پانزدهمین سمینار احتمال
و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان

15th Seminar On Probability and Stochastic Processes
August 30-31, 2025
University of Kurdistan

انجمن آمار ایران

On the double-log skew-normal distribution and its application

Abbas Mahdavi*, Hossein Negarestani

Department of Statistics, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran.

Abstract

Data confined to the interval $(0, 1)$ are frequently analyzed using various distributions in a range of practical applications. However, many existing studies fail to adequately consider aspects such as skewness, kurtosis, and heavy-tailed behavior in this data. This research introduces an alternative skew-normal type distribution that operates within a bounded interval, created by merging the characteristics of skew-normal distributions with the double-log transformation. The resulting model, with its enhanced capabilities for skewness and boundedness, offers a flexible and effective framework for modeling rates and proportions. We have implemented an EM-type algorithm to reliably estimate the parameters of the proposed model. To demonstrate the efficacy of our method, we carried out an experiment that included an analysis of one actual dataset. The findings emphasize the adaptability and precision of our proposed model when compared to conventional models.

Keywords: EM-type algorithms; Mixture models; Double-log-normal distribution, Skew-normal distribution

Mathematics Subject Classification (2020): 62E10, 62F10.

1 Introduction

The beta distribution is widely used for modeling percentages and proportions within the interval $(0, 1)$. However, various alternative distributions have been suggested in the literature to provide more flexible modeling options than the beta distribution. Among these alternatives are the Topp–Leone distribution, Kumaraswamy distribution, arcsine distribution, generalized beta type distribution, standard two-sided power distribution, McDonald arcsine distribution, exponentiated Topp-Leone distribution, generalized beta distribution, and unit Lindley distribution. The skew-normal (SN) distribution, denoted by $SN(\mu, \sigma, \lambda)$ and introduced by Azzalini (1985), has shown its great flexibility in modeling data with asymmetric behaviors by the use of fewer components in the fitting of mixture models. The probability density function (pdf) of SN distribution is expressed by

$$f_{SN}(x; \mu, \sigma, \lambda) = \frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda\frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R}, \tag{1}$$

*Corresponding author, a.mahdavi@vru.ac.ir

where $\phi(\cdot)$ and $\Phi(\cdot)$ denote the pdf and cumulative density function (cdf) of standard normal distribution, $\mu \in \mathbb{R}$, $\sigma > 0$ and $\lambda \in \mathbb{R}$ represent the location, scale and shape parameters, respectively. Setting $\lambda = 0$, the classical normal distribution with mean μ and variance σ^2 is obtained. It is noteworthy to mention that the SN distribution can be expressed by the following hierarchical stochastic representation (Lin *et al.* , 2007):

$$\begin{aligned} Y|\tau &\sim N\big(\mu + \delta(\lambda)\tau, (1-\delta^2(\lambda))\sigma^2\big), \\ \tau &\sim TN(0,\sigma^2)I_{(0,\infty)}, \end{aligned} \tag{2}$$

where $\delta(\lambda) = \lambda/\sqrt{1+\lambda^2}$. The notation $TN(\mu,\sigma^2)I_{\mathbb{A}}$ represents a doubly truncated normal distribution confined within the interval $\mathbb{A} = \{a_1 < x < a_2\}$, and $I_{\mathbb{A}}$ is an indicator function of set $\mathbb{A}$.

The main objective of this study is to introduce a new statistical distribution, called as the double-log skew-normal (DLSN) distribution, which is defined on the interval $(0,1)$ and generated using the double-log transformation, defined as $Y \overset{d}{=} \exp(-\exp(X))$, where $X \sim SN(\mu,\sigma,\lambda)$. The rest of this paper is organized as follows. In Section 2, we introduce with the DLSN distribution. Next, in Section 3, we adopt an EM-type algorithm for the fitting of the DLSN model. A real data analysis is illustrated in Section 4.

