Table of Contents

1.0 INTRODUCTION I

1.1 Variability and Uncertainty..... 1
1.2 Examples of Probabilistic Analysis 7
1.3 Key Insights from Probabilistic Analysis 10
1.4 SHEDS Model, Project Objectives, and Software Needs, ... 13
1.5 Available Software..... 14
1.6 Need for New Software..... 16
1.7 Overview of this Report..... 17

2.0 ALGORITHMS USED IN AUVTOOL.....19

2.1 Visualizing Data Using Empirical Distribution..... 19
2.2 Definitions of Probability Distributions..... 22
2.2.1 Definition of Parametric Probability Distributions 23
2.2.2 Empirical Distribution..... 24
2.3 Parameter Estimation of Probability Distributions 24
2.3.1 Method of Matching Moments 27
2.3.2 Maximum Likelihood Estimation (MLE)..... 31
2.3.3 Parameter Estimation Method Availability for Probability ..... Distributions..... 34
2.4 Algorithms for Generating Random Samples from Probability Distributions..... 34
2.4.1 Pseudo Random Number Generator 35
2.4.2 Normal Distribution..... 36
2.4.3 Lognormal Distribution 38
2.4.4 Beta Distribution..... 38
2.4.5 Gamma Distribution..... 38
2.4.6 Weibull Distribution 40
2.4.7 Uniform distribution 40
2.4.8 Symmetric Triangle Distribution..... 40
2.4.9 Empirical Distribution..... 41
2.5 Evaluation of Goodness-of-Fit of a Probability Distribution Model..... 42
2.5.1 Graphical Comparison of CDF of Fitted Distribution to the Data 45
2.5.2 Kolmogorov-Smirnov Test..... 46
2.5.3 Anderson-Darling Test..... 48
2.5.4 Graphical Comparison of Confidence Intervals for CDF of Fitted Distribution to the Data 50
2.5.5 Criteria for Automatically Seeking a Best Distribution Model in Batch Mode Analysis 51
2.5.6 Summary of Methods for Evaluating Goodness-of-Fit 53
2.6 Characterization of Variability and Uncertainty..... 54
2.6.1 Bootstrap Method..... 55

2.6.2 Methods of Generating Bootstrap Samples 57
2.6.3 Methods of Forming Bootstrap Confidence Intervals 57
2.6.4 Two-dimensional Simulation of Variability and Uncertainty ... 59
2.7 Evaluation of Dependence or Correlation between Statistics of Interest 61
2.8 Summary 62

3.0 AUVTOOL SYSTEM DEVELOPMENT AND IMPLEMENTATION 65

3.1 AuvTool Software Design Considerations 65
3.2 Development Environment and Tools 66
3.3 Structure Design of the AuvTool System 66
3.4 AuvTool Main Modules 66
3.4.1 Data Entry, Importing and Exporting Module 68
3.4.2 Loading Distribution Information Module 68
3.4.3 Random Seed Setting and Random Sample Generator Modules 69
3.4.4 Variability Analysis-Fitting Distribution Dataset by Dataset Module 69
3.4.5 Batch Analysis Module 70
3.4.6 Uncertainty Analysis-Bootstrap Simulation Module 72
3.4.7 Analyzing the Sampling Data of Statistics of Interest Module 73
3.4.8 Variability and Uncertainty Reporting Analysis Modules 74

4.0 VERIFCATION OF AUVTOOL 77

4.1 Verification of Random Number Generation for Probability Distribution Models 78
4.1.1 Specifying Parameters and Generating Random Numbers..... 78
4.1.2 Graphical Comparison of Random Numbers with Specified Distribution 81
4.1.3 Goodness-of-Fit Tests for Random Numbers 82
4.1.4 Summary of Results for Verification of the Random Number Generator 86
4.2 Verification of Parameter Estimation Algorithms 86
4.2.1 Method for Verification and Comparison of Parameter Estimates 87
4.2.2 Results of Parameter Calculations 91
4.2.3 Summary of Parameter Estimation Verification..... 97
4.3 Verification of Confidence Intervals 99
4.3.1 Verification of the Confidence Interval for Mean 100
4.3.2 Verification of the Confidence Interval for the Standard Deviation of the Normal Distribution 107
4.3.3 Summary of Results for Verification of Confidence Intervals 109
4.4 Verification of Algorithms for Goodness-of-Fit Tests 109

4.5 Evaluation of the Stability of Bootstrap Simulation Results 113
4.6 Overall Findings Regarding Verification of AuvTool..... 115

5.0 AN ILLUSTRATIVE CASE STUDY USING AUVTOOL.....117

5.1 Introduction to Case Study..... 117
5.2 AuvTool Analysis Results 117

6.0 CONCLUSIONS AND RECOMMENDATIONS 121

6.1 Conclusions 121
6.2 Recommendations 122

7.0 REFERENCES 127

APPENDIX A. GRAPHICAL COMPARISON FOR RANDOM SAMPLE GENERATION.....135

APPENDIX B: DATA USED IN VERIFICATION TESTS 143

List of Tables

Table 2-1. The Definitions of Parametric Probability Distributions 23

Table 2-2. Parameter Estimation Method Availability for Parametric Probability Distributions 35

Table 2-3. Citical Value of D_n^a the Kolmogorov-Smirnov Test..... 48

Table 2-4. The Critical Values for Anderson-Darling test for Normal, Lognormal and Weibull distributions 50

Table 2-5. The Critical Values for Anderson-Darling test for the Gamma Distribution..... 50

Table 3-1. AuvTool Function Module Summarization Table 67

Table 4-1. Parameters of the Tested Distributions..... 80

Table 4-2. Summary of the Results of K-S Test with critical value of $D_{1000}^{0.05}$ = 0.0430..... 83

Table 4-3. Procedure of the Chi-Squared test..... 85

Table 4-4. Chi-Squared Test for Random Samples and Specified Empirical Distribution 86

Table 4-5. Estimated Normal Parameters 92

Table 4-6. Estimated Lognormal Parameters..... 93

Table 4-7. Estimated Gamma Parameters..... 94

Table 4-8. Estimated Weibull Parameters..... 95

Table 4-9. Estimated Beta Parameters 96

Table 4-10. Estimated Uniform Parameters..... 97

Table 4-11. Estimated Symmetric Triangle Parameters 97

Table 4-12. Absolute 95% Confidence Interval of Mean for Normal Distribution..... 103

Table 4-13. Absolute 95% Confidence Interval of Mean for Lognormal Distribution 104

Table 4-14. Absolute 95% Confidence Interval of Mean for Gamma Distribution 104

Table 4-15. Absolute 95% Confidence Interval of Mean for Weibull Distribution..... 105

Table 4-16. Absolute 95% Confidence Interval of Mean for Beta Distribution..... 105

Table 4-17. Absolute 95% Confidence Interval of Mean for Uniform Distribution 105

Table 4-18. Absolute 95% Confidence Interval of Mean for Symmetric Triangle Distribution 106

Table 4-19. Absolute 95% Confidence Interval of Mean for Empirical Distribution[a] 106

