COMBINING THE BOX-COX POWER AND GENERALISED LOG TRANSFORMATIONS TO ACCOMMODATE NONPOSITIVE RESPONSES IN LINEAR AND MIXED-EFFECTS LINEAR MODELS

D. M. Hawkins [1]
University of Minnesota
e-mail: *dhawkins@umn.edu*

S. Weisberg
University of Minnesota
e-mail: *sandy@umn.edu*

Key words: Homoscedasticity, Multivariate analysis, Normality

Abstract: Transformation of a response variable can greatly expand the class of problems for which the linear regression model or linear mixed-model is appropriate. Beginning with the fundamental work of Box and Cox, maximum-likelihood-like estimation has been applied to select a transformation from among a family of transformations, with the possible goals of achieving approximate normality, removing nonlinearity in a mean function, or stabilizing variance. The Box-Cox power family (BC) of transformations is by far the most common with the Box-Cox methodology, and it requires a strictly positive response. In this article we introduce a new family of transformations that we call the Box-Cox power with nonpositives (BCN) family that allows inclusion of a few nonpositive values. The BCN family is a modification of the basic power family that is inspired by the generalised log, or glog transformation, proposed for use with the more limited goals of stabilizing variance or achieving approximate normality. The glog transformation is itself a special case of the Johnson S_U transformation, and we show that the BCN family derived from it is in turn a simple modification of the BC family. Computer code for implementing this family is included in the `car` package in R (Fox and Weisberg, 2011). The methodology is illustrated using a problem in clinical chemistry.

1. Introduction

Based on the fundamental work of Box and Cox (1964), allowing transformation of a response variable or variables greatly expands the problems for which the linear model can be an appropriate

[1] Corresponding author.

AMS: 62J20, 62J05

method of analysis. Suppose we have a response or dependent variable Y which we assume for now is univariate, and a set of k predictors. From the k predictors we can compute p regressors that can include a constant for an intercept, indicator variables for factors and interactions, powers or a spline basis of continuous predictors, nonlinear transformations of the predictors, and possibly others (Weisberg, 2014, Sec. 3.3). Let $\mathbf{x}_i$ be the vector of p regressors for the ith observation in the data, $i = 1, \ldots, n$. Using the methodology proposed by Box and Cox (1964), we seek a parametric transformation $z_t(Y, \theta)$ of the realizations of Y such that there is a possibly vector-valued parameter θ such that the conditional distribution

$$z_t(y, \theta)|\mathbf{x}_i \sim \mathrm{N}(\boldsymbol{\beta}'\mathbf{x}_i, \sigma^2), \tag{1}$$

where σ^2 is an unknown constant variance and $\boldsymbol{\beta}$ is an unknown coefficient vector. The transformation could be used to achieve three possible goals of achieving linearity in the mean function, also called removable non-additivity (Tukey, 1949), constant variance, and normality. Of course these three goals may be in conflict, and we cannot expect all three to be obtainable with the same transformation in every problem.

Given data $(y_i, \mathbf{x}_i), i = 1, \ldots, n$ and following Box and Cox (1964), the transformation parameter can be estimated by maximising a pseudo-likelihood function,

$$L(\theta, \sigma, \boldsymbol{\beta}; y_1, \ldots, y_n) = J \prod_{i=1}^{n} \frac{1}{\sigma} \phi \left[\frac{z_t(y_i, \theta) - \boldsymbol{\beta}'\mathbf{x}_i}{\sigma} \right], \tag{2}$$

where $\phi(z)$ is the standard normal density, and $J = \prod |(dz_t(y, \theta)/dy)|$ is the Jacobian of the transformation. This procedure is not exactly the same as maximum likelihood estimation because normality can hold in (1) for at most one value of θ, but Box and Cox (1964) show that the properties of the estimator of $(\theta, \sigma, \boldsymbol{\beta})$ that maximises (2) are similar to the properties of maximum likelihood estimates.

