بعض تطبيقات توزيع "عائلة توكي لامبدا" الموسعة في العلوم المالية والموثوقية

لحسن عبدالله باشيوة
قسم الرياضيات / جامعة حائل/المملكة العربية السعودية
bachiouala@hotmail.com

الملخص

يستخدم توزع "عائلة توكي لامبدا" المعمم وفق التصميم الواسع لمجموعة فروق البيانات المختلفة، وهو توزيع متعدد الجوانب يمكن أن يستوعب مجموعة واسعة من الحالات والأشكال. ويعرف من قبل وظيفته المستهدفة خلال تحديد المعايير والثوابت والمعلمات التي تعبر عن موقع ومقياس نطاق الوسيط وبين مؤشرات الأرباع التي تصيغ شكل للتوزيع، ويمكن إستخدام النموذج الموسع في دراسة محاكاة بعض التطبيقات المالية، كما يفتح فرص الإستخدام في مواضيع إدارة المخاطر والتحكم في حالات الفشل لتوزيع الوقت من عنصر لأخر والنظم الفرعية له أو أي نظام على مستوى عال من التعقيد لتتبع توقعاته بإستخدام أسلوب المحاكاة. والشكل العام لتوزيع الإحتمالات يمكن التعبير عنه من حيث معياري التوقع والتباين، ويتم من خلالها موائمته بأحد نماذج توزيع "عائلة توكي لامبدا"، ويتم تعميم الصيغ الموسعة اللاحقة للنموذج القياسي حسب الوظيفة التي يؤديها النموذج المستهدف، وعلاوة على ذلك تم تحديد المعايير والثوابت الجديدة التي يمكن إستخدامها للمقارنة بين مجموعات البيانات في شكل ملائم في مقاييس التباين أو الإنحدار وغيرها.

Some Applications of Lambda Distribution Tukey's Family in Finance and Reliability

Abstract

The Generalized Lambda Distribution (GLD) family is used to fit and model awide range of difference data. It is defined by its quantile function. We introduce a more intuitive parameterization of the GLD that expresses the location and scale parameters directly as the median and inter-quartile range of the distribution. A simulation study performed for fitting Extended Generalized Lambda Distribution (EGLD) for financial applications, also EGLD was used for risk management and failure time distribution of component, subsystems and top-level system. Since the general form of probability distribution can be expressed in terms of the standard distribution, the subsequent formulas are given for the standard form of the function. Moreover, the new parameterization can be used to compare data sets in a convenient asymmetry and steepness shape plot.

Key word: Generalized lambda distribution, Monte Carlo, simulation, Reliability, Inverse function.

1-Introduction

The class family of lambda distributions (LDs) was first proposed by Tukey in special casethen it had been shown to have some usefulapplications [12]. The flexibility of Lambda Distributions assuming a wide variety of shapes has being used extensively to fit and model a wide range of differing phenomena to continuous probability distributions, suchasits applications in meteorology and modeling financial data in addition to the Monte Carlo simulation studies. In fact, the inverse problem of finding the probability distribution from a given moment sequence is hard as shown first by Stieltjes. The first Generalized Lambda Distribution (GLD) was presented by Tukey (1960) [27]. TheTukey's _ family of distribution is defined by the quantilefunction $F^{-1}(u)$:

$$F^{-1}(u)=\begin{cases}\lambda^{-1}(u^{\lambda}-(1-u)^{\lambda}) & ,\lambda\neq 0\\ (1-u)^{-1}\log(u), & \lambda=0\end{cases}$$

where u is uniformly distributed random variable on $(0,1)$. It can be shown that the rectangular and logistic distributions are also members of the above family of curves. So, this family has the ability to assume a wide variety of shapes including the standard distribution types of exponential, normal, χ^2, uniform, log- normal etc. Also, GLD is useful and convenient in dealing with the percentile, (e.g. risk management).

