

基于 Rényi 熵的 q -指数分布及其 可靠性分析应用



谢 暄¹, 王敏夷¹, 白颖利², 汪东敏³, 李西峰^{1*}, 谢永乐¹

(1. 电子科技大学自动化工程学院 成都 611731; 2. 中国空间技术研究院通信卫星事业部 北京 海淀区 100094;
3. 四川慧龙科技有限责任公司 成都 610041)

【摘要】基于最大 Rényi 熵原理, 在归一化和均值约束下, 提出了一种具有封闭表达式的双参数广义指数分布, 记为 q -指数分布。该文研究了该分布的统计性质, 指出可以分别利用极大似然法和信息似然法估计 q -指数分布参数, 并将该分布用于可靠性分析。利用两个已知的数据集进行了验证, 实验结果表明, 所提出的 q -指数分布比其他常用分布, 如韦伯分布和线性失效率分布, 能够更好地拟合数据集。此外, 锂电池剩余寿命估计实验表明, 采用 q -指数分布比采用传统指数分布, 估计精度至少提高 17.857%。

关键词 指数分布; 风险函数; q -指数分布; Rényi 熵; 韦伯分布
中图分类号 TP202.1 文献标志码 A doi:10.12178/1001-0548.2020449

q -Exponential Distribution Based on Rényi Entropy and Its Application on Reliability Analysis

XIE Xuan¹, WANG Min-yi¹, BAI Yin-li², WANG Dong-min³, LI Xi-feng^{1*}, and XIE Yong-le¹

(1. School of Automation Engineering, University of Electronic Science and Technology of China Chengdu 611731;
2. Communication Satellite Division, China Academy of Space Technology Haidian Beijing 100094;
3. Sichuan Huilong Technology LLC Chengdu 610041)

Abstract We propose a two-parameter generalized exponential distribution with closed-form expression based on the maximum Rényi entropy principle under the normalization and mean constraints, which is referred as the q -exponential distribution. The statistical properties of this distribution are investigated. The maximum likelihood method and information likelihood method are used to estimate the parameters of the proposed distribution, respectively. Two well-known data sets are employed to evaluate the q -exponential distribution, and the experimental results demonstrate that the proposed distribution can fit the data sets better than other well-known distributions, such as Weibull distribution and linear failure rate distribution. Additionally, the experiment results of life estimation of the Li-ion batteries prove that compared with the exponential distribution, the proposed distribution can give more accurate prediction. In the last experiment, the estimation accuracy is improved by at least 17.857%.

Key words exponential distribution; hazard function; q -exponential distribution; Rényi entropy; Weibull distribution

随着集成系统复杂性的日益增加, 系统建模和可靠性保证面临越来越大的挑战。在可靠性分析的各种场景中, 虽然指数分布模型已被较广泛地用于建立系统寿命模型^[1], 但由于指数分布的风险函数是恒定的, 因此直接将指数分布用于描述

系统寿命, 存在无法描述损伤过程和无法准确反映故障累积效果的问题, 最典型的例子是将指数分布简单应用于描述人类死亡率和电子设备生命周期, 效果不够理想。对这类过程, 通常需要采用具备浴盆特征的风险函数所对应的寿命分布来

收稿日期: 2020-12-28; 修回日期: 2021-03-24

基金项目: 国家自然科学基金(61701095, 61601096, 61801089, 61971111, 62027803); 四川省科学技术项目(2020YFG0044, 2020YFG0046, 2021YFG0200); 国防基础科研计划(JCKY2020110C041)

作者简介: 谢暄(1983-), 女, 博士, 主要从事测试信息处理方面的研究。

通信作者: 李西峰, E-mail: xifengli@uestc.edu.cn

准确描述。韦伯分布作为指数分布的概括,以及它的带有浴盆型风险函数的扩展得到了重视,并被广泛应用于许多领域^[2-4]。

目前关于指数分布推广的研究,主要集中在广义韦伯分布的累积分布函数的参数加法上。大多数情况下,这些韦伯类型的一般化是通过简单的参数加法技术来实现的。此方法一般除保留了原有的参数,还引入了一些新的参数,使得模型对寿命数据的拟合,多数情况下优于没有新参数的模型。

基于参数加法,文献[5]提出了四参数广义指数模型,称为广义指数化线性指数分布。结果表明,这一扩展可以导出一系列指数型分布,如指数化韦伯分布。结合韦伯分布和改进的五参数韦伯分布,文献[6]提出了一种修正的韦伯分布扩展。基于此,文献[7]又提出了一个离散五参数修正型的韦伯分布,并发现基于这种分布的离散数据拟合胜过其他 3 个修正的韦伯模型。

尽管韦伯推广在寿命模型中具有良好的效果,但由于参数估计过程复杂,参数物理意义不明确(至少这些参数在这些广义的分布中,不能直接指示系统的当前状态),限制了韦伯推广的应用^[8-10]。

本文提出了一种广义双参数指数分布,它可以作为可靠性分析的基础,例如浴盆型风险函数的构造和有用寿命预测。与传统的参数加法不同,本文用最大熵原理得到广义指数分布。从物理角度看,这种广义分布的主要优点是两个参数都具有明显的物理意义。一个参数具有分形意义,表示系统的稳定程度;另一个参数则与系统的平均行为密切相关。两个参数均可以用来刻画系统的性能。

1 q -指数分布模型

1.1 一般化的指数分布

基于 Rényi 熵,可以通过最大熵方法,得到广义指数分布:

$$f_q(x) = \frac{1}{\lambda} \left(1 + \frac{q-1}{q}\right)^{q/(1-q)} \left(1 - \frac{1}{\lambda} \frac{q-1}{q} (x-\lambda)\right)^{1/(q-1)} \quad (1)$$