When the response is strictly positive, the most commonly used family of transformations is the power family, $z_p(y, \lambda) = y^\lambda$. The more elaborate unnormalised Box-Cox power family (BC),

$$z_{bc,u}(y, \lambda) = \begin{cases} \frac{1}{\lambda}\left(y^\lambda - 1\right) & \lambda \neq 0 \\ \log(y) & \lambda = 0 \end{cases}$$

is continuous as a function of λ, and includes logarithms when $\lambda = 0$. The normalised transformation is $z_{bc,n}(y, \lambda) = z_{bc,u}(y, \lambda)/J^{1/n}$, where $J^{1/n} = \dot{g}(y)^{\lambda - 1}$ and $\dot{g}(y) = \exp[(\sum \log(y_i))/n]$ is the geometric mean of y. Using the normalised family, the Jacobian term vanishes from (2), simplifying computations, as the log of the profile log-likelihood for θ evaluated at any fixed value θ_0 is simply -0.5 times the log of the residual sum of squares from the ordinary least squares regression of $z_{bc,n}(y, \theta_0)$ on $\mathbf{x}$. If θ is one-dimensional, the estimate $\hat{\theta}$ of θ is easily computed using a one-dimensional maximisation method (Brent, 2013).

The behaviour of the estimates obtained in this way was explored by Hernandez and Johnson (1980). They showed that the estimates $(\hat{\lambda}, \hat{\sigma}, \hat{\boldsymbol{\beta}})$ are, under regularity conditions, Fisher consistent for $(\lambda, \sigma, \boldsymbol{\beta})$ and asymptotically normal. They also show in their Theorem 2 that the conditional distribution of $z_{bc,n}(y, \hat{\lambda}|\mathbf{x})$ is as close to normally distributed as possible, in the sense of minimising the Kullback-Liebler divergence.

Once an estimator $\hat{\lambda}$ is obtained the response is taken to be $z_p(y,\lambda)$ or $\log(y)$ if $\hat{\lambda}$ is close enough to zero, as would be motivated by an analyst's preference for relatively simple, explainable, results.

In some problems a response that is mostly positive may have some zero or negative realised values. Zero values of the dependent are common, but strictly negative values also arise sometimes. For example, in a bioassay problem the concentration of a substance may be adjusted by subtracting a background reading, sometimes giving a negative response. In an econometric study, a response may be quarterly profit for dealers of a certain product. Some dealers may have incurred losses rather than profits, leading again to a few negative values for the response. We would like to be able to interpret a fitted model as if no nonpositive responses were observed.

Box and Cox recognised the possibility of nonpositive values by introducing a shifted power family, adding a location parameter δ to get the shifted power family (BCS),

$$z_{bcs,n}(y,\lambda,\delta) = z_{bc,n}(y+\delta,\lambda),$$

where $\delta > -\min(y)$ is called a *start* by Mosteller and Tukey (1977). The Jacobian for BCS is $J^{1/n} = \dot{g}(y+\delta)^{\lambda-1}$. Simultaneous estimation of (λ,δ) is commonly avoided because the log-likelihood profile for δ is often nearly flat, providing little information in the data about δ. The usual procedure is to select δ to be slightly larger than $-\min(y)$, and then estimate λ as if δ were known. For fixed δ the asymptotic results of Henandez and Johnson continue to apply, with the conditional distribution of $z_{bc,n}(y+\delta,\hat{\lambda}|\mathbf{x})$ as close to normal as possible, and parameter estimates are asymptotically normal.

Figure 1 shows the effect of changing δ in the important logarithmic case of $\lambda = 0$ for the start $\delta \in -\min(y) + (.01,.05,.10,.25,.50)$, with $y \in (-0.1,1)$. For values of the start close to $-\min(y)$ the transformations of the nonpositive y's change quickly as y decreases, making these potential outliers and/or influential. For relatively large and positive values of y, the shifted Box-Cox transformation increases at a rate that depends on the value of the start. This will introduce bias in estimates of coefficients relative to the estimates that would be obtained using the usual Box-Cox family applied to the positive response data only. Thus interpretation of results will depend on the arbitrary value of the start. This complicates routine use of the shifted power family.

In practice, once λ and δ are determined the power transformation $z_p(y+\gamma,\delta)$ would be used in place of y. For any $y_0 > 0$, let $k = \delta/y_0$. It is instructive to examine the ratio $z_p(y_0+\delta,\lambda)/z_p(y_0,\lambda)$ in Taylor series up to terms of order k^3,

$$\begin{aligned}\frac{z_p(y_0+\delta,\lambda)}{z_p(y_0,\lambda)} &= (1+k)^{\lambda} \\ &\approx 1+\lambda k+(\lambda-1)\lambda k^2/2+(\lambda-2)(\lambda-1)\lambda k^3/6. \qquad (3)\end{aligned}$$

Even for y_0 large enough that $k < 1$ the shifted power can differ from the unshifted power by a substantial amount due to the linear term λk, and thus interpretation of models using the shifted power transformations cannot rely on intuition about usual power transformations. This disagreement can be exacerbated if $-\min(y)$ is large, requiring using a large value of δ.

