Eur. J. Stat. 6 (2026) 2
doi: 10.28924/ada/stat.6.2

A Note on Local Maxima in Quadratic Transmuted Distributions Likelihoods

Edoh Katchekpele[1,*], Issa Cherif Geraldo[2], Tchilabalo Abozou Kpanzou[1]

[1] *Laboratoire de Modélisation Mathématique et d'Analyse Statistique Décisionnelle (LaMMASD), Département de Mathématiques, Faculté des Sciences et Techniques, Université de Kara, Kara, Togo*
edohkatchekpele@gmail.com, kpanzout@gmail.com

[2] *Laboratoire d'Analyse, de Modélisations Mathématiques et Applications (LAMMA), Département de Mathématiques, Faculté des Sciences, Université de Lomé, 1 B.P. 1515 Lomé 1, Togo*
cherifgera@gmail.com

[*] *Correspondence: edohkatchekpele@gmail.com*

Abstract. Transmutation is a widely used technique for generalizing probability distributions to improve data fitting. Its implementation often relies on maximum likelihood estimation, which reduces to a box-constrained numerical optimization problem. Despite this, many studies overlook the crucial role of the initial values required to start the optimization algorithm. In this paper, we demonstrate through two case studies on real data that improper parameter initialization can lead to convergence toward local maxima, ultimately resulting in biased estimates and incorrect conclusions. We show that the choice of starting values can significantly affect both the convergence behavior and the reliability of the final results. This study highlights the need for greater methodological rigor and increased awareness regarding parameter initialization in iterative estimation procedures, particularly within the context of transmuted distributions in order to avoid erroneous conclusions.

1. Introduction

In parametric statistics, the validity and reliability of inference depend critically on both the adequacy of the statistical model and the accurate estimation of its parameters. Over the past two decades, the search for more flexible models capable of capturing complex data patterns has led to numerous proposals for extending classical distributions. Quadratic transmutation [13] is one of the most popular of such techniquess, which transforms a baseline distribution into a more flexible form by introducing a transmutation parameter. The cumulative distribution function (cdf) of the transmuted distribution with baseline $G(x;\xi)$ is defined as

$$F(x;\theta) = (1+\lambda)G(x;\xi) - \lambda(G(x;\xi))^2, \tag{1}$$

Received: 8 Nov 2025.
Key words and phrases. numerical optimization; iterative method; maximum likelihood; exponentiated exponential.

where ξ is a real or vector parameter of the baseline distribution, $\lambda \in [-1, 1]$ is the transmutation parameter and $\theta = (\xi, \lambda)$ is the full parameter vector.

While this construction initially appeared to be a heuristic extension, recent developments have provided a deeper theoretical understanding. Bourguignon et al. [2] proved that transmuted distributions can be written as linear mixtures of the baseline distribution and the corresponding exponentiated distribution with power parameter equal to 2. Kozubowski and Podgorski [8] proved that transmuted distributions can be interpreted as extremal distributions, specifically, the distribution of the maximum (or minimum) of a random number N of i.i.d. variables following the baseline distribution, where N follows a Bernoulli distribution shifted by one. Granzotto et al. [5] proved that transmuted distributions are also mixtures of order statistics. All these results situate transmuted distributions within the broader class of distributions defined through random extremes, thereby justifying their use beyond simple empirical generalizations.

