Journal of the Korean Data & Information Science Society
2011, **22**(5), 967–976

한국데이터정보과학회지

A note on Box-Cox transformation and application in microarray data

Mezbahur Rahman[1] · Namyong Lee[2]

[12]Department of Mathematics and Statistics, Minnesota State University

Received 24 April 2011, revised 23 June 2011, accepted 21 August 2011

Abstract

The Box-Cox transformation is a well known family of power transformations that brings a set of data into agreement with the normality assumption of the residuals and hence the response variable of a postulated model in regression analysis. Normalization (studentization) of the regressors is a common practice in analyzing microarray data. Here, we implement Box-Cox transformation in normalizing regressors in microarray data. Pridictabilty of the model can be improved using data transformation compared to studentization.

Keywords: Maximum likelihood estimates, moments for the ordered standard normal variates, normality tests, Shapiro-Wilk W statistic.

1. Scientific background: Colon cancer

Colon cancer is one of the most common types of cancer with over 90,000 people diagnosed every year in USA. The cause of the onset of colon cancer is unknown although tumors are known to develop from polyps, which are extra tissue growths. Polyps maybe present in the colon for years prior to the evolution into cancer. Identification (via screening) and removal of polyps can prevent the onset of colon cancer although small polyps present in the colon do not always cause individuals any problems. Symptoms of large polyps and colon cancer include bleeding from the lower GI (gastrointestinal) tract and changes in bowel habits.

2. Discovery and prediction of tumors using microarry analysis

Current microarray technology is able to take a single tissue sample to construct an Affymetrix oglionucleotide array containing (estimated) expression levels of thousands of different genes for that tissue. General objective is to develop a more systematic approach to cancer classification based on the simulataneous expression monitoring of thousands of genes using Affymetrix oglionucleotide microarrays.

[1] Corresponding author: Professor, Department of Mathematics and Statistics, Minnesota State University, Mankato, MN 56001, USA. E-mail: mezbahur.rahman@mnsu.edu.
[2] Professor, Department of Mathematics and Statistics, Minnesota State University, Mankato, MN 56001, USA.

Gene expression patterns for cancer and normal tissues maybe substantially different, so microarry analysis could be an effective diagnostic tool based on genes. There may also be differences within cancerous samples allowing patient prognosis according to the type, stage, or location of the colon cancer.

3. Normalization

Microarray data have a large number of predictors compared to the sample sizes. A normalization approach is implemented for computational conveniences and to bring the data in compliance with usual regression analysis assumptions as much as possible. Often, in microarray data analysis, studentization ($Z = (X - \bar{X})/S$, where $\bar{X}$ is the sample mean and S is the sample standard deviation) is used (see, Yang and Throne, 2003) to normalize the independent variables (regressors) prior to fitting the model. Another method of 'normalization' is orthogonalization. Use of *log-ratios* are also termed as 'normalization' in some instances. Giles and Kipling (2003) used Box-Cox power transformation for a restricted parameter space between -3 and +3. Ekstrom *et al.* (2004) used a modified Box-Cox power transformation. Here, we consider Box-Cox power transformation in normalizing the data for unrestricted parameter space and parameter estimation is done using several competitive methods. It is to be noted that in all the referenced papers above and in this study even though we are in multiple regression situation, transformations are considered for individual regressors independently.

The remaining of the paper is arranged as follows: In Section 4 we introduce the Box-Cox transformation for a response variable in a linear regression with one response variable followed by three competing methods of transformation parameter estimation, iterative Newton-Raphson method (Section 4.1), adaptive Newton-Raphson method (Section 4.2), and Shapiro-Wilk maximization of W statistic method. In Section 5 we present a short simulation study to show the performances of the competitive estimates of the transformation parameter. In Section 6 we implement all the transformation methods discussed in Section 3 and in Section 4 for the regressors individually for a microarray data. In Section 7 we make some observations about normalization of a microarray data in light of the transformation methods used in this study which is the key focus of this paper.