Special attention is also paid to methods that allow asymmetric source distributions. Asymmetric distributions occur in important applications such as communications and biomedical signal processing. Adaptive techniques are proposed for the modeling of score functions or estimating functions. For this reason, Friemer, Mudholkar, Kollia, and Lin developed an alternative parameterization, referred to as the FMKL generalized Tukey-Lambda distribution which has the following percent point function:

$$F^{-1}(u) = \lambda_1 + \lambda_2^{-1}(\lambda_3^{-1}\left(u^{\lambda_3} - 1\right) - \lambda^{-1}{}_4\left((1-u)^{\lambda_4} - 1\right), \quad \lambda_2 \neq 0, \lambda_4 \neq 0,\ \lambda_1, \lambda_3, \lambda_4 \in IR$$

where $\lambda_1, \lambda_2, \lambda_3$ and λ_4 denote the location, the scale, and the two shape parameters, respectively[8].

TheGLDis summarized by the inverse distribution functionwith fourparameters. Another generalization of Tukey's Lambda family that assume a wide variety of shape is given in [24]. It offers risk management flexibility in modeling a broad range of probability distributions, and it has also been mainly used in fitting a distribution to the empirical data. Also,the quantile function is given in another form [19]:

$$F^{-1}(u) = \lambda_1 + \lambda_2^{-1}\left(u^{\lambda_3} - (1-u)^{\lambda_4}\right), \quad \lambda_2 \neq 0, \lambda_1, \lambda_3, \lambda_4 \in IR$$

where λ_1, λ_2, λ_3, and λ_4 denote the location, the scale, and the two shape parameters, respectively. The latest extension of generalized lambda distribution (EGLD) is given byKarian, Dudewicz and McDonald in 1996. In 2006, Bachioua & Shaker presented an extended quantile function with five parameters [2],and its general form is given by:

$$F^{-1}(u) = \lambda_1 + \lambda_2^{-1}u^{\lambda_3} - \lambda^{-1}{}_5(1-u)^{\lambda_4}, \quad \lambda_2 \neq 0, \lambda_5 \neq 0,\ \lambda_1, \lambda_3, \lambda_4 \in IR$$

where λ_2 and λ_5 represent the scale, λ_1 represent the location parameter and λ_3, λ_4 denote the two shape parameters. Moreover, the author is presenting the new extended general form with eight parameters as follows:

$$F^{-1}(u) = \lambda_1 + \lambda_2^{-1}u^{\lambda_3} - \lambda^{-1}{}_5(1-u)^{\lambda_4} + \lambda_6 u^{\lambda_7}(1-u)^{\lambda_8};$$

such that $\lambda_2 \neq 0, \lambda_5 \neq 0,\ \lambda_1, \lambda_3, \lambda_4, \lambda_6, \lambda_7, \lambda_8 \in IR$.

In particular we are looking for an application of this new form to approximate response time densities which had been suggested by Guatema and Van Gemundin simple case in the year 2003 [9]. Indeed, response time densities and quantiles have important performance and is used to control quality of service metrics data, but their analytical derivation is, in general, very expensive.

Thus, in depth discussion on the Extended Generalized Lambda family of distributions can control the range of shapes of parameters by offering the method of fitting data using the solution of system equationmoments.

The generalized lambda distribution family fits the probability distributions of a wide variety of data sets, including the most important distributions encountered in the measurement applications.

Due to its versatility, however, obtaining general formula of approximate parameters for the EGLD and a family of distributions that can take on a very wide range of shapes, very challenging with a general form. The generalized lambda distribution is proposed as a useful model for security price distributions.

When fitting the GLD to data by fixing some parameters that represent this case, the phenomena is controlled by model formula equation. It is important that the distributions found by methods such as moment matching or least squares must be properly evaluated to determine the value of each parameter. It is evident in our examples, such distributions may have a support that does not adequately span the observed data, or they may fail to satisfy a goodness-of-fit test. Extensive simulation examples showing the

reliable performance of the EGLD methods are presented in some publications related with this field. For some new estimation procedures of EGLD parameters and for further references, see [18].

The method of moments has been used to determine the parameters of the GLD that fit a given data set, negating the possibility of applying those members of the GLD family that do not possess the first equations moments and yet may provide superior fits to the data. Inthis paper, we develop a method for fitting aEGLD distribution to data that is based on percentiles rather than moments.

This approach makes a larger portion of the GLD family accessible for data fitting and also simplifies some of the computational difficulties encountered in the method of moments. The density and the distribution function of EGLD are not available directly except in a few special cases,so it must be determined by numerical methods.

