Analysis of Censored Environmental Data with Box-Cox Transformations

Jade Freeman[1,2] and Reza Modarres[2]

[1] *Office of Water, U.S. Environmental Protection Agency*
E-Mail: lee-freeman.jade@epa.gov

[2] *Department of Statistics, George Washington University*
E-Mail: reza@gwu.edu

Abstract

We present a method for estimating the mean vector from a multivariate skew distribution that includes some unobserved data below the detection limits. The method uses a Box-Cox transformation, of which the parameters are found by maximizing the likelihood function over a fixed power transformation set. To estimate the mean vector and the covariance matrix we develop an E-M algorithm solution. Given a transformation, we obtain expressions for the mean vector, covariance matrix, and the asymptotic covariance of the vector of means in the original scale. Expressions are obtained for a confidence region for the vector of means. The performance of the MLE method in selecting the correct power transformation and the coverage rate of the confidence region under several conditions are investigated in a simulation study. This method gives reliable results for finding effective transformations and the coverage rate for highly skew data sets. The method is applied to water quality monitoring data.

Key Words: Box-Cox Transformation, Censored Data, E-M Algorithm, Maximum Likelihood, Delta-Method, Confidence Region.

1 Introduction

The analysis of environmental data frequently centers on the mean concentration of pollutant of interest. For example, industry compliance with wastewater discharge regulations is usually determined by comparing the mean concentration of discharged wastewater with a legal standard. In human health and ecological risk assessment, risk is defined as the product of dose and the probability of an adverse effect per dose unit. Here, the mean amount of chemicals or microbes in an environment (e.g., air, water, soil) is often used as a direct measure of dose. However, in most cases, environmental data are not well suited for routine statistical analysis. The data are skew, usually to the right, and include several extreme observations at both low and high levels. Also, the observations at low levels are often censored, i.e., only reported as being less than some detection limit. U.S. EPA (1992) defines a detection limit as the lowest concentration level that can be determined to be different from a blank sample.

To model right skew data, one can assume that the density of the variable of interest belongs to a parametric class of distributions such as lognormal or gamma. See (McCullagh and Nelder, 1989). For example, in an interesting article, Joe (1994) develops several multivariate extreme-value distributions and fits them to environmental multivariate extreme data sets. Usually, a parametric distributional model is chosen based on physical or biological grounds. For example, Ott (1995, chapter 8) argues that repeated dilution of a contaminant with water results in a gamma distribution. The lognormal distribution arises as the product of many independent random factors (Aitchison and Brown, 1973). In such cases, there is little uncertainty about the underlying distribution, and a suitable model can be identified a priori. In many environmental applications, however, we do not have adequate physical, biological, or empirical knowledge to suggest a functional form for the underlying distribution, i.e., specify the model. However, we may have enough knowledge to suggest that the true distribution belongs to one of certain specific families. In such cases, one may assume that the distribution function F is a member of a set of parametric families $\mathcal{F}_i,\ i = 1, \ldots, k$.

Another approach would seek a transformation of the data. Tukey (1957) regards transformations as re-expressions of the data and basic tools of data analysis and statistical inference. Three well-known classes of transformations are rank transformations (Lehmann, 1975), transformations to symmetry (Hinkley, 1975), and transformations to normality (Box and Cox, 1964). Hernandez and Johnson (1980) note that the scale on which a variable is measured may not be the best for statistical analysis. For example, the logarithmic transformation is often used to analyze the data on a log-scale. Chen and Loh (1992) and Chen (1995) study the gain in the efficiency of tests following a Box-Cox transformation to a more suitable scale. Transformations to normality have played a prominent role in the analysis of non-normal data. Statisticians have an arsenal of statistical methods at their disposal when the underlying distribution of the observations is normal. Such powerful normal theory based procedures can be applied to the transformed data.

Miller (1986) discusses transformations along with nonparametric techniques and robust estimation as methods of correcting for non-normality. Hinkley and Runger (1984) provide an overview of the analysis of transformed data. Chen and Lockhart (1997) find the Fisher information and its inverse and consider the question of how to make inference in a linear model whose response has been subject to a Box-Cox transformation. Chen, Lockhart and Stephens (2002) discuss Box-Cox transformations in linear models and develop large sample theory and tests of normality. In this paper, we investigate a correlation model where both the response and the predictors are subject to variation and transformed to provide a joint confidence interval for the mean vector in the original scale. We assume that the response variable is subject to censoring on the left as in a detection limit setting.

Existing statistical procedures have been successful in dealing with censored observations. However, one needs to note that such procedures are model-based. To avoid model specification in case of model uncertainty, one can transform the data, including the censored observations, and carry out the analysis under normality in the transformed scale. We are often interested in making inferences in the original scale (Shumway, Azari and Johnson, 1989). After the parameters of interest are estimated in the transformed scale, a method of back-transformation is used to obtain the estimates in the original scale. Much of the preceding discussion concerns univariate analysis. Analysis of multivariate data offers additional challenges. Model selection and verification are more difficult, as is acquiring insight about the joint distribution of variables of interest.

In the next sections, we present a method for estimating the mean vector from a multivariate skew distribution that includes some unobserved data below the detection limits. The following section discusses Box-Cox transformations with censored observations. The E-M algorithm is discussed in Section 3, where it is used to obtain MLE's for the mean vector and covariance matrix. In Section 4, we obtain the covariance of the mean estimates and a confidence region for the mean vector. In Section 5, the performance of the method in selecting the correct power transformation and the coverage rates for the confidence region are studied. Finally, we apply this method to a data set of water quality measurements in Section 6.

2 Box-Cox Transformation on Censored Data

The Box-Cox transformation (Box and Cox, 1964) is defined as

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

where X is a positive random variable and λ is the transformation parameter. The literature on the Box-Cox transformation (Sakia, 1992) assumes that there exists a value λ such that Y has a normal distribution. However, X> 0 implies Y> $-1/\lambda$ for $\lambda > 0$ and Y< $-1/\lambda$ for $\lambda < 0$. Thus, for $\lambda \neq 0$ the domain of Y is not the entire real line. It is only when $\lambda = 0$ that Y can have an exact normal distribution. Researchers have generally assumed that Y has an approximate normal distribution. For example, one may sidestep the issue by assuming that μ is large and the coefficient of variation $\kappa = \sigma/\mu$ is small for $\lambda > 0$ so that $P(Y < -1/\lambda) = P[Z < -(\kappa^{-1} + 1/(\lambda\sigma))] < \epsilon$, where ϵ is a small value (e.g., $\epsilon \leq 10^{-6}$). Another strategy is to transform $X + c$ instead of X where c is a sufficiently large constant (Gnanadesikan, 1977, p. 143). We will assume that X is lognormal or transformed to an approximate normal distribution.

To decide which Box-Cox transformation produces data that are the most difficult to distinguish from a normal distribution, the method of maximum likelihood and the tests of normality are standard techniques. (See Johnson and Wichern, 1998, p. 204 - 214.) Miller (1986) notes that for any given type of data it is important to find a transformation that yields normally distributed observations in the majority of cases. For unimodal and right skew data, Shumway, Azari and Johnson (1989) assume $\lambda \in \Lambda = \{0, 1/4, 1/2, 1\}$ and select the best fitting normal model by maximizing the likelihood function of the sample over Λ. Here, X with parameters $(\mu_i, \sigma_i^2, \lambda_i)$ is represented by a set of parametric families $\mathcal{F}_i,\ i = 1, \ldots, k$.