4. Box-Cox transformation

In regression analysis, often the key assumption regarding normality of the error variable and hence the response variable are violated. The commonly used remedy is the Box-Cox family of power transformations (Box and Cox, 1964). The process is to select a parameter in the Box-Cox transformation which maximizes the normal likelihood using the data at hand and then apply regression analysis on the transformed response variable. There is no role of the estimates of the location and the scale parameters which were derived in the process of estimating the power transformation parameter in regression analysis. In practice, the regression model parameters are usually estimated separately after the necessary Box-Cox power transformation parameter is selected.

In literature, the estimation procedures of the Box-Cox power transformation parameter are considered by many authors. The notable ones are the normal likelihood method of Box and Cox (1964), the robustified version of the normal likelihood method of Carroll (1980) and

of Bickel and Doksum (1981), the transformation to symmetry method of Hinkley (1975), the quick estimate of Hinkley (1977) and of Taylor (1985). Lin and Vonesh (1989) constructed a nonlinear regression model which is used to estimate the transformation parameter such that the normal probability plot of the data on the transformed scale is as close to linearity as possible. Following the footsteps of Box and Cox (1982) and Lin and Vonesh (1989), Halawa (1996) considered the power transformation parameter estimation procedure using an artificial regression model which gives the estimates with very small variabilities compared to the normal likelihood procedure. Halawa (1996) conducted an exhaustive comparative study with normal likelihood procedure. In that study, he also considered estimation procedures of the location and the scale parameters in the likelihood.

Rahman (1999) introduced a method of estimating the Box-Cox power transformation parameter using maximization of the Shapiro-Wilk W (Shapiro and Wilk, 1965) statistic along with a comparison study of the normal likelihood method (Carroll, 1980), and of the artificial regression model method (Halawa, 1996). In this paper the estimation procedure for the Box-Cox power transformation parameter is considered using maximization of the normal likelihood along with the Newton-Raphson root finding method and maximization of the Shapiro-Wilk W statistic for the microarray data. Some simulation results are also presented to show the performances of the estimates.

Let $Y_1, Y_2, \cdots, Y_n$ be a random sample of size n from a population whose functional form is unknown. Box and Cox (1964) suggested that if the transformation

$$X = \begin{cases} \frac{Y^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \ln(Y), & \lambda = 0 \end{cases} \tag{4.1}$$

is performed on the data then X will have an approximate normal distribution with mean μ and variance σ^2. In equation (4.1), λ is unknown and considered as the Box-Cox power transformation parameter and 'ln' represents the natural logarithm.

4.1. Iterative Newton-Raphson method

After applying the transformation mentioned in equation (4.1), the density function of the data can be written as

$$f(y;\lambda,\mu,\sigma^2) \doteq \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}\left(\frac{y^\lambda - 1}{\lambda} - \mu\right)^2} \cdot y^{\lambda-1}, & \lambda \neq 0, \\ \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(\ln(y)-\mu)^2} \cdot \frac{1}{y}, & \lambda = 0. \end{cases} \tag{4.2}$$

The log-likelihood function $\ell_I = \ell(\lambda, \mu, \sigma^2; y_1, y_2, \cdots, y_n)$, where the subscript 'I' stands for iterative, can be written as

$$\ell_I = \begin{cases} -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n \left(\frac{y_i^\lambda - 1}{\lambda} - \mu\right)^2 + (\lambda - 1)\sum_{i=1}^n \ln(y_i), & \lambda \neq 0, \\ -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (\ln(y_i) - \mu)^2 - \sum_{i=1}^n \ln(y_i), & \lambda = 0. \end{cases} \tag{4.3}$$

Equation (4.3) is maximized when the partial derivatives of (4.3) with respect to μ, σ^2 and λ are equated to zero and the solution to the corresponding system is found. This leads to solving the following three equations simultaneously,

$$\frac{\partial \ell_I}{\partial \mu}\Big|_{\mu=\hat{\mu}_I, \sigma^2=\hat{\sigma^2}_I, \lambda=\hat{\lambda}_I} = 0 \Longrightarrow \hat{\mu}_I = \frac{1}{n}\sum_{i=1}^n \frac{y_i^{\hat{\lambda}_I} - 1}{\hat{\lambda}_I}, \tag{4.4}$$