2 The model

The random variable Y is said to follow a DLSN distribution, denoted by $DLSN(\mu,\sigma,\lambda)$, with support $(0,1)$ and shape parameters $\mu \in \mathbb{R}$, $\sigma > 0$ and $\lambda \in \mathbb{R}$, if its pdf is

$$f_{DLSN}(y;\mu,\sigma^2,\lambda) = \frac{-2}{\sigma y \log y}\phi\left(\frac{\log(-\log y)-\mu}{\sigma}\right)\Phi\left(\lambda\frac{\log(-\log y)-\mu}{\sigma}\right), \quad 0<y<1. \tag{3}$$

Note that double-log normal (DLN) distribution is recovered from (3) if $\lambda = 0$ as follows:

$$f_{DLN}(y;\mu,\sigma^2) = \frac{-1}{y\log y\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}\big(\log(-\log y)-\mu\big)^2\right\}, \quad 0<y<1, \tag{4}$$

Figure 1 visually illustrates the shapes of several DLSN densities alongside their corresponding DLN densities, offering insight into the diverse characteristics that this family of distributions can exhibit. The plots vividly demonstrate that the DLSN distribution encompasses a wide range of shapes characterized by distinct directional skewness and kurtosis properties, which are dictated by the specific parameter values chosen for each density.

It is straightforward to demonstrate that the DLSN distribution admits the following stochastic representation:

$$\begin{aligned} Y|\tau &\sim DLN\big(\mu+\delta(\lambda)\tau, \big(1-\delta^2(\lambda)\big)\sigma^2\big), \\ \tau &\sim TN(0,\sigma^2)I_{(0,\infty)}. \end{aligned} \tag{5}$$

From (5) the joint pdf of Y and τ is given by

$$\begin{aligned} f_{Y,\tau}(y,\tau) =& f_\tau(\tau)f_{Y|\tau}(y) \\ =& \frac{-1}{\pi\sigma^2 y\log y\sqrt{1-\delta^2(\lambda)}}\exp\left\{\frac{-1}{2\big(1-\delta^2(\lambda)\big)\sigma^2}\left(\log(-\log y)-\mu-\delta(\lambda)\tau\right)^2\right. \\ & \left.-\frac{1}{2\sigma^2}\tau^2\right\}, \quad 0<y<1, \tau>0. \end{aligned} \tag{6}$$

It follows from (6) that the conditional distribution of τ given Y is

$$\begin{aligned} f_{\tau|Y=y}(\tau) &= \frac{f_{Y,\tau}(y,\tau)}{f_Y(y)} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}\sqrt{1-\delta^2(\lambda)}}\frac{\exp\left\{\frac{-1}{2\left(1-\delta^2(\lambda)\right)\sigma^2}\left(\tau-\delta(\lambda)\big(\log(-\log y)-\mu\big)\right)^2\right\}}{\Phi\left(\lambda\frac{\log(-\log y)-\mu}{\sigma}\right)}, \end{aligned}$$