In Section 2 we introduce a new two-parameter family of transformations that we call the *Box-Cox allowing nonpositive values* family combining the generalised log or *glog* family used in chemometrics with the power transform. An example with both a univariate and multivariate responses is

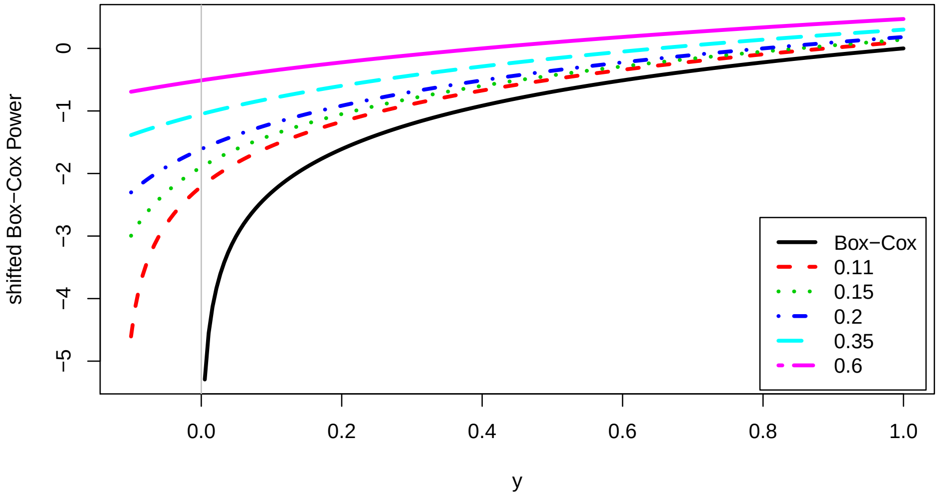

Figure 1: Plot of the unnormalised shifted Box-Cox power transform for $\lambda = 0$ and various values of the start δ. The solid black line corresponds to the usual Box-Cox transformation applied to the positive data only.

given in Section 3. Transformation families can also be used in linear mixed-models, and this extension is briefly presented in Section 4, along with a brief example illustrating the use of both Box-Cox and and the power transformations allowing for nonpositive values. A discussion concludes the paper in Section 5.

2. Box-Cox Power Family Allowing for Nonpositive Values

We have seen that a linear shift from y to $y+\delta$ can cause drastic changes in transformed values when the power is fixed. We seek a shift that is adequate to remove nonpositive values, but has a smaller effect on larger values of y. Let $s(y,\gamma) = \sqrt{y^2+\gamma^2}$, and define a new family of transformations

$$z_{bcn,u}(y,\lambda,\gamma) = z_{bc,u}\left\{.5[y+s(y,\gamma)],\lambda\right\}.$$

We call this the Box-Cox family with nonpositives allowed (BCN), as it is the BC family applied to $0.5[y+s(y,\gamma)]$ rather than to y.

Analogously to the BC methodology, the normalised BCN family is defined as $z_{bcn,n}(y,\lambda,\gamma) = z_{bcn,u}(y,\lambda,\gamma)/J_{bcn}^{1/n}$ where the Jacobian of the unnormalised transformation is

$$J_{bcn}^{1/n} = (0.5)^{\lambda}\, \dot{g}\left\{[y+s(y,\gamma)]^{\lambda-1}[1+y/s(y,\gamma)]\right\}.$$

Many of the properties of the BCN follow directly from the fact that BCN arises conceptually from a two-stage operation:

- Transform y to $w = 0.5(y+\sqrt{y^2+\gamma^2})$

- Apply the conventional BC transformation to w.

Thus the geometric properties of the BC transformation sketched in Lemma 1 of Hernandez and Johnson (1980), for example, carry over unchanged, applied to w instead of y.

The transformation from y to w is convex – its second derivative is $\gamma^2/s^3 > 0$ – but rapidly approaches the line of identity as y increases.

Some more specific properties of the transformation are:

- When $\gamma = 0$ and y is always positive, the transformation family reduces to BC.
- If $\min(y) = 0$ and γ is small, then the BCN comes down to replacing the zero values by γ and moving the nonzero values up by a value smaller than γ and decreasing with y. This is akin, but not identical, to the BCS method with a small start. The situation with $\min(y) < 0$ is qualitatively similar.
- When $\lambda = 0$, $z_{bcn,u}(y, \lambda = 0, \gamma)$ reduces to $\log(1/2) + \log(y + \sqrt{y^2 + \gamma^2})$. The second term in this expression is a simple reparameterisation of the *generalised log*, or glog, transformation of y (Rocke and Durbin, 2001; Durbin, Hardin, Hawkins and Rocke, 2002; Hawkins, 2013; Hawkins, 2014). The glog transformation is also a variant of the Johnson S_U transformation (Johnson, 1949) which has been proposed as a transformation to normality.