$$\frac{\partial \ell_I}{\partial \sigma^2}\Big|_{\mu=\hat{\mu}_I,\sigma^2=\hat{\sigma^2}_I,\lambda=\hat{\lambda}_I} = 0 \Longrightarrow \hat{\sigma^2}_I = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i^{\hat{\lambda}_I}-1}{\hat{\lambda}_I} - \hat{\mu}_I\right)^2 \tag{4.5}$$

and

$$\frac{\partial \ell_I}{\partial \lambda}\Big|_{\mu=\hat{\mu}_I,\sigma^2=\hat{\sigma^2}_I,\lambda=\hat{\lambda}_I} = 0$$

$$\Longrightarrow g(\hat{\lambda}_I) = -\frac{1}{\hat{\sigma^2}_I}\sum_{i=1}^{n}\left(\frac{y_i^{\hat{\lambda}_I}-1}{\hat{\lambda}_I} - \hat{\mu}_I\right)\frac{\hat{\lambda}_I(\ln y_i)y_i^{\hat{\lambda}_I} - y_i^{\hat{\lambda}_I}+1}{\hat{\lambda}_I^2} + \sum_{i=1}^{n}\ln y_i = 0. \tag{4.6}$$

The computation algorithm starts with an initial value of $\hat{\lambda}_I$ (an obvious choice is 1), compute $\hat{\mu}$ using equation (4.4), compute $\hat{\sigma^2}_I$ using equation (4.5) and then solve equation (4.6) using the Newton-Raphson iterations

$$\hat{\lambda}_I^{(t+1)} = \hat{\lambda}_I^{(t)} - \frac{g(\hat{\lambda}_I^{(t)})}{g'(\hat{\lambda}_I^{(t)})}, \tag{4.7}$$

where $t = 0, 1, 2, \ldots$ are the iteration steps,

$$g'(\hat{\lambda}_I) = -\frac{1}{\hat{\sigma^2}_I}\left[\frac{\sum_{i=1}^{n}\left(\hat{\lambda}_I(\ln y_i)y_i^{\hat{\lambda}_I} - y_i^{\hat{\lambda}_I}+1\right)^2}{\hat{\lambda}_I^4}\right.$$

$$\left.+\sum_{i=1}^{n}\left(\frac{y_i^{\hat{\lambda}_I}-1}{\hat{\lambda}_I} - \hat{\mu}_I\right)\frac{\left(\hat{\lambda}_I^3(\ln y_i)^2 y_i^{\hat{\lambda}_I} - 2\hat{\lambda}_I^2(\ln y_i)y_i^{\hat{\lambda}_I} + 2\hat{\lambda}_I y_i^{\hat{\lambda}_I} - 2\hat{\lambda}_I\right)}{\hat{\lambda}_I^4}\right] \tag{4.8}$$

is the derivative of $g(\hat{\lambda}_I)$. Note that equations (4.4) and (4.5) are also updated in each Newton-Raphson iteration.

It is to be noted that this algorithm is not robust against the initial value and hence the following modification is used. For details, the readers are referred to Rahman and Pearson (2008).

4.2. Adaptive Newton-Raphson method

For a fixed λ, the log-likelihood function $\ell_A = \ell(\mu, \sigma^2; y_1, y_2, \cdots, y_n)$, where the subscript "A" stands for adaptive, can be written as

$$\ell_A = \begin{cases} -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\frac{y_i^{\lambda}-1}{\lambda} - \mu\right)^2 + (\lambda-1)\sum_{i=1}^{n}\ln(y_i), & \lambda \neq 0, \\ -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(\ln(y_i) - \mu\right)^2 - \sum_{i=1}^{n}\ln(y_i), & \lambda = 0. \end{cases} \tag{4.9}$$