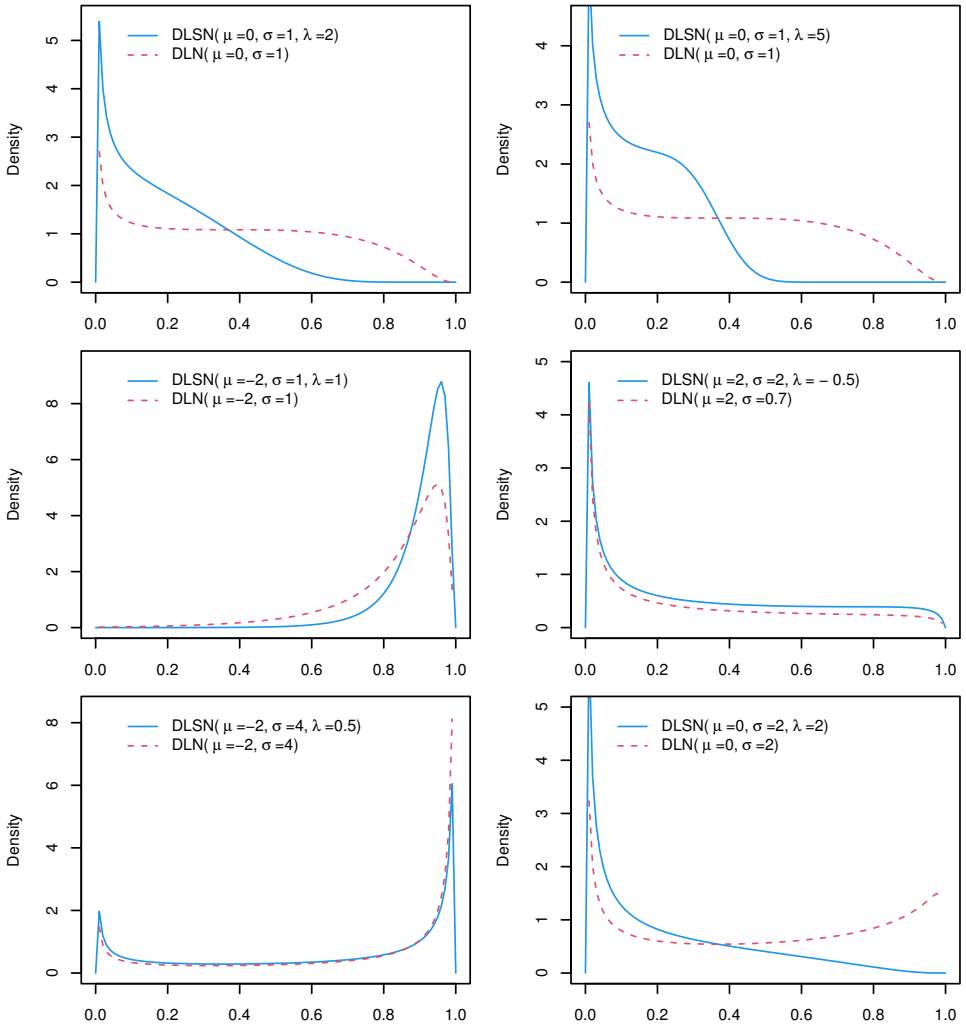


Figure 1: Different density shapes of DLSN distribution.

and using the fact that $\lambda = \delta(\lambda)/\sqrt{1-\delta^2(\lambda)}$, it yields to

$$\tau \mid Y = y \sim TN\Big(\delta(\lambda)\big(\log(-\log y) - \mu\big), \sigma\sqrt{1-\delta^2(\lambda)}\Big). \tag{7}$$

In the following subsection, we propose an EM-type algorithm for determining the ML estimates of the parameters of $DLSN(\mu, \sigma, \lambda)$ based on the joint density defined in (6).

3 Parameter estimation

Let $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$ denote the observed data, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)^\top$ be the latent variable and $\boldsymbol{y}_c = (\boldsymbol{y}^\top, \boldsymbol{\tau}^\top)^\top$ be the complete data. From (6), the log-likelihood function of $\boldsymbol{\theta} = (\mu, \sigma, \lambda)^\top$ corresponding to the complete-data $\boldsymbol{y}_c$, excluding additive constants and terms that do not involve the parameters of the model, is given by

$$\begin{aligned} & \ell_c(\boldsymbol{\theta} \mid \boldsymbol{y}_c) \\ = \; & -\frac{1}{2}\sum_{i=1}^{n}\Bigg\{\frac{1}{\sigma^2\big(1-\delta^2(\lambda)\big)}\Big(\tau_i^2 - 2\delta(\lambda)\tau_i\big(\log(-\log y_i) - \mu\big) + \big(\log(-\log y_i) - \mu\big)^2\Big) \\ & + 2\log\sigma^2 + \log\big(1-\delta^2(\lambda)\big)\Bigg\}. \end{aligned} \tag{8}$$

