RRB 89-77

STATISTICAL INFERENCE FROM MULTIPLY CENSORED DATA

by

A.H. El-Shaarawi and A. Naderisamani

National Water Research Institute
Canada Centre for Inland Waters
Burlington, Ontario, L7R 4A6

November 1989
NWRI Contribution #89-175

ABSTRACT

Maximum likelihood equations for a multiply censored normal sample are modified so that explicit estimates for the mean and standard deviation can be obtained. The estimate of the mean of a multiply censored sample is then obtained by normalizing the data using the Box and Cox transformation. An approximate confidence interval for the mean is also obtained and the probability of coverage for the lognormal mean is discussed in detail. Simulation results and applications are provided.

RESUME

Les équations de maximum de vraisemblance pour un échantillon normal multicensuré sont modifiées pour obtenir des estimations explicites de la moyenne et de l'écart-type. L'estimation de la moyenne d'un échantillon multicensuré est alors obtenue en normalisant les données à l'aide de la transformation de Box et Cox. Un intervalle de confiance approché est aussi obtenu pour la moyenne, et le domaine de probabilité pour la moyenne lognormale est analysé en détail. Des résultats de simulation et des applications sont présentés.

MANAGEMENT PERSPECTIVE

It frequently happens that a certain portion of the contaminants in water quality monitoring data sets have concentrations that cannot be measured. It is possible, however, to determine that those concentrations fall within certain intervals whose end points, called detection limits, are determined by analytical methods. This report examines the statistical analysis of a general set of data in the presence of several detection limits. The report gives an explicit formula for the estimation of the mean from a multiply censored sample and also provides an approximate confidence interval for the mean. Simulation results are also provided and the probability coverage for the confidence interval of the lognormal mean is discussed in detail. The report concludes with an application using concentrations (nanograms per litre) of Fluoranthene in water samples from the Niagara River.

PERSPECTIVE DE GESTION

Il arrive souvent qu'une partie des contaminants dans des ensembles de données sur la surveillance de la qualité des eaux ont des concentrations impossibles à mesurer. Il est toutefois possible d'établir que ces concentrations se situent dans des intervalles dont les extrémités, appelées limites de détection, sont déterminés par des méthodes analytiques. Le présent rapport traite de l'analyse statistique d'un ensemble général de données sujettes à plusieurs limites de détection. Il renferme une formule explicite d'estimation de la moyenne d'un échantillon multicensuré avec un intervalle de confiance approché. Des résultats de simulation sont aussi fournis, et le domaine de probabilité pour l'intervalle de confiance de la moyenne lognormale est analysé en détail. Le rapport conclut par une application portant sur des concentrations (nanogrammes par litre) de Fluoranthène dans des échantillons d'eau de la rivière Niagara.

INTRODUCTION

In routine water and air quality monitoring of toxic contaminants and trace metals, it frequently happens that a certain portion of the observations examined, have concentrations that cannot be measured. It is only possible to determine that the concentrations for those observations fall within certain intervals. The endpoints of these intervals are detection limits determined by analytical methods. If $D_1<\ldots<D_{k-1}<D_k$ are such detection limits, then a censored observation occurs when its value falls below D_k. Approaches adopted by environmental scientists for estimating the mean and standard deviation in the presence of a single censoring limit D_1, ranges from assigning a value to an observation reported as less than D_1, to the use of the log regression method (Gilliom and Helsel, 1986). Assuming the normal or lognormal distribution for the observations, El-Shaarawi (1989) and El-Shaarawi and Dolan (1989) discussed the use of the method of maximum likelihood for estimating the mean and standard deviation when $k = 1$. In addition, Shumway et al. (1989) considered the possibility of using the Box and Cox (1964) transformation to normalize the data. The general problem of maximum likelihood estimation of the parameters of a censored normal sample has been considered by many authors. Cohen (1950) used the maximum likelihood method to estimate the parameters of type I singly and doubly censored normal samples. Gupta (1952) found maximum likelihood equations to estimate the parameters of type II censored normal samples. Cohen (1950) and Gupta (1952) also formulated the

asymptotic variances and covariances. Harter and Moore (1966) and Harter (1970) considered the maximum likelihood for type II censoring and performed a simulation study which showed that maximum likelihood estimators had mean square errors smaller than the variances of the best linear unbiased estimators for $n \geq 10$. Tiku (1967) modified the maximum likelihood equations from a type II censored normal sample so that an explicit formula for the estimators could be obtained. The general results concerning censored normal samples have been summarized and extensively studied by Schneider (1986). Progressively censored samples from normal, exponential, Weibull and lognormal distributions have also received previous attention from Herd, Robert, Cohen and Ringer and Sprinkle, (Cohen, 1976).