Equation (4.9) is maximized when the partial derivatives of (4.9) with respect to μ and σ^2 are equated to zero and the solution to the corresponding system is found. This leads to solving the following two equations,

$$\frac{\partial \ell_A}{\partial \mu}\Big|_{\mu=\hat{\mu}_A,\sigma^2=\hat{\sigma^2}_A} = 0 \Longrightarrow \hat{\mu}_A = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i^{\lambda}-1}{\lambda} \tag{4.10}$$

and

$$\frac{\partial \ell_A}{\partial \sigma^2}\Big|_{\mu=\hat{\mu}_A,\sigma^2=\hat{\sigma^2}_A}=0\Longrightarrow \hat{\sigma^2}_A=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i^{\lambda}-1}{\lambda}-\frac{1}{n}\sum_{j=1}^{n}\left(\frac{y_j^{\lambda}-1}{\lambda}\right)\right)^2. \tag{4.11}$$

Then the pseudo log-likelihood $\ell_A^*=\ell(\lambda;y_1,y_2,\cdots,y_n)$ can be written as

$$\ell_A^*=-\frac{n}{2}\left[\ln(2\pi)+\ln\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i^{\lambda}-1}{\lambda}-\frac{1}{n}\sum_{j=1}^{n}\left(\frac{y_j^{\lambda}-1}{\lambda}\right)\right)^2\right\}+1\right]$$
$$+(\lambda-1)\sum_{i=1}^{n}\ln y_i. \tag{4.12}$$

The steps for maximizing ℓ_A^* using the Newton-Raphson method are as follows:

$$\frac{\partial \ell_A^*}{\partial \lambda}\Big|_{\lambda=\hat{\lambda}_A}=0\Longrightarrow h(\hat{\lambda}_A)=-n\frac{\sum_{i=1}^{n}\left(P_i-\bar{P}\right)\left(Q_i-\bar{Q}\right)}{\sum_{i=1}^{n}\left(P_i-\bar{P}\right)^2}+\sum_{i=1}^{n}\ln y_i=0 \tag{4.13}$$

and

$$\hat{\lambda}_A^{(t+1)}=\hat{\lambda}_A^{(t)}-\frac{h(\hat{\lambda}_A^{(t)})}{h'(\hat{\lambda}_A^{(t)})}, \tag{4.14}$$

where

$$h'(\hat{\lambda}_A)=-n\left[\frac{\sum_{i=1}^{n}\left(P_i-\bar{P}\right)\left(R_i-\bar{R}\right)+\sum_{i=1}^{n}\left(Q_i-\bar{Q}\right)^2}{\sum_{i=1}^{n}\left(P_i-\bar{P}\right)^2}\right.$$
$$\left.-2\frac{\left\{\sum_{i=1}^{n}\left(P_i-\bar{P}\right)\left(Q_i-\bar{Q}\right)\right\}^2}{\left\{\sum_{i=1}^{n}\left(P_i-\bar{P}\right)^2\right\}^2}\right],$$

$$P_i=\left(\frac{y_i^{\hat{\lambda}_A}-1}{\hat{\lambda}_A}\right),\ \bar{P}=\frac{1}{n}\sum_{j=1}^{n}\left(\frac{y_j^{\hat{\lambda}_A}-1}{\hat{\lambda}_A}\right),\ Q_i=\left(\frac{\hat{\lambda}_A(\ln y_i)y_i^{\hat{\lambda}_A}-y_i^{\hat{\lambda}_A}+1}{\hat{\lambda}_A^2}\right),$$

$\bar{Q}=\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{\lambda}_A(\ln y_j)y_j^{\hat{\lambda}_A}-y_j^{\hat{\lambda}_A}+1}{\hat{\lambda}_A^2}\right),\ R_i=\left(\frac{\hat{\lambda}_A^3(\ln y_i)^2y_i^{\hat{\lambda}_A}-2\hat{\lambda}_A^2(\ln y_i)y_i^{\hat{\lambda}_A}+2\hat{\lambda}_Ay_i^{\hat{\lambda}_A}-2\hat{\lambda}_A}{\hat{\lambda}_A^4}\right),$

and

$$\bar{R}=\frac{1}{n}\sum_{j=1}^{n}\left(\frac{\hat{\lambda}_A^3(\ln y_j)^2y_j^{\hat{\lambda}_A}-2\hat{\lambda}_A^2(\ln y_j)y_j^{\hat{\lambda}_A}+2\hat{\lambda}_Ay_j^{\hat{\lambda}_A}-2\hat{\lambda}_A}{\hat{\lambda}_A^4}\right).$$

The computation algorithm starts with an initial value of $\hat{\lambda}_A$ (an obvious choice is 1) iteratively using (4.14) and then $\hat{\mu}_A$ and $\hat{\sigma^2}_A$ are obtained using (4.10) and (4.11) by substituting $\hat{\lambda}_A$ for λ, if desired. Since this algorithm is robust against initial value (Rahman and Pearson, 2008) is used for this study.