Many environmental variables such as body weights, concentration levels, ingestion and inhalation rates are positive. Hence, our interest centers on positive random variables with

Table 1: Expected Values and Variances of X

λ	E(X)	Var(X)
0	$\exp(\mu + \frac{1}{2}\sigma^2)$	$\exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$
$\frac{1}{4}$	$(\frac{1}{4}\mu + 1)^4 + \frac{3}{8}\sigma^2(\frac{1}{4}\mu + 1)^2 + \frac{3}{256}\sigma^4$	$\frac{8}{2048}\sigma^8 + \frac{3}{32}\sigma^6(\frac{1}{4}\mu + 1)^2 + \frac{21}{32}\sigma^4\left(\frac{1}{4}\mu + 1\right)^4 + \sigma^2\left(\frac{1}{4}\mu + 1\right)^6$
$\frac{1}{3}$	$\left(\frac{1}{3}\mu + 1\right)^3 + \frac{1}{3}\sigma^2\left(\frac{1}{3}\mu + 1\right)$	$\frac{5}{243}\sigma^6 + \frac{4}{9}\sigma^4\left(\frac{1}{3}\mu + 1\right)^2 + \sigma^2\left(\frac{1}{3}\mu + 1\right)^4$
$\frac{1}{2}$	$(\frac{1}{2}\mu + 1)^2 + \frac{1}{4}\sigma^2$	$\frac{1}{8}\sigma^4 + \sigma^2\left(\frac{1}{2}\mu + 1\right)^2$
1	$\mu + 1$	σ^2

a right skew distribution. We will assume that $0 \leq \lambda \leq 1$, which includes the logarithmic transformation (i.e., the log-scale model), one fourth, cube root, square root and no transformation (i.e., the original model) among others. We can estimate the quantile function with $\hat{Q}_X(p) = (\lambda(\hat{\mu} + \hat{\sigma}\Phi^{-1}(p)) + 1)^{1/\lambda}$ for $\lambda \neq 0$ or $\hat{Q}_X(p) = \exp(\hat{\mu} + \hat{\sigma}\Phi^{-1}(p))$ for $\lambda = 0$ in the original scale, where $\hat{\mu}$ and $\hat{\sigma}^2$ are the maximum likelihood estimates. However, inference about the mean or other moments of the original scale is not always straightforward. For example, Table 1 shows that the mean in the original scale can be a non-linear function of μ and σ^2, the mean and the variance in the transformed scale.

Several different techniques for parameter estimation for censored data have appeared in the literature. In particular, most parametric techniques are based on the method of maximum likelihood estimation (MLE). For example, the E-M algorithm developed by Dempster, Laird and Rubin (1977) is shown to produce good estimates for large sample sizes (Shumway, Azari, and Johnson, 1989). The E-M algorithm approaches the problem of solving the incomplete-data log-likelihood equation, i.e., the log-likelihood equation with unobserved data, by indirectly proceeding iteratively in terms of the complete-data log-likelihood function (McLachlan and Krishnan, 1997). The complete-data log-likelihood is written as if the data were completely observed. During the expectation step (E-step), any censored value is replaced by its conditional expectation given the observed data using the current fit for the parameters of the likelihood. In the maximization step (M-step), the likelihood function is maximized over the parameter space to obtain the parameter estimates for the next iteration of the E-step. The maximum likelihood estimates are obtained when the iterations of alternating E-step and M-step converge.

3 E-M Algorithm to Estimate the Mean and Covariance

Let $(X_{1j},...,X_{(p-1)j}, X_{pj})$, $j = 1, ..., N$ be a random sample from $(X_1, ..., X_{(p-1)}, X_p)$ where $(X_1, ..., X_{(p-1)})$ are completely observed, but X_p is censored. We assume Type I Censoring on X_{pj} with the detection limit D_j for $j = n+1, ..., N$. Let $\vec{\mu}$ and Σ denote the mean vector and the covariance matrix of $\vec{Y}$ with components

$$Y_{ij} = p_{\lambda_i}(X_{ij}) = \begin{cases} \frac{X_{ij}^{\lambda_i}-1}{\lambda_i}, & \lambda_i \neq 0, \\ \ln X_{ij}, & \lambda_i = 0, \end{cases} \quad X_{ij} > 0, \quad i = 1, ..., p, \quad j = 1, ..., N,$$

where X_{ij} denotes the jth observation on the ith variable. Let $\vec{\lambda}$ be the vector of transformations. We assume that $\vec{Y} = (\vec{Y}^{(1)}, Y_p)$ has a p-variate normal distribution with mean $\vec{\mu} = (\vec{\mu}^{(1)}, \mu_p)$ and covariance $\Sigma = \begin{bmatrix} \Sigma_{11} & \vec{\sigma}^{(1)} \\ \vec{\sigma}^{(1)\prime} & \sigma_p^2 \end{bmatrix}$ where $\vec{Y}^{(1)} = (Y_1, ..., Y_{p-1})$ and $\vec{\sigma}^{(1)} = (\sigma_{1p}, \sigma_{2p}, ..., \sigma_{(p-1)p})'$. The joint density can be written as $f(\vec{y}) = f(y_p \mid \vec{y}^{(1)})g(\vec{y}^{(1)})$ where given $\vec{y}^{(1)}$, the censored and transformed Y_p has a normal distribution with mean $\mu^* = \mu_p + \vec{\sigma}^{(1)\prime}\Sigma_{11}^{-1}(\vec{y}^{(1)} - \vec{\mu}^{(1)})$ and variance $\sigma^{*2} = \sigma_p^2 - \vec{\sigma}^{(1)\prime}\Sigma_{11}^{-1}\vec{\sigma}^{(1)}$. Here, we have $\vec{Y}^{(1)} \sim N_{(p-1)}(\vec{\mu}^{(1)}, \Sigma_{11})$. The likelihood of the multivariate data in the original scale is

$$\begin{aligned} \ln L(\vec{\mu}, \Sigma \mid \vec{\lambda}, \vec{x}) &= -\frac{N}{2} \ln \mid \Sigma_{11} \mid -\frac{n}{2} \ln \sigma^{*2} + \sum_{i=1}^{p-1}(\lambda_i - 1) \sum_{j=1}^{N} \ln x_{ij} \\ &\quad -\frac{1}{2}\sum_{j=1}^{N}\left[p_{\vec{\lambda}^{(1)}}(\vec{x}_j^{(1)}) - \vec{\mu}^{(1)}\right]' \Sigma_{11}^{-1}\left[p_{\vec{\lambda}^{(1)}}(\vec{x}_j^{(1)}) - \vec{\mu}^{(1)}\right] + \sum_{x_{pj} \leq D_j} \ln \Phi(w_j) \\ &\quad +(\lambda_p - 1)\sum_{x_{pj} > D_j} \ln X_{pj} - \sum_{x_{pj} > D_j} \frac{1}{2\sigma^{*2}}(p_{\lambda_p}(X_{pj}) - \mu^*)^2 + Constant \end{aligned}$$