The present paper first examines the estimation of the mean and standard deviation from a multiply censored ($k>1$) normal sample. The estimates are obtained by the modified maximum likelihood equations and converge to the exact maximum likelihood solutions. The paper then studies the estimation of the mean from a multiply censored sample by normalizing the data using the Box and Cox (1964) transformation. An approximate confidence interval for the mean is also obtained and the probability of coverage for the lognormal mean is discussed in detail. The simulation results are provided and an application using the concentrations (nanograms per litre) of Fluoranthene in water samples from the Niagara River is presented.

1. ESTIMATION OF THE MEAN AND STANDARD DEVIATION FROM TYPE I MULTIPLY CENSORED NORMAL SAMPLES

Let $D_0 = -\infty$, and consider k detection limits $D_1, \ldots, D_k$. Let the random variable N_i ($i = 0,1,\ldots, k-1$) denote the number of observations that fall in the interval $(D_i, D_{i+1}]$. Furthermore, let the random variables $X_1,\ldots,X_n$ represent the n uncensored observations ($X_i > D_k$, $i = 1,\ldots,n$). The observed values of X_i and N_i are denoted respectively, by x_i and n_i. Under the assumption that X_i's are independent and normally distributed with mean μ and variance σ^2, the likelihood function is:

$$L = C_0 \, \sigma^{-n} \prod_{i=0}^{k-1} \left(\Phi(\eta_{i+1}) - \Phi(\eta_i)\right)^{n_i} \prod_{i=1}^{n} \phi\left(\frac{x_i-\mu}{\sigma}\right) ,$$

where C_0 is a constant, $\eta_i = (D_i-\mu)/\sigma$ ($i=1,\ldots,k$),

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2\sigma^2}\right), \text{ and } \Phi(x) = \int_{-\infty}^{x} \phi(t)dt.$$

By the mean value theorem, L can be written as:

$$L = C_0 \, \sigma^{-(n+\sum_{i=1}^{k-1} n_i)} (\Phi(\eta_1))^{n_0} \prod_{i=1}^{k-1} \left\{\phi\left(\frac{\zeta_i-\mu}{\sigma}\right)(D_{i+1} - D_i)\right\}^{n_i} \prod_{i=1}^{n} \phi\left(\frac{x_i-\mu}{\sigma}\right) ,(1)$$

where ζ_i ($i=1,\ldots, k-1$) is between D_i and D_{i+1}. Let

$$N = n + \sum_{i=0}^{k-1} n_i, \quad \bar{x} = \sum_{i=1}^{n} x_i/n,$$

and define

$$M = \left(\sum_{i=1}^{k-1} n_i \zeta_i + n\bar{x}\right)/(N-n_0) ,$$

$$S^2 = \left\{\sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{k-1} n_i \zeta_i^2 - (N-n_0)M^2\right\}/(N-n_0),$$

and $g(x) = \phi(x)/\Phi(x)$, where x is a real number. From (1) it follows that the maximum likelihood estimates of μ and σ satisfy the following equations:

$$\mu = M - \sigma \frac{n_0}{N-n_0} g(\eta_1) , \tag{2}$$

$$\sigma^2 = S^2 + (M-\mu)(M-D_1). \tag{3}$$

Replacing $g(\eta_1)$ by Tiku's (1967) linear approximation $\alpha+\beta\eta_1$, where

$$\beta = \{ g(t_2) - g(t_1)\}/(t_2-t_1), \quad \alpha = g(t_1)-t_1\beta,$$

$$t_1 = \Phi^{-1}\{q-\sqrt{q(1-q)/N}\}, \quad t_2 = \Phi^{-1}\{q + \sqrt{q(1-q)/N}\}, \quad q = n_0/N,$$

(2) becomes

$$\mu = \{M(N-n_0) - n_0\alpha\sigma-n_0\beta D_1\}/\{N-n_0(1+\beta)\}. \tag{4}$$

For small values of $\Delta_i = D_{i+1} - D_i$ $(i=1,\ldots,k-1)$, ζ_i can be approximated by the midpoint $x_{m_i} = (D_i + D_{i+1})/2$. Setting $\zeta_i = x_{m_i}$, equations (3) and (4) provide approximate explicit solutions $\hat{\mu}$ and $\hat{\sigma}$ for μ and σ. As the total number of observations tends to infinity, and as Δ_i $(i=1,\ldots,k-1)$ approaches zero, these estimates approach the maximumlikelihood estimates for μ and σ.