Table 4-20. Absolute 95% Confidence Interval of Standard Deviation for Normal Distribution 108
Table 4-21. Verification of K-S Test Results for Normal Distribution 110
Table 4-22. Verification of K-S Test Results for Logormal Distribution 110
Table 4-23. Verification of K-S Test Results for Gamma Distribution 110
Table 4-24. Verification of K-S Test Results for Weibull Distribution 111
Table 4-25. Verification of K-S Test Results for Beta Distribution 111
Table 4-26. Verification of K-S Test Results for Uniform Distribution 111
Table 4-27. Verification of K-S Test Results for Symmetric Triangle Distribution 112
Table 4-28. Stability of Bootstrap Simulation Results for the Mean and 95 Percent Confidence Intervals of the Mean, Standard Deviation, and Distribution Parameters[a] 114
Table 5- 1. Summarization Table of Input Distribution Information for the Variables without Original Data 118
Table 5-2. AuvTool Variability Analysis Result Summarization Table 118
Table 5-3. AuvTool Summarization Table for Confidence Intervals of Mean and Standard Deviation 119
Table 5-4. AuvTool Summarization Table for the Fitted Distribution to the Sampling Data of Mean Statistic 119
Table B-1. Data Sets for Normal Distribution for Parameter Estimation and K-S Test 143
Table B-2. Data Sets of Normal Distribution for Confidence Interval Test 145
Table B-3. Data Sets for Lognormal Distribution for Parameter Estimation and K-S Test 147
Table B-4. Data Sets of Lognormal Distribution for Confidence Interval Test 149
Table B-5. Data Sets for Beta, Uniform and Symmetric Triangle Distribution for Parameter Estimation and K-S Test 151
Table B-6. Data Sets of Beta Distribution for Confidence Interval Test 153
Table B-7. Data Sets of Uniform Distribution for Confidence Interval Test 155
Table B-8. Data Sets of Symmetric Triangle Distribution for Confidence Interval Test 157
Table B-9. Data Sets for Gamma and Weibull Distributions for Parameter Estimation and K-S Test 159
Table B-10. Data Sets of Gamma Distribution for Confidence Interval Test for Gamma and Empirical Distributions 161
Table B-11. Data Sets of Weibull Distribution for Confidence Interval Test 163
Table B-12. Random Samples Used for Reliability Test 164

List of Figures

Figure 2-1. Plot Illustrating the 95 Percent Probability Range on a Cumulative Distribution Function. 20

Figure 2-2. Example Graph of Visualizing Data Using the Hazen's Plotting Position Method (n=10) 21

Figure 2-3. An example of an Empirical Distribution Represented a Step Function (n=10) 24

Figure 2- 4. Example of a Probability Plot for a Weibull Distribution (n=50) 30

Figure 2-5. Comparison of Fitted Beta Distribution to an Example Dataset 46

Figure 2-6. An Illustrative Example of Graphical Comparison of Confidence Intervals for CDF of Fitted Distribution to the Data 51

Figure 2-7. Flow Diagram For Bootstrap Simulation and Two-Dimensional Simulation of Variability and Uncertainty. 60

Figure 3-1. The Conceptual Structure Design of AuvTool System 67

Figure 4-1. Comparison of Random Samples and Specified Normal Distribution, μ=1, $\acute{o}^2$=0.25 81

Figure A-1. Comparison of Random Samples and Specified Normal Distribution, μ=1,σ^2=1 135

Figure A-2. Comparison of Random Samples and Specified Normal Distribution, μ=1,σ^2=4 135

Figure A-3. Comparison of Rand. Samp. and Specified Lognormal Dist., μ_{lnx}=-0.112, σ_{lnx}=0.472 136

Figure A-4. Comparison of Rand. Samp. and Specified Lognormal Dist., μ_{lnx}=-0.347, σ_{lnx}=0.833 136

igure A-5. Comparison of Rand. Samp. and Specified Lognormal Dist., μ_{lnx}=-0.804, σ_{lnx}=1.269 136

Figure A-6. Comparison of Random Samples and Specified Gamma Distribution, α=4, β=0.25. 137

Figure A-7. Comparison of Random Samples and Specified Gamma Distribution, α=1, β=1. 137

Figure A-8. Comparison of Random Samples and Specified Gamma Distribution, α=0.25, β=4. 137

Figure A-9. Comparison of Random Samples and Specified Weibull Dist., α=2.101, β=1.129. 138

Figure A-10. Comparison of Random Samples and Specified Weibull Distribution, α=1, β=1 138

Figure A-11. Comparison of Random Samples and Specified Weibull Dist., α=0.543, β=0.575. 138

Figure A-12. Comparison of Random Samples and Specified Beta Distribution, α=12, β=12. 139

Figure A-13. Comparison of Random Samples and Specified Beta Distribution, α=1.5, β=1.5. 139

Figure A-14. Comparison of Random Samples and Specified Beta Dist., α=0.281, β=0.281. 139

Figure A-15. Comparison of Random Samples and Specified Uniform Dist., a=-0.5, b=2.5. 140

Figure A-16. Comparison of Random Samples and Specified Uniform Distribution, a=-2, b=4. 140

Figure A-17. Comparison of Random Samples and Specified Uniform Distribution, a=-5, b=7. 140

Figure A-18. Comparison of Rand. Samp. and Specified Symmtr. Triangle Dist., a=1, b=1.225. 141

Figure A-19. Comparison of Rand. Samp. and Specified Symmtr. Triangle Dist., a=1, b=2.450. 141

Figure A-20. Comparison of Rand. Samp. and Specified Symmtr. Triangle Dist., a=1, b=4.899. 141

Figure A-21. Comparison of Empirical Distribution of Randomly Generated Data and Analytical CDF of Assumed Population Distribution for the Empirical Distribution. 142

1.0 INTRODUCTION

The need for the use of quantitative methods for characterizing variability and uncertainty in exposure and risk assessment has received increasing attention in recent years (e.g., Bogen and Spear, 1997; Morgan and Henrion, 1990; Frey, 1992; Hoffman and Hammonds, 1994; Frey and Rhodes, 1996; Helton, 1996; NCRP, 1996; EPA, 1997). This report documents the technical basis for a new software tool, AuvTool, that enables an analyst to quantify variability in a data set and to quantify uncertainty in key statistics of the data set. The software was developed specifically to support the Stochastic Human Exposure and Dose Simulation (SHEDS) model that is being developed by the U.S. Environmental Protection Agency (EPA). However, AuvTool was developed as a stand-alone module. Therefore, it can be used for other analyses as well.

In the next section, the concepts of variability and uncertainty are presented. A brief review of some illustrative examples of probabilistic analyses, in which the distinction between variability and uncertainty was made, is provided in Section 1.2. Key insights and benefits that accrue as a result of probabilistic analysis are identified in Section 1.3. The SHEDS model is briefly described in Section 1.4, including the need for a new software tool to quantify both variability and uncertainty in the inputs to the model. Available software tools are reviewed in Section 1.5, including both commercial and research programs. The need for new software to support SHEDS is established in Section 1.6. The objectives of this project, and an overview of this report, are given in Sections 1.7 and 1.8, respectively.

1.1 Variability and Uncertainty

The distinction between variability and uncertainty in the context of human exposure and risk assessment has been described by Bogen and Spear (1987), Frey (1992), Hoffman and

Hammonds (1994), NAS (1994), Frey and Rhodes (1996), NCRP (1996), EPA (1996, 1997), Cullen and Frey (1999), and others.

Variability refers to real differences in the values of a quantity from one individual to another or over some other population. For example, a given human individual has a body weight, intake rate, lifetime, exposure duration, and activity patterns that are different from that of other individuals. Uncertainty refers to lack of knowledge regarding the true value of a quantity. As noted in NAS (1994), the implication of variability in risk assessment is captured by the notion that there is a certainty that different people have different exposures and different risks. Therefore, in developing risk management strategies that are intended to be protective of a substantial portion of the population, it is important to understand the variability of risk within the population. For example, the notion of a high end exposure as embodied in the 1992 Exposure Assessment Guidelines (EPA, 1992), implies the need to quantify exposures to the 90th percentile or higher of the population. In order to quantify exposures for a given percentile of the population, one must account for inter-individual variability.