where $w_j = (D_j^* - \mu^*)/\sigma^*$ and X_{pj}, j=1,...,n are the observed values of X_p. Since $\vec{X}^{(1)}$ is completely observed, we use $\hat{\vec{\mu}}^{(1)} = \frac{1}{N}\sum_{j=1}^{N} p_{\vec{\lambda}^{(1)}}(\vec{X}_j^{(1)})$ and $\hat{\Sigma}_{11} = \frac{1}{N}\sum_{j=1}^{N}\left[p_{\vec{\lambda}^{(1)}}(\vec{X}_j^{(1)}) - \hat{\vec{\mu}}^{(1)}\right]' \cdot \left[p_{\vec{\lambda}^{(1)}}(\vec{X}_j^{(1)}) - \hat{\vec{\mu}}^{(1)}\right]$. For censored X_p, the E-M algorithm operates on the log-likelihood to obtain

$$\hat{\mu}_{p,k+1} = \frac{1}{N}\left(\sum_{j=1}^{n} Y_{pj} + \sum_{X_{pj} \leq D_j} E_k(Y_{pj} \mid Y_{pj} \leq D_j^*, \vec{Y}^{(1)} = \vec{y}^{(1)})\right),$$

$$\hat{\sigma}_{p,k+1}^2 = \frac{1}{N}\left(\sum_{j=1}^{n} Y_{pj}^2 + \sum_{X_{pj} \leq D_j} E_k(Y_{pj}^2 \mid Y_{pj} \leq D_j^*, \vec{Y}^{(1)} = \vec{y}^{(1)})\right) - \mu_{p,k+1}^2,$$

and ith element of $\hat{\vec{\sigma}}^{(1)}$ for $i = 1, ..., p-1$, is estimated with

$$\hat{\sigma}_{ip,k+1} = \frac{1}{N}\left(\sum_{j=1}^{n} Y_{ij}Y_{pj} + \sum_{X_{pj} \leq D_j} y_{ij}E_k\left(Y_{pj} \mid Y_{pj} \leq D_j^*, \vec{Y}^{(1)} = \vec{y}^{(1)}\right)\right).$$

Define A_k as

$$A_k = E_k(Y_{pj} \mid Y_{pj} \le D_j^*, \vec{Y}^{(1)} = \vec{y}^{(1)}) = \mu_{p,k} + \hat{\vec{\sigma}}^{(1)'} \hat{\Sigma}_{11}^{-1}(\vec{y}^{(1)} - \hat{\vec{\mu}}^{(1)}) - \hat{\sigma}_k^* \frac{\phi(w_j)}{\Phi(w_j)}.$$

One can show the conditional moment of censored Y_p is

$$E_k(Y_{pj}^2 \mid Y_{pj} \le D_j^*, \vec{Y}^{(1)} = \vec{y}^{(1)}) \quad = \quad A_k^2 - 2A_k\sigma_k^* \frac{\phi(w_j)}{\Phi(w_j)} + \sigma_k^{*2}\left(1 - w_j \frac{\phi(w_j)}{\Phi(w_j)}\right).$$

Note that $\vec{\mu}^{(1)}$ and Σ_{11} are only estimated once whereas $(\mu_p, \sigma_p^2, \mu^*, \sigma^{*2})$ and the vector of estimated covariances at the kth stage, $\vec{\sigma}^{(1)}$, are estimated iteratively. Initial estimates can be based on the observed data. The values of the conditional expectations are affected by selecting an incorrect transformation. For example when $p = 2$, selecting $\vec{\lambda} = (1, 1)$, i.e., no transformation, for bivariate lognormal variables $\vec{X}$ results in the conditional expectation $E(p_{\lambda_2}(X_2) \mid p_{\lambda_1}(X_1) = p_{\lambda_1}(x_1))$ in the form of $\exp(\mu_2 + \beta(\exp(y_1) - 1 - \mu_1) + \sigma^{*2}/2) - 1$ rather than the correct form of $\mu_2 + \beta(y_1 - \mu_1)$. Here, $\beta = \rho\sigma_2/\sigma_1$ and $\sigma^{*2} = \sigma_p^2(1 - \rho^2)$. Similarly, one obtains $\exp(\mu_2 + \beta(y_1 - \mu_1) + \frac{1}{2}\sigma^{*2}) - 1$ and $\mu_2 + \beta(\exp(y_1) - 1 - \mu_1)$ for the conditional expected value when $\vec{\lambda} = (0, 1)$ and (1,0), respectively.

4 Confidence Region for $E(\vec{X})$

To compute $Cov(\hat{E}(\vec{X}))$, we need $Cov(\hat{\vec{\theta}}) = I^{-1}(\hat{\vec{\theta}})$ where $I(\vec{\theta})$ is the information matrix. Here, $Cov(\hat{\vec{\theta}})$ is the covariance matrix of p means and $p(p+1)/2$ covariances. Let $\vec{\theta} = (\vec{\mu}, \Sigma)$ and $\frac{\partial \ln L}{\partial \vec{\theta}} = (L_\mu, L_\Sigma)$. In the Appendix, we show that the score functions can be written as sums of non-iid zero-mean random vectors such that $E(L_{\vec{\mu}})$ and $E(L_\Sigma)$ are both zero. We also prove the asymptotic normality of this vector. To obtain the information matrix of the parameters that concern the observed variates, let $\vec{\psi}_1 = (\vec{\mu}^{(1)}, \Sigma_{11})$ and note that $\ln L(\vec{\mu}, \Sigma \mid \vec{\lambda}, \vec{x}) = \ln L_1 + \ln L_2$. One can show

$$\frac{\partial \ln L_1}{\partial \vec{\psi}_{1r}} = -\frac{N}{2}(\ln \mid \Sigma_{11} \mid)_r + \sum_{j=1}^{N} \vec{\mu}_r^{(1)'}\Sigma_{11}^{-1}\left[p_{\vec{\lambda}^{(1)}}(\vec{x}_j^{(1)}) - \vec{\mu}^{(1)}\right]$$
$$-\frac{1}{2}\sum_{j=1}^{N}\left[p_{\vec{\lambda}^{(1)}}(\vec{x}_j^{(1)}) - \vec{\mu}^{(1)}\right]'(\Sigma_{11}^{-1})_r\left[p_{\vec{\lambda}^{(1)}}(\vec{x}_j^{(1)}) - \vec{\mu}^{(1)}\right],$$

where the subscript r on the left hand side of the equality stands for partial differentiation with respect to $\vec{\psi}_{1r}$, the rth element of $\vec{\psi}_1$. One can show that the information matrix is

$$\begin{aligned} I(r, s) &= Cov(\ln L_1(\vec{\psi}_1)_r, \ln L_1(\vec{\psi}_1)_s) \\ &= \vec{\mu}_r^{(1)'}\Sigma_{11}^{-1}\vec{\mu}_s^{(1)} + \frac{1}{2}tr\left((\Sigma_{11}^{-1})_r(\Sigma_{11})(\Sigma_{11}^{-1})_s\Sigma_{11}\right); \end{aligned}$$

see Magnus and Neudecker (1988, chapter 15). Since $(\Sigma_{11}^{-1})_r = -\Sigma_{11}^{-1}(\Sigma_{11})_r\Sigma_{11}^{-1}$, we have

$$I(r,s) = \vec{\mu}_r^{(1)\prime}\Sigma_{11}^{-1}\vec{\mu}_s^{(1)} + \frac{1}{2}tr\left((\Sigma_{11})_r\Sigma_{11}^{-1}(\Sigma_{11})_s\Sigma_{11}^{-1}\right).$$