The asymptotic variance - covariance matrix of $\hat{\mu}$ and $\hat{\sigma}$ is denoted by the matrix $\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}$ and is obtained by noting that:

$$E\left(\frac{\partial^2 \ln L}{\partial \mu^2}\right) = -\frac{N}{\sigma^2}\{1 - (1+\beta)\Phi(\eta_1)\},$$

$$E\left(\frac{\partial^2 \ln L}{\partial \mu \partial \sigma}\right) = E\left(\frac{\partial^2 \ln L}{\partial \sigma \partial \mu}\right) = \frac{2N}{\sigma^3}[(\mu + \sigma g(-\eta_k))\Phi(\eta_k) +$$

$$\{(\beta\eta_1 + \frac{\alpha}{2})\sigma - \mu\}\Phi(\eta_1) - (\sigma g(-\eta_k) + A)],$$

where $A = \sum_{i=1}^{k-1} x_{m_i}(\Phi(\eta_{i+1}) - \Phi(\eta_i))$,

and

$$E\left(\frac{\partial^2 \ln L}{\partial \sigma^2}\right) = \frac{N}{\sigma^4}[3\{\sigma^2(1 + \eta_k g(-\eta_k)) - \mu^2\}\Phi(\eta_k) +$$

$$\{3\mu^2 + \sigma^2(2\alpha\eta_1 + 3\beta\eta_1^2 - 1)\}\Phi(\eta_1) - \sigma^2(2 + 3\eta_k g(-\eta_k)) - 3(B - 2\mu A)],$$

where $B = \sum_{i=1}^{k-1} x_{m_i}^2(\Phi(\eta_{i+1}) - \Phi(\eta_i))$.

Let $\hat{L}$ be obtained by substituting $\hat{\mu}$ and $\hat{\sigma}$ directly into L and let $R(\mu,\sigma)=\ln L-\ln\hat{L}$. Then, by noting that $-2R(\mu,\sigma)$ is approximately $\chi^2(2)$, an approximate a% confidence region for μ and σ may be obtained as

$$\{(\mu,\sigma): -2R(\mu,\sigma) \le b_a\} = \{(\mu,\sigma): R(\mu,\sigma) \ge -\frac{b_a}{2}\},$$

where b_a is determined from $P(\chi^2(2) \le b_a) = a$.

We now proceed to obtain the conditional bias in the estimators. Consider the Taylor expansion:

$$\left[\frac{\partial \ln L}{\partial \mu}\right]_{\hat{\mu}} = \left[\frac{\partial \ln L}{\partial \mu}\right]_{\mu} + (\hat{\mu} - \mu)\left[\frac{\partial^2 \ln L}{\partial \mu^2}\right]_{\overset{*}{\mu}},$$

where $\overset{*}{\mu}$ is between $\hat{\mu}$ and μ. We have

$$E(\hat{\mu}|\sigma) - \mu = -E\left(\frac{\partial \ln L}{\partial \mu}\Big|\sigma\right)_{\mu} \Big/ E\left(\frac{\partial^2 \ln L}{\partial \mu^2}\Big|\sigma\right)_{\overset{*}{\mu}}.$$

Hence, the conditional bias for $\hat{\mu}$ is

$$E(\hat{\mu}|\sigma) - \mu \simeq \frac{-(\sigma g(-\eta_k)+\mu)\Phi(\eta_k)+\{\mu-\sigma(\alpha+\beta\eta_1)\}\Phi(\eta_1) + A + \sigma\, g(-\eta_k)}{1-(1+\beta)\Phi(\eta_1)}. \quad (5)$$

