Quantification of Uncertainty and Variability for Censored Data Sets in Air Toxics

H. Christopher Frey and Yuchao Zhao
Department of Civil Engineering
Campus Box 7908
North Carolina State University
Raleigh, NC 27695-7908
frey@eos.ncsu.edu

ABSTRACT

Two-dimensional Monte Carlo simulation can be used to estimate the variability and uncertainty of emissions of urban air toxics for use in human exposure and risk analysis. The key steps in the two-dimensional approach include fitting a parametric distribution to data representing variability in emissions, and to use a method such as bootstrap simulation to estimate uncertainty in average emissions and regarding other statistics of the distribution. Many emission factors for urban air toxics contain several values reported only as below detection limit. Such data sets are referred to as "censored." To analyze the variability and uncertainty of censored data sets, empirical bootstrap simulation was used first to get bootstrap samples from the censored data sets. In the bootstrap simulation, randomly samples values from the original data set that were below the detection limit are treated as non-detected values in the bootstrap samples. A parametric probability distribution was fit to each bootstrap sample. The parameters of the distribution were estimated using Maximum Likelihood Estimation (MLE), which is asymptotically unbiased when applied to censored data sets. The use of MLE allows for a distribution to be fit to the entire domain of the emission factor, including estimation of the portion of the distribution that is below the detection limit. Typically, 500 bootstrap samples and distributions were generated. From the 500 alternative distributions, sampling distributions for uncertainty in any statistic can be calculated, such as for the mean. The method is illustrated with two case studies. In one case study, the detection limit was varied and the sensitivity of the results was evaluated. In a second case study, the method was applied to an empirical data set for arsenic emissions from a combustion source. The results show that with more censoring, the uncertainty of the non-detected portion of the fitted distribution increases but the range of uncertainty in the mean does not increase as much. For the empirical data set, the data contain multiple detection limits. The uncertainty of the mean for the arsenic emission factors was found to be -91% to +260%. Future work to further demonstrate this technique for dealing with censored data is recommended.

Keywords: Urban air toxics, Censored data sets, Maximum likelihood estimation, Empirical bootstrap simulation

INTRODUCTION

Toxic air pollutants pose human health risks in urban areas. The U.S. Environmental Protection Agency has developed an Integrated Urban Air Toxic Strategy, which includes a framework for addressing urban air toxics emissions. EPA developed a list of 33 urban air toxics, which represent the priority for additional assessment of the health effects of air toxics in urban areas.[1]

Air toxic emissions are subject to both variability and uncertainty. Variability refers to diversity or heterogeneity among members of population. For example, emissions of a particular pollutant typically differ from one specific emission source to another within a given source category. Uncertainty arises due to lack of knowledge regarding the true value of a quantity or regarding the true distribution for variability. Variability and uncertainty can both be represented as probability distributions in a two-dimensional modeling framework.[2] For example, confidence intervals can be constructed with respect to a best estimate of the cumulative distribution function (CDF) for variability.

The range of the confidence intervals for the CDF represent uncertainty. Information regarding variability in urban air toxic emissions is needed to identify high emitters or highly exposed populations. Information regarding uncertainty is needed to characterize the quality of an emissions inventory and to target data collection to reduce uncertainty.

Because of inherent limitations of chemical/analytical measurement methods, emissions data sets often contain several observations reported as below detection limit, which are usually referred to as "censored".[3] For example, data sets for emission factors of urban air toxics are often reported as censored with a single detection limit or with multiple detection limits. Multiple detection limits arise when data are collected by different sampling and analytical procedures or when data are collected with different gas sampling volumes.

Commonly used methods for dealing with environmental data sets that contain detection limits are statistically biased and are limited in their usefulness. Such methods are typically aimed only at developing an estimate of the mean of the data set. The several methods most often used include ignoring the non-detected values, replacing non-detected values with zero, replacing non-detected values with the detection limit (DL) or replacing non-detected values with one-half of the detection limit. All these methods cause biased estimation of the mean. Furthermore, they do not provide insight regarding the population distribution from which the measured data are a sample.

In contrast to commonly used methods, Maximum Likelihood Estimation (MLE) can be used to fit parametric distribution to censored data sets, including to the portion of the distribution that is below detection limit. Asymptotically unbiased estimates of statistics, such as the mean, can be estimated based upon the distribution fitted using MLE to censored data. The MLE approach enables a point estimate of a statistic such as the mean. In order to estimate uncertainty in the statistic, the method of bootstrap simulation is employed in this work.

The objectives of this paper are: (1) to fit parametric distributions using MLE to censored data sets; and (2) to quantify variability and uncertainty for censored data sets using empirical bootstrap simulation. The methods are tested based upon synthetic data sets, and then are applied to an example emission factor data set. A comparison of the results for the mean based upon the MLE estimate of the fitted distribution with conventional approaches for dealing with non-detects when estimating the mean is provided. Recommendations are made regarding future application of the MLE-based technique for dealing with censored data sets.

