Journal of Statistical Research *ISSN 0256 - 422 X*
2008, Vol. 42, No. 2, pp. 21-39
Bangladesh

ON THE LINEARLY TRUNCATED BIVARIATE NORMAL DISTRIBUTION

David Bellhouse

Department of Statistical and Actuarial Sciences
University of Western Ontario, London, Ontario N6A 5B7, Canada
Email: bellhouse@stats.uwo.ca

Li Wang

Department of Statistical and Actuarial Sciences
University of Western Ontario, London, Ontario N6A 5B7, Canada
Email:

SUMMARY

Motivated by a problem in archaeological data analysis, we examine through simulation the effect of linear truncation on density estimation in the bivariate normal distribution. It is shown that the estimation procedure performs poorly when the truncation line is parallel to the minor axis of the elliptical contours of the distribution. The estimation also worsens with increasing correlation and increasing disparity in the values of the variances in the distribution. Reconstruction of the bivariate normal distribution is best achieved when truncation is parallel to the major axis and there is only about 10% of the data that is missing.

Keywords and phrases: archaeological data analysis, bivariate normal, distribution estimation, linear truncation.

AMS Classification: Place Classification here. Leave as is, if there is no classification

1 Introduction

Consider bivariate data denoted by the pair of observations (x, y). The most common form of truncation of data of this type is when x or y or both are each restricted to an interval of values of on the real line. This situation and the related estimation problems are described in detail in Schneider (1986), for example, and elsewhere. Here we consider a less common form of truncation, that of linear truncation. In this situation the parent distribution is defined as usual for $-\infty < x < \infty$ and $-\infty < y < \infty$, but the truncation restricts the observable data to the subset of points in which $y \geq a + bx$. The line $y = a + bx$ is called the truncation line.

Assume that (x, y) are bivariate normal with parameters μ_1, μ_2, σ_1, σ_2 and ρ. The likelihood for linearly truncated data with n independent sets of observations (x_i, y_i) for $i = 1, \ldots, n$ is given by

$$-n\ln(\sigma_1) - n\ln(\sigma_2) - \frac{n}{2}\ln(1-\rho^2) - n\ln(G) - \frac{1}{2(1-\rho^2)}\times$$
$$\left(\frac{n(\bar{x}-\mu_1)^2 + ns_1^2}{\sigma_1^2} - 2\rho\frac{n(\bar{x}-\mu_1)(\bar{y}-\mu_2) + nrs_1s_2}{\sigma_1\sigma_2} + \frac{n(\bar{y}-\mu_2)^2 + ns_2^2}{\sigma_2^2}\right) \quad (1.1)$$

where $\bar{x}$ and $\bar{y}$ are sample means, ns_1^2, ns_2^2 and nrs_1s_2 are sample sums of squares and cross products, and

$$G = \int_\zeta^\infty \frac{1}{\sqrt{2\pi}}\exp(-t^2/2)dt = 1 - \Phi(\zeta). \quad (1.2)$$

In (1.2), $\Phi(\zeta)$ is the cumulative distribution function for the standard normal density $\phi(\zeta)$, and $\zeta = (a-\mu)/\sigma$ is the standardized value with $\mu = \mu_2 - b\mu_1$ and

$$\sigma^2 = b^2\sigma_1^2 + \sigma_2^2 - 2b\rho\sigma_1\sigma_2. \quad (1.3)$$

Nath (1971) has obtained iterative equations to calculate the maximum likelihood estimates in the special case for $b = -1$. His equations contain some minor errors in them. Nath (1972) has obtained moment estimates in the general case for linear truncation. These are equivalent to the maximum likelihood estimates and the minor errors that were in Nath (1971) have been corrected.

The major question that remains is: how good are these maximum likelihood parameter estimates for estimation of the bivariate density?

2 Motivation : Reconstruction of a Pit Feature at an Archaeological Site

The question of the goodness of the maximum likelihood estimates under truncation arose from a project in archaeological data analysis. One possible feature of archaeological sites are pits dug into the ground that contain some artifacts. We obtained a datafile for one such pit from Professor Chris Ellis of the Department of Anthropology at the University of Western Ontario. The pit, from a ten-thousand-year-old site in Ontario, contained 1339 fragments of heat fractured tool fragments that had been purposely burned. The artifacts in the pit had been disturbed in two ways: a tree root had grown through the site, and close to the time of excavation some children had dug a trench through part of the pit.

Figure 1 shows the scatter of artifacts in the pit as seen from above. The empty spot in the middle of the data cloud shows the trench area dug by the children who removed the artifacts from this area. Figure 2 shows the scatter of artifacts as seen from one side. The scatter on the left side of the figure shows the effect of the tree root. In this case there is more variation in the location of the artifacts and they have settled deeper into the soil.

The purpose of our initial data analysis was to reconstruct the form of the pit prior to its disturbance. With this in mind we carried out a simple cluster analysis to differentiate between the disturbed and undisturbed parts of the pit. The result of the cluster analysis is shown in Figure 3. The data neatly divided into two parts with an imaginary line separating the two. As a first step we decided to use the data from the undisturbed section, truncated data, to try to reconstruct the shape of the entire pit.