Any $\gamma > 0$ can accommodate non-positive y, so this family may be applied to general y, although we anticipate that it will be most useful in problems with mostly positive response and a few zero or negative values.

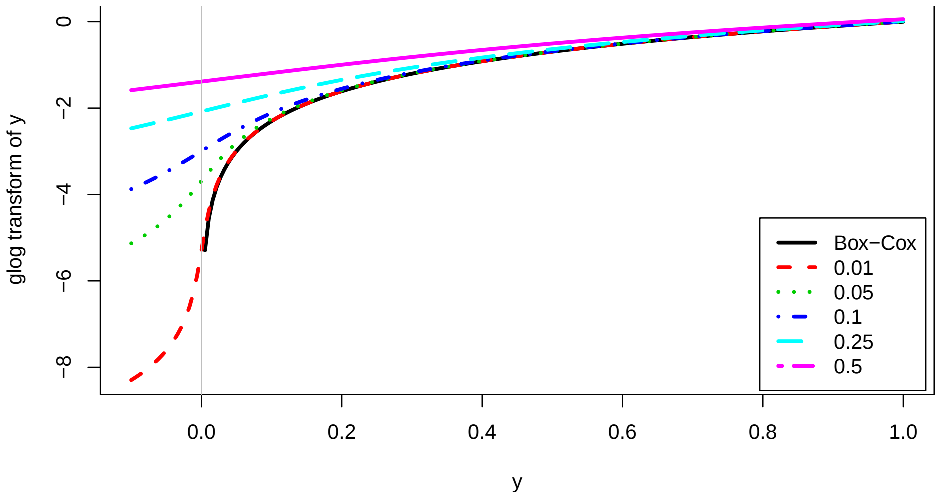

Figure 2: Plot of the unnormalised glog transform for various values of γ.

In Figure 2 we display the unnormalised BCN transform $\lambda = 0$, corresponding to the glog transformation, and for $\gamma \in (.01, .05, .10, .25, .50)$; unlike the shifted Box-Cox transformation, the permissible range of γ does not depend on the minimum value of y. When $\gamma = 0.01$ or 0.05, glog closely matches the Box-Cox power applied to the positive values only, shown by a solid black line on the graph. For all γ, the transformation flattens out for larger negative values of the response. For y greater than about 0.4, the transformed values are nearly independent of γ. Compared to the shifted Box-Cox transformation in Figure 1, the glog transformation behaves closer to the usual Box-Cox power applied to positive data only, and is less likely to make cases with nonpositive responses outliers or influential in the estimation. Thus analysis after transformation with the BCN power family can be interpreted similarly to transformations with the Box-Cox family, with only minor bias added. In particular, should λ be close to zero, estimation with the response in the glog scale can be interpreted similarly to regression with strictly positive y using a log transformation.

In practice, the simple power transformation $z_p(0.5(y+s(y,\gamma)),\lambda)$ would be used once the parameters of the transformation are determined. For a fixed $y_0 > 0$, define $k = \gamma/y_0$, and expanding in Taylor series up to terms in k^3,

$$
\begin{aligned}
\frac{z_p(0.5(y_0+s(y,\gamma)),\lambda)}{z_p(y_0,\lambda)} &= \left\{0.5(1+\sqrt{1+k^2})\right\}^{\lambda} \\
&\approx 1+\lambda k^2/4,
\end{aligned}
$$

as the terms in k and k^3 both have zero coefficients. When compared to (3) we see that the power transformation based on BCN is much closer to the standard power transformation as the leading term is k^2 rather than k, and usual intuition about power transformations can be applied.

The asymptotics of λ are quite straightforward. For fixed γ, BCN is simply the conventional BC applied to the dependent $0.5[y+s(y,\gamma)]$. Thus the estimates are asymptotically normal, and the conditional distribution of the transformed response given the predictors is as close to normal as possible in the sense of Kullback-Liebler divergence. When γ is estimated, in large samples the estimate of γ will converge to a limit, and the same results will then hold in the limit as if γ were fixed.