COMPARISON OF METHODS USED TO ANALYZE ENVIRONMENTAL CENSORED DATA SETS

Bharvirkar[4] reported that there have studies of methods for estimating the parameters of probability distributions fitted to a censored data set. However, there have been few studies for the application of these methods to environmental data for which parametric distributions are unknown and for which sample sizes are typically small.

Hass and Scheff[5] have compared various methods described in the environmental science and engineering literature for estimation of means of censored data belonging to normal distributions. The following methods were indicated as the most commonly used ones:

- Using values only above DL to calculate a mean value, and ignoring information regarding non-detects;
- Replacing values below DL with zero, which leads to an underestimate of the true mean;
- Replacing values below DL by DL/2, which leads to an approximate but biased estimate of the true mean;

Replacing values below DL by DL, which leads to an approximate but biased estimate of the true mean;

Maximum likelihood estimation, which is bias-corrected.

Newman *et al.*[6] compared the five methods mentioned above. They conclude that the first four techniques, listed above, produce biased estimate of both the mean and variance. The bias worsens as the amount of censoring increases. MLE estimates were found to be easier to develop and more accurate compared to other methods. Similar comparative studies have also been conducted by Gilliom and Helsel[7] with similar results regarding the methods listed above.

MLE PARAMETER ESTIMATES FOR CENSORED LOGNORMAL AND GAMMA DISTRIBUTIONS

For environmental data sets, such as concentrations or emission factors, lognormal and gamma distribution are often chosen as parametric distribution to represent variability in data. Therefore, in this paper, lognormal and gamma distribution are used to illustrate and evaluate the MLE-based method for fitting parametric distributions to censored data.

The lognormal distribution is defined by the following probability distribution function[9]

$$f(x \mid m, s^2) = \frac{1}{s\sqrt{2p}} \frac{e^{-(\log x - m)^2/(2s^2)}}{x} \tag{1}$$

where $0 \le x < \infty,\ -\infty < m < +\infty,\ s > 0$. In this paper, m is defined as parameter1 and s is defined as parameter 2.

The gamma distribution is defined by the following probability distribution function[9]

$$f(x \mid a, b) = \frac{1}{\Gamma(a) b^a} x^{a-1} e^{-x/b} \tag{2}$$

where $0 \le x < \infty,\ a, b > 0$. In this paper, a is defined as parameter 1 and b is defined as parameter 2.

The MLE technique is applicable to data sampled from various distributions and it is easily implemented in a computer program. The most general formulation of the likelihood function for censored data sets having detection limits is:[4]

$$L(q_1, q_2, \ldots, q_k) = \prod_{i=1}^{n} f(x_i | q_1, q_2, \cdots, q_k)\{\prod_{m=1}^{p}(\prod_{j=1}^{ND_m} F(DL_m | q_1, q_2, \cdots, q_k))\} \tag{3}$$

where,

x_i = Detected data point, where, i = 1, 2, …, n

$q, q_2, \ldots, q_k$ = Parameters of the distribution

ND_m = Number of non-detects corresponding to detection limit DL_m, where, m = 1, 2, …, P.

p = Number of detection limits

f = Probability density function

F = Cumulative distribution function

In essence, the likelihood of each detected data point is included in the likelihood function. For data that are below detection, the cumulative probability of the detection limit is used in lieu of the likelihood.

For computational convenience, it is more common to work with the log-likelihood function instead of the likelihood function itself. For the lognormal distribution, the log-likelihood function including censored data is:

$$J(m, s) = -n \ln s - \frac{n}{2}\ln(2p) - \sum_{i=1}^{n}\left\{\frac{(x_i - m)^2}{2s^2}\right\} + \sum_{m=1}^{P} ND_m \ln\left\{0.5\left[1 + erf\left(\frac{DL_m - m}{s\sqrt{2}}\right)\right]\right\} \tag{4}$$

where m = mean, s = standard deviation of the log-transformed data.

For the gamma distribution, the log-likelihood function including censoring is given by:

$$J(a,b) = -n\{a\ln(b) + \ln[\Gamma(a)]\} + \sum_{i=1}^{n}\left\{(a-1)\ln(x_i) - \frac{x_i}{b}\right\} + \sum_{m=1}^{p} ND_m \ln\left\{\frac{\int_0^{DL_m/b} e^{-t}t^{a-1}dt}{\Gamma(a)}\right\} \quad (5)$$

where a = shape parameter, b =scale parameter.