The Pearson system and adaptively combined fixed estimating functions present some cases of EGLD in whichwe solve thePearson system of differential equations. So the EGLD equations are parametric distributions and they are used to model the distributions of the mixtures [4].

The strength of these parametric equations represents wide classes of distributions, including asymmetric distributions with positive and negative kurtosis, while the estimation of the parameters is still a relatively simple procedure. The Extended Generalized Lambda Distribution (EGLD) is employed for modeling source distributions. The major benefit of the EGLD is that it also takes into account the skewness of the distributions. We briefly review maximum likelihood approach in ICA and study how the parameters of EGLD may be estimated [13].

2-Goodness- of- fit test

It is often necessary to perform multiple searches in the parameter space before finding an acceptable distribution. We have proposed a procedure that automatically restarts the downhill simplex algorithm by generating a quasi-random initial simplex at each iteration. Sonbol sequences were used to ensure uniform coverage of the parameter space. The procedure is computationally efficient; it performs fewer goodness-of-fit tests than the starship method, for example, which makes it attractive when dealing with large samples.Known best as a facile model for generating random variables with a broad range of skewness and kurtosis values, the generalized EGLD has potential financial applications, including Monte Carlo simulations, estimations of option-implied state price densities, and almost any situation requiring a flexible density shape[3].

The EGLD is a class of distributions often used for parameter estimation, fitting distributions to data, has been used in studies that have included such topics or techniques as: data mining, independent component analysis [9]. An example is used to illustrate the calculations that can be easily performed using spreadsheets [25].

Consequently, it is no longer necessary to assume normality and it is now possible to make direct inference on the actual data rather than the mean of the data.

The generalized lambda distribution family fits the probability distributions of a wide variety of data sets, including the most important distributions encountered in the measurement applications (normal, uniform, Student's t, U-shaped, exponential), the shape parameters are obtained using Tukey Lambda PPCC calculator [20].

3-General Quantile Approach

Some of those defined by their quantile function are the Johnson system, with a base of normal or logistic, the Tukey lambda and its various asymmetric generalizations, with a uniform base, and Hoaglin's g-and-h, with anormal base in distribution function whose random variable is uniformly distributed on (0, 1). It can be shown that the rectangular and logistic distributions are also members of the above family[18].

A new quantile distribution can be obtained byusing the inverse function of EGLD, but it is not generally recognized how difficult this is to achieve in asymmetric distributions, whether they have one or more than one shape parameters.

The percentileof distribution function has warning limits lies betweenQ(0.995) and Q(0.005), and the actions limits lies between Q(0.999) and Q(0.001). The central point lie at Q(0.5). The chart of this distribution displays the control limits. Here, a quantile control chart has been constructed using Logistic Quantile Distribution (LQD).

It may be plausible to conjecture that Closed-Form (QF) expressions for Commutative distribution function (CDF) and the Probability Density Function (PDF) of the mirror-exponential do not exist. If this turns out to be the case, then the new mirror-exponential joins Tukey's lambda distribution in this somewhat exclusive club. It is more common for the QF to defy analytic expression than for the CDF, and the Erlang is an example. Of course, knowledge of any one of the three basic associated functions (PDF, CDF, QF) is sufficient to represent the distribution and obtain its main parameters (albeit numericallyin some cases).

4-Size and power in finite samples

The random variable X has a standardized Tukey Lambda distribution if the random variable U is uniform between 0 and 1, and the estimated sources are modeled using EGLD. The modeling scheme provides a useful connection between practical estimator and theoretical measure of independence. The score function of the EGLD is used as an objective function of fitting continuous distributionsthat we maximize. Natural gradient and fixed-point algorithms are used to maximize the EGLD model. Also the class of GLDs is primarily used for modeling univaried real–word data.

5-Generating Random Numbers

Tukey(1960) [27] introduced a family of random variables defined by the transformation X = F(U), where U is uniformly distributed on [0,1]. Some of its properties are described with emphasis on properties of the sample range. The rectangular and logistic distributions are members of this family and distributions corresponding to certain values of parameters give useful approximations to the normal and t distributions [24].

Extended Generalized Lambda DistributionEGLD is very excellent method but we still have to generate uniform distribution to start the Processof best method. We found that the Random function from Mathematicaor from other software calculation packages were deficient. It is important to test the randomness of the generator before proceeding the generation of random numbers which is done by LET subcommands. For example, the statementLET X = UNIFORM RANDOM NUMBERS FOR I = 1, 125generates a variable X that contains uniform random numbers in the first 25 rows.