Similarly, the conditional bias for σ is obtained by

$$E(\hat{\sigma}|\mu)-\sigma=$$

$$\frac{\sigma[\{\mu^2-\sigma^2(1+\eta_k g(-\eta_k))\}\Phi(\eta_k)-\{\mu^2+\sigma^2(\alpha\eta_1+\beta\eta_1^2-1)\}\Phi(\eta_1)+\sigma^2\eta_k g(-\eta_k)}{3\{\mu^2-\sigma^2(1+\eta_k g(-\eta_k))\}\Phi(\eta_k)-\{3\mu^2+\sigma^2(2\alpha\eta_1+3\beta\eta_1^2-1)\}\Phi(\eta_1)+3(B-2\mu A)}$$

$$\frac{+(B-2\mu A)]}{+\sigma^2(2+3\eta_k g(-\eta_k))} .$$

For the special case of single censoring (k=1), (5) and (6) become

$$E(\hat{\mu}|\sigma)-\mu = \frac{-\sigma\{(\alpha+\beta\eta_1+g(-\eta_1))\Phi(\eta_1)-g(-\eta_1)\}}{1-(1+\beta)\Phi(\eta_1)} .$$

and

$$E(\hat{\sigma}|\mu)-\sigma = \frac{\sigma\eta_1\{(\alpha+\beta\eta_1+g(-\eta_1))\Phi(\eta_1)-g(-\eta_1)\}}{\eta_1(2\alpha+3\beta\eta_1)\Phi(\eta_1)-(2+3\eta_1 g(-\eta_1))\Phi(-\eta_1)} .$$

But

$$(\alpha+\beta\eta_1+g(-\eta_1))\Phi(\eta_1)-g(-\eta_1)=(\alpha+\beta\eta_1)\Phi(\eta_1)-g(-\eta_1)(1-\Phi(\eta_1))$$

$$= \Phi(\eta_1)\{\alpha+\beta\eta_1-\frac{g(-\eta_1)(1-\Phi(\eta_1))}{\Phi(\eta_1)}\}= \Phi(\eta_1)(\alpha+\beta\eta_1-g(\eta_1))=0 .$$

Hence, $E(\hat{\mu}|\sigma)-\mu=0$ and $E(\hat{\sigma}|\mu)-\sigma=0$.

Thus the estimators $\hat{\mu}$ and $\hat{\sigma}$ are nearly unbiased. In this case, the consistency of the estimators follows from the fact that $V(\hat{\mu}) = V_{11}=0$ and $V(\hat{\sigma}) = V_{22}=0$.

2. ESTIMATION UNDER TRANSFORMATIONS

When the normality assumption of the X_i's cannot be justified, it is appropriate to find a suitable transformation so that the transformed data satisfy the normality assumption. Box and Cox (1964) suggested the use of the transformation,

$$g_\lambda(x) = \begin{cases} (x^\lambda - 1)/\lambda & \lambda \neq 0 \\ \ln x & \lambda = 0 \end{cases} \tag{7}$$

with λ chosen so that the distribution of $g_\lambda(X_1)$ is normal with mean μ_λ and variance $\sigma_\lambda{}^2$.

Given detection limits $D_1,\ldots D_k$ and observations $x_1,\ldots,x_n$, $(x_i > D_k, i=1,\ldots,n)$, we may obtain the estimates $\hat{\mu}_\lambda$ and $\hat{\sigma}_\lambda$ by the methods of the previous section. The transformation parameter λ is then chosen as the value $\hat{\lambda}$ that maximizes:

$$h(\lambda,\hat{\mu}_\lambda,\hat{\sigma}_\lambda) = -n \ln \hat{\sigma}_\lambda +$$

$$\sum_{i=0}^{k-1} n_i \ln (F(D_{i+1}) - F(D_i)) + \lambda \sum_{i=1}^{n} \ln (x_i) - \frac{1}{2} \sum_{i=1}^{n} \left(\frac{g_\lambda(x_i) - \hat{\mu}_\lambda}{\hat{\sigma}_\lambda}\right)^2 ,$$

where

$$F(y) = \frac{1}{\sqrt{2\pi}\hat{\sigma}_\lambda} \int_0^y x^{\lambda-1} \exp\left\{\frac{-(g_\lambda(x)-\hat{\mu}_\lambda)^2}{2\hat{\sigma}_\lambda{}^2}\right\} dx.$$

Let

$$A_i = \frac{1}{\sqrt{2\pi}\sigma_{\hat{\lambda}}} \int_{-\frac{1}{\hat{\lambda}}}^{\infty} x^i (1+\hat{\lambda}x)^{1/\hat{\lambda}} \exp\left\{\frac{-(x-\mu_{\hat{\lambda}})^2}{2\sigma^2_{\hat{\lambda}}}\right\} dx, \ i = 0,1,2. \qquad (8)$$