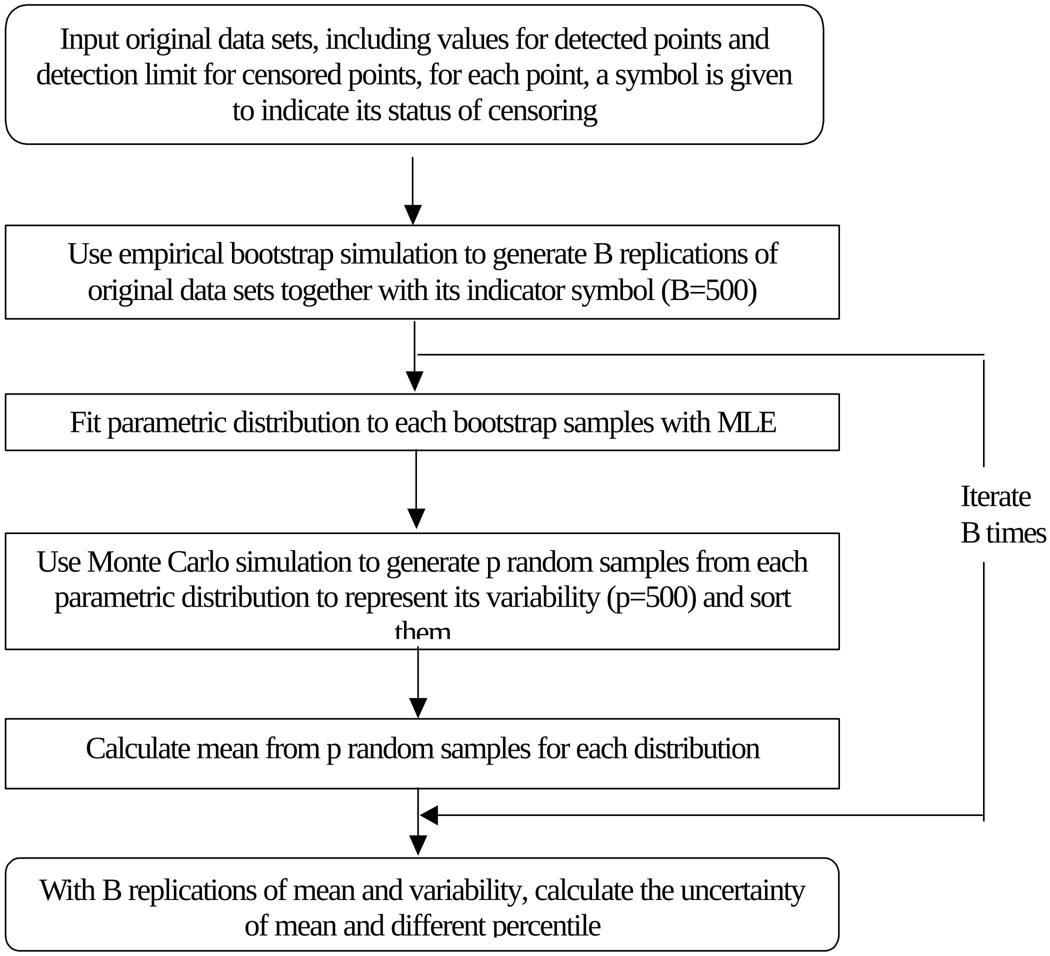
A non-linear programming optimization algorithm was used to maximize the log-likelihood functions of (4) and (5). Computer subroutines by Press *et al.* were used for optimization and for evaluation of the various special functions required in some of the log-likelihood functions, such as the gamma function, error function, and beta function.[4,10]

ESTIMATION OF UNCERTAINTY IN STATISTICS USING BOOTSTRAP SIMULATION

Parametric bootstrap simulation is widely used to estimate confidence intervals for statistics of data sets or parameters of fitted distributions in cases without censoring.[8] In conventional parametric bootstrap simulation, a parametric probability distribution is fit to the original data set, which has a sample size of *n*. Monte Carlo simulation is used to randomly simulate synthetic data sets, referred to as bootstrap samples, each of sample size *n*. Typically, *B* bootstrap samples are simulated. For each bootstrap sample, a replication of a given statistic is calculated. For example, one can obtain *B* estimates of the mean, standard deviation, or distribution parameters. The numerical value of each replication of a statistic will differ from that of other replications because of the effect of random sampling when comparing one bootstrap sample with another. The B estimates of a statistic are used to describe a probability distribution for the statistic. A probability distribution for a statistic is also referred to as a sampling distribution. The sampling distribution is used as the basis for estimating confidence intervals. For example, the 95 percent confidence interval for the mean is based upon the interval between the 2.5th and 97.5th percentiles of the sampling distribution for the mean.

In the case of a censored data set, the conventional approach to parametric bootstrap simulation cannot be directly applied. Specifically, it is necessary to generate bootstrap samples so that there can be random variation in the number of data points that are below detection limit. In order to do this, an empirical bootstrap approach is used. In empirical bootstrap simulation, each of the original *n* data points are sampled with replacement and with equal probability of being sampled. In the original data set, either the value of the data point is given for detected data or the detection limit is given for the censored data. Therefore, for each point in the data sets, an indicator symbol d is given to indicate whether it is a detected value or below a detection limit. A value of d equal to 1 was used to represent a data point below a detection limit and d equal to 0 was used to represent a detected data point. In the case of non-detected data, the numerical value of the data point used in the bootstrap simulation was the detection limit itself. When generating bootstrap samples from the original censored data set, both the data point value and its indicator symbol were sampled together. Therefore, for each bootstrap sample, it is known as to which data points are detected and which data points are censored. For each sampled bootstrap sample, MLE was used to fit a parametric distribution. Thus, B estimates of the distribution parameters and of the fitted distributions were developed. In this analysis, B=500. For each alternative fitted distribution, 500 values of the mean and standard deviation are simulated by generating 500 random samples from each of the 500 distributions fitted to the 500 bootstrap samples. The mean and

Figure 1. Scheme of Quantification of Uncertainty and Variability for Censored Data Sets



standard deviation are estimated from the 500 random samples generated from each fitted distribution. The overall scheme of the bootstrap simulation method for censored data is described in Figure 1.