式中, q 是分形参数; λ 是分布的期望值; q 和 λ 都是非负的。根据参数 q 和 λ 的值,概率密度函数 f_q 会有不同的形态。图 1 所示为不同的参数 q 时固定均值 λ 下的 f_q 形态。可以看出 q -指数分布是单峰分布,也是右偏分布。因此, q -指数分布可以被考虑用于对寿命相关的可靠性分析。

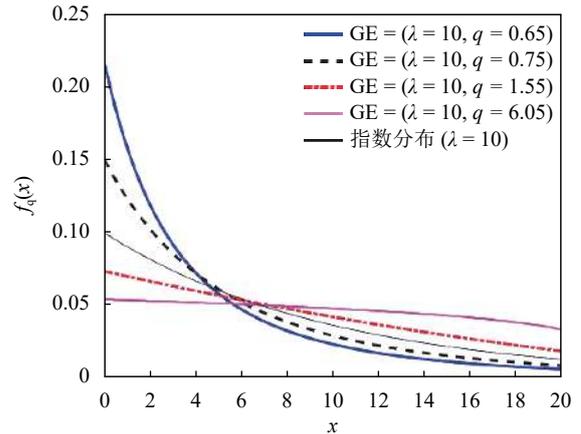


图 1 q -指数分布 $f_q(x)$

计算可得, q -指数分布的累积分布函数是:

$$F_q(x) = \frac{qx}{\lambda(2q-1)} \left[{}_2F_1 \left(\frac{1}{1-q}, 1, 2, \frac{(q-1)x}{\lambda(2q-1)} \right) \right] \quad (2)$$

式中, ${}_2F_1$ 是高斯超几何函数。

1.2 特殊衍生分布

下面将给出 q -指数分布的两个主要的衍生分布,它们在生存分析、压缩感知、剩余使用寿命预测和其他的可靠性相关领域有潜在实用价值。

首先,当 $q \rightarrow 1$ 时,该分布简化为均值为 λ 的指数分布:

$$\lim_{q \rightarrow 1} f_q(x) = \frac{1}{\lambda} e^{-x/\lambda} \quad \lambda > 0, x > 0$$

其次,使用式(1)可以得到一个广义的 q -拉普拉斯分布:

$$g_q(x) = \frac{1}{2\lambda} \left(1 + \frac{q-1}{q}\right)^{q/(1-q)} \left[1 - \frac{1}{\lambda} \frac{q-1}{q} (|x| - \lambda) \right]^{1/(q-1)} \quad (3)$$

图 2 显示了不同参数值 q 和固定 λ 值 ($\lambda=10$) 的广义拉普拉斯分布。拉普拉斯分布在压缩感知领域起到了重要作用^[11-13],这里衍生出的拉普拉斯分布在统计信号处理和机器学习领域具有潜在应用价值。

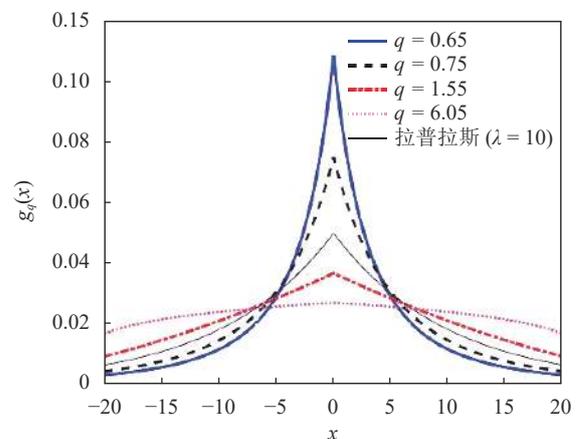


图 2 不同参数 q 下的 q -拉普拉斯分布 $g_q(x)$

2 模型性质

2.1 q -指数分布的风险函数

生存函数是表示一系列事件的随机变量函数, 通常用于表示一些基于时间的系统失败或死亡概率。假设 T 表示产品使用寿命, 其分布函数为 $F(t)$, 那么该产品寿命大于 t 的概率为:

$$S(t) = P(T > t) = 1 - F(t)$$

式中, $S(t)$ 被称为生存函数。在此基础上, 可用风险函数刻画已有效使用到 t 时刻的产品, 在 $[t, t + \Delta t]$ 极短时间内“死亡”的风险:

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t)}{S(t)\Delta t}$$

据此计算可得 q -指数分布的生存函数:

$$S(x) = 1 - {}_2F_1\left(1 - \frac{1}{1-q}, 1, 2, \frac{(q-1)x}{\lambda(2q-1)}\right) \frac{qx}{\lambda(2q-1)} \quad (4)$$

风险函数:

$$h(x) = \frac{\frac{1}{\lambda} \left(1 + \frac{q-1}{q}\right)^{q/(1-q)}}{1 - F \frac{qx}{\lambda(2q-1)} \left(1 - \frac{q-1}{\lambda q} (x-\lambda)\right)^{1/(q-1)}} \quad (5)$$

式中, $F = {}_2F_1\left(\frac{1}{1-q}, 1, 2, \frac{(q-1)x}{\lambda(2q-1)}\right)$ 。

图3为具有不同参数值的风险函数图像, 可以看到风险函数具有许多不同的形状, q -指数分布中当 $q > 1$, 显示递增特性; 当 $q < 1$, 显示递减性。