The estimation of parameters in transmuted models, particularly via the maximum likelihood (ML) method, is not straightforward. In most cases, the estimation of θ can only be done by using numerical optimization algorithms. Whatever the algorithm used, it requires important choices such that the initial estimate $\theta^{(0)}$ needed to start the algorithm. However, the associated likelihood functions are often non-convex and may exhibit multiple local maxima, especially due to the structure introduced by the transmutation term. In maximization problems, most algorithms are able to find only a local maximum [10]. So, if the log-likelihood has many local maxima, algorithms may converge to local maxima that are not the maximum likelihood (ML) estimate (MLE) and thus lead to erroneous conclusion. As emphasized by [12], local maxima in likelihood-based estimation can lead to incorrect and misleading inference. His analysis of Stata's estimation routines (using models such as the Heckman selection model and the Zinb model) demonstrates that common optimization algorithms may converge to sub-optimal local maxima, yielding results that are statistically meaningless if the global maximum is missed. Despite the seriousness of this issue, it is often overlooked by practitioners, possibly because, as noted by Hoeschele [6], it is commonly assumed that the likelihood function has a unique maximum. Consequently, when the iterative process converges to a value within the parameter space, this value is typically regarded as the final estimate. Using the examples of the transmuted exponentiated gamma and the transmuted power function distributions, Geraldo [3] and Geraldo [4] proved the possible existence of local maxima away from the maximum likelihood estimator (MLE) in the log-likelihoods of transmuted probability distributions. These local maxima are a serious concern because optimization algorithms may converge to them, resulting in incorrect parameter estimates.

This note aims to enhance the understanding of this problem in the context of transmuted distributions. We argue that particular care must be taken when implementing maximum likelihood (ML) estimation, especially by using multiple starting values to reduce the risk of convergence to a non-global solution. To illustrate this issue, we focus on the Transmuted Exponentiated Exponential (TEE) distribution introduced by Merovci [9] and the Transmuted Gompertz (TGo) distribution developed by Abdul-Momiem and Seham [1]. Using the same real dataset as in their studies, we show that different initial values lead to different parameter estimates and, importantly, we demonstrate that the original estimates reported in their works are not the true maximum likelihood estimates (MLEs). We then provide the corrected estimates and discuss their implications.

The remainder of this paper is organized as follows. Section 2 presents the TEE and TGo distributions along with their corresponding ML estimation frameworks. Section 3 reports the main findings from our numerical analyses conducted in R software [11] based on real data. Finally, Section 4 concludes with some remarks.

2. Presentation of the selected distributions

2.1. The TEE distribution. The probability density function (pdf) of the TEE distribution is defined by

$$f(x) = \alpha\beta e^{-\beta x}\left(1-e^{-\beta x}\right)^{\alpha-1}\left[1+\lambda-2\lambda\left(1-e^{-\beta x}\right)^{\alpha}\right], \quad x \geq 0$$

with vector parameter $\theta = (\alpha, \beta, \lambda) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \times [-1, 1]$. The log-likelihood function linked to a random sample $x_1, \dots, x_n$ of size n is given by

$$\ell(\theta) = n\log(\alpha\beta) - \beta\sum_{i=1}^{n} x_i + (\alpha-1)\sum_{i=1}^{n}\log\left(1-e^{-\beta x_i}\right) + \sum_{i=1}^{n}\log\left[1+\lambda-2\lambda\left(1-e^{-\beta x_i}\right)^{\alpha}\right].$$

Therefore, MLE of θ, if it exists, is solution to the following non-linear box-constrained optimization problem:

$$\hat{\theta} = \underset{\theta\in\mathbb{R}_+^*\times\mathbb{R}_+^*\times[-1,1]}{\operatorname{argmax}} \ell(\theta). \tag{2}$$

Remark 1. *The resolution of Equation (2) requires a numerical optimization algorithm. As most algorithms are able to find only a local maximum [10], it is clear that, if $\ell(\theta)$ has many local maxima, algorithms may converge to a local maximum that is not the MLE. And if one is satisfied only with testing a single starting point for the maximization of $\ell(\theta)$, it is possible to find local maxima different from the MLE and draw erroneous conclusions.*