TESTING AND EVALUATION OF THE BOOTSTRAP METHOD FOR QUANTIFYING UNCERTAINTY IN THE STATISTICS OF CENSORED DATA SETS

The bootstrap sampling method was tested by application to synthetic data sets in order to evaluate how the method performed. For this purpose, 20 data points were randomly generated from a gamma distribution with a mean equal to 1 and a standard deviation equal to 1. The samples in the randomly generated data set were ranked in ascending order. Different detection limits were assigned to the synthetic data set in order to achieve different amounts of censoring. For example, a detection limit of 0.4375 was assigned in order to obtain 30% censoring, resulting in 6 points out of 20 below the detection limit. A detection limit of 0.8127 was assigned in order to get 60% censoring in which 12 points out of 20 data points were below the detection limit. However, all of the synthetic data sets have the same numerical values of the detected data.

Table 1 summarizes the results for fitting of Gamma and Lognormal distributions to the uncensored and censored data sets based upon the synthetic data set. The parameters were estimated using MLE based upon Equations (4) and (5) for the lognormal and gamma distributions, respectively. For both the lognormal and gamma distributions fitted to the data, the parameter values change as the amount of censoring changes. This is because with additional censoring, there are fewer data points available from which to estimate the likelihood function.

Table 1. Results of the Parameter Estimation and the Statistics of the Fitted Distribution for Synthetic Data Sets

Distribution	Gamma Distribution			Lognormal Distribution		
Censoring Percentage	0%	30%	60%	0%	30%	60%
Parameter 1	1.253	1.097	0.814	-.434	-.390	-.355
Parameter 2	0.812	0.922	1.204	0.998	0.936	0.955

The results for fitting of parametric distributions to the synthetic data set are shown in Figure 2 for the case of a fitted gamma distribution and no censoring. The fitted distribution is depicted as a dashed line. The apparent clustering of the synthetic data is a random artifact. The confidence intervals for the fitted distribution were estimated using bootstrap simulation. Of the 20 data points, 13 are enclosed by the 50 percent confidence interval, 19 are enclosed by the 90 percent confidence interval, and all 20 are enclosed by the 95 percent confidence interval. Therefore, the gamma distribution is a reasonable fit to the data in this case.

The results for fitting of a gamma distribution to the synthetic data set with 30 percent censoring are shown in Figure 3. The detection limit is shown by a vertical dashed line. Below the detection limit, no data values are shown. Above the detection limit, the data values are the same as in Figure 2. The fitted distribution in Figure 3 appears to be very similar to that as in Figure 2 even though the parameter values are somewhat different as shown in Table 1. However, it is also noticeable that the range of uncertainty, as reflected by the confidence interval of the fitted distribution, is larger for values of the random variable below the detection limit, as revealed by a comparison of the lower end of the distribution in Figure 3 with that in Figure 2. Thus, as expected, there is more uncertainty regarding the estimate of the portion of the distribution that is below detection than there would be if all of the data had been detected.

Figure 4 further emphasizes the increase in uncertainty in the fitted distribution associated with an increase in the amount of censoring. With 60 percent of the 20 samples below detection, the confidence intervals for uncertainty in the portion of the distribution that is below detection are noticeably wider than for Figure 3 or for the uncensored case in Figure 2. It is also apparent that the fitted distribution agrees well with the data that are above detection limit.

Similar results were obtained based upon fitting of a lognormal distribution to the 0%, 30%, and 60% censoring cases as shown in Figures 5, 6, and 7, respectively. Because all of the data are enclosed by the 95 percent confidence interval, and more than half of the data are enclosed by the 50% confidence interval, it can be concluded from Figure 5 that the lognormal distribution is a reasonable fit to this data set. For the censored cases, the fitted distribution agrees well with the data above the detection limit, and the range of uncertainty in the cumulative distribution function at the lower end of the distribution increases with more censoring.

The mean values and the 95 percent confidence intervals for the mean are compared for 0%, 30%, and 60% censoring for both the gamma and lognormal distributions in Table 2. For the gamma distribution, the estimated mean value is approximately 1.1 for all three cases. There is relatively good agreement regarding the mean value because the mean is influenced far more by the upper tail of the distribution, which is above the detection limit in all cases, than by the lower tail of the distribution. The 95 percent confidence interval of the mean is approximately 0.6 to 1.7 in all three cases. Although there is some variation in the estimate of the 95 percent confidence interval of the mean among the three cases, when rounded to one or two significant figures the results are comparable. These results illustrate that it is possible to obtain robust estimates of the mean and of the 95 percent confidence interval for the mean using the MLE/bootstrap technique even when there is substantial censoring.