Then $\gamma = E(X_i) = A_0$. The estimate $\hat{\gamma}$ of γ may be obtained by substituting $\hat{\mu}_{\hat{\lambda}}$ and $\hat{\sigma}_{\hat{\lambda}}$ directly into γ above. Note that using the Taylor expansion, $\hat{\gamma}$ may be approximated by,

$$\hat{\gamma} \simeq \gamma + (\hat{\mu}_{\hat{\lambda}} - \mu_{\hat{\lambda}})\gamma_{\mu} + (\hat{\sigma}_{\hat{\lambda}} - \sigma_{\hat{\lambda}})\gamma_{\sigma},$$

where $\gamma_{\mu} = \frac{1}{\sigma^2_{\hat{\lambda}}}(A_1 - \mu_{\hat{\lambda}} A_0)$, and

$$\gamma_{\sigma} = \frac{1}{\sigma^3_{\hat{\lambda}}}\{A_2 + (\mu^2_{\hat{\lambda}} - \sigma^2_{\hat{\lambda}}) A_0 - 2\mu_{\hat{\lambda}} A_1\}.$$

Let $\tilde{V}_{11}$, $\tilde{V}_{22}$, and $\tilde{V}_{12}$ be the asymptotic variances of $\hat{\mu}_{\hat{\lambda}}$ and $\hat{\sigma}_{\hat{\lambda}}$, and the covariance of $\hat{\mu}_{\hat{\lambda}}$ and $\hat{\sigma}_{\hat{\lambda}}$, respectively. Then, $\hat{\gamma}$ is approximately normal with mean γ and variance $V^2_{\gamma} = \gamma^2_{\mu} \tilde{V}_{11} + 2\gamma_{\mu} \gamma_{\sigma} \tilde{V}_{12} + \gamma^2_{\sigma} \tilde{V}_{22}$ where V^2_{γ} may be approximated by $\hat{V}^2_{\gamma}$ obtained by substituting $\hat{\mu}_{\hat{\lambda}}$ and $\hat{\sigma}_{\hat{\lambda}}$ directly into V^2_{γ}.

An approximate (1-a)% confidence interval for γ may then be obtained by $(\hat{\gamma} - c_a \hat{V}_{\gamma}, \hat{\gamma} + C_a \hat{V}_{\gamma})$, where C_a is determined from $\Phi(-c_a) = \frac{a}{2}$.

3. THE LOGNORMAL CASE

The special case of $\lambda = 0$ in transformation (7) leads to the assumption that $X_1,\ldots,X_n$ have lognormal distributions. Since the lognormal distribution is frequently used as a model for the environmental data, and since the numerical integrations (8) can be avoided for this case, this section is devoted to the lognormal case. Note that the estimation results for $\hat{\lambda}$ close to zero can be approximated by the methods of this section.

Let the mean and variance of $g_0\,(X_1)$ be $\mu_\circ$ and $\sigma_\circ^2$, respectively. Then the mean $\gamma_\circ$ of X_1 is given by,

$$\gamma_\circ = \exp\,(\mu_\circ + \sigma_\circ^2/2).$$

In order to estimate the mean of a multiply censored lognormal sample, one may obtain the estimates $\hat{\mu}_\circ$ and $\hat{\sigma}_\circ$ of the mean and the standard deviation of the corresponding normal sample and then substitute these estimates directly into the expression for $\gamma_\circ$ to obtain the estimator $\hat{\gamma}_\circ$. This estimator, however, is biased as the following argument shows:

Let $\mathring{V}_{11}$, $\mathring{V}_{22}$, and $\mathring{V}_{12}$ be the asymptotic variances of $\hat{\mu}_\circ$ and $\hat{\sigma}_\circ$ and the covariance of $\hat{\mu}_\circ$ and $\hat{\sigma}_\circ$, respectively. Then, since $a\hat{\mu}_\circ + b\hat{\sigma}_\circ^2$ is approximately normal with mean $a\mu_\circ + b\sigma_\circ^2$ and variance $a^2\mathring{V}_{11} + 4ab\sigma_\circ\mathring{V}_{12} + 4b^2\sigma_\circ^2\mathring{V}_{22}$,

$$E(\exp(a\hat{\mu}_\circ + b\hat{\sigma}_\circ^2)) = \exp\{(a\mu_\circ + b\sigma_\circ^2) + \tfrac{1}{2}\,(a^2\mathring{V}_{11} + 4ab\sigma_\circ\mathring{V}_{12} + 4b^2\sigma_\circ^2\mathring{V}_{22})\}$$

$$= \exp(a\mu_\circ + b\sigma_\circ^2)\; h(a,b).$$

In particular,

$$E(\hat{\gamma}_\circ) = E(\exp(\hat{\mu}_\circ + \frac{\hat{\sigma}_\circ^2}{2})) = \gamma_\circ h(1,\tfrac{1}{2}) = \gamma_\circ \tau \ .$$