2.2. The TGo distribution. The pdf of the TGo distribution is defined by

$$f(x) = \alpha\beta e^{\alpha x}e^{-\beta(e^{\alpha x}-1)}\left[1+\lambda-2\lambda\left(1-e^{-\beta(e^{\alpha x}-1)}\right)\right], \quad x \geq 0$$

with vector parameter $\theta = (\alpha, \beta, \lambda) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \times [-1, 1]$. The log-likelihood function linked to a random sample $x_1, \ldots, x_n$ of size n is given by

$$\ell(\theta) = n \log(\alpha\beta) + \alpha \sum_{i=1}^{n} x_i - \beta \sum_{i=1}^{n} \left(e^{-\alpha x_i} - 1\right) + \sum_{i=1}^{n} \log \left[1 + \lambda - 2\lambda \left(1 - e^{-\beta(e^{\alpha x_i} - 1)}\right)\right].$$

Therefore, the MLE of θ, if it exists, is solution to the following non-linear box-constrained optimization problem:

$$\hat{\theta} = \underset{\theta \in \mathbb{R}_+^* \times \mathbb{R}_+^* \times [-1,1]}{\operatorname{argmax}} \ell(\theta). \tag{3}$$

Remark 1 also applies to the resolution of Equation (3).

3. Main result from numerical study of real data

3.1. **Example for the TEE distribution.** We illustrate the existence of local maxima with the example of the glass fibres dataset taken from [9]:

Table 1. Glass fibres dataset

0.55	0.93	1.25	1.36	1.49	1.52	1.58	1.61
1.64	1.68	1.73	1.81	2.00	0.74	1.04	1.27
1.39	1.49	1.53	1.59	1.61	1.66	1.68	1.76
1.82	2.01	0.77	1.11	1.28	1.42	1.50	1.54
1.60	1.62	1.66	1.69	1.76	1.84	2.24	0.81
1.13	1.29	1.48	1.50	1.55	1.61	1.62	1.66
1.70	1.77	1.84	0.84	1.24	1.30	1.48	1.51
1.55	1.61	1.63	1.67	1.70	1.78	1.89	

Taking into account the availability in R software of optimization algorithms that can handle box and inequality constraints and other factors such as computational time, we retained the quasi-Newton algorithm BFGS (see [10] for a detailed description) and we implemented it using the "alabama" package [14]. Table 2 gives the final estimates for one thousand different starting points randomly chosen in the parameter space $\mathbb{R}_+^* \times \mathbb{R}_+^* \times [-1, 1]$.

Table 2. Results of estimation for glass fibres dataset

Final estimates	Frequency	Log-likelihood
$\theta_1 = (31.153, 2.910, -0.696)$	737	-28.475
$\theta_2 = (21.617, 1.965, 1.000)$	263	-26.631

From Table 2, we notice that for 73.7% of the starting points, the BFGS has converged to $\theta_1 = (31.153, 2.910, -0.696)$ and for the others (26.3%), it has converged to $\theta_2 = (21.617, 1.965, 1.000)$. The value θ_1 is thus the most probable value and it is the value found by Merovci [9] as the MLE. But, it is clear that this result is erroneous and the MLE is $\theta_2 = (21.617, 1.965, 1.000)$

because $\ell(\theta_2) > \ell(\theta_1)$. As noted by Jin et al. [7], the mere existence of local maxima is not a concern unless optimization algorithms are frequently trapped in them. This, however, is not the case here, since the most likely value corresponds to a local maximum.

3.2. **Example for the TGo distribution.** We illustrate the existence of local maxima with the example of the life of fatigue fracture of Kevlar 373/epoxy subject to constant pressure at the 90% stress level [1]:

Table 3. Observed fatigue life of Kevlar 373/epoxy under constant 90% stress

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				

Table 4 gives the final estimates for one thousand different starting points randomly chosen in the parameter space $\mathbb{R}_+^* \times \mathbb{R}_+^* \times [-1, 1]$.