Table 2. Results of the Uncertainty of the Mean

Distribution	Gamma Distribution			Lognormal Distribution		
Censoring Percentage	0%	30%	60%	0%	30%	60%
Mean of Bootstrap Sample Mean	1.065	1.063	1.100	1.013	1.003	0.972
2.5 Percentile of Mean	0.641	0.652	0.609	0.646	0.625	0.509
97.5 Percentile of Mean	1.735	1.692	1.714	1.511	1.527	1.516
Width of 95% of C.I.	1.094	1.040	1.104	0.866	0.902	1.007

Figure 2. Variability and Uncertainty for Gamma Distribution Fitted to Synthetic Data Set with No Censoring

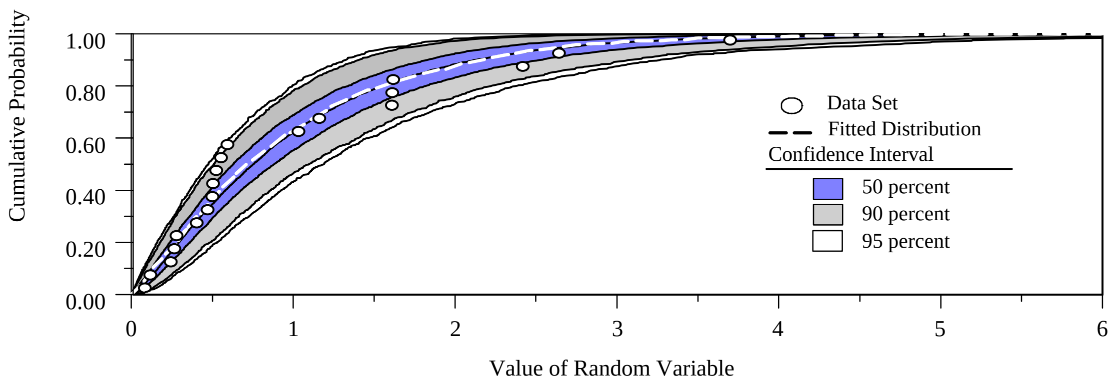


Figure 3. Variability and Uncertainty for Gamma Distribution Fitted to Synthetic Data Set with 30 Percent Censoring

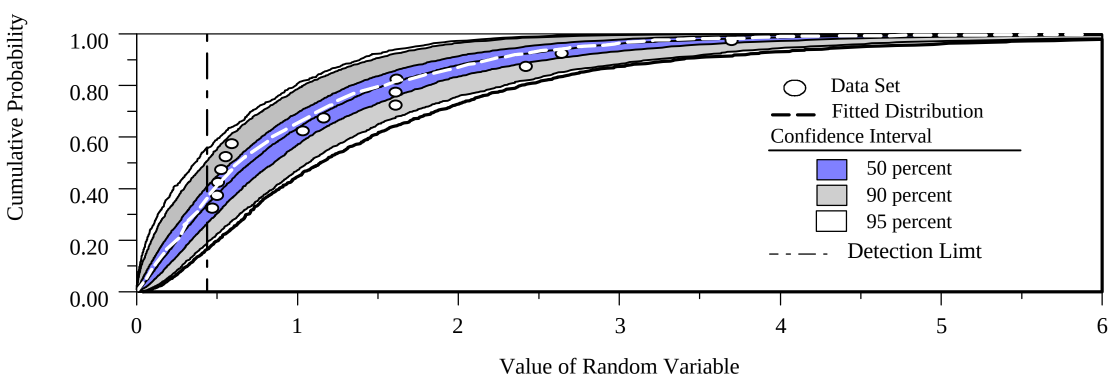


Figure 4. Variability and Uncertainty for Gamma Distribution Fitted to Synthetic Data Set with 60 Percent Censoring

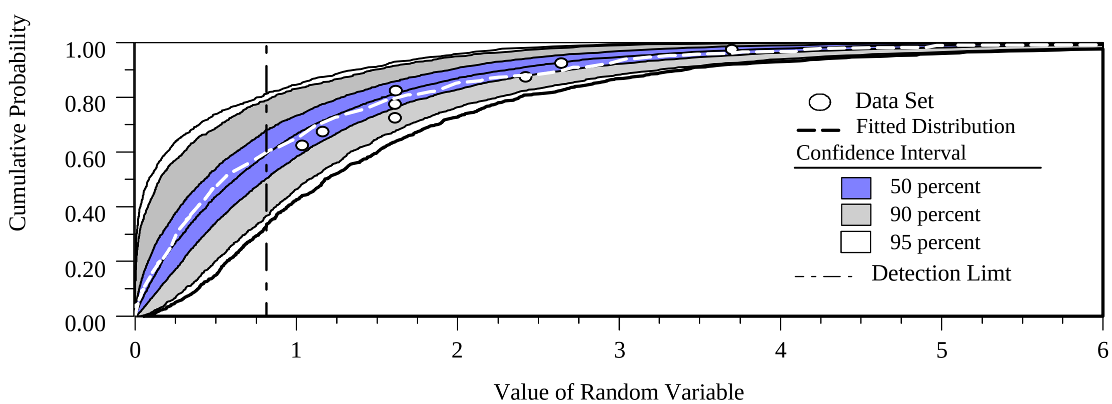


Figure 5. Variability and Uncertainty for Lognormal Distribution Fitted to Synthetic Data Set with No Censoring

