2nd National Conference on Soft Computing of Engineering Science in Industry and Society (ASEIS 2023)

همایش ملی
محاسبات نرم علوم مهندسی
در صنعت و جامعه

A new family of distribution for modeling engineering data

Ali Khosravi Tanak[1,*] and Marziyeh Najafi[2]

[1]Department of Statistics, Velayat University, Iranshahr; a.khosravi@velayat.ac.ir
[2]Department of Mathematics, Velayat University, Iranshahr; m.najafi@velayat.ac.ir

---

Abstract. In this paper, we introduce a new family of distributions called the NOTX-G family of distributions. Some properties of the this family are studied. The model parameters are estimated by the maximum likelihood method. Then, we focus our attention on a special member of this family called the NOTX-E distribution and three applications to real engineering data in order to illustrate the usefulness of the proposed family of distributions are presented.

Keywords: T-X family, Maximum likelihood, Exponential distribution.

AMS Mathematics Subject Classification [2010]: 60E05, 65C20, 62N05

---

1. Introduction

Classical distributions do not provide adequate fits to real data in several areas such as engineering, medical and biological sciences, life testing problems, demography, actuarial and economics. Hence, many works have been done by researchers to define new families to extend classical distributions and introduce flexible families for modeling data.

Khosravi et al. [3] used the maximum entropy principle to determine a distribution for modeling income distribution. The cumulative distribution function (cdf) of the proposed model is

$$U(x;\alpha,\beta,\lambda) = 1-\Big[1+\alpha(\mathrm{e}^{\beta x}-1)\Big]^{-\lambda},\ x>0, \tag{1}$$

where α, β and λ are positive-valued parameters. The probability density function (pdf) corresponding to Equation (1) is given by

$$u(x;\alpha,\beta,\lambda) = \alpha\beta\lambda\ \mathrm{e}^{\beta x}\Big[1+\alpha(\mathrm{e}^{\beta x}-1)\Big]^{-\lambda-1},\ x>0. \tag{2}$$

Consider a random variable $T \in [p,q]$ for $-\infty < p < q < \infty$ and let $O[G(x;\boldsymbol{\kappa})]$ be a function of the cdf of a random variable X such that $O[G(x;\boldsymbol{\kappa})]$ satisfies the following conditions:

(1) $O[G(x;\boldsymbol{\kappa})] \in [p,q]$ where $\boldsymbol{\kappa}$ represents a parameter vector $(1\times k)$;
(2) $O[G(x;\boldsymbol{\kappa})]$ is differentiable and monotonically non-decreasing;
(3) $O[G(x;\boldsymbol{\kappa})] \to p$ as $x\to-\infty$; and $O[G(x;\boldsymbol{\kappa})] \to q$ as $x\to\infty$.

*Speaker

Under the above conditions, Alzaatreh et al. [1] defined the T-X-G family of distribution whose cdf is given by

$$F(x;\alpha,\beta,\lambda,\boldsymbol{\kappa})=\int_{p}^{O[G(x;\boldsymbol{\kappa})]}u(t;\alpha,\beta,\lambda)\mathrm{d}t. \tag{3}$$

The corresponding pdf is given by

$$f(x;\alpha,\beta,\lambda,\boldsymbol{\kappa})=\frac{\mathrm{d}}{\mathrm{d}x}(O[G(x;\boldsymbol{\kappa})])u(O[G(x;\boldsymbol{\kappa})];\alpha,\beta,\lambda).$$

Suppose $O[G(x;\boldsymbol{\kappa})]=G(x;\boldsymbol{\kappa})/(1-G(x;\boldsymbol{\kappa}))$ be the odd ratio of a baseline distribution $G(x;\boldsymbol{\kappa})$ with parameter vector $\boldsymbol{\kappa}$. Considering T as a random variable with the cdf (1), we introduce a new family of distributions as

$$\begin{aligned}F(x;\alpha,\beta,\lambda,\boldsymbol{\kappa})&=\int_{0}^{\frac{G(x;\boldsymbol{\kappa})}{1-G(x;\boldsymbol{\kappa})}}u(t;\alpha,\beta,\theta)\mathrm{d}t\\&=U\left(\frac{G(x;\boldsymbol{\kappa})}{1-G(x;\boldsymbol{\kappa})};\alpha,\beta,\lambda\right)\\&=1-\left[1+\alpha(\mathrm{e}^{\frac{\beta G(x;\boldsymbol{\kappa})}{1-G(x;\boldsymbol{\kappa})}}-1)\right]^{-\lambda}.\end{aligned}$$

We shall refer to this family as the New Odd T-X-G (NOTX-G) family of distributions.