The situation with γ is less straightforward. Often, the minimizing γ will lie well above zero and conventional confidence intervals based on the Hessian of the pseudo-likelihood will provide good information. If all y are strictly positive, then γ may be unnecessary, as indicated either by the maximum occurring at $\gamma = 0$ or by a likelihood ratio test indicating that a nonzero γ is not significantly better, in which case parsimony is likely to lead one to setting it to zero and using conventional BC.

If the y include any nonpositive values, then a strictly positive γ is required. This may lead to a regular situation with the maximising value lying well above zero. On occasion a degenerate situation arises in which the pseudo-likelihood increases without limit as $\gamma \to 0$. In this situation, a non-trivial γ can not be estimated, and instead a small standard value such as 0.01 may be used. This approach is used in the software.

While the parameter γ has a physical interpretation as the ratio of two variances in the chemometric glog setting, in most other applications it is not of particular interest in its own right, but is more like a nuisance parameter that must be estimated but not interpreted.

3. Examples

3.1. Univariate

Important characteristics of a clinical chemistry assay are its limit of blank (LoB), and its limit of detection (LoD) (CLSI, 2012). The LoB, conceptually the highest reading likely to be obtained from a zero-concentration sample, is defined operationally by the upper 95% point of readings obtained from samples that do not contain the analyte. The LoD, conceptually the lowest level of analyte that can be reliably determined not to be blank, is defined operationally as the true value at which there is a 95% chance of the reading being above the LoB.

The ideal situation is one in which the assay of an analyte with true concentration ξ follows a normal distribution with mean $\mu(\xi)$ and standard deviation $\sigma(\xi)$. In this idealised situation, let a_i be the level of analyte added, and define $\mu(a_i)$ and $\sigma(a_i)$ to be, respectively, the mean and standard deviation when the analyte concentration is a_i. For a zero-concentration sample, $a_i = 0$, the LoB is defined in CLSI (2012) by

$$\mathtt{LoB} = \mu(0) + 1.645\sigma(0) \tag{4}$$

and the LoD is defined by

$$\mathtt{LoD} = \mathtt{LoB} + 1.645\sigma(\mathtt{LoD}) \tag{5}$$

In the constant variance case, σ is the same for all $\mu(a_i)$, and the defining equations simplify to

$$\mathtt{LoB} = \mu(0) + 1.645\sigma$$
$$\mathtt{LoD} = \mathtt{LoB} + 1.645\sigma$$

The LoB and LoD are explored using a series of pools going from blank with $a_i = 0$, no analyte present, to a few successively larger values of a_i. Replicate assays are made on each pool. If the normal distribution model holds, the spread of the replicate readings is used to study σ, and the blank samples are used to estimate $\mu(0)$. This can typically be done with moderately-sized samples, particularly if σ is constant.

If the normal model does not hold however, a considerably more complex non-parametric approach is required, leading to sample sizes numbering in the hundreds. This creates a substantial incentive to find a good way to transform non-normal or heteroscedastic assays to equivalents following normal distributions with constant variance.

The illustrative data set we discuss is a portion of a LoB/D study of an assay for a drug used to treat certain cancers. Twelve pools were used, four of them blanks of different types, and eight with successively increasing drug levels. The data are given in the data file LoBD in the car package in R (Fox and Weisberg, 2011). For this example the variable labelled I1L1 is taken as the response; it is in the units of pmol/L. We will fit one-way analysis of variance models, replacing the generally unknown concentrations a by p indicator variables $\mathbf{x}$ indicating the p panels used in the study. The residuals from the one-way analysis of variance are shown in Figure 3, plotted against the panel mean. The plot casts doubt on a constant variance assumption, and this heteroscedasticity makes checking for normality difficult. Some of the lower-number groups include nonpositive values, so the standard Box-Cox transformation cannot be used, making the BCN family an attractive possibility.

The estimates for the BCN family are computed with the bcnPower family argument to the powerTransform function in the car package to be $(\hat{\lambda}, \hat{\gamma}) = (0.54, 16.24)$. The standard errors of

