistical Methodology, 2009, 6(4): 344-362.

[4] CAKMAKYAPAN S, OZEL G. The Lindley Family of Distributions: Properties and Applications [J]. Hacettepe Journal of Mathematics and Statistics, 2016, 46(116): 1-11.

[5] LINDLEY D V. Fiducial Distributions and Bayes' Theorem [J]. Journal of the Royal Statistical Society Series B (Methodological), 1958, 20(1): 102-107.

[6] GHITANY M E, ATIEH B, NADARAJAH S. Lindley Distribution and Its Application [J]. Mathematics and Computers in Simulation, 2008, 78(4): 493-506.

[7] JIMMY R, YURI A I, PEDRO J, HÉCTOR W G. The Sslash Lindley-Weibull Distribution [J]. Methodology and Computing in Applied Probability, 2019, 21(1): 235-251.

[8] JUAN M A, YURI A I, HÉCTOR W G, et al. Modified Slashed Generalized Exponential Distribution [DB/OL]. (2019-04-26)[2019-07-05]. https://doi.org/10.1080/03610926.2019.1604959.

[9] JIMMY R, HÉCTOR W G, HELENO B [J]. Modified Slash Distribution. Statistics, 2013, 47(5): 929-941.

[10] YURI A I, NABOR O C, HELENO B, et al. Modified Slashed-Rayleigh Distribution [J]. Communication in Statistics-Theory and Methods, 2017, 47(13): 3220-3233.

[11] BHATI D, MALIK M A, VAMAN H J. Lindley-Exponential Distribution: Properties and Applications [J]. Metron, 2015, 73(3): 335-357.

Modified Slashed Lindley-Weibull Distribution

ZHANG　Hong,　LI Ting-ting

School of Mathematics and Statistics, Southwest University, Chongqing 400715, China

Abstract: A new distribution, namely, the modified slashed Lindley-Weibull distribution, has been introduced in this paper. The new distribution has been expressed as the quotient of two independent random variables, with the Lindley-Weibull distribution in the numerator and the power of an exponential distribution in the denominator. The probability density function, hazard rate function and moments of the new distribution have been studied, and the maximum likelihood estimation method for the parameters in the distribution also been discussed. A practical dataset has been used to demonstrate the effectiveness of the new distribution in fitting the positive data with high kurtosis.

Key words: Lindley-Weibull distribution; Kurtosis; maximum likelihood estimator; slashed Lindley-Weibull distribution

责任编辑　张　枸