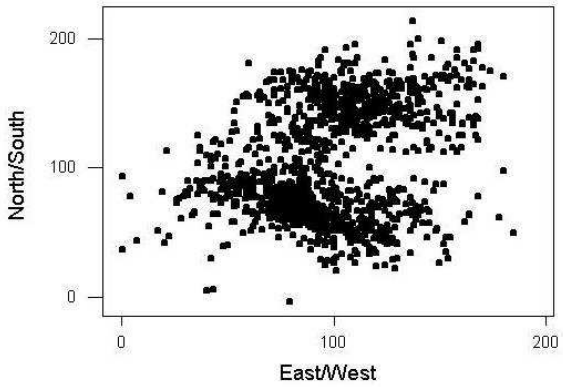

Figure 1: Artifact Scatter as Seen from Above

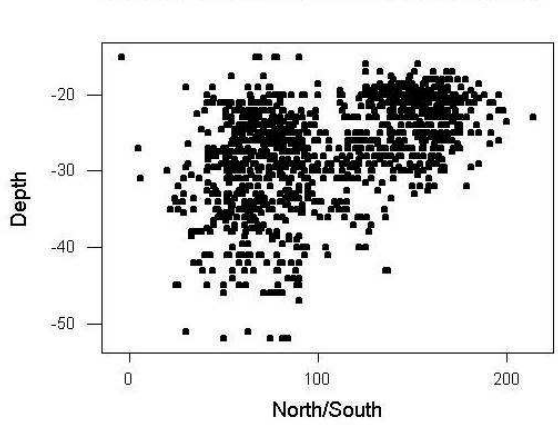

Figure 2: Artifact Scatter as Seen from One Side

From Figure 3 the undisturbed section of the artifact scatter appears to be in the form of a truncated ellipse. One way to reconstruct the pit, at least as viewed from above, is to

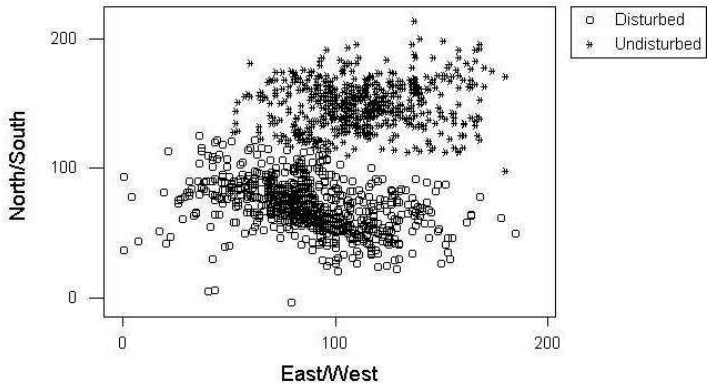

Figure 3: Artifact Scatter as Seen from Above by Disturbed and Undisturbed Sections

reconstruct the full ellipse from the truncated one. Assuming that the (x, y) position of an artifact follows a bivariate normal distribution, then the ellipse given by the equation

$$\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} = -2(1-\rho^2)\ln(1-\alpha) \tag{2.1}$$

contains 100 α% of the distribution; see, for example Kotz, Balakrishnan and Johnson (2000, p. 254). If we can estimate the parameters, then we can reconstruct the ellipse. A bivariate normal distribution may be a reasonable assumption as a working model. When artifacts are thrown or placed in a pit there are typically more artifacts near the center of the pit and fewer by the edges. The problem with our estimation procedure on this particular set of data was that the iterative procedure obtained from Nath (1971, 1972) did not converge. It was then necessary to study the conditions under which the convergence can be obtained. This issue is the focus of the current study.

3 The Iterative Equations for Maximum Likelihood Estimation

On using Nath's (1971) notation, the likelihood equations for iteration obtained from (1.1) are given by:

$$\begin{aligned}
\hat{\mu}_1 &= \bar{x} - (\hat{P} + \hat{\rho}\hat{Q})\hat{\sigma}_1 \\
\hat{\mu}_2 &= \bar{y} - (\hat{Q} + \hat{\rho}\hat{P})\hat{\sigma}_2 \\
\hat{\sigma}_1 &= s_1[(1 + \hat{\xi}\hat{P}\hat{G}/\hat{\sigma}) + \hat{\rho}^2(\hat{A} + \hat{\eta}\hat{Q}\hat{G}/\hat{\sigma}) - (\hat{P} + \hat{\rho}\hat{Q})^2]^{-1/2} \\
\hat{\sigma}_2 &= s_2[(1 + \hat{\eta}\hat{Q}\hat{G}/\hat{\sigma}) + \hat{\rho}^2(\hat{A} + \hat{\xi}\hat{P}\hat{G}/\hat{\sigma}) - (\hat{Q} + \hat{\rho}\hat{P})^2]^{-1/2} \\
\hat{\rho} &= rs_1s_2(\hat{\sigma}_1\hat{\sigma}_2)^{-1}[(1 + \hat{\xi}\hat{P}\hat{G}/\hat{\sigma} + \hat{\eta}\hat{Q}\hat{G}/\hat{\sigma} + \hat{A}) - (\hat{P} + \hat{\rho}\hat{Q})(\hat{Q} + \hat{\rho}\hat{P})/\hat{\rho}]^{-1}
\end{aligned} \tag{3.1}$$