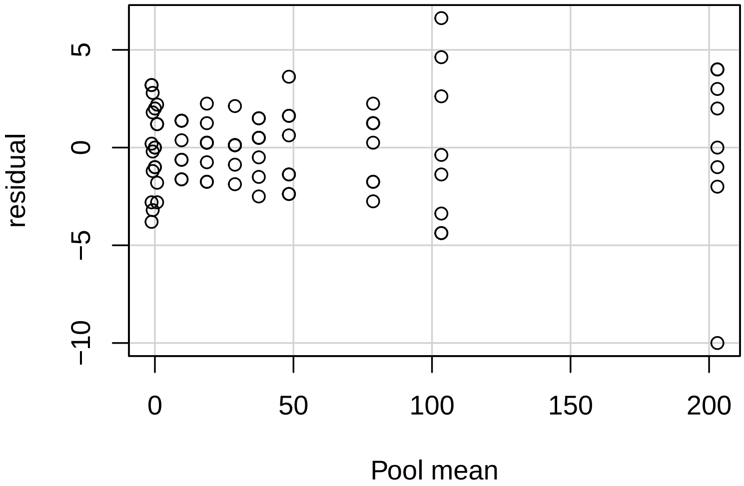

Figure 3: Plot of the data for the LoBD example

the estimates are computed from the inverse of the Hessian to be $(0.14, 11.97)$, respectively. Wald-type confidence intervals can be based on asymptotic normality of the estimates, except that the lower bound for γ is never less than 0.

If we regard γ as a nuisance parameter and simply maximise over it, likelihood ratio tests for specific values λ of the transformation parameters can be based on the profile log-likelihood for λ only. The test with null hypothesis $\lambda = 0$ versus a general alternative has p-value close to zero, while the test for $\lambda = 1$ also has p-value close to zero. These tests are implemented automatically in the `powerTransform` function; the manual page for the function give more information.

Figure 4a provides a contour plot for the profile log-likelihood for (λ, γ), while Figure 4b provides the profile log-likelihood for λ only. This latter plot is a modest generalisation of the `boxcox` method in the `MASS` package (Venables and Ripley, 2002).

In this example both the Wald interval for λ with γ fixed at its estimate and the tests based on profile log-likelihood argue against using either untransformed ($\lambda = 1$) or log-transformed ($\lambda = 0$) data. Since the contour log-likelihood shown in Figure 4a is so clearly non-elliptical, inference based based on the profile log-likelihood for λ are likely to be more accurate in this problem.

Using the estimates $(\hat{\lambda},\ \hat{\gamma}) = (0.54, 16.24)$, the transformation appears to be successful in achieving both normality and constant variance of the transforms, as both the Levene and Shapiro-Wilk tests return large P values. This allows us to estimate the `LoB` and `LoD`. On the $z_{bcn,u}$ scale, the four blank pools give a mean of 3.85 and the pooled estimate of σ is 0.36. From this the `LoB` and `LoD` in $z_{bcn,u}$-scale are estimated using (4) and (5) to be 4.44 and 5.03, respectively. Converting these back to the original measurement scale using (6) in the Appendix then gives the `LoB` as 2.76 pmol/L and the `LoD` as 5.54 pmol/L.

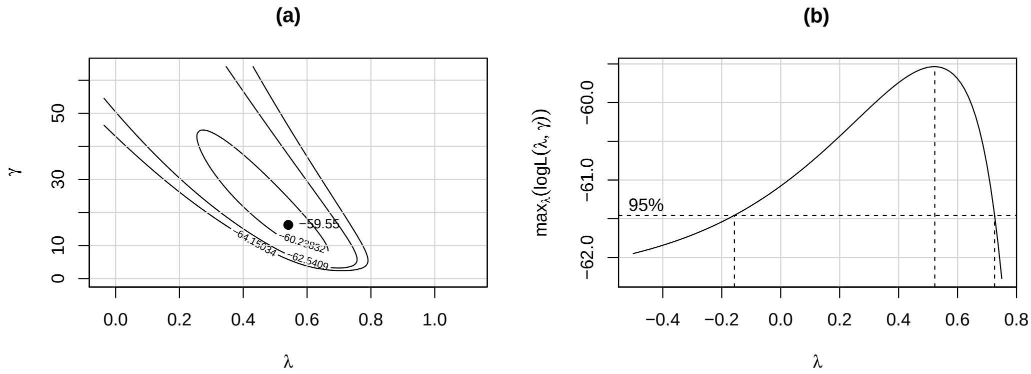

Figure 4: (a) Contour plot of the profile log-likelihood for (λ, γ) in the univariate example. The displayed contours are for 50%, 95% and 99% likelihood based confidence regions. (b) Profile log-likelihood for λ alone maximising over γ.

3.2. Multivariate extension

Velilla (1993) provided the framework for a multivariate linear model extension of the Box-Cox methodology in which separate λ and γ are estimated for each response. Suppose we have r responses, $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ are r-vectors of parameters, and $\mathbf{Z}(\boldsymbol{\lambda}, \boldsymbol{\gamma})$ is an $n \times r$ matrix whose jth column is the normalised BCN transformation with parameters (λ_j, γ_j) of the jth response variable. Then $(\boldsymbol{\lambda}, \boldsymbol{\gamma})$ are estimated by minimising the log-determinant of the matrix of residuals from the multivariate regression of $\mathbf{Z}(\boldsymbol{\lambda}, \boldsymbol{\gamma})$ on the regressors. Details of the “likelihood" function, and an outline of the computational algorithm, are described in Weisberg (2014, Appendix A.12).