Given the current estimate of θ in the kth iteration, $\hat{\theta}^{(k)} = \big(\hat{\mu}^{(k)}, \hat{\sigma}^{(k)}, \hat{\lambda}^{(k)}\big)^\top$, the E-step of the EM algorithm involves the evaluation of the Q-function, a conditional expectation of (8) given the observed data. It requires the conditional expectations derived from the conditional distribution as expressed in (7). By Lemma 2 of Lin *et al.* (2007), we obtain

$$\begin{aligned}\hat{\tau}_i^{(k)} &= E(\tau_i \mid y_i, \hat{\boldsymbol{\theta}}^{(k)}) \\ &= \delta(\hat{\lambda}^{(k)})\big(\log(-\log y_i) - \hat{\mu}^{(k)}\big) + \hat{\sigma}^{(k)}\sqrt{1-\delta^2(\hat{\lambda}^{(k)})}\,\frac{\phi\big(\hat{\lambda}^{(k)}\hat{\eta}_i^{(k)}\big)}{\Phi\big(\hat{\lambda}^{(k)}\hat{\eta}_i^{(k)}\big)}, \end{aligned} \tag{9}$$

and

$$\begin{aligned}\hat{s}_i^{(k)} &= E(\tau_i^2 \mid y_i, \hat{\boldsymbol{\theta}}^{(k)}) \\ &= \delta^2(\hat{\lambda}^{(k)})\big(\log(-\log y_i) - \hat{\mu}^{(k)}\big)^2 + \hat{\sigma}^{2(k)}\big(1-\delta^2(\hat{\lambda}^{(k)})\big) \\ &\quad +\delta(\hat{\lambda}^{(k)})\big(\log(-\log y_i) - \hat{\mu}^{(k)}\big)\hat{\sigma}^{(k)}\sqrt{1-\delta^2(\hat{\lambda}^{(k)})}\,\frac{\phi\big(\hat{\lambda}^{(k)}\hat{\eta}_i^{(k)}\big)}{\Phi\big(\hat{\lambda}^{(k)}\hat{\eta}_i^{(k)}\big)}, \end{aligned} \tag{10}$$

where $\hat{\eta}_i^{(k)} = (\log(-\log y_i) - \hat{\mu}^{(k)})/\hat{\sigma}^{(k)}$ and $\delta(\hat{\lambda}^{(k)}) = \hat{\lambda}^{(k)}/\sqrt{1+\hat{\lambda}^{2(k)}}$.

Then the objective function $Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)})$ is

$$\begin{aligned} Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)}) &= -\frac{1}{2}\sum_{i=1}^{n}\Bigg\{\frac{1}{\hat{\sigma}^{2(k)}\big(1-\delta^2(\hat{\lambda}^{(k)})\big)}\Bigg(\hat{s}_i^{(k)} - 2\delta(\hat{\lambda}^{(k)})\hat{\tau}_i^{(k)}\big(\log(-\log y_i) - \hat{\mu}^{(k)}\big) \\ &+ \big(\log(-\log y_i) - \hat{\mu}^{(k)}\big)^2\Bigg) + 2\log\hat{\sigma}^{2(k)} + \log\big(1-\delta^2(\hat{\lambda}^{(k)})\big)\Bigg\}. \end{aligned} \tag{11}$$

The CM-step then conditionally maximizes $Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}^{(k)})$ with respect to θ, resulting a new estimate $\hat{\theta}^{(k+1)}$, as outlined below:

CMQ-step 1: Update $\hat{\mu}^{(k+1)}$ by

$$\hat{\mu}^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\Big(\log(-\log y_i) - \delta(\hat{\lambda}^{(k)})\hat{\tau}_i^{(k)}\Big); \tag{12}$$

CMQ-step 2: Fixing $\mu = \hat{\mu}^{(k+1)}$, we can then update $\hat{\sigma}^{2(k)}$ by

$$\begin{aligned}\hat{\sigma}^{2(k+1)} =& \frac{1}{2n\big(1-\delta^2(\lambda)\big)}\Bigg\{\sum_{i=1}^{n}\hat{s}_i^{(k)} - 2\delta(\lambda)\sum_{i=1}^{n}\big(\log(-\log y_i) - \hat{\mu}^{(k+1)}\big)\hat{\tau}_i^{(k)} \\ &+ \sum_{i=1}^{n}\big(\log(-\log y_i) - \hat{\mu}^{(k+1)}\big)^2\Bigg\}; \end{aligned} \tag{13}$$