This paper is organized as follows. In Section 2, we present some mathematical properties of NOTX-G family including hazard rate function, quantile, skewness and kurtosis. In Section 3, estimation of the parameters of NOTX-G distributions by the method of maximum likelihood are discussed. An special cases of the proposed family are described in Section 4. Section 5 is devoted to applications of NOTX-G distributions against to other distributions in modeling real engineering data.

2. Probability density and hazard rate functions

The pdf of the NOTX-G family takes the form

$$f(x;\alpha,\beta,\lambda,\boldsymbol{\kappa})=\frac{\alpha\beta\lambda g(x;\boldsymbol{\kappa})}{1-G(x;\boldsymbol{\kappa})}\mathrm{e}^{\frac{\beta G(x;\boldsymbol{\kappa})}{1-G(x;\boldsymbol{\kappa})}}\left[1+\alpha(\mathrm{e}^{\frac{\beta G(x;\boldsymbol{\kappa})}{1-G(x;\boldsymbol{\kappa})}}-1)\right]^{-\lambda-1},\quad x\in\mathbb{R}, \tag{4}$$

where $g(x;\boldsymbol{\kappa})$ represents the pdf of the baseline model. Hereafter, a random variable X with pdf (4) is denoted by $X\sim$ NOTX-G$(\alpha,\beta,\lambda,\boldsymbol{\kappa})$. The reliability function of the EAS-G family can be expressed as follows

$$R(x;\alpha,\beta,\lambda,\boldsymbol{\kappa})=\left[1+\alpha(\mathrm{e}^{\frac{\beta G(x;\boldsymbol{\kappa})}{1-G(x;\boldsymbol{\kappa})}}-1)\right]^{-\lambda},\quad x\in\mathbb{R}. \tag{5}$$

The Hazard rate function (hrf) of a random variable X with pdf f and reliability function R is defined as

$$h(x)=\frac{f(x)}{R(x)}.$$

Considering the random variable X as a lifetime random variable, the hazard rate $h(x)$ represents the likelihood that X be realized right after time x, given that it was not realized up to time x. If $X\sim$ NOTX-G$(\alpha,\beta,\lambda,\boldsymbol{\kappa})$, then the hrf of X is given by

$$h(x;\alpha,\beta,\lambda,\boldsymbol{\kappa})=\frac{\alpha\beta\lambda g(x;\boldsymbol{\kappa})}{1-G(x;\boldsymbol{\kappa})}\mathrm{e}^{\frac{\beta G(x;\boldsymbol{\kappa})}{1-G(x;\boldsymbol{\kappa})}}\left[1+\alpha(\mathrm{e}^{\frac{\beta G(x;\boldsymbol{\kappa})}{1-G(x;\boldsymbol{\kappa})}}-1)\right]^{-1},\quad x\in\mathbb{R}. \tag{6}$$

The quantile function of the NOTX-G family can be obtained from the inverse of the cdf in (1) as follows

$$Q(u)=G^{-1}\left(\frac{1}{1+\left\{\log\left[\frac{1}{\alpha}\left(u^{-\frac{1}{\lambda}}-1\right)+1\right]\right\}^{-1}}\right),\ u\in(0,1), \tag{7}$$

where G^{-1} represents the baseline quantile function. In addition, $Q(u)$ gives a trivial random variable generation: if $U\sim\mathcal{U}(0,1)$, then

$$X=G^{-1}\left(\frac{1}{1+\left\{\log\left[\frac{1}{\alpha}\left(U^{-\frac{1}{\lambda}}-1\right)+1\right]\right\}^{-1}}\right),$$

follows the NOTX-G$(\alpha,\beta,\lambda,\boldsymbol{\kappa})$. The median of the NOTX-G family can be derived from (7) by setting $u=0.5$. The effect of the shape parameters on the skewness and kurtosis of distributions can be studied by quantile-based measures.