As a result the estimator $\hat{\gamma}_\circ$ can be modified to yield approximately an unbiased estimator for $\gamma_\circ$ as follows:

$$\tilde{\gamma}_\circ = \hat{\gamma}_\circ/\tilde{\tau},$$

where $\tilde{\tau}$ is obtained by replacing the parameters $\mu_\circ$ and $\sigma_\circ$ in τ by $\hat{\mu}_\circ$ and $\hat{\sigma}_\circ$ respectively.

An approximate confidence interval for $\gamma_\circ$ may be obtained based on the fact that $\hat{\mu}_\circ + \frac{1}{2}\hat{\sigma}_\circ^2$ is approximately normal with mean $\mu_\circ + \frac{1}{2}\sigma_\circ^2$ and variance $\hat{V}_{\gamma_\circ}{}^2 = \mathring{V}_{11} + 2\sigma_\circ\mathring{V}_{12} + \sigma_\circ^2\mathring{V}_{22}$. By a similar argument used in Land (1972) for complete samples, an approximate (1-a)% confidence interval for $\hat{\gamma}_\circ$ is directly obtained by

$$(\hat{\gamma}_\circ \exp(-c_a \hat{V}_{\gamma_\circ}),\ \hat{\gamma}_\circ \exp(c_a \hat{V}_{\gamma_\circ})),$$

where $\hat{V}_{\gamma_\circ}$ is the value of $V_{\gamma_\circ}$ evaluated at $\hat{\mu}_\circ$ and $\hat{\sigma}_\circ$, and c_a is as defined earlier.

4. SIMULATION RESULTS AND APPLICATIONS

Simulation experiments were conducted to evaluate the performance of the methods of this paper and their sensitivity to small-sample effects. For a given sample size N and two detection limits D_1 and D_2, samples from the standard normal distribution and from the lognormal distribution with mean 7.389 and standard deviation 54.096 were generated using the International Mathematical and Statistical Libraries (IMSL, 1987). The values of the detection limits reflect both low and high levels of censoring. The results summarized in Tables 1a and 1b are the averages over 1000 repetitions. The estimates of the mean and the standard deviation for the normal samples along with their biases and asymptotic variance-covariance are listed in Table 1a. These estimates, as is expected, appear to be uncorrelated for low levels of censoring. Table 1a shows that the elements of the asymptotic variance-covariance matrix decline with the increase in sample sizes. Table 2 gives the estimates of the lognormal means along with their 95% confidence interval as well as the probabilities of coverage for both low and high levels of censoring. It can be seen that the width of the confidence intervals decreases as the number of observations increases. Figure 1 also presents the probabilities of coverage for both low and high levels of censoring using both direct and Taylor expansion methods. This figure also reflects the dependency of both methods on the standard deviations of the corresponding normal distributions. The results indicate that both methods provide good probabilities of coverage for small values of the standard deviations. For moderate and large values of the standard deviations, however, the Taylor expansion method is less satisfactory and the direct method performs noticeably better. The results for both methods become more satisfactory as the number of observations increases, and confirm the earlier results obtained by Land (1972) for complete samples. The methods of this paper were also applied to the concentrations (nanograms per litre) of Fluoranthene in water samples from the Niagara River collected by Environment Canada at the Niagara-on-the-Lake station (Data Interpretation Group, 1989). The values for the number of observations and detection limits as well as the estimation results are presented in Table 2.

REFERENCES

Box, G.E.P., and Cox, D.R. (1964). An analysis of transformations (with discussion). Journal of the Royal Statistical Society, Ser. B, 39, 211-252.

Cohen, A.C. (1950). Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples. Annals of Mathematical Statistics, 21,557-569.

Cohen, A.C. (1976). Progressively censored sampling in the three parameter lognormal distribution. Technometrics, 18,99-103.

Data Interpretation Group (1989). Joint Evaluation of Upstream/Downstream Niagara River Monitoring Data, 1987-1988. A Joint Publication of EC, USEPA, MOE and HYSDEC.