The LoBD data provides a multivariate application. The data set used in the univariate setting was part of a larger exercise involving four instruments and two reagent lots. For simplicity here we will consider $r = 2$ responses obtained using two reagent lots with the same instrument. In the multivariate fit, we get $(\hat{\lambda}_1, \hat{\gamma}_1) = (0.56, 13.43)$ for the estimates for the response `I1L2` obtained using the second reagent lot, and $(\hat{\lambda}_2, \hat{\gamma}_2) = (0.52, 17.27)$ for the response `I1L1` for the data set discussed earlier. The estimates for `I1L1` with bivariate fitting are a little different from the estimates with univariate fitting because the multivariate computation includes estimating the covariance between the responses.

We can formally test for equality by stacking the two responses into one column, and fitting a two-way analysis of variance, with one factor for the pool of values and a second factor for the response number, The value of the test statistic is 0.93, and it can be compared to χ^2_2 to get a significance level, and clearly suggests no evidence of a need for different transformations for the two responses, and allows us to move forward using the same transformation for both reagent lots.

4. Transformations with Linear Mixed-Effects Models

The Box-Cox approach to transformations has generally been applied in the context of a linear model. Generalisation to mixed models is straightforward. Similar to the notation in Section 3.2, let $\mathbf{Z}(\theta)$ be a $n \times 1$ vector of normalised transformed values of a response y, with θ the parameter or parameters of the transformation. If X is an $n \times p$ design matrix of fixed effects and $\mathbf{W}$ is an $n \times q$ design matrix of q random effects $\mathbf{b}$, then the linear mixed-model can be written as (Pinheiro and Bates, 2006, Sec. 2.1)

$$\mathbf{Z}(\theta)|\mathbf{b} \sim \mathrm{N}\left(X\boldsymbol{\beta} + \mathbf{W}\mathbf{b}, \sigma^2 I\right)$$
$$\mathbf{b} \sim \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Psi}\right),$$

where typically $\boldsymbol{\Psi}$ depends on a small number of parameters. The log-likelihood for $(\theta, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\Psi})$ is equal to the log-density of the assumed multivariate normal distribution obtained by integrating the joint distribution of $(\mathbf{Z}(\theta), \mathbf{b})$ over the distribution of $\mathbf{b}$. The value of the profile log-likelihood for fixed θ but maximising over all other parameters is returned by software for fitting linear mixed models, such as the `lmer` function in the `lme4` package (Bates, Mächler, Bolker and Walker, 2015). When θ is one-dimensional as for the Box-Cox power family, we use the `optimise` function in the `stats` package in R to maximise over θ. When θ is two-dimensional, we use an alternating algorithm, first optimising over λ for fixed γ, and then γ for fixed λ.

As a simple example, we revisit the LoBD data, this time considering only the four blank pools, treating `pool` as a random, rather than a fixed, effect. The BCN transformation can be estimated by reformatting the data to stack all the responses in the same column, and adding a variable for the pool names. This result is very similar to the fixed-effects analysis presented earlier, with $(\hat{\lambda}, \hat{\gamma}) = (0.45, 6.25)$, again suggesting a square root transformation.

5. Discussion

The use of the BCN power family is recommended for problems where selecting a transformation of one or more responses is at issue, but the response includes a few zero or negative values. The BCN transformed responses can be interpreted similarly to transformed values using the BC family with no nonpositive responses. The shifted Box-Cox BCS family that adds a constant to all values to make them positive produce results that can depend heavily on the essentially arbitrary start. Another alternative family proposed by Yeo and Johnson (2000) uses different powers depending on the sign of the value to be transformed, leading to difficulty in interpretation. While not avoiding these problems entirely, the BCN transformations produce results that are less dependent on the added parameter γ, akin to the start for the BCS family, and also allow for interpretation of results in the same way they would be interpreted for a strictly positive response. In particular, if a log transform is supported by the data, the usual interpretation of $\exp(\hat{\beta})$ as an approximate percent change in a response for a unit change in a corresponding regression (Weisberg, 2014, Sec. 4.1.7) can be used, provided as usual that $\hat{\beta}$ is not too large, and adding the restriction of applying the results to observations with fitted mean not too close to zero.

