Computational Statistics and Data Analysis 115 (2017) 250–266

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

# On hyperbolic transformations to normality

Arthur C. Tsai *, Michelle Liou *, Maria Simak, Philip E. Cheng

*Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan*

## HIGHLIGHTS

- A family of hyperbolic transformations towards normality is proposed/constructed.
- The family is effective in transforming skewed/platykurtic distributions to normal.
- The matching quantile approach is used for initial parameter estimates.
- The new family outperforms other well-known transformations in a simulation.
- Data examples on mathematics test scores and DNA microarrays are illustrated.

## ARTICLE INFO

*Article history:*
Received 19 December 2014
Received in revised form 30 March 2017
Accepted 5 June 2017
Available online 27 June 2017

*Keywords:*
Behrens–Fisher problem
Bimodal distribution
Box–Cox transformation
Levene test
Welch test

## ABSTRACT

In biological and social sciences, it is essential to consider data transformations to normality for detecting structural effects and for better data representation and interpretation. An array of transformations to normality has been derived for data exhibiting skewed, leptokurtic and unimodal shapes, but is less amenable to data exhibiting platykurtic shapes, such as a nearly bimodal distribution. This study proposes and constructs a new family of hyperbolic power transformations for improving normality of raw data with varying degrees of skewness and kurtosis. An advantage this new family has is its effectiveness in transforming platykurtic or bimodal data distributions to normal. A simulation study and a real data example on mathematics achievement test scores are used to illustrate the wide-ranging applications of the proposed family of transformations. As a cautionary note, usefulness and limitations of the proposed method will be discussed for stabilizing the variance of DNA microarray data and for symmetrizing the data distribution towards normality. The empirical applications also illustrate an example of conservative $t$- and ANOVA $F$-tests when the assumption of normality is violated.



## 1. Introduction

In biological and social sciences, researchers can be concerned by the presence of nonnormally distributed variables since commonly employed parametric modeling and analysis methods are derived under the assumption of population normality. Empirical and Monte Carlo studies have provided evidence in support of the robustness of parametric inference in small to large samples under the violation of the normality assumption (Sawilowsky and Blair, 1992; Schmider et al., 2010; Rasch et al., 2011). Nevertheless, this does not preclude a usable alternative, namely, data transformation to normality. There are many valid reasons for utilizing data transformations, including improvement of normality, variance stabilization, and conversion of scales to interval measurement (Osborne, 2002; Liermann et al., 2004; Greenacre, 2009; D'Haese et al., 2011;

* Corresponding authors.
*E-mail addresses:* arthur@stat.sinica.edu.tw (A.C. Tsai), mliou@stat.sinica.edu.tw (M. Liou).

http://dx.doi.org/10.1016/j.csda.2017.06.001

Hou et al., 2011; Pattyn et al., 2011). An array of transformations to normality has been derived from mixing concave and convex functions in order to adjust for both kurtosis and skewness in the data (see Sakia, 1992 for a review of the early literature). In these transformations, a location parameter is defined to adjust for varied skewness in the two tails of a data distribution. The transformations are suitable for data exhibiting skewed, leptokurtic and unimodal shapes as they are similar mixtures of concave and convex functions. It is, however, unclear whether they are practical in application to data exhibiting platykurtic shapes, including the commonly encountered bimodal distribution, which is itself often a mixture of two normal distributions.

This study contributes to the important literature on transformations to normality by introducing a new family of hyperbolic power transformations, hereafter referred to as the HP family. The HP transformation is constructed by implementing a pair of power and scale parameters in a product of hyperbolic functions. Similar to a few existing transformations, two pairs of these parameters are used to adjust for varied shapes of skewness and kurtosis appropriate for general data distributions, but the usual location parameter is not needed. Thus, the HP family incorporates four essential types of transformations in a single formula, which applies concave and convex functions simultaneously to both sides of the sample median.