where

$$P = -b\frac{\sigma_1\phi(\zeta)}{\sigma G}, \quad Q = \frac{\sigma_2\phi(\zeta)}{\sigma G}, \qquad A = -b\frac{\sigma_1\sigma_2\zeta(1-\rho^2),\phi(\zeta)}{\sigma^2\rho G}$$
$$\xi = \frac{\zeta(\rho\sigma_2 - b\sigma_1)}{G}, \quad \eta = \frac{\zeta(\sigma_2 - b\rho\sigma_1)}{G},$$

with G given by (1.2) and σ by (1.3). These equations, in a different form, were obtained correctly by Nath (1972). From the information matrix for μ_1, μ_2, σ_1, σ_2 and ρ, the variance-covariance matrix for the estimates of these parameters is of the form

$$\frac{1}{n}\begin{bmatrix} \mathbf{B} & \mathbf{C} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1}$$

where $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are 2×2, 2×3 and 3×3 matrices respectively. The forms of $\mathbf{C}$ and $\mathbf{D}$ are quite complicated and can be obtained via computer algebra. Moreover, $\mathbf{C} \neq \mathbf{0}$ so that, unlike the situation in which there is no truncation, $\hat{\mu}_1$ and $\hat{\mu}_2$ are not independent of $\hat{\sigma}_1$, $\hat{\sigma}_2$ and $\hat{\rho}$. The matrix $\mathbf{B}$ can be easily expressed as

$$\mathbf{B} = \begin{bmatrix} \frac{1}{(1-\rho^2)\sigma_1^2} + b^2\alpha & -\frac{\rho}{(1-\rho^2)\sigma_1\sigma_2} - b\alpha \\ -\frac{\rho}{(1-\rho^2)\sigma_1\sigma_2} - b\alpha & \frac{1}{(1-\rho^2)\sigma_2^2} + \alpha \end{bmatrix},$$

where

$$\alpha = \frac{\phi(\zeta)}{G\sigma^2}\left(\zeta - \frac{\phi(\zeta)}{G}\right).$$

If we treat σ_1, σ_2 and ρ as known constants, then $\mathbf{B}^{-1}/n$ may be used as the variance-covariance matrix of $\hat{\mu}_1$ and $\hat{\mu}_2$. Further,

$$\mathbf{B}^{-1} = \frac{1}{1+\alpha\sigma^2}\begin{bmatrix} \sigma_1^2 + \alpha\sigma_1^2\sigma_2^2(1-\rho^2) & \rho\sigma_1\sigma_2 + b\alpha\sigma_1^2\sigma_2^2(1-\rho^2) \\ \rho\sigma_1\sigma_2 + b\alpha\sigma_1^2\sigma_2^2(1-\rho^2) & \sigma_2^2 + b^2\alpha\sigma_1^2\sigma_2^2(1-\rho^2) \end{bmatrix}.$$

Note that when the truncation line is removed, i.e. $a \to -\infty$, then $\alpha \to 0$. Also, in the same situation $\mathbf{C} \to \mathbf{0}$ so that $\hat{\mu}_1$ and $\hat{\mu}_2$ are independent of $\hat{\sigma}_1$, $\hat{\sigma}_2$ and $\hat{\rho}$, as expected.

4 Simulation Study

For the purposes of our simulation study, the location of the ellipse is irrelevant. It is rather the shape of the ellipse that is important. Consequently, without loss of generality we assume that $\mu_1 = \mu_2 = 0$ in (1.1). A measure of the shape of the ellipse may be obtained from the ratio of the lengths of the major to the minor axes of the ellipse. The lengths may be obtained from the eigenvectors of

$$\begin{bmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{bmatrix},$$

which are

$$\left(\frac{\sigma_1^2 - \sigma_2^2 + \tau}{2\rho\sigma_1\sigma_2}, 1\right) \text{ and } \left(\frac{\sigma_1^2 - \sigma_2^2 - \tau}{2\rho\sigma_1\sigma_2}, 1\right),$$

where

$$\tau = \sqrt{\sigma_1^4 - 2\sigma_1^2\sigma_2^2 + 4\rho^2\sigma_1^2\sigma_2^2 + \sigma_2^4} \tag{4.1}$$

See, for example, Zwillinger (2003, pp. 330 - 331). After some algebra, it may be shown that the ratio of the lengths of the eigenvectors depend only on ρ^2 and the variance ratio σ_1^2/σ_2^2. We consider three types of truncation lines: (1) lines parallel to major axis, (2) lines parallel to minor axis (3) and lines at a 45° angle to the major and minor axes. We refer to the latter axis as the median axis. These are shown in Figure 4.