For example, covariance of $\vec{\psi}_1 = (\mu_i, \sigma_i, \mu_j, \sigma_j, \rho_{ij})$ for $i, j = 1, ..., p-1$ is given by

$$Cov(\vec{\psi}_1) = \frac{1}{N}\begin{pmatrix} \sigma_i^2 & 0 & \rho_{ij}\sigma_i\sigma_j & 0 & 0 \\ 0 & \frac{\sigma_i^2}{2} & 0 & \frac{\rho_{ij}{}^2\sigma_i\sigma_j}{2} & \frac{\rho_{ij}(1-\rho_{ij}{}^2)\sigma_i}{2} \\ \rho_{ij}\sigma_i\sigma_j & 0 & \sigma_j^2 & 0 & 0 \\ 0 & \frac{\rho_{ij}{}^2\sigma_i\sigma_j}{2} & 0 & \frac{\sigma_j^2}{2} & \frac{\rho_{ij}(1-\rho_{ij}{}^2)\sigma_j}{2} \\ 0 & \frac{\rho_{ij}(1-\rho_{ij}{}^2)\sigma_i}{2} & 0 & \frac{\rho_{ij}(1-\rho_{ij}{}^2)\sigma_j}{2} & (1-\rho_{ij}{}^2)^2 \end{pmatrix}.$$

Let $\vec{\psi}_2 = (\mu_p, \sigma_p^2, \vec{\sigma}^{(1)})$ be the parameters concerning the censored variable X_p. Based on the form of the likelihood function, we will first find the information matrix of $\vec{\theta} = (\alpha, \vec{\beta}, \sigma^*)$ where $\alpha = \mu_p - \vec{\beta}'\vec{\mu}^{(1)}$ and $\vec{\beta}' = \vec{\sigma}^{(1)\prime}\Sigma_{11}^{-1}$ with $\sigma^{*2} = \sigma_p^2 - \vec{\sigma}^{(1)\prime}\Sigma_{11}^{-1}\vec{\sigma}^{(1)}$. Note that

$$\ln L_2(\vec{\theta}) = -\frac{n}{2}\ln\sigma^{*2} + \sum_{x_{pj}\le D_j}\ln\Phi(w_j) - \frac{1}{2\sigma^{*2}}\sum_{x_{pj}>D_j}(y_{pj} - \alpha - \vec{\beta}'\vec{y}_j^{(1)})^2 + Constant,$$

where $\vec{Y}_j^{(1)}$ is the vector of the jth observation on the first $p-1$ variables. The elements of $\partial^2 \ln L_2(\vec{\theta})/\partial\vec{\theta}\partial\vec{\theta}'$ are

$$\frac{\partial^2 \ln L_2}{\partial\alpha^2} = -\frac{n}{\sigma^{*2}} - \frac{1}{\sigma^{*2}}\sum_{x_{pj}\le D_j}\left(R^2(w_j) + w_jR(w_j)\right),$$

$$\frac{\partial^2 \ln L_2}{\partial\alpha\partial\vec{\beta}_i} = -\frac{1}{\sigma^{*2}}\sum_{x_{pj}>D_j} y_{ij} - \frac{1}{\sigma^{*2}}\sum_{x_{pj}\le D_j} y_{ij}\left(w_jR(w_j) + R^2(w_j)\right),$$

$$\frac{\partial^2 \ln L_2}{\partial\alpha\partial\sigma^*} = -\frac{2}{\sigma^{*3}}\sum_{x_{pj}>D_j}(y_{pj} - \alpha - \vec{\beta}'\vec{y}_j^{(1)}) - \frac{1}{\sigma^{*2}}\sum_{x_{pj}\le D_j}\left(w_jR^2(w_j) + w_j^2R(w_j) - R(w_j)\right),$$

$$\frac{\partial^2 \ln L_2}{\partial\vec{\beta}_i^2} = -\frac{1}{\sigma^{*2}}\sum_{x_{pj}>D_j} y_{ij}^2 - \frac{1}{\sigma^{*2}}\sum_{x_{pj}\le D_j} y_{ij}^2\left(w_jR(w_j) + R^2(w_j)\right),$$

$$\frac{\partial^2 \ln L_2}{\partial\vec{\beta}_i\partial\sigma^*} = -\frac{2}{\sigma^{*3}}\sum_{x_{pj}>D_j}(y_{pj} - \alpha - \vec{\beta}'\vec{y}_j^{(1)})y_{ij} - \frac{1}{\sigma^{*2}}\sum_{x_{pj}\le D_j} y_{ij}\left(w_jR^2(w_j) + w_j^2R(w_j) - R(w_j)\right),$$

and

$$\frac{\partial^2 \ln L_2}{\partial \sigma^{*2}} = \frac{n}{\sigma^{*2}} - \frac{3}{\sigma^{*4}} \sum_{x_{pj} > D_j} (y_{pj} - \alpha - \vec{\beta}' \vec{y}_j^{(1)})^2 - \frac{1}{\sigma^{*2}} \sum_{x_{pj} \le D_j} \left(w_j^2 R^2(w_j) + w_j^3 R(w_j) - 2 w_j R(w_j) \right),$$

where $R(w) = \phi(w)/\Phi(w)$. Let $J = (\partial \vec{\theta}_j / \partial \vec{\psi}_{2i})$ and note that $I(\vec{\psi}_2) = J I(\vec{\theta}) J'$. By applying Cramér's delta-method (Cramér, 1946; Serfling, 1980, Section 3.3), the large sample covariance matrix of $\hat{E}(\vec{X})$ is computed from

$$Cov\left(\hat{E}(\vec{X})\right) = \left(\frac{\partial E(\vec{X})}{\partial \vec{\gamma}}\right)'_{\hat{\vec{\gamma}}} Cov\left(\hat{\vec{\gamma}}\right) \left(\frac{\partial E(\vec{X})}{\partial \vec{\gamma}}\right)_{\hat{\vec{\gamma}}},$$

where $Cov(\hat{\vec{\gamma}}) = I^{-1}(\hat{\vec{\gamma}})$ and $I(\hat{\vec{\gamma}})$ is the observed information matrix of $\vec{\gamma} = (\vec{\psi}_1, \vec{\psi}_2)$.