3. Maximum likelihood estimation

In this section, the estimation of the parameters of the NOTX-G distributions by the method of maximum likelihood is considered. Let $X_1,X_2,\dots,X_n$ be a random sample of size n of the NOTX-G family with unknown parameter vector α,β,λ and $\boldsymbol{\kappa}$. The log-likelihood function for the parameters based on a given random sample can be expressed as

$$\begin{aligned}\ell(\alpha,\beta,\lambda,\boldsymbol{\kappa})=&n\log(\alpha\beta\lambda)+\sum_{i=1}^{n}\log\left[g\left(x_i;\boldsymbol{\kappa}\right)\right]+\sum_{i=1}^{n}\log\left[1-G\left(x_i;\boldsymbol{\kappa}\right)\right]\\&+\beta\sum_{i=1}^{n}\log\left[\frac{G\left(x_i;\boldsymbol{\kappa}\right)}{1-G\left(x_i;\boldsymbol{\kappa}\right)}\right]-\sum_{i=1}^{n}\log\left[1+\alpha(\mathrm{e}^{\frac{\beta G(x_i;\boldsymbol{\kappa})}{1-G(x_i;\boldsymbol{\kappa})}}-1)\right].\end{aligned}$$

The maximum likelihood estimates (MLEs) of the unknown parameters are obtained by maximizing $\ell(\alpha,\beta,\lambda,\boldsymbol{\kappa})$ with respect to the parameters. The first partial derivatives of the log-likelihood function with respect to the parameters are given by

$$\begin{aligned}\frac{\partial\ell}{\partial\alpha}=&\frac{n}{\alpha}+\sum_{i=1}^{n}\frac{\mathrm{e}^{\frac{\beta G(x_i;\boldsymbol{\kappa})}{1-G(x_i;\boldsymbol{\kappa})}}-1}{1+\alpha(\mathrm{e}^{\frac{\beta G(x_i;\boldsymbol{\kappa})}{1-G(x_i;\boldsymbol{\kappa})}}-1)},\\
\frac{\partial\ell}{\partial\beta}=&\frac{n}{\beta}+\sum_{i=1}^{n}\log\left[\frac{G\left(x_i;\boldsymbol{\kappa}\right)}{1-G\left(x_i;\boldsymbol{\kappa}\right)}\right]+\sum_{i=1}^{n}\frac{\frac{\alpha G(x_i;\boldsymbol{\kappa})}{1-G(x_i;\boldsymbol{\kappa})}\mathrm{e}^{\frac{\beta G(x_i;\boldsymbol{\kappa})}{1-G(x_i;\boldsymbol{\kappa})}}}{1+\alpha(\mathrm{e}^{\frac{\beta G(x_i;\boldsymbol{\kappa})}{1-G(x_i;\boldsymbol{\kappa})}}-1)},\\
\frac{\partial\ell}{\partial\lambda}=&\frac{n}{\lambda},\\
\frac{\partial\ell}{\partial\kappa_j}=&\sum_{i=1}^{n}\frac{\left[g'\left(x_i;\boldsymbol{\kappa}\right)\right]_{\kappa_j}}{g\left(x_i;\boldsymbol{\kappa}\right)}-\sum_{i=1}^{n}\frac{\left[G'\left(x_i;\boldsymbol{\kappa}\right)\right]_{\kappa_j}}{1-G\left(x_i;\boldsymbol{\kappa}\right)}+\beta\sum_{i=1}^{n}\frac{\left[G'\left(x_i;\boldsymbol{\kappa}\right)\right]_{\kappa_j}}{G\left(x_i;\boldsymbol{\kappa}\right)-G\left(x_i;\boldsymbol{\kappa}\right)^2}\\
&-\sum_{i=1}^{n}\frac{\frac{\alpha\beta[G'(x_i;\boldsymbol{\kappa})]_{\kappa_j}}{[1-G(x_i;\boldsymbol{\kappa})]^2}\mathrm{e}^{\frac{\beta G(x_i;\boldsymbol{\kappa})}{1-G(x_i;\boldsymbol{\kappa})}}}{1+\alpha(\mathrm{e}^{\frac{\beta G(x_i;\boldsymbol{\kappa})}{1-G(x_i;\boldsymbol{\kappa})}}-1)},\ j=1,2,\dots,k,\end{aligned}$$