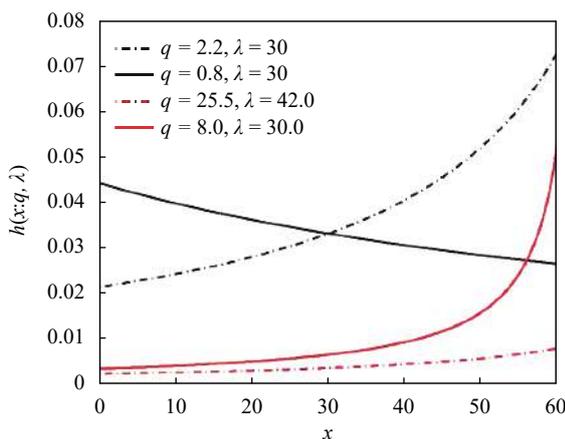


图3 q -指数分布的风险函数

注意到 q -指数分布的风险函数呈现多态性, 其中部分具有澡盆特征。可根据实际使用环境, 选择恰当的 q -指数建模风险过程, 从而可以提高寿命估计精度。

2.2 q -指数分布的矩

q -指数分布 f_q 的均值是 λ , 方差是:

$$\mu^2 = D(x) = \int_0^\infty x^2 f_q(x) dx = \frac{(2-1/q)^{q/(1-q)} q^{1/(1-q)} (2q-1)^{q/(q-1)}}{3q-2} \lambda^2$$

可以得到分布的 k 阶矩为:

$$m\mu^k = \int_0^\infty x^k dF(x) = \int_0^\infty x^k f_q(x) dx = \int_0^\infty x^k \frac{1}{\lambda} \left(1 + \frac{q-1}{q}\right)^{q/(1-q)} \left(1 - \frac{1}{\lambda} \frac{q-1}{q} (x-\lambda)\right)^{1/(q-1)} dx = \frac{1}{\lambda} \frac{q}{2q-1} \frac{x^{k+1}}{k+1} \left[{}_2F_1\left(\frac{1}{1-q}, 1+k, 2+k, \frac{(q-1)x}{\lambda(2q-1)}\right) \right]$$

式中, k 表示分布的 k 阶矩。

3 参数估计

为了在不同场景下选择恰当的参数, 这里给出两种估计 q -指数分布中参数的方法: 最大似然估计法 (maximum likelihood estimation, MLE) 和信息似然估计法。最大似然估计法适用于满足高斯分布的数据集, 信息似然估计法适用于高斯和非高斯分布的数据集。

3.1 最大似然估计

根据式 (1), 建立对数似然函数:

$$\psi_L = \log L = n \ln(1/\lambda) + \frac{nq}{1-q} \ln\left(1 + \frac{q-1}{q}\right) +$$

$$\frac{1}{q-1} \sum_{i=1}^n \ln\left(1 - \frac{1}{\lambda} \frac{q-1}{q} (x_i - \lambda)\right)$$

然后有:

$$\frac{\partial \psi_L}{\partial q} = \frac{1}{(q-1)^2(2q-1)} \times$$

$$\left[(2q^2 - 3q + 1) / q^2 \sum_{i=1}^n - \frac{x_i/\lambda - 1}{x_i/q(\lambda-1) - x_i/\lambda + 2} \right] +$$

$$\frac{1}{(q-1)^2(2q-1)} (1-2q) \sum_{i=1}^n \log(x_i/q(\lambda-1) - x_i/\lambda + 2) +$$

$$\frac{1}{(q-1)^2(2q-1)} n(1-q + (2q-1)\log(2-1/q))$$

$$\frac{\partial \psi_L}{\partial \lambda} = -\frac{n}{\lambda} + \sum_{i=1}^n \frac{x_i}{\lambda^2 q - \lambda(q-1)(x_i - \lambda)}$$

q 和 λ 可以通过求解方程组 (6) 获得:

$$\begin{cases} \frac{\partial \psi_L}{\partial \lambda} = 0 \\ \frac{\partial \psi_L}{\partial q} = 0 \end{cases} \quad (6)$$

3.2 信息似然估计

另一个估计 Rényi 信息未知参数的方法, 将信息论和谱估计相结合, 可在最小先验的条件下, 取得经验风险最小的参数估计值^[14]。具体步骤如下:

首先, 已知 Rényi 信息频谱定义为:

$$\mathcal{L}_R(\lambda) = \frac{1}{1-\lambda} \log \left(\int f^\lambda(x) dx \right) \quad (7)$$

当 $\lambda \rightarrow 1$,

$$\lim_{\lambda \rightarrow 1} \mathcal{L}_R(\lambda) = \int f(x) \log f(x) dx \quad (8)$$

Rényi 信息频谱趋向于香农熵。

同时, Rényi 信息的频谱梯度被定义为:

$$\mathcal{L}_R(1) = -\frac{1}{2} D(\log(f(X))) \quad (9)$$

式中, $D(\cdot)$ 是随机变量 x 的方差。

令 $\phi_f := -2\mathcal{L}_R(1)$

一方面, 将式 (1) 中的 f_q 代入式 (9) 计算 ϕ_f , 根据文献 [13] 可得:

$$\phi_f = -2 \lim_{\lambda \rightarrow 1} \mathcal{L}_R(\lambda) = \frac{1}{q^2} \quad (10)$$

另一方面, 使用核方法估计 ϕ_f 。具体而言, ϕ_f 可以通过核方法计算如下:

$$\hat{\phi}_f = \hat{\Delta}_2 - \hat{\Delta}_1^2 \quad (11)$$

式中,

$$\hat{\Delta}_i = \int f_n(x) \log^i f_n(x) dx \quad i = 1, 2 \quad (12)$$

$f_n(x)$ 是为无参数内核密度估计量, 定义为:

$$f_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{b_n} K\left(\frac{x-X_i}{b_n}\right)$$