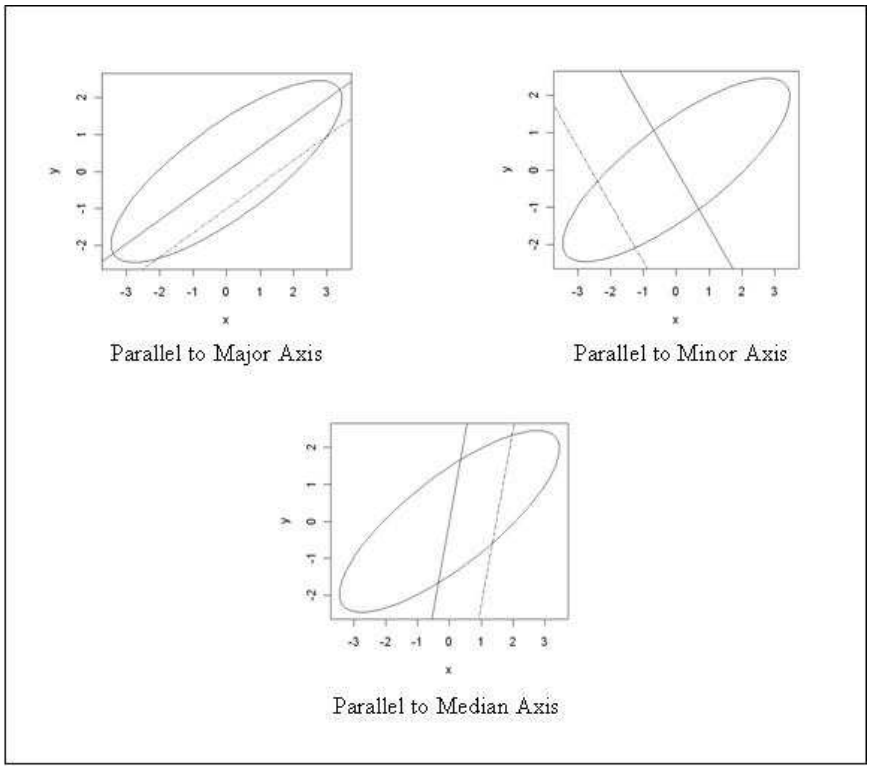

Figure 4: Truncation Types

Since $\mu_1 = \mu_2 = 0$ the major and minor axes of the ellipse, as well as the line between the two (say the median axis) fall on the lines $y = b_{ma}x$, $y = b_{mi}x$ and $y = b_{me}x$. If $\sigma_1/\sigma_2 \geq 1$ and $\rho > 0$, then

$$b_{ma} = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2 + \tau}$$

and

$$b_{mi} = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2 - \tau}.$$

The slope of the line describing the median axis may be obtained from the formula for the tangent of the sum of two angles (45°, and the angle between the x-axis and the major axis of the ellipse), and is given by

$$b_{me} = \frac{1 + b_{ma}}{1 - b_{ma}}.$$

When $\sigma_1 = \sigma_2$, then $b_{ma} = 1$ and the truncation type reduce to truncation on x alone. Consequently, we have not considered this case in the simulation study.

We considered two types of simulation study. In the first, 1000 sets of samples of truncated bivariate normals were used to examine the frequency of convergence of the maximum likelihood estimates as well as the bias of the estimates when convergence was obtained. In the second study, 1000 full samples, with truncated subsamples, were to used to assess how well the full data set is described by the ellipse estimated from the truncated sample.

In the first study truncated samples of size 500 were generated and truncation amounts of 10%, 30% or 50% were chosen. The coefficients in the truncation line $y = a + bx$ were obtained from the parameters. The term b is one of b_{ma}, b_{mi} or b_{me}, while a depends on the truncation: for 50%, $a = 0$; 30%, $a = -0.5244\sigma$; and 10% truncation $a = -1/28155\sigma$. The parameter settings were: $\rho = 0.1, 0.3, 0.5, 0.7, 0.9$; $\sigma_2 = 1$; and $\sigma_1 = 1, 2, 10$. Tables 1 through 18 show the results of this study. They are divided into three groups of six tables, one group for each truncation amount. Within a group the first table shows the likelihood of convergence and the remaining tables show the absolute biases of the parameters.

Table 1: Percentage of Successes at Convergence for Truncation Parallel to the Major Axis

Amount of		ρ				
Truncation	σ_1^2/σ_2^2	0.1	0.3	0.5	0.7	0.9
10%	1	100	100	100	100	100
	2	100	100	100	100	100
	10	100	100	100	100	100
30%	1	100	100	100	100	100
	2	100	100	100	100	100
	10	100	100	100	100	100
50%	1	100	100	100	100	100
	2	100	100	100	100	100
	10	100	100	100	100	100

Table 2: Absolute Bias of the Estimate of the Mean of x, μ_1, for Truncation Parallel to the Major Axis

Amount of Truncation	σ_1^2/σ_2^2	ρ 0.1	0.3	0.5	0.7	0.9
10%	1	0.155	0.025	0.020	0.010	0.000
	2	0.025	0.020	0.020	0.010	0.003
	10	0.010	0.010	0.036	0.000	0.006
30%	1	0.287	0.168	0.134	0.096	0.037
	2	0.053	0.100	0.109	0.088	0.040
	10	0.017	0.020	0.044	0.042	0.017
50%	1	0.403	0.358	0.298	0.232	0.134
	2	0.079	0.187	0.224	0.198	0.123
	10	0.014	0.062	0.085	0.096	0.071

Some observations can be made from this study as presented in Tables 1 through 18. When truncation is parallel to the major axis the iterative solutions to the maximum likelihood estimates almost always converge as shown in Table 1. The few times when convergence was not attained are in the second study given by Tables 19 through 21. When truncation is parallel to the minor axis the estimates often fail to converge. The failure to converge occurred especially at high correlation values and as the amount of truncation increased. In the intermediate case of truncation at 45° to the major axis, the estimates failed to converge occurs at high values of the correlation.