4.3. Shapiro-Wilk W statistic

The Shapiro-Wilk W test statistic (Shapiro and Wilk, 1965) is obtained by dividing the square of an appropriate linear combination of the sample order statistics by the usual symmetric estimate of the variance.

Let $(X_1, X_2, \cdots, X_n)$ be a random sample to be tested for normality, ordered $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$. Define

$$W = \frac{\left(\sum_{i=1}^{n} a_i X_{(i)}\right)^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

where the vector $\mathbf{a} = \frac{\mathbf{m}' V^{-1}}{(\mathbf{m}'\mathbf{V}^{-1}\mathbf{V}^{-1}\mathbf{m})^{1/2}}$, $\mathbf{m}$ is the vector of the expected values and $\mathbf{V}$ is the variance covariance matrix of the standard normal order statistics.

The value of W is closer to 1 means that the data is closer to normality and the maximum value of W is 1. The Shapiro and Wilk (1965) W statistic has been shown to yield a powerful test of normality for a variety of nonnormal distributions (Pearson *et al.*, 1977; Shapiro *et al.*, 1968).

The values of the a_i's are tabulated in Shapiro and Wilk (1965) for $n = 2, 3, \ldots, 50$. For other sample sizes, the a_i's can be estimated using the following suggested approximations:

$$\hat{a}_i^* = 2m_i, \quad i = 2, 3, \cdots, n-1, \quad \text{and}$$

$$\hat{a}_1^2 = \hat{a}_n^2 = \begin{cases} \frac{\Gamma\left(\frac{1}{2}n\right)}{\sqrt{2}\Gamma\left(\frac{1}{2}(n+1)\right)}, & n \leq 20, \\ \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\sqrt{2}\Gamma\left(\frac{1}{2}n+1\right)}, & n > 20, \end{cases}$$

then a_i^* for $i = 2, 3, \ldots, n-1$ are normalized by dividing by $C = \sqrt{-2.722 + 4.0832n}$ as suggested by Shapiro and Wilk (1965).

The values of the m_i's are tabulated in Harter (1961) for n = 2(1)100, 125(25)250, 300, 350, and 400. More accurate values of the $\mathbf{m}_i$'s and $\mathbf{V}$ are also given in Parish (1992a, 1992b) for n= 2(1)50, using Legendre Polynomial. Rahman and Pearson (2000) showed that Monte-Carlo simulation can also be effectively used to generate the coefficients (a_i's).

5. Simulation study

This section presents simulation results in which there exist values of the transformation parameter that satisfy model (4.2) under transformation (4.1). The choices of λ values, in combination with certain choices of the other model parameters are used to generate data as

$$Y = (1 + \lambda(\mu + \sigma\epsilon))^{\frac{1}{\lambda}}$$

where ϵ is a pseudo $N(0, 1)$ random vector and $\lambda \neq 0$. The choices of parameters are made in such a way that the Y vector is always positive, even in cases when $\lambda \neq 0$.

A non-symmetric distribution such as shifted exponential distribution

$$f(y; \alpha, \theta) = \frac{1}{\theta} e^{-(y-\alpha)/\theta}, \quad y > \alpha$$

is also considered in generating random samples.

The simulation study was performed using MATLAB software. The MATLAB codes are available from the authors. For each parameter configuration in Table 5.1, 1000 samples were generated.