CMQ-step 3: Update $\hat{\lambda}^{(k+1)}$ by obtaining the root of the following equation respect to the λ:

$$\begin{aligned} &n\hat{\sigma}^{2(k+1)}\delta(\lambda)\big(1-\delta^2(\lambda)\big) + \big(1+\delta^2(\lambda)\big)\sum_{i=1}^{n}\Big(\log(-\log y_i) - \hat{\mu}^{(k+1)}\Big)\hat{\tau}_i^{(k)} \\ &- \delta(\lambda)\sum_{i=1}^{n}\Big(\hat{s}_i^{(k)} + \big(\log(-\log y_i) - \hat{\mu}^{(k+1)}\big)^2\Big) = 0. \end{aligned} \tag{14}$$

Updating $\hat{\lambda}^{(k+1)}$ requires performing a one-dimensional search to find the root of λ. An alternative strategy involves maximizing the restricted actual log-likelihood function, leading to the next step in the ECME algorithm known as the CML step.

CML-step: Update $\hat{\lambda}^{(k+1)}$ by optimizing the following constrained log-likelihood function:

$$\hat{\lambda}^{(k+1)} = \arg\max_{\lambda}\sum_{i=1}^{n}\log f_{DLSN}\big(y_i; \hat{\mu}^{(k+1)}, \hat{\sigma}^{(k+1)}, \lambda\big).$$

Table 1: Summary results from fitting various models to the Anxiety data.

Parameter	Beta	KU	GBI	McA	SN	DLN	DLSN
$\hat{\mu}(\hat{\alpha})$	6.01	4.64	3.55	10.40	1.16	1.21	1.28
$\hat{\sigma}(\hat{\beta})$	3.28	4.28	13.24	0.55	1.22	1.53	0.45
$\hat{\delta}$	–	–	15.17	41.11	–	–0.87	0.51
$\hat{\nu}$	–	–	–	–	–	–	1.75
$\hat{\ell}$	94.23	97.32	101.11	101.07	99.91	93.42	102.71
AIC	–184.45	–190.64	–196.21	–196.13	–193.81	–182.84	–199.42
BIC	–178.01	–184.19	–186.55	–186.47	-184.15	–176.40	–189.76
K-S statistic	0.079	0.064	0.038	0.034	0.044	0.040	0.029

The iterative nature of the algorithm used above necessitates the repetition of iterations until a convergence criterion is satisfied. This convergence criterion is typically based on the relative change in the log-likelihood function, such as

$$\left\| \frac{\ell(\hat{\boldsymbol{\theta}}^{(k+1)})}{\ell(\hat{\boldsymbol{\theta}}^{(k)})} - 1 \right\|,$$

where $\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f_{DLSN}(y_i; \boldsymbol{\theta})$. To achieve convergence, it is necessary to set a threshold for this relative change that is considered small enough to indicate convergence. In this case, a threshold of 10^{-6} is chosen as the desired level of relative change.

4 Data analysis

The Better Life Index (BLI) data set, measured in the year 2015, is used to demonstrate the usefulness of DLSN distribution. The data set can be found in https://stats.oecd.org/index.aspx?DataSetCode=BLI2015. The BLI data set consists of 11 indicator and 24 variables and it is used to classify the OECD (Organisation for Economic Co-operation and Development) countries as well as Brazil and Russia. Here, we use an indicator that is entitled "Employment rate" as data set.

To facilitate a comprehensive comparison, we fit several statistical distributions to these data sets, including

a) Beta distribution :

$$f(y:\alpha,\beta) = \frac{1}{B(\alpha,\beta)} y^{\alpha-1}(1-y)^{\beta-1}, \quad \alpha>0,\ \beta>0,$$

where $B(\alpha,\beta)$ is the beta function.

b) Kumaraswamy (KU) distribution:

$$f_{KU}(y:\alpha,\beta) = \alpha\beta y^{\alpha-1}(1-y^{\alpha})^{\beta-1}, \quad \alpha>0,\ \beta>0.$$

c) Generalized beta of the first kind (GBI):

$$f_{GBI}(y:\alpha,\beta,\delta) = \frac{\delta}{B(\alpha\delta^{-1},\beta)} y^{\alpha-1}(1-y^{\delta})^{\beta-1}, \quad \alpha>0,\ \beta>0,\ \delta>0.$$

d) McDonald arcsine (McA) distribution :

$$f_{McA}(y:\alpha,\beta,\delta) = \frac{\delta}{B(\alpha\delta^{-1},\beta)\pi\sqrt{y-y^2}} \left[\frac{2}{\pi} arcsin(\sqrt{y})\right]^{\alpha-1} \times \left(1-\left[\frac{2}{\pi} arcsin(\sqrt{y})\right]^{\delta}\right)^{\beta-1}, \quad \alpha>0,\ \beta>0,\ \delta>0.$$