In contrast to the notion of variability as defined in the exposure and risk assessment community, uncertainty refers to lack of knowledge regarding the true but unknown value of a quantity. The simplest example of uncertainty arises when an attempt is made to measure a quantity whose true value is a single unchanging value. The measurement instrument is typically imperfect, perhaps because of errors in calibration, spectral interference, inability to fully control all other factors aside from the one being measured, difficulties in obtaining a representative sample, or for a variety of other reasons (e.g., Mandel, 1969; Morgan and Henrion, 1990; Cullen and Frey, 1999).

If measurements are made repeatedly with the same instrument, it is typically the case that there will be some random variation in the observed values even if the true but unknown value is a constant (e.g., Henrion and Fischoff, 1986). The random variation in the measured values can be used to infer a probability distribution that represents the range and relative likelihood of the measurements. If the measurement method is known to be unbiased, then the average of the observations will converge to the true but unknown point value as the number of measurements becomes large. Therefore, an analyst may infer that the true but unknown value of the quantity being measured is enclosed by the range of values observed in repeated measurements. If the measurement method is known to be biased, then the method is said to be inaccurate. In this case, the average of repeated measurements does not converge to the true but unknown value. The difference between the mean of many measurements and of the true, but unknown value, is the bias. If the bias is known, such as based upon repeated measurements of a known sample, then a bias correction can be applied to the observations to produce a more accurate estimate of the true, but unknown, value.

The dispersion of repeated observations above or below the mean observation is related to the precision of the measurements. A highly precise measurement method will produce a tight cluster of observations close to the mean observation. An imprecise method will have a large deviation of individual observations with respect to the average observation. A measurement method may be very precise but inaccurate, in that there may be a tight clustering of observations about the mean of the observations, but the mean of the observations may differ substantially from the true but known value. Alternatively, a measurement method may be very accurate but highly imprecise, in which the mean of the observations is equal to the true but unknown value but where there is a large deviance of values with respect to the mean. Other combinations, such

as accurate and precise, and inaccurate and imprecise, are also possible. Mandel (1969), Cullen and Frey (1999), and others discuss in more detail the notions of precision and accuracy.

Uncertainty arises for reasons other than imperfections in making measurements. Another key source of uncertainty is random sampling error, which refers to the random variation in estimates of a statistic calculated from a finite random sample of data. For example, the sample mean is itself a random variable because it is estimated from a random sample of data. A key assumption is that the data are a sample from the true but unknown population distribution of interest. One random sample of a given sample size (e.g., n=10) will result in one estimate of the sample mean. Another random sample of the same sample size will result in a different estimate of the sample mean because of random differences between the two samples. A confidence interval for the mean calculated based upon a standard error is an indication of lack of knowledge regarding the sample mean based upon the variation in estimates of the mean that would be obtained from repeated random sampling with a given sample size for each random sample.

There are differing interpretations of confidence intervals depending on whether one takes a strictly frequentist view of probability or a Bayesian view. The distinction between these views is described in Morgan and Henrion (1990) and Cullen and Frey (1999). A confidence interval is specified as a range such that it encloses the true but unknown population point value of the statistic with a confidence (or probability) of á. As the sample size of the random sample of the data increase, and if all other factors are the same (e.g., the variance of the random sample of the data) the width of the confidence interval will decrease. Therefore, more data typically translate into more knowledge relevant to inferring the true but unknown quantity. In this work,

confidence intervals are interpreted as in indication of the precision with which a statistic is known.

Another key source of uncertainty is the potential lack of representativeness of data. Very often in human exposure and risk assessment, data for the quantities of direct interest are not available, and surrogate data are used instead. For example, data regarding dietary consumption patterns of specific foods may not be available for specific subpopulations, such as fish consumption for subsistence fisherman. In such cases, an analyst might use fish consumption data for the general population as a basis upon which to make a judgment regarding the presumably higher fish consumption rate of the specific subpopulation. The most common concern with the use of surrogate data is that it may be biased. EPA (1999a) addresses issues and methods associated with the use of surrogate data as a basis for developing input distributions for probabilistic assessment.

Lack of data is an obvious source of uncertainty. In some cases, there may be little or no data available for the quantity of interest or for sufficiently relevant surrogates. Data analysis is an insufficient approach for quantifying either variability and/or uncertainty in such a quantity. However, analogies with other data sets or judgments based upon established theory may serve as a basis for making bounding assumptions or even probabilistic assumptions regarding such quantities. In this situation, methods for eliciting expert judgment and for encoding the judgment in the form of subjective probability distributions can be used (e.g., Cooke, 1991; Kaheman and Tverski, 1982; Morgan and Henrion, 1990).

Finally, other sources of uncertainty include model formulation and specification of the scenarios to be analyzed. A model is a simplified representation of a real system. Simplifications used in model development include aggregation and exclusion. Aggregation

refers to lumping details of the real world system into a single quantity. For example, in an air quality model several chemical species with common characteristics, such as aldehydes, may be simulated as if they were a single "lumped" chemical that has important characteristics considered common to all aldehydes, but that may not capture differences considered less important among the aldehydes. Exclusion refers to ignoring some aspects of the real world system in developing the model. For example, in an air quality model, some chemical compounds that may be potentially important to the formation of secondary pollutants may be excluded, either intentionally or unintentionally, from the model. Cullen and Frey (1999) introduce many concepts relevant to modeling and model uncertainty. One approach for addressing model uncertainty is to compare predictions made with alternative models. For example, Evans et al. (1994) present a probability tree in which alternative conceptual models are included.

Scenario uncertainty refers to the possible failure to specify a scenario or a set of scenario that actually captures the real world problem that an analyst wishes to address. For example, if the intention is to estimate risk to humans because of human exposure to hazardous air pollutants, a scenario based only upon direct inhalation exposure to pollutants emitted from nearby sources may fail to capture the most important situations. Perhaps long-range transport is more important than short-range transport, implying that the analysis should have included a larger geographic area of emission sources. Perhaps indirect exposure pathways, such as food ingestion, are really the means by which humans are exposed to some of the HAPs, which may have undergone deposition and uptake by biological receptors (e.g., fish, plants). A key consideration in reducing scenario uncertainty is to perform a screening analysis to identify which exposure pathways are most important.

This report focuses on methods for characterizing uncertainty based upon random sampling error, which yields insight into the precision of the estimate for a statistic such as the mean, standard deviation, or parameters of a distribution fit to a data set describing inter-individual variability. In the analysis of such data, an implicit assumption is that the data are an unbiased, random, representative sample of the quantity of interest. EPA (1999a) suggests methods for making adjustments to distributions for inter-individual variability if non-representativeness or other sources of bias are believed to be present. While uncertainty associated with measurement error can be significant in some cases, this source of uncertainty is not addressed at this time in the development of AuvTool. While expert judgment is not explicitly addressed by AuvTool as a basis for specifying subjective probability distributions, many of the parametric distributions that are commonly used to represent expert judgments, such as the uniform, symmetric triangle, beta, normal, and lognormal, are included in the framework. It is assumed that the analyst has a relevant model and has developed appropriate scenarios. AuvTool can be used to support development of probabilistic input assumptions for multiple models and scenarios.

1.2 Examples of Probabilistic Analysis

The use of probabilistic analysis methods for dealing with variability and uncertainty is becoming more widely recognized and recommended for environmental modeling and assessment applications. The National Research Council and others have recommend that EPA use quantitative probabilistic analysis methods that distinguish between variability and uncertainty (NAS, 1994). One of the recommendations of the Emission Inventory Improvement Program (EIIP), which is jointly sponsored by EPA and other organizations, is to encourage the use of quantitative methods to characterize variability and uncertainties in emission inventories (Radian, 1996).

EPA has been responsive to these recommendations. For example, EPA has sponsored workshops regarding Monte Carlo simulation methods, has developed a guidance document on Monte Carlo methods, and has included guidance regarding probabilistic analysis in its most recent draft of Risk Assessment Guidance for Superfund (EPA, 1996; EPA, 1997; EPA, 1999a; EPA, 1999b). Uncertainty analysis is now part of the planning process for major assessments performed by EPA, such as the National Air Toxics Assessment.