The output from therandom number generation is always a variable, never a parameter or function. The term (or terms) beforeRANDOM NUMBERS identifies the distribution for which random numbers are being generated.

6-Monte Carlo Simulations

In the Monte Carlo simulations, the fire heat release rates are sampled from the generalized extended general Tukey distribution using the rejection sampling method. These models return the smoke layer height, X, at a specified critical time. We have arbitrarily chosen the critical time to be 120s. Our objective in performing Monte Carlo simulations is to determine the accuracy of the results obtained using moment methodsand therefore we do not consider issues on the computational efficiency of our Monte Carlo simulations [25].

7-Fitting Methods

The fitting EGLD to empirical data always requires numerical optimization; due that are only defined by the inverse of the quantile functions. It involves a careful consideration of initial values for optimization, a suitable criterion to optimize and ensuring the eventual solution remains a valid probability distribution[24].The Maximum Likelihood (ML) fits the mean, variances, skewness and kurtosis resemble more closely to that of empirical data, in addition to satisfying the bootstrap Kolgomorov-Smirnoff (KS) goodness of fit better or equally well to that of starship method, which is a very flexible method in which the same data can be fitted by several different credible distributions, depending on the bin-widths representation of the histogram. The advantage of this method is that it allows the users to conduct sensitivity analysis by capturing the features of the empirical distribution of their interest[26].

The ML and starship methods are the definitive fits to the data set and will be used to obtain good fit. It provides the generally acceptable fits to empirical data using standard statistical criteria. As there is currently no single method that would work the best in every situation, the best fitted distribution chosen for a data set can be determined by fitted EGLD which best matches the moments, i.e. the mean, variance, skewness and kurtosis of the empirical data. The mean and variance of the data are often quite crucial [4].

For this reason this criterion is used rather than the alternative bootstrap KS test, which is a general purpose goodness of fit procedure. This is useful as a second stage selection as sometimes the goodness of fit results is very similar between the different fits. At this point in time, there are no definitive diagnostic tools in statistical literature to assess the quality of fit for quantile distributions such as the EGLD[16].

8-Simulations

In this section, we use Monte Carlo simulations to assess the size and power of the proposed tests; we consider well known distributions such as the normal[26]. The generalizedlambda distribution, denoted F_s, is defined in terms of the inverse of the cumulative distribution $F^{-1}(u) = \hat{\lambda}_1 + \hat{\lambda}_2 \, u^{\hat{\lambda}_3} - \hat{\lambda}_5 \, (1-u)^{\hat{\lambda}_4} + \hat{\lambda}_6 \, u^{\hat{\lambda}_7} (1-u)^{\hat{\lambda}_8}$, $0 < u < 1$. This family nests a wide range of symmetric and asymmetric distributions[18].

9-Reliability and Risk Management

To make Risk Management a regular part of actuarial practice: the development and improvement of educational materials for actuaries wishing to practice in the risk management area. The advancement of professional recognition and opportunities for actuaries in the arena of risk management. Ultimately risk management practices of actuaries will be as well documented as the practices for calculation of actuarial reserves [22].

In many practical engineering circumstances, systems reliability analysis is complicated by the fact that the failure time distributions of the constituent subsystems cannot be accurately modeled by standard distributions. In this paper, we present a low-cost, compositional approach based on the use of the first four statistical moments to characterize the failure time distributions of the constituent components, subsystems, and top-level system [19].

The approach is based on the use of Pearson distributions as an intermediate analytical vehicle, in terms of which the constituent failure time distributions are approximated. The analysis technique is presented for -out-of- systems with identical subsystems, series systems with different subsystems, and systems exploiting standby redundancy [20].

The moment-in-moment-out approach allows for the analysis of systems with arbitrary hierarchy and arbitrary (unimodal) failure time distributions, provided the subsystems are independent such that the resulting failure time can be expressed in terms of sums or order statistics. The technique consistently exhibits very good accuracy average, much less than 1 percenterror at very modest computing cost [6].

