Estimation of q-Exponential Distribution Parameters through Kullback-Leibler Divergence of Survival Functions

Pedro Augusto Pereira Lucchetti[1], Bruno da Fonte Leite[2], Lavínia Maria Mendes Araújo[3], Márcio José das Chagas Moura[4], Isis Didier Lins[5]

1. INTRODUCTION

The q-exponential is a biparametric distribution derived from the maximization of the Tsallis entropy under defined constraints [1]. Tsallis distribution has been observed as a good alternative to describe systems in which nonequilibrium is dominant [2,3]. For example, Barbosa *et al.* [4] has proposed the application of Tsallis distribution to describe the Sun-Earth system during geomagnetic storms. Several other applications can be found in fields such as geology, anatomy, and economics.

The specific case of the q-exponential distribution has several other applications in the field of reliability engineering. Sales Filho *et al.* [5] have successfully applied the q-exponential to compute Reliability $R = P(Y < X)$ in a mechanical problem of stress–strength, where stress Y and strength X are independent q-exponential random variables. Since fatigue life data can involve extremely large values, other distributions such as Weibull may not be ideal to model phenomena in the tail of the distribution. As well as obtained by Sales Filho et al. [5], q-Exponential distribution provided better results when fitting the data analyzed by Lins *et al.* [6] when comparing a q-Exponential based generalized renewal process (GRP) with a q-Weibull-GRP.

Indeed, the q-exponential distribution is very useful in the field of reliability since it can represent all three phases of the bathtub-shaped hazard rate function: improvement when $1 < q < 2$, useful life when q > 1, and wear-out when $q < 1$ [7]. As observed by Sales Filho et al. [5], it can model the power-law behavior with a heavy-tailed probability density function (PDF), which can be illustrated by the realization of rare events – failure of equipment working for a long time with low hazard rates. However, Sales Filho et al. [5] have observed that for the wear-out phase ($q<1$) the estimation of the q-exponential parameters presented convergence problems when applying the traditional Maximum Likelihood method due to the “monotone likelihood” problem, which occurs when the log-likelihood obtain its maximum for infinite parameter values [8]. De Negreiros *et al.* [7] have tackled such a problem by evaluating corrections to the original q-exponential distribution with Firth’s method and optimization using the Nelder-Mead method, which led to satisfactory results in terms of parameters estimation.

To avoid the handling of the “monotone” likelihood problem, the proposition of the current work is to use an alternative method to estimate q-Exponential parameters, by using the Kullback-Liebler divergence. It can be understood as a relative entropy between two probability distributions [9]. Since the evaluation of KL divergence for PDF may present numerical convergence issues and in reliability in general the main goal is to analyze the survival function of a given equipment, the Kullback-Liebler divergence of Survival functions “KLS” is herein considered. As applied by Liu [10], it will be measured the divergence between the modeling of a set of sample data with q-exponential and an empirical survival function. By including a new mathematical term so that KLS remains positive during the calculations, the KLS function can be minimized to obtain estimations for the q-exponential two parameters q (shape) and η (scale).

2. APPROACH FOR APPLICATION OF KLS TO ESTIMATE *q*-EXPONENTIAL DISTRIBUTION PARAMETERS

2.1 q-Exponential Distribution

The q-exponential function $\exp_q(t)$ is defined as [5]:

1 B.Sc, Engenheiro Químico – a
2, 3 B.Sc, Engenheiro de Produção – b
4, 5 PhD, Engenheiro de Produção - b
a DNV
b CEERMA/Universidade Federal de Pernambuco

$$\exp_q(t) = \begin{cases} [1+(1-q)t]^{1/(1-q)}, \text{if } [1+(1-q)t] \geq 0 \\ 0, \quad \text{otherwise} \end{cases} \tag{1}$$

where,
q = shape parameter (also known as entropic index).

The cumulative distribution function (CDF) $F_q(t)$ for the q-Exponential distribution is defined by the following expression, and presents different supports for t depending on the parameter q. For the limit $q \to 1$ the q-Exponential reaches the traditional Exponential distribution [5].

$$F_q(t) = \begin{cases} 1 - [\exp_q\left(\frac{-t(2-q)}{\eta}\right)], \text{if } t \geq 0 \\ 0, \text{ otherwise} \end{cases} \qquad t \in \begin{cases} [0; \infty), \text{if } 1 \geq q > 2 \\ [0; \frac{\eta}{(1-q)}), \text{ if } q < 1 \end{cases} \tag{2}$$

Where,
η = scale parameter.