In Tables 5.1, $\hat{\lambda}_L$ indicates the maximum likelihood estimate of λ described in Section 4.2, $\hat{\lambda}_W$ indicates the maximization of the Shapiro-Wilk W statistic estimate of λ described in Section 4.3. W_0 indicates the Shapiro-Wilk W statistic for the data at hand, $W_{\hat{\lambda}_L}$ indicates the Shapiro-Wilk W statistic for the transformed data using the maximum likelihood method, $W_{\hat{\lambda}_W}$ indicates the Shapiro-Wilk W statistic for the transformed data using the maximization of the Shapiro-Wilk W statistic method, and W_{LN} indicates the Shapiro-Wilk W statistic for the transformed data using the logarithmic ('ln') transformation. Then the means ('mean') and the standard deviations ('st.error') are displayed.

For a large sample, in computing the Shapiro-Wilk W statistic, Rahman and Pearson (2000) simulation coefficients are also used to show the performance compared to using more exact readily available coefficients.

In Table 5.1, for normal samples when λ is known, biases are smaller for $\hat{\lambda}_W$ compared to $\hat{\lambda}_L$. Standard errors for $\hat{\lambda}_L$ are smaller except for sample size 40 and when simulation coefficients are used in computing $\hat{\lambda}_W$. W_{LN} values are the lowest for normal samples, that is, a log transformation for a normal sample makes it worse. For exponential samples, W_0 is the lowest as expected. $W_{\hat{\lambda}_W}$ values are slightly higher compared to $W_{\hat{\lambda}_L}$. In all cases W_{LN} values are lower compared to $W_{\hat{\lambda}_L}$ and $W_{\hat{\lambda}_W}$. W values are very close for exact and simulated coefficients. In Table 5.2, for exponential samples, similar properties are pertained as in normal samples.

Table 5.1 Normal samples

	$\hat{\lambda}_L$	$\hat{\lambda}_W$	W_0	$W_{\hat{\lambda}_L}$	$W_{\hat{\lambda}_W}$	W_{LN}
		$n = 20 \quad N(\mu = 3, \sigma^2 = 4) \quad \lambda = 2.0$				
mean	1.3387	1.6366	0.9391	0.9551	0.9584	0.8786
st.error	0.5741	0.7498	0.0307	0.0230	0.0202	0.0538
		$n = 40 \quad N(\mu = 5, \sigma^2 = 4) \quad \lambda = 2.0$				
mean	1.7537	1.8758	0.9545	0.9787	0.9791	0.8871
st.error	0.6588	0.7593	0.0278	0.0096	0.0092	0.0730
	$n = 40$	$N(\mu = 5, \sigma^2 = 4)$	$\lambda = 2.0$	Simulated Coefficients		
mean	1.7069	2.1045	0.9223	0.9610	0.9636	0.8539
st.error	0.6413	0.6308	0.0272	0.0119	0.0107	0.0705
	$n = 100$	$N(\mu = 5, \sigma^2 = 4)$	$\lambda = 2.0$	Simulated Coefficients		
mean	1.8521	1.9488	0.9663	0.9903	0.9906	0.8800
st.error	0.3360	0.3864	0.0179	0.0045	0.0045	0.0521

6. Microarray data

The Colon Cancer Data set (Alon *et al.*, 1999) consists of 62 tissue samples, 40 of which are taken from colon tumors and 22 are normal samples. For each of the samples an Affymetrix oglinucleotide array consisting of 6817 Expression Sequence Tags was constructed and our data set contains the 2000 ESTs with the highest minimal intensity across the 62 tissue samples. Since each of the 2000 ESTs can be mapped 1-1 to a gene each EST can be referred to as a gene. The data is already summarized with one intensity (formed from 20 probe pairs) for each gene and each sample.

Table 5.2 Exponential samples

	$\hat{\lambda}_L$	$\hat{\lambda}_W$	W_0	$W_{\hat{\lambda}_L}$	$W_{\hat{\lambda}_W}$	W_{LN}
	$n = 20$ $EXP(\theta = 3) + 3$					
mean	-0.7425	-0.8298	0.8273	0.9486	0.9499	0.9166
st.error	0.4876	0.6712	0.0815	0.0225	0.0217	0.0370
	$n = 40$ $EXP(\theta = 3) + 3$					
mean	-0.7884	-0.8342	0.8208	0.9588	0.9593	0.9287
st.error	0.3101	0.4186	0.0676	0.0152	0.0148	0.0268
	$n = 40$ $EXP(\theta = 3) + 3$ Simulated Coefficients					
mean	-0.7884	-0.8860	0.8178	0.9687	0.9694	0.9341
st.error	0.3101	0.4130	0.0732	0.0134	0.0131	0.0283
	$n = 100$ $EXP(\theta = 3) + 3$ Simulated Coefficients					
mean	-0.7900	-0.8218	0.8175	0.9726	0.9728	0.9431
st.error	0.1907	0.2497	0.0544	0.0090	0.0089	0.0173