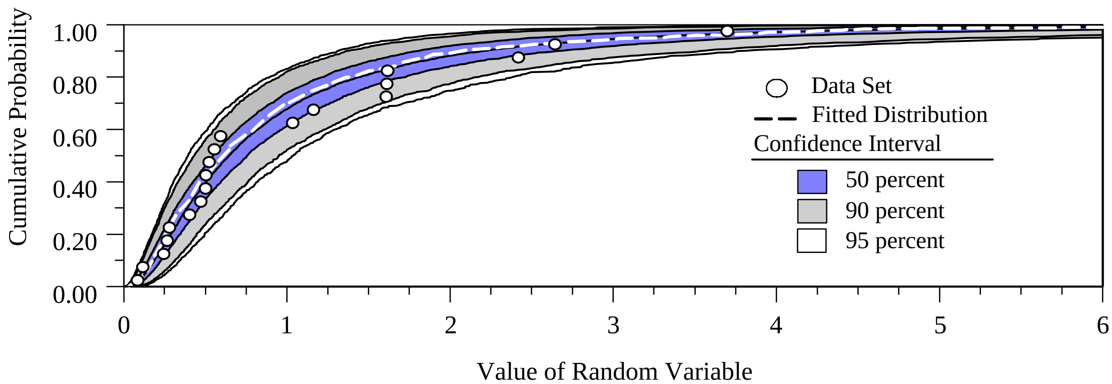


Figure 6. Variability and Uncertainty for Lognormal Distribution Fitted to Synthetic Data Set with 30 Percent Censoring

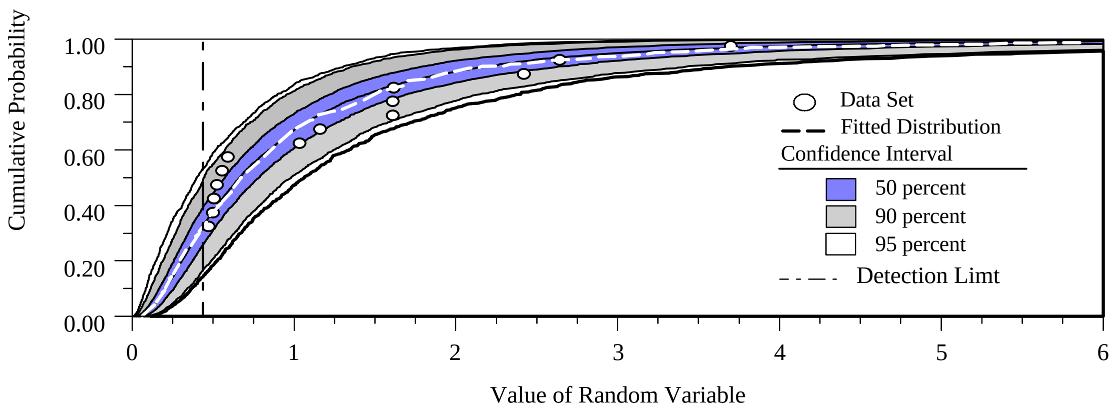
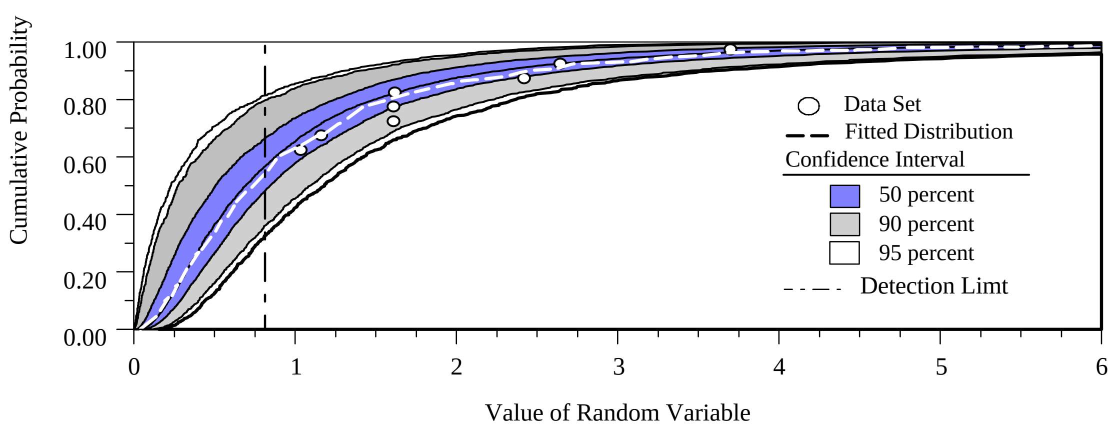