The E-M algorithm coupled with the invariance property of the MLE's provide us with maximum likelihood estimates for the mean vector. After estimating $\vec{\mu}$ and Σ, the maximum likelihood estimate of the mean vector in the original scale is obtained using appropriate forms from Table 1. Further, one can obtain a large sample confidence region for the mean vector in the original scale. Using the asymptotic normality result in the appendix, one can show that given $\vec{\lambda}$, a 100(1-α)% confidence region for $E(\vec{X})$ is obtained from $(\hat{E}(\vec{X}) - E(\vec{X}))' \hat{Cov}(\hat{E}(\vec{X}))^{-1} (\hat{E}(\vec{X}) - E(\vec{X})) \le \chi_p^2(1-\alpha)$. (See Anderson, 1984, p. 163 - 165.) For example, when $p = 2$ and $\vec{\lambda} = (0, 0)$ we obtain $\vec{d}' \hat{\Gamma}^{-1} \vec{d} \le \chi_2^2(1-\alpha)$ where $\vec{d}' = (\hat{E}_1 - E_1, \hat{E}_2 - E_2)$, $E_i = E(X_i) = \exp(\mu + \sigma_i^2/2)$, and

$$Cov(\hat{E}(X_1), \hat{E}(X_2)) = \frac{1}{N} \begin{pmatrix} E_1^2 \sigma_1^2 (1 + \sigma_1^2/2) & E_1 E_2 \sigma_{12} (1 + \sigma_{12}/2) \\ E_1 E_2 \sigma_{12} (1 + \sigma_{12}/2) & E_2^2 \sigma_2^2 (1 + \sigma_2^2/2) \end{pmatrix}$$

for the confidence region of the mean vector.

Suppose we use

$$\hat{E}(X_i) = \begin{cases} \hat{E}(\lambda Y_i + 1)^{\frac{1}{\lambda}}, & \lambda \neq 0, \\ \exp(\hat{\mu}_i + \hat{\sigma}_i^2/2), & \lambda = 0 \end{cases}$$

to estimate the mean of the ith margin $E(X_i)$. One can show for an integer $m = 1/\lambda$,

$$E(X) = \sum_{\text{Even } j=0}^{m} \begin{pmatrix} m \\ j \end{pmatrix} \lambda^j (\lambda\mu + 1)^{(m-j)} \frac{\sigma^j j!.}{2^{j/2}(j/2)!}$$

More generally, when $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$ one can show

$$Cov(\hat{E}(X_1), \hat{E}(X_2)) = \frac{1}{N} \begin{bmatrix} a_1^2 \sigma_1^2 + b_1^2 \sigma_1^2/2 & \sigma_{12}(a_1 a_2 + b_1 b_2 \rho) \\ \sigma_{12}(a_1 a_2 + b_1 b_2 \rho) & a_2^2 \sigma_2^2 + b_2^2 \sigma_2^2/2 \end{bmatrix}$$

where $a_i = \partial E(X_i)/\partial\mu_i$ and $b_i = \partial E(X_i)/\partial\sigma_i$, for $i = 1, 2$. In particular, we use

$$N \cdot Var(\hat{E}(X)) = \begin{cases} \sigma^2(\partial E(X)/\partial\mu)^2 + \frac{1}{2}\sigma^2(\partial E(X)/\partial\sigma)^2, & \lambda \neq 0, \\ \sigma^2(1 + \sigma^2/2)\exp(2\mu + \sigma^2), & \lambda = 0. \end{cases}$$

5 Simulation Study

A simulation study was conducted to determine if the method described in the previous section provides a reliable confidence region for the mean vector of censored and right skew vectors. Furthermore, the performance of the maximum likelihood method in selecting a power transformation as a function of censoring proportion, the correlation between the two variables, and sample size was investigated. It is also of interest to study the coverage rates of univariate confidence intervals and bivariate confidence regions based on a selected transformation. The bivariate data sets produced for the simulations have 30, 50, and 100 observations with 20%, 10% or no censoring. Several classes of Applied-Selected rules (AS) were implemented. Since the findings were similar we only report the results for the following case when $\rho = 0.9$. We consider the rule AS where we apply a transformation from $\Lambda_A = \{(\lambda_1, \lambda_2) : (0, 0), (1/2, 1/2), (1, 1)\}$ to a data set X to produce a bivariate normal data set with $\vec{\mu} = (\mu_1, \mu_2) = (7, 9)$, unit variances, and the correlation of coefficient ρ=0.9, and select a transformation (λ_1, λ_2) from $\Lambda_S = \{(\lambda_1, \lambda_2) : (0, 0), (0, 1), (1/2, 1/2), (1, 0), (1, 1)\}$ for which the log-likelihood is maximized.

To estimate the mean vector in the original scale, the variables were transformed first to achieve normality. Transformation parameters were selected from a list of the candidates in Λ_S using the MLE method. In the transformed scale, $\vec{\mu}$ and Σ were estimated under normality via E-M algorithm. At each iteration k, the updated log-likelihood, say L_k, were monitored and the iterations were determined to have converged when $| (L_{k+1} - L_k)/L_k | < 10^{-6}$. If the iterations did not converge within 30 stages, then the algorithm was terminated. In general, the algorithm rarely failed to converge within 30 iterations. When a correct transformation was chosen, the process mostly converged within 10 iterations.

Table 2 shows the performance of MLE method for choosing an effective transformation and the estimated coverage rates of 90% a confidence region and marginal confidence intervals based on the selected transformation with 1,000 replications of the experiment. The findings are summarized below. We note that in selecting the normalizing transformation, the method performed better given less number of choices. A shorter candidate list of power transformations reflect a priori knowledge of an effective transformation. Hence, one may consider placing prior probabilities on a longer list. The method also worked better for larger samples. The results show that it was more likely to select an appropriate transformation for highly skew data (lognormal data) than for moderately skew or normal data. At the same time, when data were skew, it was more likely to select a transformation than no transformation. On the other hand, when data were normal, selecting a transformation was as likely as selecting no transformation. We note that for data sets with a high correlation ($\rho = 0.9$), the transformation selection method performed better than the data sets with a low correlation. There were slight degradation in the selection rates of correct transformations for censored data sets. The selection rate of the correct transformation was the best for highly skew data sets with a high correlation and a shorter list of candidate power transformations.

6 Application

The Garden State Race Track, Inc. (G.S.R.T.) owns and operates a horse race facility located in Camden County, New Jersey. In 1990, as a result of complaints, representatives of the Department of Environmental Protection inspected the facility and observed that there was a discharge of pollutants from horse washdown into the waters of the state. During the investigation, the representatives collected samples of the discharge and found significant amount of contamination. At the time of the discharge, G.S.R.T. did not possess a Pollution Discharge Elimination System permit authorizing the discharge. Therefore, G.S.R.T. was ordered to cease immediately all discharge of pollutants at the facility and begin conducting a water quality monitoring program until a permit was final and effective. We analyzed 39 weekly measurements of Biochemical Oxygen Demand (BOD) and Total Suspended Solids (TSS) from the water quality monitoring samples in Table 3.