假设 $K(\cdot)$ 为有界变分的概率密度函数 (核函数), 其支撑集位于部分有限区间。 X_i 是随机变量 X 的样本。设 b_n 为满足以下收敛条件的序列:

$$\lim_{n \rightarrow \infty} b_n = 0, \quad \lim_{n \rightarrow \infty} n b_n = \infty, \quad \lim_{n \rightarrow \infty} \frac{\log b_n^{-1}}{n b_n} = 0$$

比较式 (10) 和式 (11), 式 (1) 中的参数 q 可以通过求解式 (13) 得到:

$$q = \frac{1}{\sqrt{\hat{\phi}_f}} \quad (13)$$

4 数据分析

本文使用了 3 个与可靠性相关的实验去评估 q -指数分布在可靠性分析中的有效性, 包括白血病病人生存期分析、设备寿命预测以及锂电池寿命预测。

第一个数据集是 40 个白血病病人的生存期数据^[15], 它具有不断增加的风险率。第二个数据集来自 50 个组件的样本^[16], 具有浴盆型风险的风险率。利用 K-S 统计量评估了 q -指数与上述两个著名数据集的拟合结果, 并计算出赤池信息准则 (AIC) 值和贝叶斯信息准则 (BIC) 值, 用于比较它和其他广义指数分布的拟合优越度。第三个实验建立了关于 3 组锂电池的容量退化模型, 与普通指数分布的结果相比, 利用 q -指数分布可以得到更准确的剩余寿命预测结果。

4.1 白血病病人生存期

该数据集由文献 [15] 给出, 数据源自沙特阿拉伯卫生医院部门, 如表 1 所示。它记录了 40 名白血病患者的生命周期。

表 1 40 名白血病患者的生存期

患者 编号	生存期 /天	患者 编号	生存期 /天	患者 编号	生存期 /天	患者 编号	生存期 /天
1	115	11	807	21	1222	31	1578
2	181	12	865	22	1251	32	1578
3	255	13	924	23	1277	33	1599
4	418	14	983	24	1290	34	1603
5	441	15	1024	25	1357	35	1605
6	461	16	1062	26	1369	36	1696
7	516	17	1063	27	1408	37	1735
8	739	18	1165	28	1455	38	1799
9	743	19	1191	29	1478	39	1815
10	789	20	1222	30	1549	40	1852

基于双参数模型的生存函数经验估计如图 4 所示, 图 5 描述了模型的经验拟合风险函数。同时, 由于该数据集所示患者死亡风险是随时间上升的, 因此, 韦伯分布、线性失效率分布 (linear failure rate distribution, LFR) 和 q -指数分布均可作为数据拟合的候选者。为了判断上述哪个密度函数更适合数据拟合, 用 MLE 方法对 3 种模型参数进行了估计, 如表 2 所示。然后利用几种不同的测试统计测度对拟合结果进行了评估。

由统计学可知, K-S 值表示分隔程度, 一般大于 0.2 即表明模型具有良好的分隔性能。由表 3 可知, q -指数分布与韦伯分布、指数分布和 LFR 分布的 K-S 均有良好分隔能力。同时, 由表 3 列出的

对数似然函数值可知, q -指数分布与韦伯分布、指数分布的最大似然函数值均在相近水平, 所以 q -指数分布可以充分利用先验信息获得对未知参数的最大似然估计。

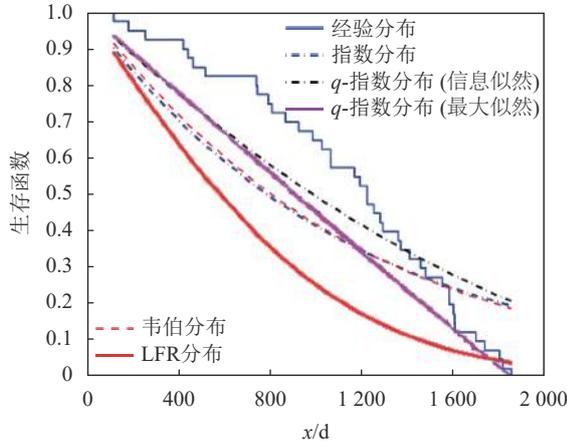


图4 白血病数据集的生存函数

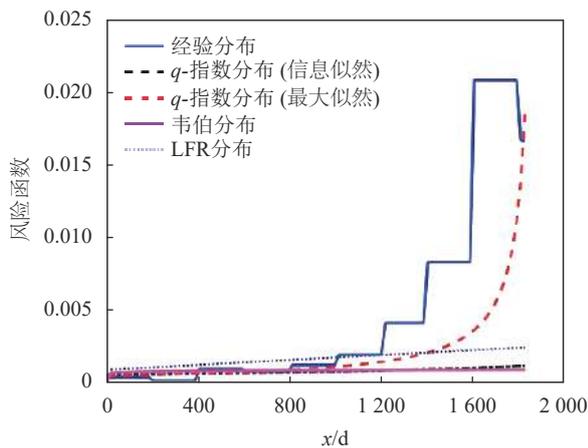


图5 白血病数据集的风险函数

表2 白血病数据集各模型参数的 MLE 值

模型	参数的最大似然估计值
LFR(a, b)	$\hat{a} = 9.501 \times 10^{-4}, \hat{b} = 4.229 \times 10^{-7}$
指数分布	$\hat{\lambda} = 1\ 137$
韦伯分布	$\hat{\sigma} = 1\ 143.3, \hat{c} = 1.055$
q -指数分布	$\hat{q} = 1.9, \hat{\lambda} = 1\ 137$