El-Shaarawi, A.H. (1989). Inferences about the mean from censored water quality Data. Water Resour. Res., 25, 685-690.

El-Shaarawi, A.H., and Dolan, D.M. (1989). Maximum likelihood estimation of water quality concentrations from censored data. Canadian Journal of Fisheries and Aquatic Sciences, 46,v 1033-1039.

Gilliom, R.J., and Helsel, D.R. (1986). Estimation of distributional paramaters for censored trace level water quality data, 1, Estimation Techniques. Water Resour. Res., 22, 135-146.

Gupta, A.K. (1952). Estimation of the mean and standard deviation of a normal population from a censored sample. Biometrika, 39,260-273.

Harter, H.L. (1970). Order statistics and their use in testing and estimation, vol. 2, Washington, U.S. Government printing office.

Harter, H.L. and Moore, A.H. (1966). Iterative maximum likelihood estimation of the parameters of normal populations from singly and doubly censored samples. Biometrika, 53,205-213.

IMSL (1987) Math/Library and Stat/Library, IMSL, Inc., Houston: Texas

Land, C.E. (1972). An evaluation of approximate confidence interval estimation methods for lognormal means. Technometrics, 14, 145-158.

Schneider, H. (1986). Truncated And Censored Samples From Normal Populations. Marcel Dekker, Inc.: New York.

Shumway, R.H., Azari, A.S. and Johnson, P. (1989). Estimating mean concentrations under transformation for environmental data with detection limits. Technometrics, 31, 347-356.

Tiku, M.L. (1967). Estimating the mean and standard deviation from a censored normal sample. Biometrika, 54, 155-165.

TABLE 1a

N	n	$\hat{\mu}$	$\hat{\sigma}$	$V(\hat{\mu})$	$Cov(\hat{\mu},\hat{\sigma})$	$V(\hat{\sigma})$	$E(\hat{\mu}\|\sigma)-\mu$	$E(\hat{\sigma}\|\mu)-\sigma$	(D_1, D_2)
30	27	-.004	.98	.033	-.001	.018	.0006	-.007	(-1.6,-1.5)
60	55	-.01	.994	.017	-.0004	.009	-.0004	-.003	(-1.6,-1.5)
120	111	.001	.997	.008	-.0002	.004	.0009	-.002	(-1.6,-1.5)
240	223	.0003	.999	.004	-.00008	.002	.0004	-.001	(-1.6,-1.5)
30	13	-.022	.989	.057	-.025	.046	-.005	.00001	(0., .1)
60	27	-.016	.996	.027	-.011	.022	-.003	-.0003	(0., .1)
120	55	-.003	.999	.013	-.005	.011	-.002	-.0001	(0., .1)
240	110	-.002	1.0003	.006	-.002	.005	-.001	-.0001	(0., .1)

Table 1a. Simulation results for the standard normal distribution.

TABLE 1b

N	n	$\hat{\mu}_\circ$	$\hat{\sigma}_\circ$	$\hat{\gamma}_\circ$	$\tilde{\gamma}_\circ$	Approximate 95% confidence interval for $\gamma_\circ$	Probability of Coverage	(D_1, D_2)
30	28	-.008	1.960	8.746	6.631	(2.211, 39.939)	91.7	(.04076, .04505)
60	56	-.020	1.989	8.016	7.109	(3.135, 21.012)	93.6	(.04076, .04505)
120	112	.002	1.994	7.788	7.364	(4.070, 15.000)	95.2	(.04076, .04505)
240	225	.0007	1.998	7.612	7.409	(4.837,11.999)	93.6	(.04076, .04505)
30	14	-.044	1.978	11.523	6.021	(1.804, 444.237)	89.2	(1., 1.10517)
60	28	-.033	1.992	8.549	7.021	(2.737, 30.722)	92.5	(1., 1.10517)
120	57	-.006	1.998	7.996	7.369	(3.723, 17.661)	94.9	(1., 1.10517)
240	115	-.004	2.0007	7.719	7.436	(4.554, 13.161)	93.8	(1., 1.10517)

Table 1b. Simulation results for the lognormal distribution with mean 7.389 and standard deviation 54.096. The mean and the standard deviation for the corresponding normal distribution are 0 and 2 respectively.