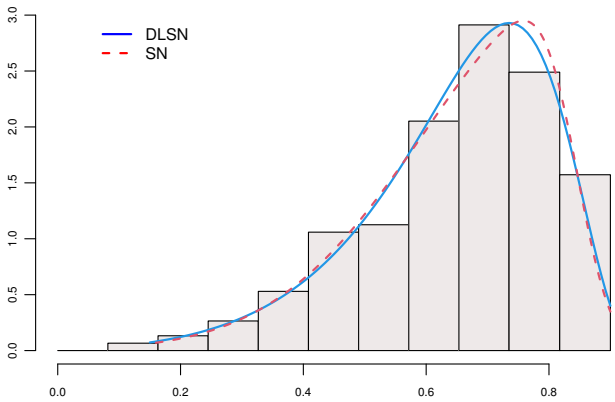


Figure 2: Histogram of the BLI samples. The lines represent the fitted DLSN and SN densities.

e) The SN distribution given in (1).

Furthermore, the reduced DLN distribution is fitted as the sub-models of the proposed DLSN distribution using the EM-algorithm given in Section 3. Given that the competing models vary in the number of parameters they encompass, we employ standard statistical criteria to assess their goodness-of-fit. Specifically, we utilize the Akaike Information Criterion (AIC) (Akaike , 1973) and the Bayesian Information Criterion (BIC) (Schwarz , 1978). Through the model comparison considering the penalized criteria, the Kolmogorov–Smirnov (K-S) test (Smirnov , 1948) is considered to evaluate the goodness of fit of all models.

The summary results of the fitted models are detailed in Table 1. The findings indicate that the proposed DLSN distribution demonstrates a significantly superior fit compared to competing models, as evidenced by its notably lower values of both AIC and BIC. Besides, the smallest K-S test statistic (0.029) suggests that the DLSN model yields the shortest distance between the empirical and parametric distributions of experimental data. To illustrate these findings, Figure 2 presents a histogram of the BLI data overlaid with the fitted DLSN and SN distributions. A closer examination reveals that the DLSN distribution more accurately captures the heavy tail and leptokurtic features inherent in the data when compared to the other models.

Discussion and Results

In this study, we introduced and examined a novel flexible bounded distribution defined on the interval $(0, 1)$, integrating features from the SN distribution and the double-log function. The DLSN distribution we proposed is particularly effective at modeling bounded data with significant skewness, making it a valuable tool for researchers dealing with such datasets. Moreover, we developed an ECME algorithm for estimating the model parameters. To evaluate the model's performance, we conducted an analysis of a real dataset, comparing our findings against several established models in the literature. The results suggest that our proposed model serves as an alternative to traditional bounded distributions, especially for modeling rates and proportions. The methodologies and EM-type algorithms developed in this study have the potential to be adapted for a multivariate version of the DLSN model. We are currently exploring this extension and look forward to sharing our findings in a future publication..

References

Akaike, H. (1998), Information theory and an extension of the maximum likelihood principle, *In Second International Symposium on Information Theory (Eds., B.N. Petrov and F. Csaki)," BNPBF Csaki Budapest: Academiai Kiado.*

Azzalini, A. (1985), A class of distributions which includes the normal ones, *Scandinavian Journal of Statistics*, 171-178.

Lin, T. I., Lee, J. C. and Yen, S. Y. (2007), Finite mixture modelling using the skew normal distribution, *Statistica Sinica*, 909-927.

Schwarz, G. (1978), Estimating the dimension of a model, *The Annals of Statistics*, 461-464.

Smirnov, N. (1948), Table for estimating the goodness of fit of empirical distributions, *The Annals of Mathematical Statistics*, 19(2), 279-281.