Total Suspended Solids (TSS) are solid matters in water that can be trapped by a 2-micron filter. High concentrations of suspended solids can cause many problems for stream health and aquatic life. The decrease in water clarity caused by TSS can affect the ability of fish to see and catch food. Suspended sediment can also clog fish gills, reduce growth rates, decrease resistance to disease, and prevent egg and larvae development. When suspended solids settle to the bottom of a water body, they can smother the eggs of fish and insects and newly hatched insect larvae. High TSS can prevent light from reaching submerged plants resulting reduced rate of photosynthesis and less dissolved oxygen to be released into the water by plants. High TSS can also increase the temperature of water from sunlight, which causes dissolved oxygen levels to fall even further, leads to fish kills, and harms aquatic life in many other ways. Biochemical Oxygen Demand (BOD) measures the amount of oxygen consumed by microorganisms in decomposing inorganic matter. Thus, BOD directly affects the amount of dissolved oxygen in rivers and streams. The greater the BOD, the more rapidly oxygen is depleted in the stream. This means less oxygen is available to higher forms of aquatic life. At High BOD values aquatic organisms become stressed, suffocate, and die. Raboud and Pintile (1992) study the effects of the environment on the growth of fish. Values of TSS and BOD are positive and skew to the right. The data set contains several extreme observations at both low and high levels. Some values of BOD are censored below the detection limit of 2 mg/l. It is of interest to model the joint distribution of BOD and TSS. Since the size of the sample is not large, both univariate and bivariate model and parameter uncertainties exist. Once a suitable models is identified, we would like to build interval estimates for the individual means and a confidence region for the joint means of BOD and TSS.

Three measurements of BOD were censored below the detection limit of 2 mg/l. The sample mean and standard deviation of BOD computed by setting the censored amount equal to the detection limit were 46.98 mg/l and 48.35 mg/l. Similarly, the sample mean and standard deviation of TSS were found to be 82.46 and 104.05 mg/l. Using univariate Q-Q plots and tests of normality, we identified $\Lambda_1 = \{0, 1/3, 1\}$ and $\Lambda_2 = \{0, 1/4, 1/3, 1/2, 1\}$ for BOD and TSS. Various Box-Cox transformations were applied to the data to find the trans-

Table 2: Transformation Selection Rates and 90% Confidence Regions and Intervals Based on 1,000 Replications and $\rho = 0.9$.

True (λ_1, λ_2)	N	Censoring	Selection Rate (λ_1, λ_2) (0,0)	(0,1)	$(\frac{1}{2},\frac{1}{2})$	(1,0)	(1,1)	Coverage Rate of $\mathrm{E}(X_1, X_2)$	Coverage Rate of $\mathrm{E}(X_1)$	Coverage Rate of $\mathrm{E}(X_2)$
(0,0)	30	0%	.983	.000	.017	.000	.000	.803	.851	.868
		10%	.962	.006	.032	.000	.000	.817	.860	.854
		20%	.929	.035	.036	.000	.000	.821	.845	.864
	50	0%	.998	.000	.002	.000	.000	.871	.897	.895
		10%	.992	.001	.007	.000	.000	.857	.880	881
		20%	.990	.002	.008	.000	.000	.864	.885	.878
	100	0%	1.000	.000	.000	.000	.000	.875	.893	.896
		10%	.999	.000	.001	.000	.000	.878	.889	900
		20%	1.000	.000	.000	.000	.000	.876	.892	.897
$(\frac{1}{2},\frac{1}{2})$	30	0%	.235	.136	.232	.159	.238	.868	.882	.882
		10%	.195	.210	.175	.164	.256	.865	.882	.867
		20%	.166	.233	.159	.171	.271	.882	886	.877
	50	0%	.239	.112	.336	.110	.203	.897	.907	.905
		10%	.187	.169	.290	.124	.230	.876	.890	.896
		20%	.186	.186	.261	.151	.216	.866	.892	.874
	100	0%	.174	.045	.587	.049	.145	.893	.898	.911
		10%	.145	.092	.493	.088	.182	.910	.913	.906
		20%	.153	.101	.465	.091	.190	.882	.903	.892
(1,1)	30	0%	.125	.201	.063	.275	.336	.862	.883	.902
		10%	.110	.203	.058	.240	.389	.857	.890	.891
		20%	.116	.264	.042	.218	.360	.861	.881	.875
	50	0%	.083	.163	.102	.207	.445	.878	.899	.881
		10%	.097	.176	.079	.214	.434	.873	.889	.881
		20%	.091	.173	.075	.206	.455	.842	.873	.870
	100	0%	.046	.083	.160	.154	.557	.884	.895	.891
		10%	.056	.102	.129	.173	.540	.899	.906	.885
		20%	.054	.108	.122	.190	.526	.874	.899	.895

Table 3: Water Quality Weekly Monitoring Data Collected at Race Track

Analyte	Amount (mg/l)									
TSS	50	120	580	26	120	66	220	82	110	26
BOD	31	10	120	26	98	150	97	170	180	50
TSS	21	28	290	120	61	170	130	57	15	40
BOD	64	70	17	10	4.6	45	74	130	22	47
TSS	4	15	210	63	99	12	62	39	12	34
BOD	21	2.6	23	42	72	15	53.9	57	24.7	33.6
TSS	36	46	64	78	35	27	6	33	9	
BOD	4.1	2.9	13.1	18.1	17.6	10	<2	<2	<2	

formation that maximizes the likelihood. P-values of the Shapiro-Wilk tests for normality on the transformed data yield $(0.997, 0.079, 0.001)$ and $(0.051, 0.297, 0.241, 0.053, 0.001)$ for λ_1 and λ_2, respectively. As potential candidates, we considered

$$\Lambda = \{(0,0),(0,1/4),(0,1/3),(0,1/2),(0,1),(1/3,0),(1/3,1/3),(1/3,1/2),(1/3,1)\}.$$

For each $\vec{\lambda} \in \Lambda$, we obtained MLE's of $(\mu_1, \sigma_2, \mu_2, \sigma_2, \rho)$ using the E-M algorithm and computed the log-likelihood. The values of the log-likelihoods are

$$(-325.08, -321.77, -321.48, -322.0, -330.85, -328.46, -324.82, -325.33, -334.12).$$

According to the marginal analysis, one should select λ_1=0 and λ_2=1/4; however, the transformation (0,1/3) results in the largest likelihood and neither (0,1/4) nor the bivariate log-normal distribution is selected. Figure 1 displays a scatter plot of BOD vs. TSS with the contours of a transformed distribution with $\vec{\lambda} = (0, 1/3)$. The contours show seven levels at every 15th percentile between the 5th and the 95th percentile. Furthermore, note that the values of the conditional expectations are affected by selecting an incorrect transformation. The selection of the optimal value for $\vec{\lambda}$ is very important as it affects the estimation of the E-M algorithm parameters and in turn leads to inaccurate estimates in the original scale.