where $[g'(x_i;\boldsymbol{\kappa})]_{\kappa_j} = \partial g(x_i;\boldsymbol{\kappa})/\partial\kappa_j$ and $[G'(x_i;\boldsymbol{\kappa})]_{\kappa_j} = \partial G(x_i;\boldsymbol{\kappa})/\partial\kappa_j$ for $j = 1, 2, \ldots, k$. The MLE $(\widehat{\alpha}, \widehat{\beta}, \widehat{\lambda}, \widehat{\boldsymbol{\kappa}})$ of $(\alpha, \beta, \lambda, \boldsymbol{\kappa})$ can be obtained by solving the following equations simultaneously:

$$\frac{\partial \ell}{\partial \alpha} = \frac{\partial \ell}{\partial \beta} = \frac{\partial \ell}{\partial \gamma} = \frac{\partial \ell(\lambda, \boldsymbol{\kappa})}{\partial \kappa_j} = 0, \; j = 1, 2, \ldots, k.$$

A nonlinear optimization algorithm such as Newton-Raphson iterative technique can be applied to solve the equations and obtain the MLEs numerically.

4. Applications

Let the baseline distribution G be the Exponential (E) distribution with parameter θ. Then, the cdf of the NOTX-E distribution is given by

$$F(x;\alpha,\beta,\lambda,\theta) = 1 - \left[1 + \alpha(\mathrm{e}^{\frac{\beta(1-\mathrm{e}^{-\theta x})}{\mathrm{e}^{-\theta x}}} - 1)\right]^{-\lambda}, \quad x > 0, \tag{8}$$

where the parameters $\alpha > 0$, $\beta > 0$ and $\lambda > 0$ control the shapes of the distribution and $\theta > 0$ is the scale parameter. In this section, the empirical importance of the NOTX-E model is studied using three applications to real engineering data. These applications will show the flexibility of the new family of distributions in modeling real data.

- Data I: This data set is given by Xu et al. [5] on the time-to-failure (10^3 h) of the turbocharger of one type of engine. The data set is: 1.6, 2.0, 2.6, 3.0, 3.5, 3.9, 4.5, 4.6, 4.8, 5.0, 5.1, 5.3, 5.4, 5.6, 5.8, 6.0, 6.0, 6.1, 6.3, 6.5, 6.5, 6.7, 7.0, 7.1, 7.3, 7.3, 7.3, 7.7, 7.7, 7.8, 7.9, 8.0, 8.1, 8.3, 8.4, 8.4, 8.5, 8.7, 8.8, 9.0.
- Data II: The second data set represents the tensile strength data measured in GPa for single carbon fibers [4]. The data are: 0.312, 0.314, 0.479, 0.552, 0.700, 0.803, 0.861, 0.865, 0.944, 0.958, 0.966, 0.997, 1.006, 1.021, 1.027, 1.055, 1.063, 1.098, 1.140, 1.179, 1.224, 1.240, 1.253, 1.270, 1.272, 1.274, 1.301, 1.301, 1.359, 1.382, 1.382, 1.426, 1.434, 1.435, 1.478, 1.490, 1.511, 1.514, 1.535, 1.554, 1.566, 1.570, 1.586, 1.629, 1.633, 1.642, 1.648, 1.684, 1.697, 1.726, 1.770, 1.773, 1.800, 1.809, 1.818, 1.821, 1.848, 1.880, 1.954, 2.012, 2.067, 2.084, 2.090, 2.096, 2.128, 2.233, 2.433, 2.585, 2.585.
- Data III: The third data set reported by Andrews and Herzberg [2] represents the life of fatigue fracture of Kevlar 373 epoxy that is subjected to constant pressure at the 90 stress level until all have failed. The measurements of this data set are: 0.0251, 0.0886, 0.0891, 0.2501, 0.3113, 0.3451, 0.4763, 0.5650, 0.5671, 0.6566, 0.6748, 0.6751, 0.6753, 0.7696, 0.8375, 0.8391, 0.8425, 0.8645, 0.8851, 0.9113, 0.9120, 0.9836, 1.0483, 1.0596, 1.0773, 1.1733, 1.2570, 1.2766, 1.2985, 1.3211, 1.3503, 1.3551, 1.4595, 1.4880, 1.5728, 1.5733, 1.7083, 1.7263, 1.7460, 1.7630, 1.7746, 1.8275, 1.8375, 1.8503, 1.8808, 1.8878, 1.8881, 1.9316, 1.9558, 2.0048, 2.0408, 2.0903, 2.1093, 2.1330, 2.2100, 2.2460, 2.2878, 2.3203, 2.3470, 2.3513, 2.4951, 2.5260, 2.9911, 3.0256, 3.2678, 3.4045, 3.4846, 3.7433, 3.7455, 3.9143, 4.8073, 5.4005, 5.4435, 5.5295, 6.5541, 9.0960.