In the case of truncation parallel to the minor axis there continues to be a problem even when the estimates converge. In many of the cases in which the estimates converge, the estimate of ρ converged to 0. This is evident in Table 18 where the absolute bias is often equal to the value of ρ. The biases of the location parameters tend to increase as the ratio of the variances increases as shown in Tables 14 and 15.

In the best scenario for convergence, the case of truncation parallel to the major axis, the location parameters can be badly biased as shown in Tables 2 and 3. The bias tends to decrease with an increase in the correlation between x and y. With the exception of a few cases, the biases in the estimates of the variance tend to be better than the location parameters (Tables 4 and 5). The bias in tends to decrease as the ratio of the variances increases (Table 6). In the intermediate position, the case of truncation at 45° to the major axis, the estimates can be badly biased in the location parameters. The bias is relatively good in the estimate of the smaller variance and in the correlation.

A possible reason that truncation along the major axis of the ellipse works best is the information provided by the curvature of the truncated ellipse. As can be seen in Figure

Table 3: Absolute Bias of the Estimate of the Mean of y, μ_2, for Truncation Parallel to the Major Axis

Amount of		ρ				
Truncation	σ_1^2/σ_2^2	0.1	0.3	0.5	0.7	0.9
10%	1	0.386	0.148	0.114	0.075	0.005
	2	0.348	0.190	0.145	0.088	0.014
	10	0.095	0.164	0.127	0.093	0.020
30%	1	0.545	0.378	0.313	0.230	0.110
	2	0.588	0.457	0.380	0.272	0.122
	10	0.409	0.466	0.403	0.308	0.145
50%	1	0.668	0.587	0.498	0.385	0.221
	2	0.798	0.723	0.609	0.467	0.267
	10	0.796	0.758	0.680	0.549	0.327

4, there is very little change in curvature on the sides of the ellipse that follow the major axis while there is substantial change in curvature at the two ends of the ellipse. Within the scenario of truncated data, truncation parallel to the major axis retains data pertinent to the areas of greatest curvature in the ellipse. Since there is very little variability in the tangent to the ellipse close to the truncation points when truncation is parallel to the minor axis, variation in the actual data could lead to poor estimates of the whole ellipse.

This reason may also explain how successful estimation would be when truncation is done in the x-direction only. Changes in the curvature where the line cuts the ellipse are greatest when the correlation is low and least when the correlation is high. It would be expected that estimation under truncation would not work well in this case when the correlation is high.

The second simulation study addressed the following question. If the estimates do converge, how good is the reconstruction of the ellipse? The answer to this question is relevant to the reconstruction of the archaeological pit data. Had our estimates converged, would the reconstruction of the pit been reasonable? The second simulation study proceeded similar to the first one with a slight twist. In this case the complete sample was generated such that the total sample size had an expected truncation number of 500. Consequently, comparisons of ellipse reconstruction would be based on the same expected sample size. For 10%, 30% and 50% truncation of the sample, the full sample sizes were 556, 715 and 1000 respectively.

After generation of the full sample, it was separated into two parts: the part in the region that did not satisfy the truncation constraint and the part in the region that did. The latter sample points were used as the truncated sample for estimation purposes. If convergence was attained the 95% ellipsoid was calculated using (2.1). Using the data points generated

Table 4: Absolute Bias of the Estimate of Variance of x, σ_1^2, for Truncation Parallel to the Major Axis

Amount of Truncation	σ_1^2/σ_2^2	ρ 0.1	0.3	0.5	0.7	0.9
10%	1	0.105	0.020	0.012	0.007	0.002
	2	0.000	0.003	0.003	0.003	0.000
	10	0.018	0.031	0.005	0.020	0.001
30%	1	0.156	0.095	0.061	0.036	0.008
	2	0.001	0.025	0.041	0.027	0.009
	10	0.004	0.032	0.000	0.016	0.016
50%	1	0.200	0.153	0.106	0.063	0.021
	2	0.010	0.044	0.056	0.046	0.016
	10	0.022	0.023	0.027	0.040	0.006

for the full sample, we calculated the fraction falling in the ellipsoid. After 1000 simulations of this sampling procedure, we calculated the fraction of times the estimates converges and conditional on convergence, the average fraction of times the data points from the full sample fell in the ellipse.

The results are shown in Tables 19, 20 and 21. The main entry in each table is the average fraction of times the data points fell in the ellipse. If the entry is 0, then none of the simulated samples had convergent estimates. If a non-zero fraction is followed by be a bracketed number then less the 100% of the simulated samples had convergent estimates. The percentage of convergent estimates is shown in the brackets. In the cases of a non-zero fraction without a bracketed number, all the estimates always converged in the simulations. In all cases the reconstruction of the ellipse, at least in terms of 95% coverage of the sample points, worsens as the amount of truncation increases. A phenomenon similar to the first simulation study was also here. As the truncation line moved away from being parallel to the major axis of the ellipse, the estimates increasingly tended not to converge.