表3 白血病数据的对数似然函数值和 K-S

模型	对数似然函数值	K-S
LFR(a, b)	-318.458	0.385
指数分布	-321.446	0.308
韦伯分布	-319.874	0.282
q -指数分布	-299.511	0.205

根据统计学原理, p 值是判断原假设是否成立的依据, 一般认为 $p > 0.05$, 说明两组样本无统计学

差异。由表4可知, q -指数分布与韦伯分布的 p 值在相近水平, q -指数分布具备描述寿命分布的能力。同时, 通过计算 AIC 或者 BIC 值, 相比指数分布、韦伯分布和 LFR 分布, 不难发现 q -指数分布具有更小的 AIC 或 BIC 值, 因此 q -指数分布具有更好的寿命数据拟合性。

表4 白血病数据集各模型的 p 值、AIC 和 BIC

模型	p 值	AIC	BIC
LFR(a, b)	0.006	640.916	644.2938
指数分布	0.050	644.892	646.5808
韦伯分布	0.090	643.747	647.1258
q -指数分布	0.085	603.022	606.3998

4.2 元件寿命数据

如表5所示, 这是由文献[17]提供的50台设备的寿命数据, 已有研究人员利用韦伯分布[18-21]、指数分布[5]、LFR分布[22]分析了这个数据集。表6给出了所使用的每个分布参数的 MLE 估计值。图6和图7分别给出了设备数据集的经验参数生存函数以及风险函数。

表5 50台设备的生存期

设备编号	生存期 /月						
1	0.1	2	0.2	3	1	4	1
5	1	6	1	7	1	8	1
9	3	10	6	11	7	12	11
13	12	14	47	15	50	16	55
17	18	18	18	19	18	20	18
21	18	22	21	23	32	24	36
25	40	26	45	27	46	28	60
29	63	30	63	31	67	32	67
33	67	34	67	35	72	36	75
37	79	38	82	39	82	40	83
41	84	42	84	43	84	44	85
45	85	46	85	47	85	48	85
49	86	50	86				

表6 设备时间数据集各模型参数的 MLE 值

模型	参数的最大似然估计值
LFR(a, b)	$\hat{a} = 0.014, \hat{b} = 2.4 \times 10^{-4}$
指数分布	$\hat{\lambda} = 45.686$
韦伯分布	$\hat{\sigma} = 44.913, \hat{c} = 0.949$
q -指数分布	$\hat{q} = 24.213\ 2, \hat{\lambda} = 42.372\ 8$

已知 Aarset 数据集的风险函数具有浴盆形状。不失一般性和为了分析简便, 本文只专注风险函数中单调递增部分, 使用双参数分布来近似风险函数。为了进行参数比较, 使用似然检验去对比原假

设和备选假设。此外,利用 AIC^[23] 在多个模型中选择最优模型。最适合数据拟合的模型应具有最低的 AIC。表 7 和表 8 给出了对数似然函数值、K-S 值、 p 值、AIC 和 BIC 值^[24]。

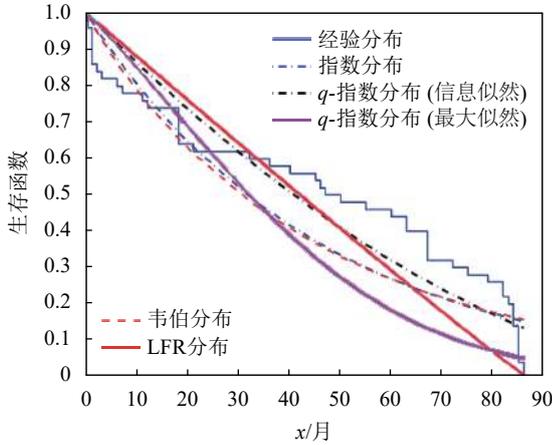


图 6 Aarset 数据集的生存函数

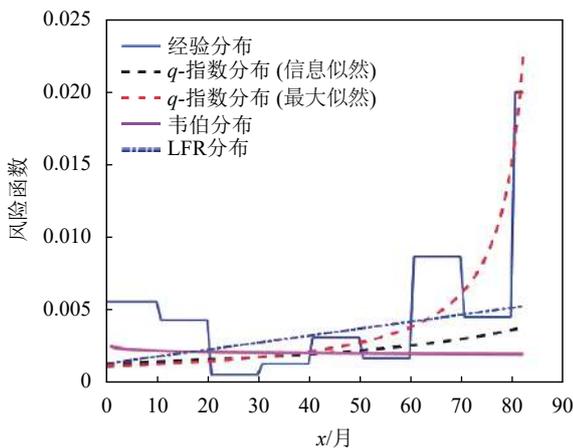


图 7 Aarset 数据集的风险函数

表 7 设备时间数据集各模型的对数似然函数值和 K-S 值

模型	对数似然函数值	K-S
LFR(a,b)	-238.0640	0.258
指数分布	-241.0896	0.226
韦伯分布	-241.0020	0.226
q -指数分布	-219.7252	0.204

表 8 设备时间数据的 P 值, AIC 和 BIC

模型	p 值	AIC	BIC
LFR(a,b)	0.253	480.128	483.9520
指数	0.408	484.179	486.0912
韦伯	0.408	486.004	489.8280
q -指数分布	0.607	443.450	459.0100

对于 Aarset 数据集,由表 7 可知 q -指数分布的 K-S 值大于 0.2, 所以具备良好的分隔能力。由

所列对数似然函数值可知, q -指数分布与韦伯分布、指数分布的对数似然函数值在相近水平, 说明 q -指数分布亦可以充分利用先验信息获得对未知参数的最大似然估计。