Hence, the survival function reliability $R_q(t)$ is structured as follows (Eq. 3). It will be further considered in the application of the Kullback-Liebler Divergence of survival functions:

$$R_q(t) = \begin{cases} \exp_q\left(\frac{-t(2-q)}{\eta}\right), \text{if } t \geq 0 \\ 0, \quad \text{otherwise} \end{cases} \tag{3}$$

2.2 Kullback-Liebler Divergence of Survival Functions

Kullback-Liebler divergence can be understood as a relative entropy between two probability distributions [9]. In mathematical terms, given two PDF f and g for the same random variable X, the Kullback-Liebler divergence of *f* in relation to g is defined by:

$$D(f||g) = \int_{\mathbb{R}} f(x) \ln \frac{f(x)}{g(x)} dx \tag{4}$$

Since the evaluation of Kullback-Liebler divergence for PDF may present numerical convergence problems, the Kullback-Liebler divergence of Survival functions “KLS” is herein considered. For such an approach, the survival functions F and G are used instead of PDFs f and g. Here, the non-parametric empirical survival function, $G_n(x)$, of a random sample of size n is defined as below:

$$G_n(x) = \sum_{i=0}^{n-1} \left(1 - \frac{i}{n}\right) I_{[X_{(i)}, X_{(i+1)}]}(x), \tag{5}$$

where I is the indicator function and $(0 = X_{(0)} \leq)\ X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)})$ are the ordered samples.

Similar as Yari *et al.* [9] has performed when applying Kullback-Liebler divergence for estimation of the Weibull parameters, after defining the empirical survival function, an adjustment to the Kullback-Liebler divergence is performed so that the function value remains positive during the parameter estimation process. Hence, the adjusted Kullback-Leibler divergence (KLS) of survival functions $G_n(x)$ and F(x) is:

$$KLS(G_n||F) = \int_0^{\infty} \left(G_n(x) \ln \frac{G_n(x)}{F(x)} - G_n(x) + F(x)\right) dx \tag{6}$$

From Eq. 6, the terms where F(x) is present are functions of q-Exponential parameters q (shape) and η (scale), which is the basis for the parameter estimation by minimizing the KLS.

3. RESULTS AND DISCUSSION

3.1 Validation Data

For validation of the application of Kullback-Liebler divergence for parameter estimation, it is herein considered data from a MRI scanner, same used by De Negreiros [7]. Table 1 presents the data used as the basis for validation of current methodology to estimate q-exponential distribution parameters.

Table 1 - Time between failures of a MRI scanner (days) [7]

#	TBF	#	TBF	#	TBF	#	TBF	#	TBF
1	99	16	77	31	19	46	26	61	47
2	38	17	24	32	47	47	135	62	26
3	109	18	66	33	14	48	44	63	87
4	10	19	25	34	53	49	59	64	6
5	35	20	4	35	14	50	11	65	13
6	42	21	8	36	35	51	18		
7	31	22	26	37	73	52	3		
8	18	23	98	38	18	53	46		
9	53	24	11	39	38	54	17		
10	3	25	87	40	140	55	7		
11	12	26	11	41	19	56	75		
12	13	27	54	42	10	57	58		
13	40	28	22	43	17	58	102		
14	6	29	13	44	4	59	6		
15	78	30	54	45	54	60	53		

3.2 Parameter Estimation Results for MRI Scanner Data

The parameter estimation for the q-Exponential is obtained by minimizing the KLS function. In the current work, the optimization method used for KLS minimization and parameter estimation was the differential evolution method [11], which can be used to optimize multidimensional functions without using gradients. The Differential Evolution method optimizes the function by a given set of candidate solutions and creating new candidate solutions by combining existing ones. The solution is the one that presents the best fitness on the optimization problem [11].