Table 5: Absolute Bias of the Estimate of Variance of y, σ_2^2, for Truncation Parallel to the Major Axis

Amount of Truncation	σ_1^2/σ_2^2	ρ				
		0.1	0.3	0.5	0.7	0.9
10%	1	0.104	0.022	0.014	0.007	0.001
	2	0.186	0.069	0.039	0.013	0.000
	10	0.001	0.073	0.052	0.022	0.002
30%	1	0.159	0.096	0.065	0.040	0.009
	2	0.330	0.216	0.138	0.072	0.014
	10	0.232	0.251	0.195	0.115	0.032
50%	1	0.197	0.155	0.110	0.065	0.021
	2	0.429	0.332	0.224	0.124	0.038
	10	0.437	0.394	0.311	0.199	0.069

Table 6: Absolute Bias of the Estimate of Correlation ρ for Truncation Parallel to the Major Axis

Amount of Truncation	σ_1^2/σ_2^2	ρ				
		0.1	0.3	0.5	0.7	0.9
10%	1	0.130	0.030	0.021	0.010	0.000
	2	0.030	0.029	0.025	0.012	0.000
	10	0.001	0.016	0.016	0.011	0.001
30%	1	0.214	0.133	0.106	0.064	0.017
	2	0.064	0.106	0.098	0.063	0.018
	10	0.020	0.056	0.068	0.056	0.017
50%	1	0.271	0.237	0.186	0.121	0.043
	2	0.084	0.178	0.168	0.117	0.043
	10	0.039	0.104	0.122	0.102	0.041

Table 7: Percentage of Successes at Convergence for Truncation at 45° to the Major and Minor Axes

Amount of		ρ				
Truncation	σ_1^2/σ_2^2	0.1	0.3	0.5	0.7	0.9
10%	2	99.8	100	100	100	9.8
	10	77.2	100	100	100	14.1
30%	2	98.9	100	100	100	0.7
	10	99.6	100	100	100	2.9
50%	2	100	100	100	99.5	0
	10	99.8	100	100	99.6	0.4

Table 8: Absolute Bias of the Estimate of the Mean of x, μ_1, for Truncation at 45° to the Major and Minor Axes

Amount of		ρ				
Truncation	σ_1^2/σ_2^2	0.1	0.3	0.5	0.7	0.9
10%	2	0.112	0.140	0.080	0.106	0.135
	10	0.010	0.069	0.078	0.048	0.217
30%	2	0.022	0.385	0.279	0.146	0.363
	10	0.698	0.622	0.498	0.353	0.614
50%	2	0.691	0.652	0.517	0.354	NA
	10	1.462	1.344	1.181	0.897	1.258

Table 9: Absolute Bias of the Estimate of the Mean of y, μ_2, for Truncation at 45° to the Major and Minor Axes

Amount of		ρ				
Truncation	σ_1^2/σ_2^2	0.1	0.3	0.5	0.7	0.9
10%	2	0.252	0.233	0.162	0.094	0.170
	10	0.090	0.134	0.129	0.082	0.034
30%	2	0.114	0.325	0.217	0.127	0.175
	10	0.399	0.307	0.214	0.114	0.113
50%	2	0.560	0.377	0.229	0.096	NA
	10	0.513	0.375	0.223	0.082	0.275

Table 10: Absolute Bias of the Estimate of Variance of x, σ_1^2, for Truncation at 45° to the Major and Minor Axes

Amount of		ρ				
Truncation	σ_1^2/σ_2^2	0.1	0.3	0.5	0.7	0.9
10%	2	0.157	0.192	0.119	0.046	0.317
	10	0.129	0.242	0.312	0.166	1.324
30%	2	0.038	0.428	0.317	0.167	0.559
	10	1.846	1.639	1.331	0.925	2.470
50%	2	0.616	0.620	0.497	0.346	NA
	10	3.176	2.940	2.604	1.915	3.993

Table 11: Absolute Bias of the Estimate of Variance of y, σ_2^2, for Truncation at 45° to the Major and Minor Axes

Amount of		ρ				
Truncation	σ_1^2/σ_2^2	0.1	0.3	0.5	0.7	0.9
10%	2	0.015	0.005	0.006	0.007	0.139
	10	0.006	0.000	0.002	0.005	0.116
30%	2	0.005	0.001	0.016	0.028	0.253
	10	0.004	0.001	0.014	0.028	0.214
50%	2	0.067	0.003	0.060	0.062	NA
	10	0.017	0.000	0.023	0.060	0.329

Table 12: Absolute Bias of the Estimate of Correlation ρ for Truncation at 45° to the Major and Minor Axes

Amount of		ρ				
Truncation	σ_1^2/σ_2^2	0.1	0.3	0.5	0.7	0.9
10%	2	0.048	0.021	0.003	0.003	0.022
	10	0.018	0.003	0.002	0.004	0.018
30%	2	0.009	0.046	0.007	0.012	0.046
	10	0.054	0.019	0.006	0.011	0.035
50%	2	0.202	0.074	0.006	0.026	NA
	10	0.101	0.039	0.008	0.024	0.069