由表 8 知, q -指数分布的 p 值大于 0.05, 因而具备描述 Aarset 数据集的能力。同时, 可以看出 q -指数分布在本文所提到的所有分布中具有最小的 AIC 和最小的 BIC 值。这说明在所列分布中, q -指数分布能够最好地拟合本数据集。

4.3 锂电池寿命预测

估计剩余使用寿命 (remaining useful life, RUL) 有助于降低实际系统中发生灾难性事件的概率^[25]。为了研究所提出的 q -指数分布的有效性, 本文采用美国宇航局 (NASA) Prognostics 公司 (PCoE) 的 3 个电池数据集来预测锂电池的剩余使用寿命。在该数据集中用电池容量来表征寿命, 容量越低, 剩余寿命越少。

3 组电池 (即蓄电池 005、蓄电池 006 和蓄电池 018) 属于同一类型, 通过在室温下工作在 3 种不同的状态下 (充电、放电和阻抗) 进行加速老化试验^[26]。这种电池的额定容量是 2 Ah, 当电池容量减少到额定容量的 70% (从 2 Ah 减少到 1.4 Ah) 时, 电池就会达到使用寿命终止 (EOL) 标准, 容量数据如图 8 所示。

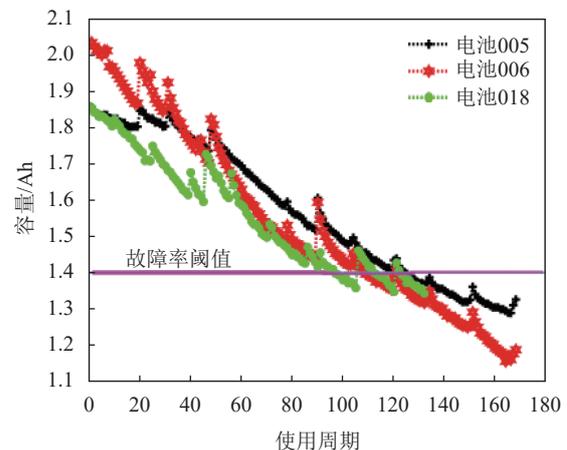


图 8 NASA PCoE 的电池容量数据

容量退化过程可用状态空间模型来描述:

$$\begin{cases} x_k = h(x_{k-1}) + w_{k-1} \\ y_k = x_k + v_k \end{cases} \quad (14)$$

式中, x_k 表示 k 周期的真实容量值; y_k 表示 k 周期的预测值; w_{k-1} 表示环境的扰动; v_k 表示观测噪声。

分别采用指数分布和 q -指数分布作为 $h(x_k)$ 来描述容量状态转换, 采用粒子滤波算法^[27] 自适应地

确定预测容量值, 并使用最大似然法确定指数分布和 q -指数分布的相应参数。

由粒子滤波基本原理可知, 按照频率派的概率观点^[24], 这里剩余寿命预测的概率分布可以通过对粒子滤波每个周期间隔内的粒子个数计数得到。间隔内粒子数越多, 说明该区间代表大多数情况下的寿命长度。

在电池 005 实验中, 共有 168 个循环样本, 分别用前 60 个、80 个和 100 个样本点训练粒子滤波器, 设定粒子滤波预测的起始时间分别为 $T=60$ 、80、100。当起始点 $T=100$ 时, RUL 的分布直方图和 RUL 预测结果如图 9、图 10 和表 9 所示。其中, 图 9b 使用参数 $q=1.001$ 。

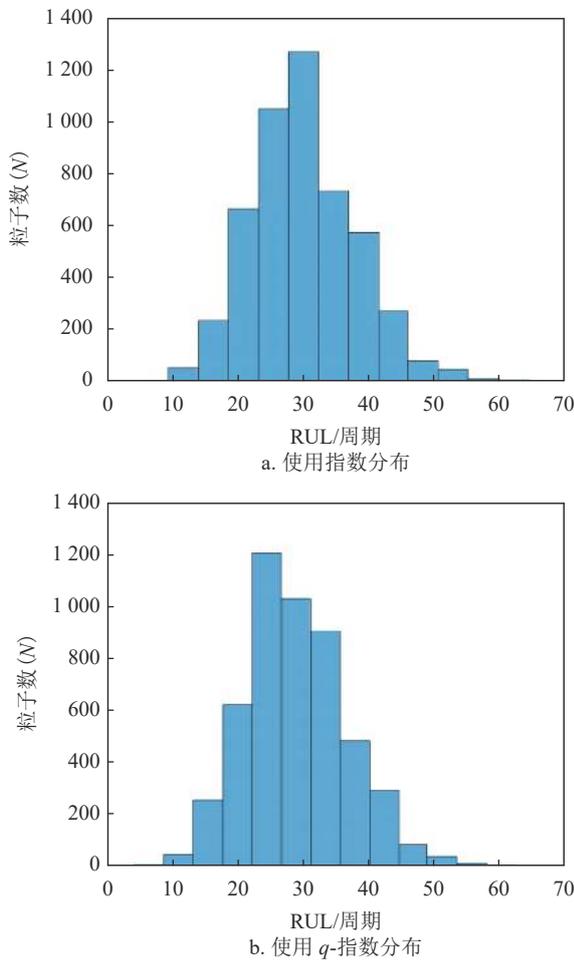


图 9 电池 005 的 RUL 直方图

此外, 为了定量评价预测精度, 将预测误差定义如下:

$$e_{RUL} = RUL_p - RUL_t \quad (15)$$

式中, RUL_p 表示预测的周期数; RUL_t 表示真实的周期数。

如图 10(其中 $q=1.001$) 所示, 电池 005 的真实寿命周期的结束数为 124, 使用 q -指数分布相对能更好地拟合实际值。根据表 9, 在指数分布假设下, 在起始预测点 $T=60$ 时, 粒子滤波算法预测的寿命周期数为 152。根据上述定义, 指数分布模型预测误差为:

$$e_{RUL} = RUL_p - RUL_t = 152 - 124 = 28$$

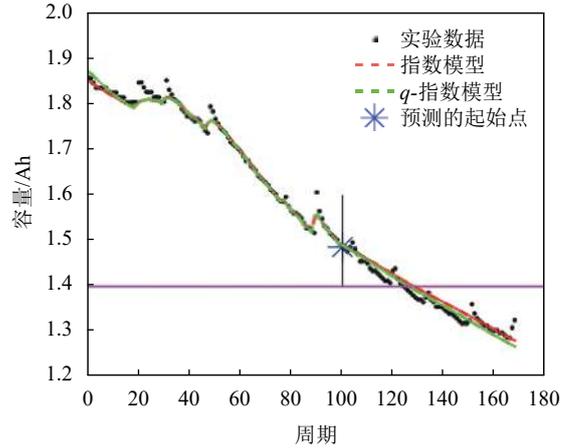


图 10 电池 005 在起始点 T 为 100 的 RUL 预测结果

表 9 不同起始点 T 下锂电池的 RUL 预测结果

电池编号	T	实际寿命 (RUL_t)	指数分布预测寿命	q -指数分布预测寿命	q	指数分布预测误差	q -指数分布预测误差	q -指数分布比指数分布预测误差缩小/%
005	60	124	152	147	1.001	28	23	17.857
	80	124	138	134	1.001	14	10	28.571
	100	124	131	128	1.001	7	4	42.857
006	60	108	106	106	1.010	-2	-2	0
	80	108	99	99	1.010	-9	-9	0
	100	108	107	107	1.010	-1	-1	0
018	60	96	92	94	0.990	-4	-2	50.000
	80	96	103	101	0.990	7	5	28.571
	100	96	94	95	0.990	-2	-1	50.000

同样以 $T=60$ 为起始预测点, q -指数分布在 $q=1.001$ 的情况下, 预测误差为 23, 小于指数分布假设的结果。因此, 提出的 q -指数分布辅助粒子滤波算法对电池 005 的剩余使用寿命有更准确的估计。

同理, 观察图 11(其中 $q=1.010$) 与图 12(其中 $q=0.990$) 中电池 006 和电池 018 的结果, 结合表 9, 发现对于电池 006 而言, 真实寿命周期的结束数

为108,而在指数分布假设下,在起始预测点 $T=60、80、100$ 时,使用 q -指数分布获得预测结果均与指数分布的相同,说明基于 q -指数分布的粒子滤波方法至少可以获得和指数分布假设下一样的估计精度。

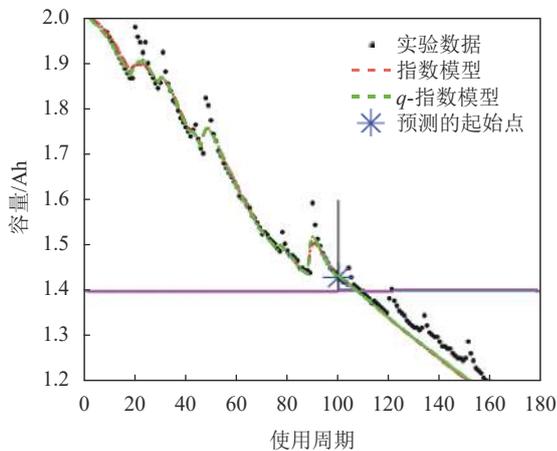


图11 电池006在起始点 $T=100$ 的RUL预测结果

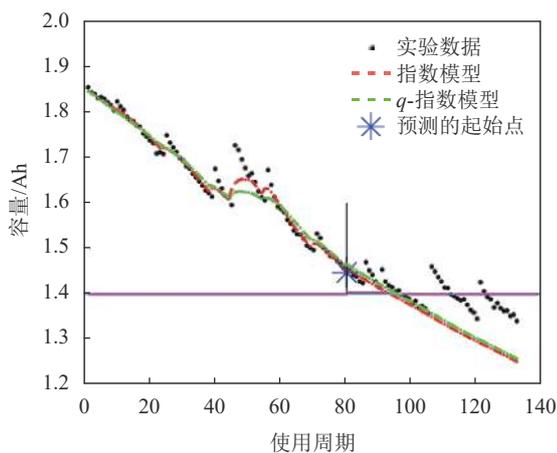


图12 电池018在起始点 $T=80$ 处的RUL预测结果

对于电池018而言,真实寿命周期的结束数为96,在指数分布假设下,在起始预测点 $T=60$ 时,预测的寿命周期数为92,预测误差为-4。而 q -指数分布在 $q=0.990$ 的情况下,如表9所示的同一点的预测误差为-2,误差预测减小了50%。对于 $T=80$ 和 $T=100$ 的起始预测点,使用 q -指数分布均可取得更准确的效果。

5 结束语

本文基于最大熵方法,通过计算均值约束下最大 Rényi 熵,得到一种新的广义指数分布: q -指数分布。本文对 q -指数分布的统计特性进行了分析,并给出了均值和各阶矩的解析公式。为了便于应用

于可靠性分析,给出了基于 q -指数分布可靠性的预测模型及对应的生存函数和风险函数的解析表达式。指出可采用了两种方法:极大似然估计法和信息似然估计法进行双参数估计。最后,结合医学白血病患者寿命数据集、设备元件寿命数据集及锂电池剩余寿命数据集进行了验证,通过与韦伯分布、指数分布等常用寿命预测分布对比,验证了 q -指数分布的有效性和估计精度的优良性。下一步将挖掘 q -指数分布在复杂系统建模中的高效应用。