Table 4 provides summaries of the analysis under $(\lambda_1, \lambda_2) = (0, 0)$ and $(0, 1/3)$. Compare the sample estimates of the mean vector (82.46, 46.98) to the delta method estimates of (84.40,46.20) with $(\lambda_1, \lambda_2) = (0, 1/3)$. Note that the percentage of censoring is about 8%. Higher percentages lead to a lower estimated mean in the original scale. While the means are comparable in this case, the delta-method estimates show smaller variability. This will lead to a smaller confidence region around the mean vector. Comparing the estimates for $(\lambda_1, \lambda_2) = (0, 0)$ and $(\lambda_1, \lambda_2) = (0, 1/3)$, the mean estimate of BOD with the latter are much lower than that with the former by at least 16 mg/l. Using $\vec{\lambda} = (0, 1/3)$, Figure 2 displays the contours of the bivariate normal with transformed observations. It is interesting to note that the point $(120, 580)$, an extreme value under bivariate log-normal or $(0, 1/3)$ transformation,

Table 4: Parameter Estimation of Race Track Data

Estimator			Estimates	
			$\vec{\lambda} = (0, 0)$	$\vec{\lambda} = (0, 1/3)$
Maximum Likelihood	$\hat{\mu}$	TSS	3.88	3.88
		BOD	3.16	6.31
	$\hat{\sigma}$	TSS	1.06	1.06
		BOD	1.40	3.96
	$\hat{\rho}$		0.46	0.47
Delta-Method	$\hat{E}(\mathrm{X})$	TSS	84.40	84.40
		BOD	62.48	46.20
	$Std.Dev(\hat{E}(\mathrm{X}))$	TSS	17.85	17.85
		BOD	19.64	8.11
	$Corr(\hat{E}(X_1), \hat{E}(X_2))$		0.35	0.42

is not longer as extreme after the transformation. It falls within the $95th$ density contour. Figure 3 shows the estimated 95% confidence region for the mean vector.

Acknowledgment: This research was supported by grants from USEPA and USGS.

APPENDIX

We sketch the asymptotic properties of the maximum likelihood estimates of $\vec{\theta}$ (Cox and Hinkley, 1974). Let $L_{\vec{\theta}} = (L_\alpha, L_{\vec{\beta}}, L_{\sigma^*})$ be the vector of the maximum likelihood equations. From the logarithm of the likelihood we obtain

$$L_\alpha = \frac{1}{\sigma^{*2}} \sum_{j=1}^{N} I_{\{x_{pj} > D_j\}} (y_{pj} - \alpha - \vec{\beta}' \vec{y}_j^{(1)}) - \frac{1}{\sigma^*} \sum_{j=1}^{N} I_{\{x_{pj} \leq D_j\}} R(w_j),$$

$$L_{\vec{\beta}_i} = \frac{1}{\sigma^{*2}} \sum_{j=1}^{N} I_{\{x_{pj} > D_j\}} (y_{pj} - \alpha - \vec{\beta}' \vec{y}_j^{(1)}) y_{ij} - \frac{1}{\sigma^*} \sum_{j=1}^{N} I_{\{x_{pj} \leq D_j\}} R(w_j) y_{ij},$$

and

$$L_{\sigma^*} = -\frac{n}{\sigma^*} + \frac{1}{\sigma^{*3}} \sum_{j=1}^{N} I_{\{x_{pj} > D_j\}} (y_{pj} - \alpha - \vec{\beta}' \vec{y}_j^{(1)})^2 - \frac{1}{\sigma^*} \sum_{j=1}^{N} I_{\{x_{pj} \leq D_j\}} R(w_j) w_j.$$

The log-likelihood equations L_α, $L_{\vec{\beta}}$, and L_σ are the sums of non-i.i.d. zero-mean random variables. For each element of $\vec{\theta}$, one can show $E(L_{\vec{\theta}_i}) = 0$. Additionally, the derivatives of $E(L_{\vec{\theta}})$ with respect to $\vec{\theta}$ can be obtained by differentiating the log-likelihood equations under the expectations (see Billingsley, 1995, p. 212). Then, the covariance matrix of the likelihood functions is $Cov(L_{\vec{\theta}}) = I_N(\vec{\theta})$ where $I_N(\vec{\theta}) = E(\partial L_{\vec{\theta}}/\partial\vec{\theta})^2 = -E(\partial^2 \ln L(\vec{\theta})/\partial\vec{\theta}\partial\vec{\theta}')$. The elements of $I_N(\vec{\theta})$ include

$$E\left(\frac{\partial^2 \ln L}{\partial \alpha^2}\right) = -\frac{1}{\sigma^{*2}}\sum_{j=1}^{N}(1 - \Phi(w_j) + \Phi(w_j)R^2(w_j) + w_j\phi(w_j)),$$

$$E\left(\frac{\partial^2 \ln L}{\partial \alpha \partial \vec{\beta}_i}\right) = -\frac{\mu_i}{\sigma^{*2}}\sum_{j=1}^{N}(1 - \Phi(w_j) + w_j\phi(w_j) + \Phi(w_j)R^2(w_j)),$$

$$E\left(\frac{\partial^2 \ln L}{\partial \alpha \partial \sigma^*}\right) = -\frac{1}{\sigma^{*2}}\sum_{j=1}^{N}(\phi(w_j) + w_j^2\phi(w_j) + w_j\Phi(w_j)R^2(w_j)),$$

$$E\left(\frac{\partial^2 \ln L}{\partial \beta_i^2}\right) = -\frac{\mu_i^2 + \sigma_i^2}{\sigma^{*2}}\sum_{j=1}^{N}(1 - \Phi(w_j) + w_j\phi(w_j) + \Phi(w_j)R^2(w_j)),$$

$$E\left(\frac{\partial^2 \ln L}{\partial \vec{\beta}_i \partial \sigma^*}\right) = -\frac{\mu_i}{\sigma^{*2}}\sum_{j=1}^{N}(\phi(w_j) + w_j^2\phi(w_j) + w_j\Phi(w_j)R^2(w_j)),$$

and

$$E\left(\frac{\partial^2 \ln L}{\partial \sigma^{*2}}\right) = -\frac{1}{\sigma^{*2}}\sum_{i=1}^{N}(2 - 2\Phi(w_i) + w_i\phi(w_i) + w_i^3\phi(w_i) + w_i^2\Phi(w_i)R^2(w_i)).$$

There exist the second derivatives of $L_{\vec{\theta}}$ (the third derivatives of $\ln L(\vec{\theta})$) with respect to $\vec{\hat{\theta}}$ and an integrable function $G(w)$ with bounded expectation satisfying $\left|\frac{\partial^2 L_\theta}{\partial\theta\partial\theta'}\right| < N^{-1}\sum_w G(w)$ for all w in the neighborhood of θ, an element of $\vec{\theta}$. Let $F_N(w)$ be an empirical distribution defined as $F_N(w) = N^{-1}I_{\{w_i<w\}}$. We see that all the terms of $E\left(\partial L_{\vec{\theta}}/\partial\vec{\theta}\right)$ divided by N are

of the form $N^{-1}\sum_{i=1}^{N} q(w_i) = \int_{-\infty}^{\infty} q(w)dF_N(w)$. Assume that $F_N(w) \to F(w)$, an integrable distribution function. Since $w^2\Phi(w)R^2(w)$ is continuous in addition to

$$\lim_{w\to\pm\infty} w^2\Phi(w)R^2(w) = \lim_{w\to\pm\infty} \frac{w^2}{\exp(w^2)\int_{-\infty}^{w}\exp(-t^2/2)dt} = 0,$$

it is bounded. Similarly, other additive components of $q(w_i)$ are also continuous and bounded. Since $q(w_i)$ are continuous and bounded and $F_N \to F$, by Helly-Bray theorem (Rao, 1973, p. 117), $\int_{-\infty}^{\infty} q(w)dF_N(w) \to \int_{-\infty}^{\infty} q(w)dF(w)$. Therefore, assuming $I(\vec{\theta})$ is positive definite, each term of $N^{-1}I_N(\vec{\theta})$ will converge to the appropriate integral expression, say $I(\theta_i)$. For example,

$$-\lim_{N\to\infty}\frac{1}{N}E\left(\frac{\partial^2 \ln L}{\partial\alpha^2}\right) = \frac{1}{\sigma^{*2}}\int_{-\infty}^{\infty}(1-\Phi(w)+R^2(w)\Phi(w)+w\phi(w))dF(w) = I(\alpha).$$