Table 13: Percentage of Successes at Convergence for Truncation Parallel to the Minor Axis

Amount of Truncation	σ_1^2/σ_2^2	ρ 0.1	0.3	0.5	0.7	0.9
10%	1	100.0	37.2	49.1	17.1	0
	2	16.1	88.5	33.3	0	0
	10	0.4	100.0	85.9	1.2	0
30%	1	33.9	86.7	0	0	0
	2	0.4	5.2	0	0	0
	10	0.4	99.8	51.8	1.2	0
50%	1	51.3	99.8	18.2	0	0
	2	0	0	0	0	0
	10	0.4	0	0	0	0

Table 14: Absolute Bias of the Estimate of the Mean of x, μ_1, for Truncation Parallel to the Minor Axis

Amount of Truncation	σ_1^2/σ_2^2	ρ 0.1	0.3	0.5	0.7	0.9
10%	1	0.002	0.088	0.008	0.026	NA
	2	0.081	0.002	0.002	NA	NA
	10	0.741	0.008	0.003	0.030	NA
30%	1	0.021	0.220	NA	NA	NA
	2	0.338	0.010	NA	NA	NA
	10	0.639	0.014	0.013	0.126	NA
50%	1	0.106	0.329	0.465	NA	NA
	2	NA	NA	NA	NA	NA
	10	0.828	NA	NA	NA	NA

Table 15: Absolute Bias of the Estimate of the Mean of y, μ_2, for Truncation Parallel to the Minor Axis

Amount of		ρ				
Truncation	σ_1^2/σ_2^2	0.1	0.3	0.5	0.7	0.9
10%	1	0.000	0.087	0.006	0.024	NA
	2	0.362	0.002	0.025	NA	NA
	10	0.478	0.000	0.022	0.009	NA
30%	1	0.020	0.220	NA	NA	NA
	2	0.421	0.078	NA	NA	NA
	10	0.495	0.092	0.156	0.070	NA
50%	1	0.119	0.333	0.468	NA	NA
	2	NA	NA	NA	NA	NA
	10	0.502	NA	NA	NA	NA

Table 16: Absolute Bias of the Estimate of Variance of x, σ_1^2, for Truncation Parallel to the Minor Axis

Amount of		ρ				
Truncation	σ_1^2/σ_2^2	0.1	0.3	0.5	0.7	0.9
10%	1	0.001	0.351	0.040	0.106	NA
	2	0.070	0.002	0.096	NA	NA
	10	2.452	0.017	0.103	1.516	NA
30%	1	0.032	0.182	NA	NA	NA
	2	0.485	0.108	NA	NA	NA
	10	1.792	0.036	0.228	1.723	NA
50%	1	0.049	0.210	0.341	NA	NA
	2	NA	NA	NA	NA	NA
	10	1.793	NA	NA	NA	NA

Table 17: Absolute Bias of the Estimate of Variance of y, σ_2^2, for Truncation Parallel to the Minor Axis

Amount of		ρ				
Truncation	σ_1^2/σ_2^2	0.1	0.3	0.5	0.7	0.9
10%	1	0.001	0.105	0.043	0.096	NA
	2	0.020	0.001	0.048	NA	NA
	10	0.010	0.003	0.013	0.151	NA
30%	1	0.025	0.177	NA	NA	NA
	2	0.034	0.057	NA	NA	NA
	10	0.006	0.002	0.013	0.165	NA
50%	1	0.093	0.218	0.340	NA	NA
	2	NA	NA	NA	NA	NA
	10	0.054	NA	NA	NA	NA

Table 18: Absolute Bias of the Estimate of Correlation ρ for Truncation Parallel to the Minor Axis

Amount of		ρ				
Truncation	σ_1^2/σ_2^2	0.1	0.3	0.5	0.7	0.9
10%	1	0.000	0.297	0.036	0.043	NA
	2	0.100	0.007	0.041	NA	NA
	10	0.100	0.003	0.009	0.053	NA
30%	1	0.100	0.298	NA	NA	NA
	2	0.100	0.088	NA	NA	NA
	10	0.100	0.002	0.033	0.082	NA
50%	1	0.100	0.300	0.500	NA	NA
	2	NA	NA	NA	NA	NA
	10	0.100	NA	NA	NA	NA

Table 19: Percentage of Points within the 95% for Truncation Parallel to the Major Axis

Amount of Truncation	σ_1^2/σ_2^2	ρ				
		0.1	0.3	0.5	0.7	0.9
10%	1	93.8 (99.9)	93.7	93.8	93.8	93.8
	2	93.8 (99.4)	93.7	93.8	93.9	93.8
	10	93.8	93.8	93.9	93.8	93.8
30%	1	88.3	88.4	88.4	88.3	88.5
	2	88.2	88.1	88.2	88.3	88.4
	10	88.5 (99.4)	88.4	88.4	88.5	88.5
50%	1	78.0	78.0	78.0	78.0	78.0
	2	77.8	77.7	77.8	77.9	78.1
	10	77.8 (99.8)	77.9	78.1	78.1	78.2