Recently, the National Research Council released a report on mobile source emissions estimation that calls for new efforts to quantify uncertainty in such emissions (NRC, 2000). The Intergovernmental Panel on Climate Change (IPCC) recently issues a good practice document regarding uncertainty analysis for greenhouse gas emission inventories (IPCC, 2000). Thus, the quantification of variability and uncertainty has become widely accepted not only in human health risk assessment but also in supporting or related areas, such as emissions estimation. In addition, there is a growing track record of the demonstrated use of quantitative methods for characterizing variability and uncertainty applied to emission factors, emission inventories, air quality modeling, exposure assessment, and risk assessment. Some examples of these are briefly mentioned here.

There have been a number of projects aimed at quantifying variability and uncertainty in highway vehicle emissions, including uncertainty estimates associated with the Mobile5a emission factor model and with the EMFAC emission factor model used in California (Kini and Frey, 1997; Frey, 1997; Frey, Bharvirkar and Zheng, 1999; Pollack *et al.*,.1999). Frey and Eichenberger (1997) and Frey *et al*. (2001) have quantified uncertainty in highway vehicle emission factors estimated based upon measured data collected using remote sensing and on-board instrumentation, respectively. Frey *et al*. (2002) have recommended modeling methods

for the New Generation Model (NGM) that will succeed the Mobile6 emission factor model. These methods include quantification of unexplained inter-vehicle variability and fleet average uncertainty.

There have been a number of efforts aimed at probabilistic analysis of various other emission sources, including power plants, non-road mobile sources, natural gas-fired engines, and specific area sources (Frey, Rhodes, 1996; Frey, Bharvirkar and Zheng, 1999; Frey and Zheng, 2000; Frey and Bammi, 2002a&b; Frey, and Zheng, 2002; Frey and Bharvirkar, 2002; Li and Frey, 2002, Abdel-Aziz and Frey, 2002). Probabilistic anlayses have also been applied to air quality models, such as the Urban Airshed Model (e.g., Hanna *et al*., 2001).

In the area of exposure and risk assessment, there have been a number of analyses in which variability and uncertainty were distinguished. These include, for example, Bogen and Spear (1987), Frey (1992), Hoffman and Hammonds (1996), Cohen *et al*. (1996), and others.

As an example of a probabilistic analysis in which variability and uncertainty were distinguished, Frey and Rhodes (1996) quantified variability and uncertainty in emissions of selected hazardous air pollutants from coal-fired power plants. Limited data were available regarding the concentration of trace species, such as arsenic, in coal, and regarding the partitioning of the trace species in the major process areas of the plant, including the boiler, particulate matter control device, and flue gas desulfurization system. Parametric distributions were fit to the available data that represented the inter-unit variability in plant performance. Bootstrap simulation was used to estimate confidence intervals for the fitted cumulative distribution function (CDF) for each input data set. Both variability and uncertainty were propagated through an emissions model to yield estimates of variability in emissions from one averaging time to another and uncertainty in emissions for any given simulated averaging period.

Two averaging times were considered: three-day averages and annual averages. Three-day averages were included because the reported measurements represented plant operation over approximately a three-day period. Annual averages were also simulated because they were of more direct policy interest, such as in comparing estimated emissions with possible emission standards. Methods for considering measurement error, and for using mixture distributions to represent variability, were addressed. In addition, methods for identify the key sources of variability and uncertainty with respect to estimated emissions were illustrated.

1.3 Key Insights from Probabilistic Analysis

As noted in EPA (1997) and Cullen and Frey (1999), there is a tiered set of quantitative analysis methods that can be used in exposure and risk assessment. A starting point for an assessment typically includes worst-case, bounding, or screening analyses based upon point estimates and perhaps simplified models. Such analyses are intentionally biased and are intended to determine whether, under worst case conditions, an exposure or risk estimate may be sufficiently small that no further action is needed. If an analysis based upon such methods implies that the exposure and risk may be high enough to warrant further attention, then a second tier of analysis will typically include the use of more realistic input assumptions and the selection of more refined models. Probabilistic methods are typically used to replace point estimates for input assumptions as the assessment becomes more refined. With probabilistic input assumptions in the form of probability distributions, an analyst is able to characterize the range of possible values and the relative likelihood of values within the range, instead of being forced to choose a single point value. In contrast, if all inputs to a model are assigned point values that represent worst-case or high-end assumptions, the resulting exposure or risk estimate may be very high compared to the actual exposures or risks faced by the high-end exposed portions of the population (e.g., Burmaster and Hattis, 1994; Cullen, 1994; Finkel, 1990).

Key limitations of point estimates are that they do not provide an indication as to what fraction of the population have exposure or risk less than or equal to that of the point estimate, or greater than or equal to the point estimate. Therefore, information regarding inter-individual variability is not adequately characterized. Furthermore, in the point estimate approach, uncertainty is not quantified. Therefore, no insight is provided regarding the magnitude of uncertainty in the estimate or regarding key sources of uncertainty. If risks are over-estimated, then resources may end up being devoted to risk management strategies that yield less significant benefits than if they had been devoted to risk management in other areas. The use of biased point estimates can lead to inefficient allocation of resources.

A probabilistic analysis approach incorporates more information into the assessment than does a point estimate approach. The relative range and likelihood of values for model inputs are characterized using probability distributions. The distributions are propagated through the model using a technique such as Monte Carlo simulation or related variations thereof. For each model output, both the range and likelihood of possible values is estimated. Therefore, decision-makers gain insight into the magnitude of inter-individual variability, including whether there are some individuals that may be subject to high risks even though many members of the population have low risk. Insight regarding the distribution of risks among members of the population is important in developing effective risk management strategies. Information regarding uncertainty in the risk estimate is useful in determining how likely it is that a given individual or portion of the population may actually face a high risk. If the range of uncertainty is large, it may be useful to identify key sources of uncertainty. In turn, additional data collection or research can be targeted to reduce uncertainty in the model inputs that most contribute to overall uncertainty in the risk estimate (Thompson and Graham, 1996).

Variability and uncertainty should be treated separately because they each have different decision-making and policy implications. Knowledge regarding variability can be used to identify subpopulations that face the greatest risks, such as children, asthmatics or individuals with activity patterns that bring them into greater contact with specific chemicals. Information regarding key sources of uncertainty can be used to prioritize additional data collection or research to improve estimates of exposure and risk. The National Research Council has recommended that the distinction between variability and uncertainty should be maintained rigorously at the level of individual components of a risk assessment as well as at the level of an integrated risk assessment (NRC, 1994).

As summarized in Cullen and Frey (1999), probabilistic analysis is useful when: (a) a screening level analysis indicates that exposure and risk may be unacceptably high; (b) there is a need to identify priorities for collecting additional information in an effort to reduce uncertainty; (c) significant equity issues are raised regarding the inter-individual distribution of exposure and risk; (d) there is a need to identify, and determine how to target resources to reduce risk to particular subpopulations of highly exposed individuals; (e) there is a need to rank exposures, pathways, sites, or contaminants taking into account both variability and uncertainty; and/or (f) when the cost of remediation or intervention is high. Conversely, probabilistic analysis may not be needed in situations where a conservative screening analysis indicates no significant problem or when the costs of intervention or remediation are sufficiently small that they outweigh the costs of analysis. Another possible but unlikely reason that a probabilistic analysis might not be needed is if the variability and uncertainty are sufficiently narrow that a single point estimate is considered to be reliable.