After Taylor's series expansion of $L_{\hat{\vec{\theta}}}$ about $\vec{\theta}$, one can show that

$$L_{\hat{\vec{\theta}}} = L_{\vec{\theta}} + (\hat{\vec{\theta}}-\vec{\theta})\frac{\partial L_{\vec{\theta}}}{\partial\vec{\theta}} + \frac{1}{2}(\hat{\vec{\theta}}-\vec{\theta})\left[\frac{\partial^2 L_{\vec{\theta}}}{\partial\vec{\theta}\partial\vec{\theta}'}\right]_{\tilde{\vec{\theta}}}(\hat{\vec{\theta}}-\vec{\theta})' = 0,$$

where $\tilde{\theta}_i$ lies between $\hat{\theta}_i$ and θ_i for each component of $\tilde{\vec{\theta}}$. Thus,

$$\sqrt{N}(\hat{\vec{\theta}}-\vec{\theta}) = -\frac{1}{\sqrt{N}}L_{\vec{\theta}}\cdot\left[\frac{1}{N}\left(\frac{\partial L_{\vec{\theta}}}{\partial\vec{\theta}}\right)+\frac{1}{2N}\left[\frac{\partial^2 L_{\vec{\theta}}}{\partial\vec{\theta}\partial\vec{\theta}}\right]'_{\tilde{\vec{\theta}}}(\hat{\vec{\theta}}-\vec{\theta})'\right]^{-1}.$$

By the multivariate central limit result from Lindeberg-Feller's theorem (Serfling, 1980, p. 30) and Helly-Bray theorem, it follows that $-N^{-1/2}L_{\vec{\theta}} \to N(\vec{0}, I(\vec{\theta}))$ in distribution and $N^{-1}\partial L_{\vec{\theta}}(\vec{\theta})/\partial\vec{\theta} \to -I(\vec{\theta})$ as $N\to\infty$. The quantity $(\hat{\vec{\theta}}-\vec{\theta}) = o_p(1)$ is multiplied by $N^{-1}\sum_w G(w)$, which is $O_p(1)$. By Slutsky's theorem, $\sqrt{N}(\hat{\vec{\theta}}-\vec{\theta}) \to I(\vec{\theta})^{-1}N(\vec{0}, I(\vec{\theta}))$. Hence, $\sqrt{N}(\hat{\vec{\theta}}-\vec{\theta}) \to N(\vec{0}, I(\vec{\theta})^{-1})$.

References

[1] Aitchison, J., and Brown, J.A.C. (1973). The Lognormal Distribution. Cambridge Press.

[2] Anderson, T. W. (1984). An Introduction to Multivariate Analysis. Second Edition. Wiely.

[3] Billingsley, P. (1995). Probability and Measure. Third Edition. Wiley Interscience.

[4] Box, G. E., and Cox, D. R. (1964). An Analysis of Transformed Data. Journal of Royal Statistical Society. B, 39, 211-252.

[5] Chen, G. and Lockhart, R. A. (1997). Box-Cox transformed linear models: a parameter-based asymptotic approach. The Canadian Journal of Statistics, vol. 25, No 4, pp. 517-529.

[6] Chen, G., Lockhart, R. A. and Stephens, M. A. (2002). Box-Cox transformations in linear models: Large sample theory and tests of normality (with discussion). The Canadian Journal of Statistics, Vol. 30, No 2, pp. 177-234

[7] Chen, H. (1995). Test Following Transformations. The Annals of Statistics. 23, 5, 1587-1593.

[8] Chen, H., and Loh, W. Y. (1992). Bounds on AREs of Tests Based on Transformed Data. Annals of Statistics. Vol. 20, No. 3, 1485-1500.

[9] Cox, D. R., and Hinkley, D. V. (1974). Theoretical Statistics. Chapman and Hall, London.

[10] Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press, Princeton, New Jersey.

[11] Demster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the E-M Algorithm. Journal of the Royal Statistical Society. B, 39, 1-38.

[12] Gnanadesikan, R. (1977). Method for Statistical Data Analysis of Multivariate Observations. John Wiley & Sons .

[13] Hernandez, F., and Johnson, R. A. (1980). The Large Sample Behavior of Transformations to Normality. Journal of American Statistical Association. Volume 75, No. 372, 855-861.

[14] Hinkley, D. V. (1975). On Power Transformations to Symmetry, Biometrika, Vol. 62, No. 1, 101-111.

[15] Hinkley, D. V., and Runger, G. (1984). Analysis of Transformed Data. Journal of American Statistical Association. Volume 79, No. 386, 302-309.

[16] Joe, H. (1994). Multivariate Extreme-value Distributions with Applications to Environmental Data. Canadian Journal of Statistics, Vol. 22, No 1, pp. 47- 64.

[17] Johnson, R. A., and Wichern, D. W. (1998). Applied Multivariate Analysis. Fourth Edition, Prentice Hall.

[18] Lehmann, E. L. (1975). Nonparametric: Statistical Methods Based on Ranks. Holden-Day, San Francisco.

[19] Magnus, J. R. and Neudecker, H. (1988). Matrix Differential Calculus. John Wiley & Sons.

[20] McCullagh, P., and Nelder, J. A. (1989). Generalized Linear Models. Chapman and Hall, New York.

[21] McLachlan, G. J., and Krishnan, T. (1997). The E-M algorithm and Extensions. John Wiley & Sons, Inc.

[22] Miller, R. G. (1986). Beyond ANOVA, Basics of Applied Statistics. Wiley.

[23] Ott, W. R. (1995). Environmental Statistics and Data Analysis. Lewis Publishers.

[24] Raboud, J. and Pintilie (1992). The Efect of Environment on the Growth of Fish. The Canadian Journal of Statistics, Vol. 20, No. 2, 233-239.

[25] Rao, C. R. (1973). Linear Statistical Inference and Its Application. Second Edition. Wiley, New York.

[26] Sakia, R. M. (1992). Box-Cox Transformation Technique: A Review. The Statistician. 42, 169-178.

[27] Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics. Wiley, New York.

[28] Shumway, R. H., Azari, A.S., and Johnson, P. (1989). Transformation for Environmental Data with Detection Limits. Technometrics. Volume 31, No. 3, 347-356.

[29] Tukey, J. W. (1957). On the Comparative Anatomy of Transformations. Annals of Mathematical Statistics. Volume 28, Issue 3, 602-632.

[30] U.S. EPA (1992). Methods for Evaluating the Attainment of Clean up Standards. Vol. 2, Groundwater, EPA/230-R-92-014, Office of Policy, Planning and Evaluation.