Table 4. Results of estimation for kevlar dataset

Final estimates	Frequency	Log-likelihood
$\theta_1 = (0.188, 1.148, 0.819)$	644	-124.819
$\theta_2 = (0.014, 51.555, -0.837)$	356	-121.616

From Table 4, we notice that for 64.4% of the starting points, the BFGS has converged to $\theta_1 = (0.188, 1.148, 0.819)$ and for the others (35.6%), it has converged to $\theta_2 = (0.014, 51.555, -0.837)$. The value θ_1 is thus the most probable value and it is the value found by Abdul-Momiem and Seham [1] as the MLE. But, it is clear that this result is erroneous and the MLE is $\theta_2 = (0.014, 51.555, -0.837)$ because $\ell(\theta_2) > \ell(\theta_1)$.

4. Conclusion

In this note, we have proven through the example of the transmuted exponentiated exponential and the transmuted Gompertz distributions, that the log-likelihood of a transmuted distribution could have several local maxima and, depending on the starting value, optimization algorithms could converge to local maxima that are not the maximum likelihood estimate (MLE). It would then be advisable to compare several starting values in order to avoid erroneous conclusion.

Competing interests: The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] I.B. Abdul-Moniem, M. Seham, Transmuted Gompertz Distribution, Comput. Appl. Math. J. 1 (2015), 88–96.

[2] M. Bourguignon, I. Ghosh, G.M. Cordeiro, General Results for the Transmuted Family of Distributions and New Models, J. Probab. Stat. 2016 (2016), 7208425. https://doi.org/10.1155/2016/7208425.

[3] I.C. Geraldo, On the Maximum Likelihood Method for the Transmuted Exponentiated Gamma Distribution, Int. J. Math. Stat. 22 (2021), 20–39.

[4] I.C. Geraldo, Transmuted Power Function Distribution Revisited: Simulation Study and Important Lessons on Starting Values and Local Maxima, J. Nonlinear Sci. Appl. 18 (2025), 250–258. https://doi.org/10.22436/jnsa.018.04.03.

[5] D.C.T. Granzotto, F. Louzada, N. Balakrishnan, Cubic Rank Transmuted Distributions: Inferential Issues and Applications, J. Stat. Comput. Simul. 87 (2017), 2760–2778. https://doi.org/10.1080/00949655.2017.1344239.

[6] I. Hoeschele, A Note on Local Maxima in Maximum Likelihood, Restricted Maximum Likelihood, and Baysian Estimation of Variance Components, J. Stat. Comput. Simul. 33 (1989), 149–160. https://doi.org/10.1080/00949658908811193.

[7] C. Jin, Y. Zhang, S. Balakrishnan, M.J. Wainwright, M.I. Jordan, Local Maxima in the Likelihood of Gaussian Mixture Models: Structural Results and Algorithmic Consequences, in: 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain, 2016.

[8] T.J. Kozubowski, K. Podgórski, Transmuted Distributions and Random Extrema, Stat. Probab. Lett. 116 (2016), 6–8. https://doi.org/10.1016/j.spl.2016.04.001.

[9] F. Merovci, Transmuted Exponentiated Exponential Distribution, Math. Sci. Appl. E-Notes, 1 (2013), 112–122.

[10] J. Nocedal, S.J. Wright, Numerical Optimization, Springer New York, 2006. https://doi.org/10.1007/978-0-387-40065-5.

[11] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2025.

[12] J.S. Silva, Local Maxima in the Estimation of the Zinb and Sample Selection Models, in: United Kingdom Stata Users' Group Meetings 2017, Stata Users Group, 2017. https://econpapers.repec.org/paper/bocusug17/07.htm

[13] W.T. Shaw, I.R.C. Buckley, The Alchemy of Probability Distributions: Beyond Gram-Charlier Expansions, and a Skew-Kurtotic-Normal Distribution from a Rank Transmutation Map, arXiv:0901.0434 (2009). https://doi.org/10.48550/ARXIV.0901.0434.

[14] R. Varadhan, Alabama: Constrained Nonlinear Optimization, R package version 2023.1.0. (2023).