1.4 SHEDS Model, Project Objectives, and Software Needs,

The SHEDS model, which is being developed by EPA's National Exposure Research Laboratory (NERL), utilizes a probabilistic approach to predicting population exposures to pollutants. Concentrations of pollutants in various exposure media, as well as the physical factors that influence exposure, are input as distributions in the models. The model uses a two-stage Monte Carlo simulation technique to produce distributions of exposure for various cohorts (e.g., age groups). Therefore, the SHEDS model requires that both variability and uncertainty in the model inputs be characterized. The SHEDS model involves in a large number of model inputs. For most of these inputs, it is necessary to quantify both variability and uncertainty. Currently, the characterization of variability and uncertainty for SHEDS model inputs must be done off-line. Therefore, there is a need for a software tool to support the development of probabilistic input assumptions for the SHEDS model, including characterization of both variability and uncertainty.

The objective of this project was to develop a stand-alone software tool that can conduct statistical analysis of variability and uncertainty associated with fitting probability distributions to data sets for use with the SHEDS modeling framework. Secondary objectives were to develop a tool that would be user-friendly, to develop a tool so that it could be used for general purpose applications, and to verify the new software through extensive testing of its algorithms.

In identifying the specific needs for a software tool, it is critically important to clearly determine the specifications for the software, including the input information that will be provided to the software and the output information that is needed from the software. The key specifications for the development of a software tool that supports SHEDS include the following:

1. For each input to SHEDS, either a random representative data set representing variability will be available, or information will be available regarding a parametric probability distribution that represents variability.
2. The software must be capable of fitting a parametric probability distribution to input variables for which random representative data are provided. These distributions will represent variability.
3. Graphical techniques based upon confidence intervals for the fitted cumulative distribution function (CDF) and quantitative statistical goodness-of-fit methods must be available for assessing the adequacy of a candidate parametric probability distribution in representing variability for a data set.
4. Uncertainty in the parameters of the distribution, and regarding the mean and standard deviation, will be estimated based upon random sampling error.
5. The software must be capable of performing batch operations in order to process information for a potentially large number of SHEDS model inputs.
6. Information regarding both variability and uncertainty must be reported in a format consistent with SHEDS model input requirements.

Prior to developing a new software tool, it is important to determine whether existing tools are capable of meeting the specific requirements of the SHEDS model as set forth above.

1.5 Available Software

A variety of programs have been developed that are capable of various types of probabilistic analysis. There are several commercially available software packages, such as Crystal Ball, @Risk, Analytica and RiskQ. Crystal Ball and @Risk both are Microsoft Excel-based add-in aprograms (Palisades,1997; Decisioneering ,2001). Analytica is a stand-alone program for creating, analyzing, and communicating probabilistic models for risk and policy

analysis (Lumina, 1996). RiskQ is implemented in Mathematica (Bogen,1992). Capabilities to address both variability and uncertainty are available in Crystal Ball and RiskQ. While RiskQ has many powerful capabilities, it requires knowledge of programming in Mathematica (Murray and Burmaster,1993). @Risk and Analytica do not provide convenient capabilities for simultaneous analysis of both variability and uncertainty. Therefore RiskQ, @Risk, and Analytica were not applicable to the identified needs for supporting the SHEDS model.

Crystal Ball uses a two-stage Monte Carlo simulation method as presented by Cohen *et al*., 1996. The method of Cohen *et al*. (1996) is very similar to that of Frey (1992) and Frey and Rhodes (1996). The primary difference is that the approach of Cohen *et al*. (1996) discards many intermediate values during the simulation. While this can reduce memory or storage requirements, it also results in the loss of useful information. Therefore, this approach was not selected.

Frey (1992) developed case studies illustrating the distinction between variability and uncertainty using an earlier version of Analytica, which was known at that time as Demos. However, Demos and Analytica are structured to work with one dimension of probabilistic information. The limitations of Demos and Analytica at that time motivated the development of a specialized software tool. Therefore, Frey and Rhodes (1996, 1998, 1999) developed a FORTRAN-based program at North Carolina State University referred to as "BOOTSIM." BOOTSIM featured two-dimensional probabilistic representations of variability and/or uncertainty for model inputs, propagation of the two-dimensional probabilistic information through a model, characterization of both variability and uncertainty in model results, and analysis of model results to identify key sources of variability and uncertainty. BOOTSIM included a technique for quantifying uncertainty in selected statistics using bootstrap simulation.

The EPA Office of Air Quality Planning and Standards (OAQPS) supported the development of a prototype software tool for Analysis of Uncertainty and Variability in Emissions Estimation (AUVEE) (Frey and Zheng, 2000). AUVEE was developed based on BOOTSIM. However, BOOTSIM did not contain a capability to fit a parametric probability distribution to a data set or to compare alternative fitted distributions to data. For AUVEE, a capability was included to fit parametric distributions to data and to compare the fitted distribution and the data using a graphical display. Thus, unlike BOOTSIM, AUVEE included a Graphical User Interface (GUI) and capabilities to fit distributions to data. However, because AUVEE was restricted to an example case study, it did not have a capable to allow users to enter their own data. Furthermore, AUVEE did not include the use of statistical goodness-of-fit tests nor a batch analysis capability for working with many variables automatically.

1.6 Need for New Software

Based upon a review of available software tools, it was established that there was not existing software available that could meet all of the needs for the SHEDS model. However, AUVEE contained many of the features that were needed. Therefore, AUVEE was selected as the basis for developing a new software tool, referred to here as "AuvTool."

In addition to the capabilities of AUVEE, AuvTool introduces new capabilities for the following: (a) accepts user input in a convenient spreadsheet format; (b) can accommodate input data for variability in the form of a data set or a specified parametric probability distribution; (c) additional options for parametric probability distribution models and the option of using an empirical distribution based upon resampling of the user-provide data set; (d) calculation of goodness-of-fit statistics; (e) a batch analysis capability for handling many input data sets or input distributions automatically; and (f) output information that meets the requirements of the SHEDS model. To facilitate a modular approach to software development, AuvTool was

developed using C++. Many of the algorithms of AUVEE were incorporated and were translated from FORTRAN into the new software tool.

1.7 Overview of this Report

The algorithms used in AuvTool are documented in Chapter 2. In Chapter 3, the system development and implementation of AuvTool are described in detail, including the design considerations, development environment, structure design and the main function modules. Chapter 4 summarizes the verifications of results with the use of AuvTool. An illustrative case study is given in Chapter 5. The case study demonstrates the use of the batch analysis feature of AuvTool and presents examples of variability and uncertainty analysis results that the new software can provide. Readers interested in more detail regarding how to use the AuvTool software are referred to the accompanying User's Guide (Zheng and Frey, 2002). Chapter 6 provides a summary, conclusions, and recommendations.

2.0 ALGORITHMS USED IN AUVTOOL

The algorithms used in the AuvTool software are presented in this section. Seven key areas for which algorithms are documented include: (1) visualization of data sets using empirical distributions; (2) definition of the parametric probabilistic distributions used in AuvTool; (3) parameter estimation for the parametric probability distributions; (4) generation of random samples from probability distribution models; (5) evaluation and selection of a fitted probability distribution model based upon statistical goodness-of-fit tests; (6) characterization of uncertainty based upon random sampling error using bootstrap simulation; and (7) evaluation of dependence or correlation between statistics of interest, including the mean, standard deviation, and distribution parameters. Each of these areas are addressed in the following subsections.

2.1 Visualizing Data Using Empirical Distribution

Some of the key purposes of visualizing data sets include: (1) evaluation of the central tendency and dispersion of the data; (2) visual inspection of the shape of the empirical distribution of the data as a potential aid in selecting parametric probability distribution models to fit to the data; and (3) identification of possible anomalies in the data set (e.g., outliers). Specific techniques for evaluating and visualizing data include calculation of summary statistics, and plotting a data set as an empirical Cumulative Distribution Function (CDF).