参考文献

- [1] BARLOW R E, PROSCHAN F. Mathematical theory of reliability[M]. [S. l.]: Society for Industrial and Applied Mathematics, 1996.
- [2] WEIBULL W. Wide applicability[J]. Journal of Applied Mechanics, 1951, 103(730): 293-297.
- [3] 邓炳杰,陈晓慧. Weibull分布下基于遗传算法的设备寿命预测[J]. 数学杂志, 2016, 36(2): 385-392.
DENG Bing-jie, CHEN Xiao-hui. Device life prediction based on genetic algorithm under Weibull distribution[J]. Journal of Mathematics, 2016, 36(2): 385-392.
- [4] 吕佳,高慧,华思蕊. Weibull分布和指数分布以及均匀分布的关系研究[J]. 甘肃科学学报, 2017, 29(3): 1-3.
LÜ Jia, GAO Hui, HUA Si-rui. Study on the relationship between Weibull distribution and exponential distribution and uniform distribution[J]. Journal of Gansu Sciences, 2017, 29(3): 1-3.
- [5] SARHAN A M, AHMAD A A, ALASBAHI I A. Exponentiated generalized linear exponential distribution[J]. Applied Mathematical Modelling, 2013, 37(5): 2838-2849.
- [6] ALMALKI S J, YUAN J. A new modified Weibull distribution[J]. Reliability Engineering & System Safety, 2013, 111: 164-170.
- [7] ALMALKI S J, NADARAJAH S. A new discrete modified Weibull distribution[J]. IEEE Trans on Reliability, 2014, 63(1): 68-80.
- [8] JIANG H, XIE M, TANG L C. On MLEs of the parameters of a modified Weibull distribution for progressively type-2 censored samples[J]. Journal of Applied Statistics, 2010, 37(4): 617-627.
- [9] SOLIMAN A A, ABDELLAH A H, ABOUELHEGGAG N A, et al. Modified Weibull model: A Bayes study using MCMC approach based on progressive censoring data[J]. Reliability Engineering & System Safety, 2012, 100: 48-57.
- [10] UPADHYAY S K, GUPTA A. A Bayes analysis of modified Weibull distribution via Markov chain Monte Carlo simulation[J]. Journal of Statistical Computation and Simulation, 2010, 80(3): 241-254.
- [11] MALLOY M L, NOWAK R D. Near-optimal adaptive compressed sensing[J]. IEEE Trans on Information Theory, 2014, 60(7): 4001-4012.
- [12] ZHANG P, GAN L, SUN S, et al. Modulated unit-norm tight frames for compressed sensing[J]. IEEE Trans on

- [Signal Processing](#), 2015, 63(15): 3974-3985.
- [13] BI D, XIE Y, LI X, et al. A sparsity basis selection method for compressed sensing[J]. [IEEE Signal Processing Letters](#), 2015, 22(10): 1738-1742.
- [14] SONG K S. Rényi information, loglikelihood and an intrinsic distribution measure[J]. [Journal of Statistical Planning and Inference](#), 2001, 93(1-2): 51-69.
- [15] MAHMOUD M A W, ALAM F M A. The generalized linear exponential distribution[J]. [Statistics & probability Letters](#), 2010, 80(11-12): 1005-1014.
- [16] ABOUAMMOH A M, ABDULGHANI S A, QAMBER I S. On partial orderings and testing of new better than renewal used classes[J]. [Reliability Engineering & System Safety](#), 1994, 43(1): 37-41.
- [17] AARSET M V. How to identify a bathtub hazard rate[J]. [IEEE Trans on Reliability](#), 1987, 36(1): 106-108.
- [18] MUDHOLKAR G S, SRIVASTAVA D K. Exponentiated Weibull family for analyzing bathtub failure-rate data[J]. [IEEE Trans on Reliability](#), 1993, 42(2): 299-302.
- [19] XIE M, LAI C D. Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function[J]. [Reliability Engineering & System Safety](#), 1996, 52(1): 87-93.
- [20] LAI C D, XIE M, MURTHY D N P. A modified Weibull distribution[J]. [IEEE Trans on Reliability](#), 2003, 52(1): 33-37.
- [21] SILVA G O, ORTEGA E M M, CORDEIRO G M. The beta modified Weibull distribution[J]. [Lifetime Data Analysis](#), 2010, 16(3): 409-430.
- [22] AKAIKE H. A new look at the statistical model identification[J]. [IEEE Trans on Automatic Control](#), 1974, 19(6): 716-723.
- [23] 朱立蓉, 徐哲峰. Z_p 上一种指数函数的分布[J]. [纯粹数学与应用数学](#), 2020, 36(2): 161-167.
ZHU Li-rong, XU Zhe-feng. The distribution of an exponential function over Z_p [J]. [Pure and Applied Mathematics](#), 2020, 36(2): 161-167.
- [24] SCHWARZ G. Estimating the dimension of a model[J]. [The Annals of Statistics](#), 1978, 6(2): 461-464.
- [25] DONG G, CHEN Z, WEI J, et al. Battery health prognosis using Brownian motion modeling and particle filtering[J]. [IEEE Trans on Industrial Electronics](#), 2018, 65(11): 8646-8655.
- [26] GOEBEL K, SAHA B, SAXENA A, et al. Prognostics in battery health management[J]. [IEEE Instrumentation & Measurement Magazine](#), 2008, 11(4): 33-40.
- [27] ARULAMPALAM M S, MASKELL S, GORDON N, et al. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking[J]. [IEEE Trans on Signal Processing](#), 2002, 50(2): 174-188.

编辑 刘飞阳