To give an impression of the data, in Table 6.2 the responses are displayeed and in Table 6.3 one of the gene sequence (predictor) is presented.

In Table 6.1, the quantile distributions of the W statistics for the transformed ($W_{\hat{\lambda}_L}$, $W_{\hat{\lambda}_W}$, W_{LN}) and non-transformed (W_0) along with simulated W values when the distribution is normal are presented. W_0 values are significantly lower compared to the simulated W values. Among the transformations, W_{LN} values are lower.

Multiple linear regression models are fitted using generalized inverses as the number of parameters are far greater than the number of data points. Then the leave-one-out approach is used to determine the rates of correct predictions and are displayed in Table 6.4. It is to be noted that in all model fittings, MSE's were very small and R^2's were very close to 1.

In Table 6.4, rates of correct predictions are displayed. X indicates without transformation, Z_X indicates studentized, $Z_{ln(X)}$ indicates studentized after log transformation, and so on. For this data, log transformation and maximum likelihood Box-Cox transformation yielded the highest rates of correct predictions. Studentizations yielded the lowest rates of correct predictions.

Table 6.1 W quantiles: Microarray data

p	Simulation	W_0	$W_{\hat{\lambda}_L}$	$W_{\hat{\lambda}_W}$	W_{LN}
0.01	0.9439	0.5001	0.9564	0.9560	0.8878
0.02	0.9521	0.6052	0.9633	0.9635	0.9028
0.05	0.9616	0.6908	0.9714	0.9715	0.9274
0.10	0.9684	0.7457	0.9763	0.9765	0.9459
0.50	0.9837	0.8821	0.9876	0.9876	0.9778
0.90	0.9913	0.9517	0.9929	0.9929	0.9893
0.95	0.9927	0.9627	0.9939	0.9939	0.9910
0.98	0.9939	0.9713	0.9948	0.9948	0.9924
0.99	0.9946	0.9762	0.9954	0.9953	0.9935

7. Concluding remarks

In applying maximum likelihood method in estimating Box-Cox power transformation parameter the adaptive approach mentioned above should be used. Maximization of the

Table 6.2 Y: Response score, negative sign indicates cancer diagnosis

-1	1	-2	2	-3	3	-4	4	-5	5	-6	6	-7	7
-8	8	-9	9	-10	10	-11	11	-12	12	-13	-14	-15	-16
-17	-18	-19	-20	-21	-22	-23	-24	-25	-26	27	-27	-28	28
29	-29	-30	-31	-32	32	-33	33	34	-34	-35	35	36	-36
-37	-38	-39	39	-40	40								

Table 6.3 X: An example of gene measurement

8589.4163	9164.2537	3825.7050	6246.4487	3230.3287	2510.3250
7126.5988	4028.7100	9330.6787	5271.5175	14876.4070	14173.0540
4469.0900	4985.2188	4913.7988	5627.2512	7144.4062	4865.2200
5382.3938	4412.4763	7434.8213	6995.4100	4214.9000	1914.6775
8865.4587	5934.8888	5821.6175	9767.0275	13324.7290	12977.7120
8753.2388	5012.0200	6904.8012	8347.9838	5100.5363	4554.5762
5466.9300	4201.5075	3656.7837	9128.1188	3799.0888	6194.2450
11447.6310	14641.9310	9443.8775	14368.2760	6951.3538	6302.7763
6357.9212	7121.1737	6870.3225	11605.9720	7666.6750	4177.2838
5777.1738	4527.0038	4653.2375	4972.1662	9112.3725	6730.6250
6234.6225	7472.0100				

