Analytical Mechanisms of Shape Formation of a Generalized Family of Distributions

T. A. Nemchenko[1], **H. Ya. Tuluchenko**[1]
[1]National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine
mailtan@ukr.net, tuluchenko@ukr.net

Modeling of empirical data often requires the use of probability distributions with complex shapes that go beyond simple unimodality. Mixtures of distributions are traditionally used to describe such complex distributions. Although this approach is widespread, it has certain drawbacks, including: an increase in the number of parameters with an increasing number of mixture components, the complexity of analytical investigation of mixture properties, and problems related to identifiability and numerical estimation of parameters. Given this, the development of flexible single-component distributions capable of modeling complex empirical data and their nontrivial shapes is a relevant direction of research.

Within the framework of this direction, we consider a generalized family of distributions proposed in [1]:

$$F(x) = (1+\nu)^{\omega} \cdot \left(\frac{G(x)}{\nu + G(x)}\right)^{\omega}, \tag{1}$$

where $\nu > 0$, $\omega > 0$; and $G(x)$ is the cumulative distribution function of a continuous distribution.

The authors conducted a systematic analysis of the analytical conditions for the existence of the PDF and the inflection points of the CDF for the family (1), allowing the identification of the mechanisms that shape its form and the limitations of its application depending on the choice of the base distribution and the values of the parameters. The analytical dependencies obtained are the key to a justified choice of the base distribution and parameter range, ensuring effective modeling of empirical data with a complex structure.

Let us elucidate the analytical mechanisms governing the shape formation of the distribution family (1), drawing on:

1) the conditions for the existence of the PDF:

$$\frac{d}{dx}F(x) = \nu\omega(1+\nu)^{\omega} \cdot \frac{(G(x))^{\omega-1}}{(\nu+G(x))^{\omega+1}} \cdot \frac{d}{dx}G(x), \tag{2}$$

2) the necessary conditions for the existence of inflection points of its CDF:

$$\frac{d^2}{dx^2}F(x) = \nu\omega(1+\nu)^{\omega} \cdot \frac{(G(x))^{\omega-2}}{(\nu+G(x))^{\omega+2}} \cdot M = 0, \tag{3}$$

where

$$M = G(x)(\nu+G(x))\frac{d^2}{dx^2}G(x) + (\nu(\omega-1) - 2G(x))\left(\frac{d}{dx}G(x)\right)^2. \tag{4}$$

When $0 < \omega \leq 1$, the factor $(G(x))^{\omega-1}$ in expression (2) and the factor $(G(x))^{\omega-2}$ in expression (3) significantly complicate the analysis of the existence conditions for the PDF (2) and the inflection points of the CDF (3). The analysis becomes analytically non-obvious and requires special approaches for each specific base distribution.

Viewed as a differential equation for the function $G(x)$, the equation $M = 0$ belongs to the class of Liouville-type equations. It has a general solution:

$$G_{\text{General}}(x, C_1, C_2) = \frac{\nu(\nu\omega(C_1 x + C_2))^{1/\omega}}{1 - (\nu\omega(C_1 x + C_2))^{1/\omega}}, \tag{5}$$

where C_1, C_2 are arbitrary constants.

Functions from the general solution family (5) do not satisfy the requirements for cumulative distribution functions, since, in particular, we have:

$$\lim_{x \to \pm\infty} G_{\text{General}}(x, C_1, C_2) = -\nu.$$

Therefore, Equation (2) can be valid for the base functions $G(x)$ only at specific points x.

We investigate the generalized family of distributions (1), when the base function $G(x)$ is the CDF of the exponential distribution with parameter $\lambda > 0$. In this case, the factor M (4) takes the form:

$$M_{\text{Exp}} = \lambda^2 \exp(-\lambda x) \cdot (\exp(-2\lambda x) + \nu\omega \exp(-\lambda x) - (1 + \nu)). \tag{6}$$

Evidently, the dependence of the second-order critical point coordinate x^*_{Exp} on the parameters of the generalized distribution (λ, ν, ω) can be explicitly expressed from Equation (6):

$$x^*_{\text{Exp}}(\lambda, \nu, \omega) = -\frac{1}{\lambda} \ln\left(\frac{1}{2}(-\nu\omega + \sqrt{D})\right),$$

where $D = \nu^2\omega^2 + 4(1 + \nu)$.

This means that the CDF of the generalized distribution (1) with the base exponential distribution can have no more than one inflection point for all admissible parameter values.

For the normal and Weibull distributions as base distributions, the existence of more than one inflection point in the CDF is shown when $0 < \omega \leq 1$, indicating a more complex shape than sigmoid one. The possibility of violating the existence conditions for the corresponding PDFs is also shown.

When $\omega > 1$, in problems of fitting probability distributions, the generalized cumulative distribution function (1) demonstrates good flexibility, retaining its sigmoid shape for all admissible parameter values. The application of formula (1) significantly increases the flexibility of the base distribution as a model function.

The inclusion of the scaling factor K in the expression of function (1):

$$F_{Scale}(x) = K \cdot F(x) \tag{7}$$

allows its use in the approximating of experimental dependencies with different saturation levels. The factor K equals the theoretical saturation level (e.g., $K = 100\%$).

Therefore, in practical problems of modeling sigmoid dependencies, it is appropriate to use the constraint $\omega > 1$. Computational experiments taking into account the specified constraint showed that models (1) and (7) provide higher accuracy in approximating experimental dependencies with asymmetric dynamics (with a significant difference in the duration and intensity of growth and slowdown phases) compared to classical logistic models.

1. Sapkota L. P., Bam N., Kumar V. A . New Exponential Family of Distributions with Applications to Engineering and Medical Data. Preprint, 2024, Jun. 26, *In Review.* doi: 10.21203/rs.3.rs-4522315/v1.