Table 20: Percentage of Points within the 95% for Truncation at 45° to the Major and Minor Axes

Amount of Truncation	σ_1^2/σ_2^2	ρ				
		0.1	0.3	0.5	0.7	0.9
10%	2	94.0 (98.9)	94.3	94.6	94.7	94.6 (35.1)
	10	94.2	94.4	94.5	94.7	94.2 (25.4)
30%	2	89.5	91.1	92.5	93.6	92.1 (1.9)
	10	90.9	91.6	92.5	93.4	0
50%	2	80.5	84.1	87.8	91.0	0
	10	83.9	85.5	87.8	90.3 (99.8)	0

5 Discussion and Conclusion

Parameter estimation from, and density reconstruction of, a truncated bivariate normal distribution is not a simple problem. A feasible solution depends on the amount of truncation and the position of the truncation line. As the truncation line moves from being parallel to the major axis to being parallel to the minor axis, the likelihood of a feasible and reasonable solution diminishes. The same is true as the amount of truncation increases. The best scenario for estimation from this truncated distribution is when the truncation line is parallel to the major axis of the elliptical contours describing this distribution and when the amount

Table 21: Percentage of Points within the 95% for Truncation Parallel to the Minor Axis

Amount of		ρ				
Truncation	σ_1^2/σ_2^2	0.1	0.3	0.5	0.7	0.9
10%	1	94.4 (68.4)	93.3 (0.2)	0	0	0
	2	94.8 (62.5)	94.7 (9.9)	0	0	0
	10	94.7 (96.9)	94.8 (52.4)	0	0	0
30%	1	93.7 (69.8)	90.4 (82.1)	0	0	0
	2	93.4 (11.2)	92.52 (2.4)	0	0	0
	10	93.3 (37.3)	0	0	0	0
50%	1	93.1 (49.1)	0	0	0	0
	2	87.7 (0.001)	0	0	0	0
	10	88.4 (0.001)	91.0 (5.5)	91.7 (3.2)	0	0

of truncation is 10% or less. Reconstruction of the bivariate distribution can be badly biased when the truncation line cuts across the ellipse at the narrower axis. The reconstruction process is much better if the truncation line cuts the ellipse at the wider axis.

Probability distributions other than the bivariate normal, such as those suggested in Azzalini and Dalla Valle (1996) as well as Azzalini and Capitanio (1999), could be tried as an alternate approach. For two reasons, we expect that similar results would occur to what we have seen in the bivariate normal. The first reason is that the resulting probability contours from which we obtained an ellipse in the normal case would be of similar, though not the same, shape. If the truncation line cuts across a region of low curvature in the contour, then we would have the same problem that we currently see. The second reason is that some non-normal distributions will come with additional parameters to account for skewness. Additional parameters may exacerbate rather than help the estimation process, since additional parameters may lead to further instability in the estimates when the sample size remains the same.

The use of the multivariate normal to obtain an estimate of the elliptical contour to describe the archaeological data was unsuccessful, yet it did provide a motivation for examining the performance of estimation under truncation. We suggest two methods to explore in the future for the reproduction of the desired ellipse for our archaeological data. One suggestion is to take the points near the edge of the data scatter in the undisturbed part of the archaeological site and, using a nonparametric regression technique, fit a smooth line running through these points. Then rotate the smooth line until the rotated line and the original line form a closed shape. Finally, fit an ellipse to the smoothed figure. A second suggestion to explore is to take a page from data sharpening techniques originally studied by Choi and Hall (1999). In our case, the points in the disturbed part of the archaeological

site are adjusted towards the major concentration of the scatter and a bivariate normal or skew normal is fit to the entire data. This variation on data sharpening is continued until there is a good fit of an elliptical, or other contour, to the undisturbed part of the site.

Acknowledgements

The authors would like to thank Professor Chris Ellis of the Department of Anthropology of the University of Western Ontario for providing us with archaeological data. This work was supported by a grant for the National Sciences and Engineering Research Council of Canada.

References

[1] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the skew normal distribution. *Journal of the Royal Statistical Society, Series B*, **61**, 579-602.

[2] Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

[3] Choi, E. and Hall, P. (1999). Data sharpening as a prelude to density estimation. *Biometrika*, **86**, 941-947.

[4] Deller, D. B. and Ellis, C. J. (1984). Crowfield: A preliminary report on a probable Paleo-Indian cremation in Southwestern Ontario. *Archaeology of Eastern North America*, **12**, 41-71.

[5] Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions, Volume I: Models and Applications.* New York: Wiley.

[6] Nath, G. B. (1971). Estimation in truncated bivariate normal distributions. *Applied Statistics*, **20**, 313-319.

[7] Nath, G. B. (1972). Moments of a linearly truncated bivariate normal distribution. *Australian Journal of Statistics*, **14**, 97-102.

[8] Schneider, H. (1986). *Truncated and Censored Samples from Normal Populations.* New York: Marcel Dekker.

[9] Zwillinger, D. (2003). *CRC Standard Mathematical Tables and Formulae, 31st Edition.* Boca Raton: Chapman & Hall/CRC.