Figure 7. Variability and Uncertainty for Lognormal Distribution Fitted to Synthetic Data Set with 60 Percent Censoring



For the lognormal distribution, the results in Table 2 also illustrate substantial agreement regarding the estimate of the mean and of the 95 percent confidence interval for the mean among the three cases considered. The mean value is approximately 1.0 in all three cases. The 95 percent confidence interval for the mean is approximately 0.6 to 1.5 in all three cases, with the exception that the range of the 95 percent confidence interval is wider for the case with the most censoring. In general, it is expected that the range of uncertainty should increase as the amount of censoring increases. The results given here illustrate that the MLE/bootstrap method for dealing with censored data sets provides reasonable and robust estimates of the mean and of uncertainty in the mean.

For the gamma distribution with 30% censoring, different methods as addressed previously were compared in terms of the estimate of the mean value. The mean value is estimated to be 1.352 if just considering detected values, 0.947 if replacing values below DL with zero, 1.012 if replacing values below DL with DL/2 and 1.078 if replacing values below DL with DL. Comparing with the asymptotically unbiased mean of 1.063 using the MLE method, the mean appears to be underestimated for all cases except when the non-detects are replaced with the DL, in which case the mean is over-estimated. However, a key point is that the approximately methods are not only biased, but they cannot be used to develop estimates of uncertainty in the mean.

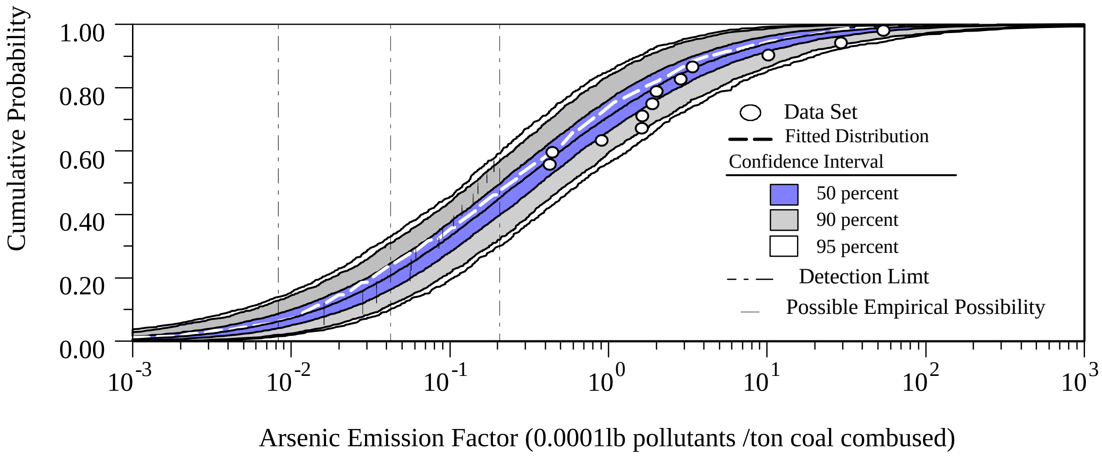
CASE STUDY FOR ESTIMATION OF VARIABILITY AND UNCERTAINTY IN AN EMISSION FACTOR BASED UPON CENSORED DATA

As an example of an emission factor data set that contains censoring, a data set for arsenic emissions from combustion sources was selected. The data set contains 29 data points, of which 3 are censored. Each of the three censored data points has a different detection limit. There are 12 data values that are greater than the largest of the three detection limits. There are no detected data points that are smaller than the smallest of the three detection limits. There are three detected data values that are greater than the smallest detection limit but smaller than the second largest detection limit. There are 11 detected data values that are larger than the second detection limit but smaller than the third detection limit. Thus, there is uncertainty regarding what value of cumulative probability to assign to the total of 14 detected data points that are less than the largest detection limit. For example, the true but unknown values of the three non-detected data points could all be less than the numerical value of the smallest detection limit, which would imply that the smallest detected data point could have a rank of 4. However, it is also possible that the smallest detected data value could have a rank of 2 because this data point has a value greater than the smallest detection limit. Therefore, the true but unknown value of the data point below the smallest detection limit must be less than the value of the smallest detected data point. The uncertainty regarding the rank of the detected data points is indicated by vertical lines representing each detected data point in Figure 10. For the 12 values that are larger than the largest detection limit, there is no uncertainty regarding the cumulative probability in these cases.

A lognormal distribution was fitted to the arsenic emission factor data set using MLE and taking into account that there were three non-detected data with three different detection limits. The fitted distribution is shown as the white dashed line in Figure 10. The white dashed line agrees reasonably well with the observed data, taking into account uncertainty regarding the cumulative probability of the detected values that have smaller values than the largest detection limit. The 95 percent confidence interval on the fitted distribution encloses all of the detected values, and approximately one half of the data are enclosed by the 50 percent confidence interval.