It approximates a component failure time distribution by a lambda distribution having four parameters, which can be determined if the input moments are known. An extension of the method is given in [9]. Unfortunately, the lambda distribution method turns out to be less accurate, and less stable in computing than the method presented in the current paper. In addition, the four Pearson characterization parameters can directly be expressed in the four given or measured moments, quite different from the method presented in [11] where a table look-up or linear programming method is required to obtain these parameters. Moreover, our method enables the calculation of the failure time pdf of the composite of components of different type [10].

10-Conclusions:

The EGLD family fits the probability distributions of a wide variety of data sets, including the most important distributions encountered in the measurement applications (normal, uniform, Student's t, U-shaped, exponential). How the eight parameters needed for such distribution can be exploited in the expression of measurement uncertainty and to extend the information related to a measurement. The obtained representation allows an immediate calculation of coverage intervals and is particularly useful to support the techniques commonly applied in the estimation of the combined uncertainty [20].

The major benefit of the EGLD employed for modeling source distributions. The EGLD is that it also takes into account the skewness of the distributions. Known best as a facile model for generating random variables with a broad range of skewness and kurtosis values, potential financial applications of the EGLD include Monte Carlo simulations, estimation of option-implied state price densities, and almost any situation requiring a

flexible density shape, is a flexible distribution capable of mimicking the shapes of many distributions and data samples including those with heavy tails.

A small simulation study of the 98^{th} percentile (far-right tail) is conducted for a heavy-tail EGLD with high-outlier contamination. The simulations show, with respect to estimation of the 98th-percent quantile, that TL-moments are less biased (more robost) in the presence of high-outlier contamination. However, the robustness comes at the expense of considerably more sampling variability [1].

Both sub Gaussian and supergaussian one-dimensional time series were employed throughout the numerical simulations [11].

Applications of Tukey's GH distribution family -which are composed by a skewness transformation G and a kurtosis transformation H - can be found, for instance, in financial, environmental or medical statistics.

Recently, alternative transformations emerged in the literature.. Within this work we present a very general kurtosis transformation which nests H-, K-and an approximation to the J-transformation and, hence, permits to discriminate between them. Applications to financial and telegraphic data are given[6].

11-Future work

The methods presented in this paper are highly reliable for all type of sources while still being relatively fast. Thus, it is natural to consider this basic separation of instantaneous mixtures of real-values sources a relatively well-understood problem with feasible solutions. However, there are relatively few ICA techniques for blind identification and even fewer for blind source extraction for the underdetermined case. These techniques should have a huge demand in some application areas, e.g. in communications.

Reference:

[1]-Asquith, William H, (2007)."**L-moments and TL-Moments of the Generalized Lambda Distribution**", Data Analysis, V(51), 9, 15 May, 4481-4496.
[2]-BachiouaLahcene Shaker, H. S.(2006)."**On Extended Tukey's Lambda Distribution Models**", Jordan Journal of Applied Science Natural Sciences,V(8), 2, 25-32.
[3]-Corrado, Charles. J, (2001)." **Option Pricing Based on the Generalized Lambda Distribution**", Journal of Futures Market, February, JEL classification: C14, C15, G13.
[4]-Fischer M., A. Horn and Klein, I., (2006)."**Power Kurtosis Transformations: Definition, Properties and Ordering**",AllgemeinesStatistischesArchiv, V(90), 3 / September, 395-401.
[5]-Johnson and Kotz, (1973)."**Extended and Multivariate Tukey Lambda Distributions",**Biometrika; V(60), 655-661.
[6]-Fischer, M., Horn, (2008)."**Generalized Tukey-type Distributions With Application to Financial and Teletraffic Data**", Springer Berlin / Heidelberg, Statistical Papers, Wednesday, January 09, 0932-5026 (Print) 1613-9798 (Online).
[7]-Fischer, M., Horn, A. & I. Klein, (2007)."**Tukey-type Distributions in the Context of Financial Data**", Communications in Statistics (Theory and Methods), V(36), 23-35.
[8]-Friemer, Mudholkar, Kollia, and Lin, (1988), "**A Study of the Generalized Lambda Family**", Communications in Statistics-Theory and Methods, V(17), 3547-3567.
[9]-Gautama H. and van Gemund A. J. C., (2003)."**Reliability Analysis of K-out-of- n Systems Using Generalized Lambda Distributions**", in Proc".of the Int'l Conference on Mathematics and Its Applications (SEAM—GMU 2003), Yogyakarta, Indonesia, 466–474.
[10]-Gautama H. ,(2004). "**A statistical Approach to the Performance Modeling of Parallel Systems",** Ph.D. dissertation, Delft University of Technology, The Netherlands.
[11]-Gautama H. and van Gemund A. J. C., (2006). "**Low-cost Static Performance Prediction of Parallel Stochastic Task Compositions**", IEEE Trans. Par. Distr. Syst., V(17), 78–91.
[12]-Hastings, C., Mosteller, F., Tukey, J. W., Winsor, C. P., (1947)."**Low Moments for Small Samples: A Comparative Study of Order Statistics".** Annals of Mathematical Statistics, V(18), 413–426.
[13]-Jan Eriksson, JuhaKarvanen, Visa Koivunen, (2000)."**Source Distribution Adaptive Maximum Likelihood Estimation of ICA Model",** in Proceedings of ICA.
[14]-Johnson and Kotz, (1973)."**Extended and Multivariate Tukey Lambda Distributions",** Biometrika; V(60), 655-661.
[15]-Jones, M. C., (2004).' **Families of Distributions Arising From Distributions of Order Statistics"**, TEST, Springer Berlin / Heidelberg, V(13), 1, June,1-43.
[16]-Karian, Z.A. and Dudewicz, E.J. and P. McDonald., (1996)."**The Extended Generalized Lambda Distribution System for Fitting Distributions to Data: History, Completion of Theory, Tables, Applications, the "Final Word" on Moment fits".** Comm. in Statistics: Simulation and Computation, V(25),3,611–642.
[17]-Karvanen ,Juha, (2008)."**The Generalized Lambda Distribution by L-Moments**", Computational Statistics & Data Analysis, Elsevier Science Publishers B. V, pp 1971-1983.
[18]-King. R.R.andMacGillivray. H. , (1999)."**A Starship Method for the Generalized Lambda Distributions",** Australian New Zealand J. Statistics, V(41), 3, 353–374.
[19]-Lakhany A., Mausser H., (2000)."**Estimating the Parameters of the Generalized Lambda Distribution",**Algo Research Quarterly, V (3), 47-58.

[20]-Lampasi, D.A.; Fabio Di Nicola; Podesta, (2006)."**Generalized Lambda Distribution for the Expression of Measurement Uncertainty",** IEEE Transactions on V(55), 4, Aug, 1281 – 1287

[21]-Ozturk and Dale, (1985). "**Least Squares Estimation of the Parameters of the Generalized Lambda Distribution**", Technometrics, V(27), 1, 81-84.

[22]-Pal, Surajit, (2005). "**Evaluation of Nonnormal Process Capability Indices Using Generalized Lambda Distribution",** Quality Engineering,V(17), 77–85.

[23]-Ramberg and Schmeiser, (1972), "**An Approximate Method for Generating Symmetric Random Variables**", Communications of the Association for Computing Machinery, V(15), pp 987-990.

[24]-Ramberg and Schmeiser, (1974)."**An Approximate Method for Generating Asymmetric Random Variables**", Communications of the Association for Computing Machinery, V(17), pp 78-82.

[25]-Surajit Pal, (2004)."**Evaluation of Nonnormal Process Capability Indices Using Generalized Lambda Distribution**", Quality Engineering, V(http://www.informaworld.com/smpp/title~content=t713597292~db=all~tab=issueslist~branches=17 - v1717), 1,December, 77 – 85.

[26]-Todd C. Headricka, Abdel Mugdadi, (2006)."**On Simulating Multivariate Non-normal Distributions From the Generalized Lambda Distribution**",Computational Statistics & Data Analysis,V(50), pp 3343 – 3353.

[27]- Tukey, J. W., (1960)."**The Practical Relationship Between the Common Transformation of Percentages of Counts and of Amounts**". Technical ReportV(36), Statistical Techniques Research Group., Princeton University.

[28]-Wessa P., (2007)."**Star Plot (v1.0.2) in Free Statistics Software (v1.1.22-r5)",** Office for Research Development and Education, URL, web site: http://www.wessa.net/rwasp_starplot.wasp/