Three key characteristics of a CDF are its central tendency, dispersion, and shape. There are several measures of central tendency, which include the mean, median, and mode. The dispersion, or the spread, of a distribution is measured by the standard deviation or the variance of the distribution. The relative standard deviation (RSD), also known as the coefficient of variation (CV), is the standard deviation divided by the mean. For a non-zero mean, the CV provides a normalized indication of the dispersion of data values, with a large CV indicating relatively large variability in the data set. The shape of the distribution is reflected by quantities

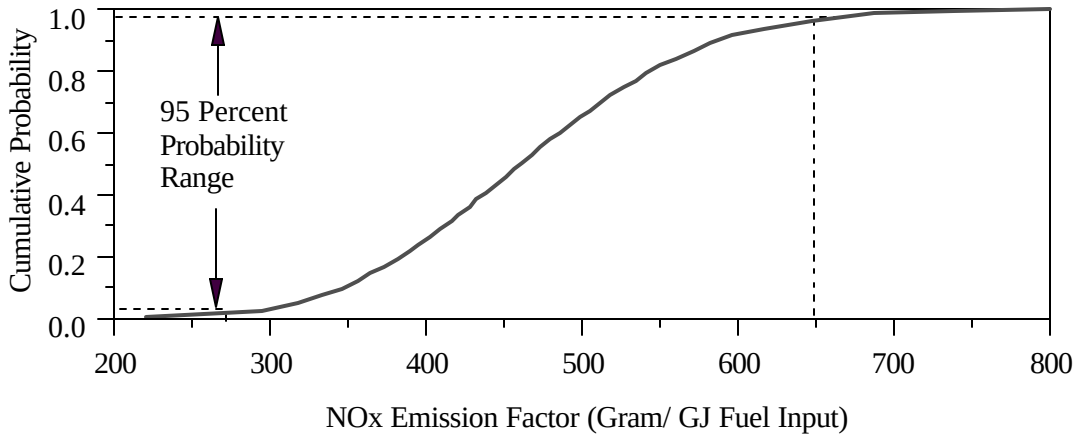


Figure 2-1. Plot Illustrating the 95 Percent Probability Range on a Cumulative Distribution Function.

such as skewness and kurtosis. The skewness is the asymmetric of a distribution, and the kurtosis refers to the peakedness of a distribution. These statistics can be used to aid in the selection of a parametric probability distribution model to fit to the data (Cullen and Frey, 1999).

A CDF is a relationship between "cumulative probability" and values of the random variable. Cumulative probability is the probability that the random variable has values less than or equal to a specific numerical value of the random variable. CDFs provide a relationship between fractiles and quantiles. A fractile is the fraction of values that are less than or equal to a specific value of a random variable. Fractiles expressed on a percentage basis are referred to as percentiles. A quantile is the value of a random variable associated with a given fractile (Frey, Bharvirkar and Zheng, 1999). For example, the range of data values enclosed by the 0.025 and 0.975 fractiles (2.5 and 97.5 percentiles) is often of particular interest, since this provides an indication of the dispersion of a distribution as reflected by the 95 percent probability range of values. An example of a CDF is illustrated in Figure 2-1

Empirical estimation of a fractile from data requires rank ordering of the data. There are several possible methods for estimating the percentile of an empirically observed data point.

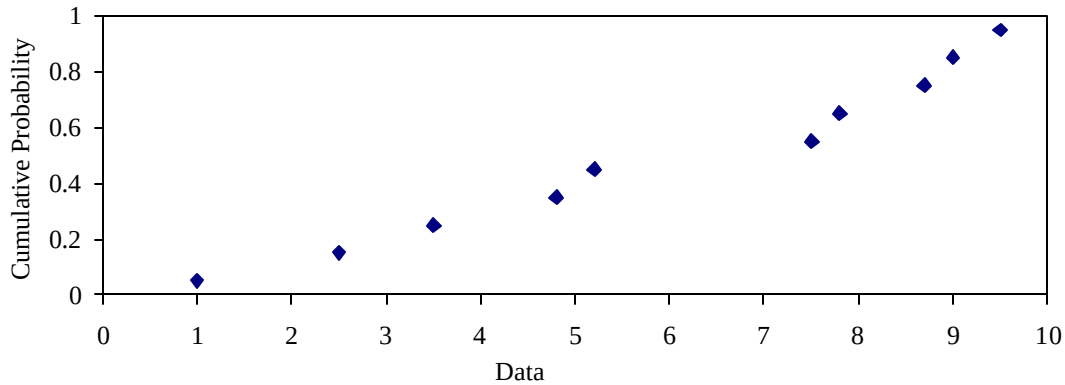


Figure 2-2. Example Graph of Visualizing Data Using the Hazen's Plotting Position Method (n=10)

These methods are referred to as "plotting positions." The plotting position is an estimate of the cumulative probability of a data point. As described by Cullen and Frey (1999), Harter (1984) provides an overview of the various types of plotting positions.

A commonly used plotting position, proposed by Hazen (1914), is used in AuvTool for displaying data points in comparison to fitted parametric distributions:

$$F_x(x_i) = \Pr(X < x_i) = \frac{i - 0.5}{n}, \text{ for } i = 1, 2, \ldots, n \text{ and } x_1 < x_2 < \ldots < x_n \qquad (2\text{-}1)$$

where,

i = Rank of the data point when the data set is arranged in an ascending order

n = number of data points

$x_1 < x_2 < \ldots < x_n$ are data points in the rank-ordered data set

$\Pr(X<x_i)$ = Cumulative probability of obtaining a data point whose value is less than x_i

An example graph of visualizing data using the Hazen's plotting position method is shown in Figure 2-2. The figure depicts the plotting position of each of 10 data points for a small data set.

2.2 Definitions of Probability Distributions

Probability distribution models used in AuvTool include the normal, lognormal, Weibull, gamma, beta, uniform, symmetric triangle parametric distributions and empirical distribution. Ang and Tang (1984), Hahn and Shapiro (1967), Morgan and Henrion (1990), Cullen and Frey (1999) and others review the theoretical basis underlying each of these distributions. The normal and lognormal distributions have an underlying theoretical basis in the central limit theorem (CLT) when applied to additive or multiplicative processes, respectively. For example, a process of pollutant dispersion generated by the sum of many random variations can be described by the Gaussian plume model (Seinfeld, 1986). Although the normal distribution is not appropriate for representing non-negative quanitities because it has an infinite negative tail, it is often used to represent non-negative quantities, such as weight or length, so long as the coefficient of variation is less than about 0.2 (Morgan and Henrion, 1990).

The lognormal, gamma and Weibull distributions are useful for representing non-negative and positively-skewed data. The two-parameter beta distribution is bounded by zero and one, and has flexibility to represent data with a variety of central tendency and skewness. The uniform and symmetric triangle distributions are most commonly used to represent expert judgments made in the absence of data. Empirical distributions can be used instead of parametric distributions. A comparison of empirical and parametric distribution is described in EPA (1999a) and in Section 2.3.

More discussion of distribution selection criteria can be found in Hahn and Shapiro (1967), Ang and Tang (1984), Morgan and Henrion (1990), Hattis and Burmaster (1994), and Alvarez (1996), and Cullen and Frey (1999), among others.