TABLE 2

Data	N	n	$\hat{\lambda}$	$\mu_{\hat{\lambda}}$	$\sigma_{\hat{\lambda}}$	$\hat{\gamma}$	Confidence Interval for γ	(D_1, D_2)
Fluoranthene	44	27	.16	-.660	.662	.618	(.480,.755)	(.35,.4)

Table 2. The results for the Fluoranthene data.

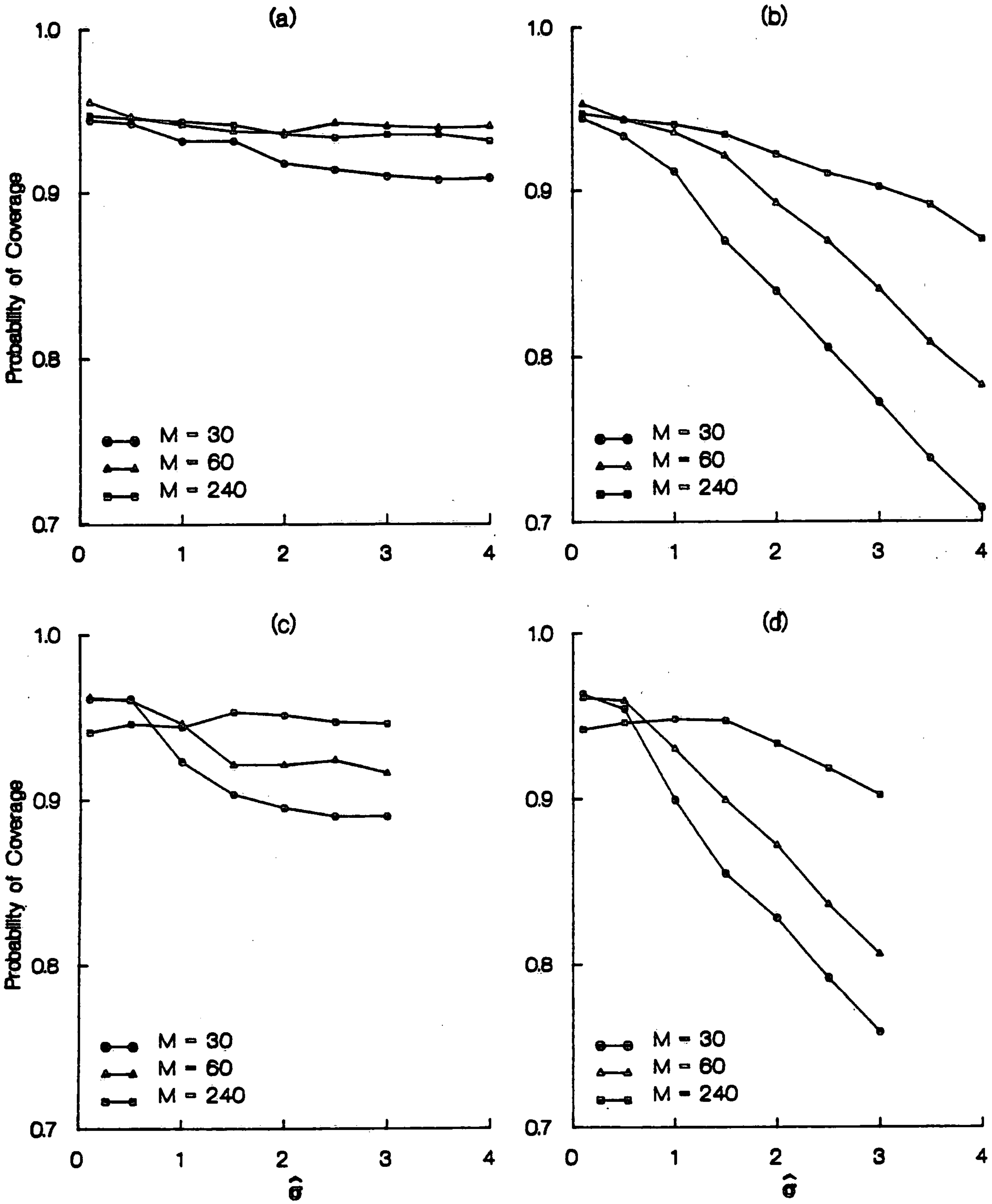

Figure 1. Probability of coverage of the 95% confidence interval for the lognormal mean :
(a) low-level censoring - direct method (b) low-level censoring - Taylor expansion method
(c) high-level censoring - direct method (d) high-level censoring - Taylor expansion method .