Computations are greatly simplified if the parameters are restricted to a range that would be of practical use, and we consider only $\lambda \in [-3, 3]$ and $\gamma \in (0, \max(y))$. When fitting in one dimension, this allows the use of line searches in an alternating algorithm, maximising first λ given γ, and then maximizing γ given λ. In the multivariate case, we obtain starting values for each of the r responses as if we had r univariate problems. We then use a similar alternating algorithm, maximising $\boldsymbol{\lambda}$ given $\boldsymbol{\gamma}$ and then maximising over $\boldsymbol{\gamma}$ given $\boldsymbol{\lambda}$. The manual pages for the `powerTransformation` and `bcnPower` functions in the `car` package for R provide further information, including the R code for most of the fitting in this paper.

Appendix

The BCN transformations are just a special case of power transformations. They are strictly monotonic; the inverse for the unnormalised BCN transformation is

$$y = (q^2 - \gamma^2)/(2q), \tag{6}$$

where, writing $z = z_{bcn,u}(y, \lambda, \gamma)$,

$$q = \begin{cases} 2(\lambda z + 1)^{1/\lambda} & \lambda \neq 0 \\ 2\exp(z) & \lambda = 0. \end{cases}$$

The inverse BCN transformation is useful for converting estimates and confidence intervals made on the z-scale to y-scale counterparts.

References

BATES, D., MÄCHLER, M., BOLKER, B., AND WALKER, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67** (1), 1–48. doi:10.18637/jss.v067.i01.

BOX, G. E. P. AND COX, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, **26** (2), 211–252.
URL: `http://www.jstor.org/stable/2984418`

BRENT, R. P. (2013). *Algorithms for Minimization without Derivatives*. Dover Publications, Inc.: Mineola, New York.

CLSI (2012). *Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures; Approved Guideline*. Clinical and Laboratory Standards Institute, Wayne PA, second edition.

DURBIN, B. P., HARDIN, J. S., HAWKINS, D. M., AND ROCKE, D. M. (2002). A variance-stabilizing transformation for gene-expression microarray data. *Bioinformatics*, **18** (suppl 1), S105–S110.
URL: `http://doi.org/10.1093/bioinformatics/18.suppl_1.S105`

FOX, J. AND WEISBERG, S. (2011). *An R Companion to Applied Regression*. Second edition. Sage: Thousand Oaks, CA.

HAWKINS, D. M. (2013). A general variance model in methods comparison. *Journal of Chemometrics*, **27** (11), 414–419.

HAWKINS, D. M. (2014). A model for assay precision. *Statistics in Biopharmaceutical Research*, **6** (3), 263–269.

HERNANDEZ, F. AND JOHNSON, R. A. (1980). The large-sample behavior of transformations to normality. *Journal of the American Statistical Association*, **75** (372), 855–861. URL: `http://doi.org/10.1080/01621459.1980.10477563`

JOHNSON, N. L. (1949). Systems of frequency curves generated by methods of translation. *Biometrika*, **36** (1-2), 149–176. doi:http://doi.org/10.1093/biomet/36.1-2.149.

MOSTELLER, F. AND TUKEY, J. W. (1977). *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley: Reading, MA.

PINHEIRO, J. AND BATES, D. (2006). *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media: New York.

ROCKE, D. M. AND DURBIN, B. (2001). A model for measurement error for gene eexpression arrays. *Journal of computational biology*, **8** (6), 557–569. URL: `http://doi.org/10.1089/106652701753307485`

TUKEY, J. W. (1949). One degree of freedom for non-additivity. *Biometrics*, **5** (3), 232–242. URL: `http://www.jstor.org/stable/3001938`

VELILLA, S. (1993). A note on the multivariate Box-Cox transformation to normality. *Statistics and Probability Letters*, **17**, 259–263. URL: `http://doi.org/10.1016/0167-7152(93)90200-3`

VENABLES, W. N. AND RIPLEY, B. D. (2002). *Modern Applied Statistics with S*. Fourth edition. Springer: New York. ISBN 0-387-95457-0. URL: `http://www.stats.ox.ac.uk/pub/MASS4`

WEISBERG, S. (2014). *Applied Linear Regression*. Fourth edition. John Wiley & Sons: Hoboken, NJ.

YEO, I.-K. AND JOHNSON, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, **87** (4), 954–959. URL: `http://www.jstor.org/stable/2673623`

Manuscript received, 2016-11-30, revised, 2017-06-09, accepted, 2017-06-28.