Table 2-1. The Definitions of Parametric Probability Distributions

Name of Distribution	Probability Density Function (PDF)	
Normal	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	
Lognormal	$f(x) = \frac{1}{x\sqrt{2\pi s_{\ln x}^2}} e^{\frac{-(\ln x - m_{\ln x})^2}{2s^2}}$	$(0 < x < \infty)$
Beta	$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$	$(0 \le x \le 1)$
Gamma	$f(x) = \frac{\beta^{-\alpha} x^{\beta-1} e^{-x/\beta}}{\Gamma(\alpha)}$	$(0 \le x < \infty)$
Weibull	$f(x) = \frac{c}{k}(x/k)^{c-1} \exp(-(x/k)^c)$	$(0 \le x < \infty)$
Uniform	$f(x) = \frac{1}{b-a}$	$(a \le x \le b)$
Symmetric Triangle	$f(x) = \frac{b-\lvert x-a \rvert}{b^2}$	$(a-b \le x \le a+b)$

2.2.1 Definition of Parametric Probability Distributions

The definitions of the seven parametric distributions included in AuvTool are presented in Table 2-1. The definitions are based upon the probability density function (PDF).

In Table 2-1, for the normal distribution, ì is the arithmetic mean, and ó is the arithmetic standard deviation. For the lognormal distribution, $ì_{lnx}$ is the mean of the *lnx*, and $ó_{lnx}$ is the standard deviation of *lnx*. For the beta distribution, á and â are shape parameters, and *B*(á ,â) is the beta function. For the gamma distribution, á is the shape parameter, â is the scale parameter, and Ã(·) is the gamma function. For the Weibull distribution, k is the scale parameter, and *c* is the shape parameter. For the uniform distribution, a and b are the smallest and largest possible values. For the symmetric triangle distributions, a and b determine the range within which the variable can vary.

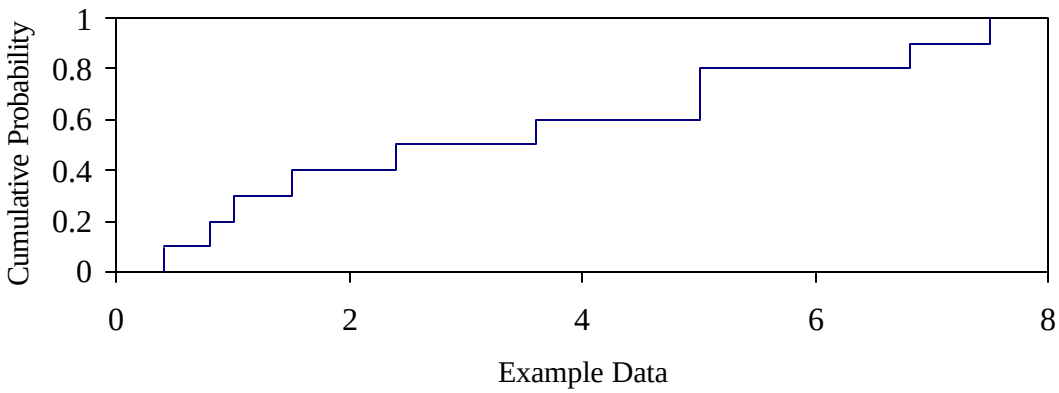


Figure 2-3. An example of an Empirical Distribution Represented a Step Function (n=10)

2.2.2 Empirical Distribution

An empirical distribution is defined as a discrete distribution, *F*, that gives equal probability, 1/n, to each value x_i in the dataset, **x** (Efron, 1979). The CDF for this function is therefore a step function of original data set, **x**, where each value x_i is assigned a cumulative probability of *i*/n for *i*= {1,2,…n}. An example of an empirical distribution representing a step function is provided in Figure 2-3.

2.3 Parameter Estimation of Probability Distributions

A probability distribution model is a description of the probabilities of all possible values in a sample space. A probability distribution model is typically represented as a PDF or a CDF for a continuous random variable. The PDF for a continuous random variable indicates the range and relative likelihood of values. The CDF is obtained by integrating the PDF (Cullen and Frey, 1999).

Probability distribution models may be empirical, parametric, or combinations of both. A parametric probability distribution model is a model described by parameters. The power of using parametric probability distribution models is that data sets, which may contain large numbers of data points, can be described in a compact manner based on a particular type of

parametric distribution function and the values of its parameters. For example, a normal distribution is fully specified if its mean and standard deviation are known. Another potential advantage of parametric probability distributions compared to empirical distributions is that it is possible to make predictions in the tails of the distribution beyond the range of observed data. In contrast, using conventional empirical distributions, the minimum and maximum values of the distribution are limited to their minimum and maximum values, respectively, of the data set. These values typically change as more data are collected. EPA (1999a) presents a discussion of the use of empirical verse parametric distributions.

Based upon visual inspection of an empirical distribution of data as described in Section 2.1, and consideration of processes that generated the data, the analyst can make a judgment regarding selection of one or more candidate parametric distributions to fit to the data set. Once a particular parametric distribution has been selected, a key step is to estimate the parameters of the distribution. The method of Maximum Likelihood Estimation (MLE) and the Method of Matching Moments (MoMM) are among the most typical techniques used for estimating the parameters.

In order to estimate values of the parameters of a parametric probability distribution, statistical estimation methods must be used. Using such estimation methods, inferences are made from an available data set regarding a single best estimate of the parameter values. Usually, there are alternative methods available to estimate parameter values. Thus, it is necessary to choose a parameter estimation method. Small (1990) has discussed the following six desirable characteristics of estimators. These characteristics are useful when comparing and selecting an estimation method:

Consistency: A consistent estimator converges to the "true" value of the parameter as the number of samples increases.

Lack of Bias: On average over many applications to many different data sets, an unbiased estimator yields an average value of the parameter estimate that is equal to that of the population value.

Efficiency: An efficient estimator has minimum variance in the sampling distribution of the estimate. A sampling distribution is a probability distribution for a statistic (e.g., mean, standard deviation, distribution parameters).

Sufficiency: An estimator that makes maximum use of information contained in a data set is said to be sufficient.

Robustness: A robust estimator is one that works well even if there are departures of the data from the underlying distribution. In other words, such as estimator will yield reasonable values of the parameters even if there are some anomalies in the data set.

Practicality: A practical estimator is one that satisfies the needs for the preceding five characteristics while remaining computationally efficient.

For small sample sizes, the MLE method does not always yield minimum variance or unbiased estimates (Holland and Fitz-Simmons, 1982). However, for larger sample sizes, the MLE method tends to better satisfy the first five criteria for statistical estimation than other methods. Compared to MLE, MoMM estimators tend to be more robust but less efficient. MLE can be extended to estimate parameters for distributions fitted to censored data. In the present study, both MLE and MoMM are included as options for estimation of parameters of parametric

probability distributions. The MoMM and MLE methods are described in more detail in the next subsections.

2.3.1 Method of Matching Moments

MoMM is based upon matching the moments or central moments of a parametric distribution (e.g., mean, variance) to the moments or central moments of the data set. MoMM estimators are often but now always easy to calculate. Therefore, this method is often the most straightforward to implement. Thus, it typically satisfies the criterion of practicality. For example, there are convenient solutions for MoMM parameter estimates for the normal, lognormal, gamma, and beta distributions (Hahn and Shapiro, 1967), as well as for the uniform and symmetric triangle distributions. However, MoMM may not fully satisfy the other criteria as previously noted. In the following sections, the MoMM estimators for each of the parametric distributions are presented

2.3.1.1 Normal Distribution

As defined in Table 2-1, the parameters for the normal distribution are the arithmetic mean, *ì*, and the arithmetic variance, *ó*^{2}. The MoMM estimator of the mean is the sample mean, $\overline{x}$. The MoMM estimator of the variance is the unbiased sample variance, s^2 (Morgan and Henrion, 1990; Casella and Berger, 1990).

$$\hat{\mu} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2-2}$$

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \tag{2-3}$$

2.3.1.2 Lognormal Distribution

The parameters of the lognormal distribution can be defined as: (1) the geometric mean, $ì_g$, and geometric standard deviation, $ó_g$, estimated by $\hat{m}_g$ and $\hat{s}_g$, respectively; or (2) the mean

and standard deviation of the logarithm of X, $\mu_{\ln(x)}$, and $\sigma_{\ln(x)}$, estimated by $\hat{\mu}_{\ln(x)}$ and $\hat{s}_{\ln(x)}$, respectively (Morgan and Henrion, 1990; Casella and Berger, 1990).

$$\hat{\mu}_{;nx} = \ln(\overline{X}) - \frac{1}{2}\hat{s}^{2}_{\ln x} \tag{2-4}$$

$$\hat{s}_{\ln x} = \sqrt{\ln(\overline{X}^{2} + s^{2}) - 2\ln(\overline{X})} \tag{2-5}$$

In AuvTool, the mean of lnx, μ_{lnx}, and the standard deviation of lnx, σ_{lnx}, are used as the parameters to define the lognormal distribution.