In addition of the proposed model, we also fit the BGW family to the data. Table 1–3 give the values of the following goodness of fit statistics for the considered models: $-\ell$ (the negative maximized log-likelihood), AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), and AICc (Akaike second-order corrected Information Criterion). The statistical packages are used by R 4.1.0 to obtain numerical results.

TABLE 1. The values of $-\ell$, AIC, BIC, and AICc for the models fitted to data I.

Model	$-\ell$	AIC	BIC	AICc
5 parameters				
BGW	77.781	165.561	174.005	167.325
4 parameters				
NOTX-E	77.890	163.780	170.535	162.699
BGE	81.154	170.307	177.062	171.449
BW	82.129	172.258	179.014	173.401
3 parameters				
GW	90.429	186.859	191.926	187.526
BE	87.465	180.929	185.995	181.595

TABLE 2. The values of $-\ell$, AIC, BIC, and AICc for the models fitted to data II.

Model	$-\ell$	AIC	BIC	AICc
5 parameters				
BGW	48.232	106.464	117.634	105.541
4 parameters				
NOTX-E	48.098	104.196	113.132	103.590
BGE	48.758	105.516	114.452	104.910
BW	48.895	105.790	114.726	105.184
3 parameters				
GW	49.540	105.080	111.782	104.722
BE	49.738	105.476	112.178	105.118

TABLE 3. The values of $-\ell$, AIC, BIC, and AICc for the models fitted to data III.

Model	$-\ell$	AIC	BIC	AICc
5 parameters				
BGW	122.019	254.039	265.692	254.896
4 parameters				
NOTX-E	120.523	249.046	258.369	248.498
BGE	122.050	252.100	261.423	252.663
BW	122.156	252.313	261.636	252.876
3 parameters				
GW	122.163	250.327	257.319	250.661
BE	122.227	250.455	257.447	250.788

From the values of these statistics, we infer that the NOTX-E distribution provides a better fit than other considered distributions for the real data set.

References

1. A. Alzaatreh, C. Lee, and F. Famoye, “A new method for generating families of continuous distributions”, Metron 71 (2013), pp. 63–79.
2. Andrews, D.F. and Herzberg, A.M. “Data: a collection of problems from many fields for the student and research worker”, Springer Science & Business Media, 2012.
3. A. Khosravi Tanak, G. R. Mohtashami Borzadaran, and J. Ahmadi, “Entropy maximization under the constraints on the generalized Gini index and its application in modeling income distributions, ”Physica A, 438 (2015), pp. 657–666.
4. M. Raqab, T. Madi, and K. Debasis, “Estimation of $P(Y < X)$ for the 3-parameter generalized exponential distribution”, Comm. Statist. Theory Methods, 37 (2008), pp. 2854–2864.
5. K. Xu, M. Xie, L.C. Tang, S.L. Ho. “Application of neural networks in forecasting engine systems reliability”, Appl Soft Comput. 2(4) (2003), pp. 255–268.