The paper proceeds as follows. Section 2 introduces the proposed HP family in detail along with a technical review on the Box–Cox transformation (hereafter, the BC family; Box and Cox, 1964) and its extended methods. To find maximum likelihood (ML) parameter estimates of the HP transformation, a method of initial parameter estimation is introduced by matching pairs of selected quantiles in the normalized raw data distribution to the corresponding pairs in the standard normal distribution. This elementary matching quantile approach may facilitate the search for the ML estimates, and often secure compatible ML estimation of the HP parameters. Section 3 provides a simulation to compare the performance of the HP transformation with the BC, gpower, modulus and *sinh–arcsinh* transformations for a range of nonnormal distributions, including the beta, Cauchy, gamma, Laplace, lognormal, Weibull, uniform, and bimodal distributions. The evaluation criteria are the skewness and kurtosis of the transformed distributions along with the results on testing the null hypothesis of normality. Section 4 contains an empirical example on mathematics achievement test scores to demonstrate that a nearly bimodal distribution can be transformed into a normal distribution with the HP transformation in the context of a conservative two-sample $t$-test and nonrobust ANOVA $F$-test under bimodality. Section 5 presents the usefulness and limitations of the HP family for stabilizing the variance of DNA microarray data as well as for symmetrizing the data distribution towards normality. Finally, a discussion is offered on further research and applications pertinent to the HP family.

## 2. The hyperbolic power transformation

The BC family of power transformations is defined on the positive real line ($x > 0$) as

$$\psi^{BC}(x, \lambda) = \begin{cases} (x^{\lambda} - 1)/\lambda & \lambda \neq 0, \\ \log(x) & \lambda = 0, \end{cases}$$

where $\lambda$ denotes the transformation power parameter. This family transforms skewed data distributions towards normality, and is defined on the positive side of the real line, as depicted in Fig. 1(a). Flexible transformations without such a domain restriction include the exponential transformations by Manly (1976; see Fig. 1(b)), and the extended power transformations by Yeo and Johnson (2000; see Fig. 1(c)). These revised transformations are monotonically concave or convex, making them particularly suitable for data exhibiting a skewed or unimodal shape, but inapplicable to data with platykurtic, leptokurtic, or bimodal shapes.

Useful transformations have also been derived through mixing concave and convex functions in order to incorporate adjustments of kurtosis in the data distribution. Some examples of these families include the signed power transformation (Fig. 1(d); Bickel and Doksum, 1981), inverse hyperbolic sine transformation (Fig. 1(e); Burbidge et al., 1988), and modulus transformation (Fig. 1(f); John and Draper, 1980). Recently, the *sinh–arcsinh* transformation (Fig. 1(g); Jones and Pewsey, 2009) and the gpower transformation (Fig. 1(h); Kelmansky et al., 2013) were proposed to treat data with a peaked sample mode and with heavier or lighter tails than the normal distribution.

As noted before, the BC power transformation has been generalized in the literature to include a variety of concave and convex functions in order to adjust for both kurtosis and skewness in data. An analogous treatment of the kurtosis of raw data distribution is the hyperbolic tangent function $\psi(x) = \tanh(x)$, which has been used as a transfer function in 'infomax' algorithms for characterizing a source density with specified kurtosis (Bell and Sejnowski, 1995; Lee et al., 1999; Hyvärinen et al., 2001). The hyperbolic tangent function is an essential mathematical tool for describing the rate of action potential firing in a neural cell, which is often applied to simulate the dynamic process of input current intensity. However, the hyperbolic tangent function ignores the treatment of data skewness. This important pitfall is the motivation behind this study, which explores the potential utility of treating both kurtosis and skewness based on the hyperbolic tangent function appropriate for general applications.

Without loss of generality, assume that the median of the raw data is located at $x = 0$. The HP transformation is defined as

$$\psi(x, \theta) = \alpha \sinh(\beta x)\mathrm{sech}^{\lambda}(\beta x)/\beta, \tag{1}$$

where $\theta = \{\alpha, \beta, \lambda\}$, $\alpha, \beta > 0$, and $\lambda \leq 1.0$. Thus, a power parameter $\lambda$, a scale parameter $\beta$, and a slope parameter $\alpha$ (dependent on the other parameters) are implemented in a product of two hyperbolic functions to yield the HP family in