The detected arsenic emission factor data vary over almost four orders-of-magnitude. Therefore, there is a large amount of uncertainty in the mean value associated with the small sample size, large amount of variability in the data, and the presence of non-detected measurements in the data set. The mean estimated based upon the lognormal distribution fitted to the data using MLE was found to be 8.24. The 95 percent confidence interval for the mean ranges from minus 91 percent to plus 264 percent of the mean value. The asymmetric of this confidence interval is based upon the large amount of variability in the data, the relatively small sample size, and the fact that an emission factor must be non-negative.

Figure 10. Variability and Uncertainty in the Arsenic Emission Factor for a Combustion Source Estimated Based Upon a Lognormal Distribution Fitted to Data Containing Three Non-Detects with Three Different Detection Limits



CONCLUSIONS

In this paper, a statistically rigorous method for fitting parametric distributions to censored data sets using Maximum Likelihood Estimation has been described that can be applied to multiply censored data and to data containing a large amount of censoring. The parametric distributions represent variability within the data set. The MLE method is asymptotically unbiased and is robust to large amounts of censoring. A bootstrap simulation method was introduced for estimating uncertainty in statistics such as the mean that are estimated from censored data. Test cases with a synthetic data set containing 0%, 30%, and 60% censoring to which both gamma and lognormal distributions were fit indicated that robust estimates can be obtained for the mean and for the 95 percent confidence interval of the mean even for as much as 60% censoring. Based upon the positive results from evaluation of the method on a synthetic data set, the method was applied to an empirical data set for an arsenic emission factor that contained three non-detected values, each with its own detection limit. The results from the arsenic emission factor data set illustrate the importance of using a method that does not impose normality assumptions on the estimate of uncertainty in the mean. The assymmetry of the uncertainty estimate appropriately reflects the large amount of variability in the data, the relatively small sample size, and that an emission factor must be non-negative. It is recommended that the MLE and bootstrap methods described here be tested more systematically on a larger number of synthetic data sets in order to characterize the robustness of the method when there are varying sample sizes, different ranges of variability in the data, different amounts of censoring, and different parametric distributions (e.g., lognormal, gamma, Weibull). After more thoroughly demonstrating and verifying the method, the method can and should be applied to empirical data sets in place of simplified biased methods in order to obtain unbiased estimates of the mean and regarding uncertainty in the mean.

8.0 ACKNOWLEDGEMENTS

The authors acknowledge the support of the Science to Achieve Results (STAR) grants program of the U.S. Environmental Protection Agency, which funded this work under Grant No. R826790. Although the research described in this paper has been funded wholly or in part by the U.S. EPA, this paper has not been subject to any EPA review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.

9.0 REFERENCES

1. Smith, R.L.; French, C.L. "Ranking and Selection of Hazardous Air Pollutants for Listing Under Section 112(k) of the Clean Air Act Amendments", EPA, 1999. http://www.epa.gov/ttn/uatw/urban/urbanpg.html (accessed by 10/1/ 2000)
2. Frey, H.C.; Bharvirkar, R. "Quantitative Analysis of Variability and Uncertainty in Emissions Estimation", Prepared for Office of Air Quality Planning and Standards U.S. Environmental Protection Agency Research Triangle Park, NC.
3. Rao, S.T.; Ku, J.Y. and Rao, K. S. "Analysis of Toxic Air Contaminant Data Containing Concentrations Below the Limit of Detection," *J. Air Waste Manage. Assoc.* **1991,** 41: 442-448
4. Bharvirkar, R. M.S. Thesis, North Carolina State University, 1999.
5. Hass, C.N., and Scheff, P.A. (1990), "Estimation of Averages in Truncated Samples," *Environmental Sciences and Technology, 24(6): 912-919.*
6. Newman, M.C., Dixon, P.M., Looney, B.B., and Pinder, J.E. (1989), "Estimating Mean and Variance for Environmental Samples with Below Detection Limit Observations", *Water Resources Bulletin*, 25(4): 905-915.
7. Gilliom, R.J., and Helsel, D.R. (1986), "Estimation of Distributional Parameters for Censored Trace Level Water Quality Data 1. Estimation Techniques," *Water Resources Research*, 22(2): 135-146.
8. Efron, B., and R.J. Tibshirani (1993), An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability 57, Chapman & Hall: New York.
9. Casella, G; Berger, R.L. *Statistical Inference*. Duxbury Press: Pacific Grove, CA.
10. Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (1992). *Numerical Recipes in Fortran: The Art of Scientific Computing, 2nd ed.*, Cambridge University Press: New York.
11. "Emission Factor Documentation for AP-42 Section 1.1 Bituminous and Subbituminous Coal Combustion" Prepared by Acurex Environmental Corporation, for Office of Air Quality Planning and Standards Office of Air and Radiation, EPA, 1993. http://www.epa.gov/ttn/chief/ap42/ch01/bgdocs/b01s01.pdf (accessed by 4/ 2001)