By using the Differential Evolution optimization method for minimizing the KLS function of q and η, the results presented in Table 2 were obtained. The ranges applied for the solution are the following:

- $0 < q < 3/2$
- $10 < \eta < 3000$ days

It can be seen that the values obtained for q and η are very similar from the ones obtained by De Negreiros using the Maximum Likelihood method. For validation purposes, the KLS function value was evaluated for the two solutions found for the different methods and it is concluded that the KLS is indeed smaller when evaluated for the parameters herein obtained $q=0.69$ and $\eta=64.68$ days.

Table 2 – Parameter estimation based on ML with Firth penalization method and based on minimization of Kullback-Liebler divergence for MRI scanner failure data

Method	$\hat{q}$	$\hat{\eta}$ (days)	KLS function value
ML+Firth [7]	0.71	60.48	0.2370
Kullback-Leibler	0.69	64.68	0.1868

4. CONCLUSIONS

The proposition of the current work is to use an alternative method to estimate q-Exponential parameters, by using the Kullback-Liebler divergence of survival functions. The optimization was performed using the Differential Evolution method, which optimizes the function by a given set of candidate solutions and creating new candidate solutions by combining existing ones. The solution is the one that presents the best fitness on the optimization problem. For the MRI Scanner failure data, the optimization of KLS function has provided a very similar solution as presented by De Negreiros *et al.* [7]. The advantage of the proposed method is that there is no need to handle the typical “monotone” likelihood problem often present when estimating parameters of probability distributions.

5. ACKNOWLEDGMENTS

The authors thank the Humans Resource Program (PRH 38.1) entitled “Risk Analysis and Environmental Modeling in the Exploration, Development and Production of Oil and Gas”, managed by the Brazilian Agency for Petroleum, Natural Gas and Biofuels (ANP) and the Brazilian Funding Authority for Studies and Projects (FINEP) for the financial support in this research. This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES), Fincance Code 001. The last author thanks the Brazilian National Agency for Research (CNPq) for the financial support through research grants.

6. REFERENCES

[1] JAYNES, E.T. "Gibbs vs Boltzmann Entropies", *Am. J. Phys*. vol. 33, 391–398, (1965). https://doi.org/10.1119/1.1971557.

[2] TSALLIS, C. "Introduction to Nonextensive Statistical Mechanics and Thermodynamics", *Springer, Berlin* (2009).

[3] TSALLIS, C. "Nonadditive entropy Sq and nonextensive statistical mechanics: Applications in geophysics and elsewhere, Acta Geophys". 502–525, (2012)

[4] BARBOSA, C.S.; CARABALLO, R.; ALVES, L.R.; HARTMANN, G.A.; BEGGAN, C.D.;. VILJANEN, A; NGWIRA, C.M.; PAPA, A.R.R.; PIRJOLA, R.J., "The Tsallis statistical distribution applied to geomagnetically induced currents", *Sp. Weather*. 15 1094–1101 (2017).

[5] SALES FILHO, R.L.M.; DROGUETT; E.L.; LINS, I.D.; MOURA, M.C; AMIRI, M.; AZEVEDO, R.V. "Stress-Strength Reliability Analysis with Extreme Values based on q-Exponential Distribution", Qual. Reliab. Eng. Int. 33 (2017) 457–477.

[6] LINS, I.D.; MOURA, M.C.; DROGUETT, E.L.; CORRÊA, T.L. "Combining generalized renewal processes with non-extensive entropy-based q-distributions for reliability applications", *Entropy*. 20 (2018).

[7] DE NEGREIROS, A.C.S.V.; LINS, I.D.; MOURA, M. C.; DROGUETT, E.L. "Reliability data analysis of systems in the wear-out phase using a (corrected) q-Exponential likelihood", *Reliab. Eng. Syst. Saf.* 197 (2020).

[8] PIANTO, D.M.; CRIBARI-NETO, F. "Dealing with monotone likelihood in a model for speckled data", *Comput. Stat. Data Anal.* 55 1394–1409 (2011).

[9] YARI, G.; MIRHABIBI, A.; SAGHAFI, A. "Estimation of the Weibull parameters by Kullback-Leibler divergence of survival functions", *Appl. Math. Inf. Sci*. 7 (2013) 187–192.

[10] LIU, J. "Information theoretic content and probability", (2007). (Ph.D. Thesis, University of Florida, USA, pp. 25-26, 2007).

[11] STORN, R.; PRICE, K."Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces", *J. Glob. Optim.* 11 284–287 (1997).