Table 6.4 Rate of correct prediction

	X	Z_X	$ln(X)$	$Z_{ln(X)}$	$ln(ln(X))$	$Z_{ln(ln(X))}$	$X_{\hat{\lambda}_L}$	$X_{\hat{\lambda}_W}$
p	0.6774	0.4516	0.7097	0.4355	0.6774	0.4355	0.7097	0.6935

Shapiro-Wilk W statistic can also be used effectively to estimate the Box-Cox transformation parameter. And both the methods can be used for unrestricted parameter space without difficulty and ensuring highest possible Shapiro-Wilk W statistic value.

In this study, at least for the data is used, studentization should be avoided when the rates of correct predictions are the objectives.

Without checking for normality, Box-Cox transformation can do no harm while log transformation might.

References

Alon, U., Barkai, N., Notterman, D. A., Gish, K., Ybarra, S., Mack, D. and Levine, A. J. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proceedings of the National Academy of Sciences USA,* **96**, 6745-6750.

Bickel, P. J. and Doksum, K. A. (1981). An analysis of transformations revisited. *Journal of the American Statistical Association,* **76**, 296-311.

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society B,* **26**, 211-252.

Box, G. E. P. and Cox, D. R. (1982). An analysis of transformations revisited (rebutted). *Journal of the American Statistical Association,* **77**, 209-210.

Carroll, R. J. (1980). A robust method for testing transformations to achieve approximate normality. *Journal of the Royal Statistical Society B,* **42**, 71-78.

Ekstrom, C. T., Bak, S., Kristensen, C. and Rudemo, M. (2004). Spot shape modelling and data transformations for microarrays. *Bioinformatics,* **20**, 2270-2278.

Giles, P. J. and Kipling, D. (2003). Normality of oligonucleotide microarray data and implications for parametric statistical analyses. *Bioinformatics,* **19**, 2254-2262.

Halawa, A. M. (1996). Estimating the Box-Cox transformation via an artificial regression model. *Communications in Statistics — Simulation and Computation,* **25**, 331-350.

Harter, H. L. (1961). Expected values of normal order statistics. *Biometrika,* **48**, 151-165.

Hinkley, D. V. (1975). On power transformation to symmetry. *Biometrika,* **62**, 101-111.

Hinkley, D. V. (1977). On quick choice of power transformation. *Applied Statistics,* **26**, 67-68.

Lin, L. I. and Vonesh, E. F. (1989). An empirical nonlinear data-fitting approach for transforming data to normality. *American Statistician,* **43**, 237-243.

Parish, R. S. (1992a). Computing expected values of normal order statistics. *Communications in Statistics - Simulation and Computation,* **21**, 57-70.

Parish, R. S. (1992b). Computing variances and covariances of normal order statistics. *Communications in Statistics - Simulation and Computation,* **21**, 71-101.

Pearson, E. S., D'Agostino, R. B. and Bowman, K. O. (1977). Tests for departure from normality: Comparison of powers. *Biometrika,* **64**, 231-246.

Rahman, M. (1999). Estimating the Box-Cox transformation via Shapiro-Wilk W statistic. *Communications in Statistics - Simulation and Computation,* **28**, 223-241.

Rahman, M. and Pearson, L. M. (2000). Shapiro-Francia W' statistic using exclusive simulation. *Journal of the Korean Data & Information Sciences Society*, **11**, 139-155.

Rahman, M. and Pearson, L. M. (2008). A note on the maximum likelihood Box-Cox transformation parameter. *Journal of Probability and Statistical Science*, **6**, 155-168.

Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality. *Biometrika,* **52**, 591-611.

Shapiro, S. S., Wilk, M. B. and Chen, H. J. (1968). A comparative study of various tests of normality. *Journal of the American Statistical Association,* **63**, 1343-1372.

Taylor, J. M. G. (1985). Power transformations to symmetry. *Annals of Mathematical Statistics,* **33**, 1-67.

Yang, Y. H. and Throne, N. P. (2003). Normalization for two-color cDNA microarray data. *Institute of Mathematical Statistics,* **40**, 403-418.