2.3.1.3 Beta Distribution

The beta distribution has two shape parameters. The parameters can be estimated through relationships with the sample mean and the unbiased sample variance, $\overline{x}$ and s^2 (Hahn and Shapiro, 1967; Morgan and Henrion, 1990):

$$\hat{\alpha} = \overline{X}\left[\overline{X}\frac{(1-\overline{X})}{s^{2}} - 1\right] \tag{2-6}$$

$$\hat{\beta} = (\overline{X} - 1)\left[\overline{X}\frac{(1-\overline{X})}{s^{2}} - 1\right] \tag{2-7}$$

2.3.1.4 Gamma Distribution

The parameters of the gamma distribution are the shape parameter α, and the scale parameter $\hat{a}$, where $\hat{a}$ is an estimate of α, and $\hat{b}$ is an estimate of $\hat{a}$. These parameters are estimated through relationships with the sample mean and unbiased sample variance, $\overline{x}$ and s^2 (Morgan and Henrion, 1990; Casella and Berger, 1990).

$$\hat{a} = \frac{\overline{X}^{2}}{s^{2}} \tag{2-8}$$

$$\hat{\beta} = \frac{s^{2}}{\overline{X}} \tag{2-9}$$

2.3.1.5 Weibull Distribution

For the Weibull distribution, the relationship between the parameters and the central moments of the data are (Morgan and Henrion, 1990):

$$\overline{X} = \hat{\beta}\Gamma\left(1 + \frac{1}{\hat{\alpha}}\right) \tag{2-10}$$

$$s^2 = \hat{\beta}^2\left[\Gamma(1 + \frac{2}{\hat{\alpha}}) - \Gamma^2(1 + \frac{1}{\hat{\alpha}})\right] \tag{2-11}$$

There is no closed form solution for the MoMM estimator of the parameters of the Weibull distribution. Therefore, as an alternative, a parameter estimation method based upon regression analysis of a probability plot is used.

In the probability plot method, if a data set is reasonably described by a Weibull distribution, then the following transformation may be used to plot the data (Cullen and Frey, 1999):

$$\ln\left\{\ln\left[\frac{1}{\overline{F}(x_i)}\right]\right\} = c\ln(x_i) - c\ln(k) \tag{2-12}$$

where,

c = shape parameter

k = scale parameter

$$\overline{F}(x_i) = 1 - F(x_i) \tag{2-13}$$

$\overline{F}(x_i)$ is the complementary CDF of x. An empirical estimate of the CDF can be obtained using Equation (2-1), presented by Hazen (1914). Thus, it is possible to plot the data set and to calculate the scale and shape parameters from the intercept and slope of a best fit

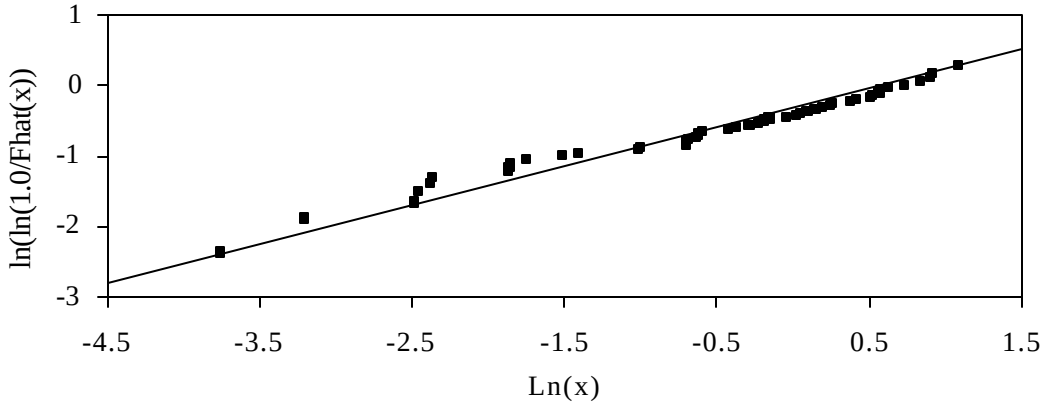


Figure 2- 4. Example of a Probability Plot for a Weibull Distribution (n=50)

regression line obtained using conventional least-squares regression. An example is shown in Figure 2-4 for n=50. In this example, the best fit equation was:

$$\ln\left\{\ln\left[\frac{1}{\overline{F}(x_i)}\right]\right\} = 0.47313\ \ln(x_i) - 0.3644 \tag{2-14}$$

Therefore, the shape parameter is c=0.47313. The scale parameter can be found by solving the expression:

$$k = \exp\left(\frac{0.3644}{c}\right) \tag{2-15}$$

From Equation (2-15), it can be inferred that k is equal to 2.17.

2.3.1.6 Uniform Distribution

The parameters of the uniform distribution are the endpoints, a and b, which are estimated by $\hat{a}$ and $\hat{b}$. The parameter estimation formulae using MoMM are as follows (Morgan and Henrion, 1990):

$$\hat{a} = \overline{X} - \sqrt{3}\,s \tag{2-16}$$

$$\hat{b} = \overline{X} + \sqrt{3}\,s \tag{2-17}$$

2.3.1.7 Symmetric Triangle Distribution

The parameters of symmetric triangle distribution are a and b, which are estimated by $\hat{a}$ and $\hat{b}$. MoMM parameter estimation formulas for these two parameters are (Morgan and Henrion, 1990):

$$\hat{a} = \overline{X} \tag{2-18}$$

$$\hat{b} = \sqrt{6}\, s \tag{2-19}$$

2.3.2 Maximum Likelihood Estimation (MLE)

The MLE methods involves the selection of parameter values that characterize a distribution which was most likely to yield the observed data set (Cohen and Whitten, 1993). A likelihood function for independent samples is defined as the product of the PDF evaluated at each of the sample values. For a continuous random variable, for which independent samples have been obtained, the likelihood function is:

$$L(\theta_1, \theta_2, ..., \theta_k) = \prod_{i=1}^{n} f(x_i \mid \theta_1, \theta_2, ..., \theta_k) \tag{2-20}$$

where,

$q_1, q_2, \ldots, q_k$ = Parameters of the parametric probability distribution model.

k = Number of parameters for the parametric probability distribution model.

x_i = Values of the random variable, for, $i = 1, 2, \ldots, n$

n = Number of data points in the data set.

f = Probability density function.

The general idea behind MLE is to choose values of the parameters of the fitted distribution so that the likelihood that the observed data is a sample from the fitted distribution is maximized. The likelihood is calculated by evaluating the probability density function for each observed data point, conditioned upon assumed values for the parameters, and multiplying the

results. The parameter values may be changed, such as by using an optimization method, to change the value of the likelihood function until a maximum is reached. More commonly, the log-transformed version of the likelihood function is used, which is based upon the sum of the natural log of the probability density function evaluated for each data point, conditioned upon assumed values or the parameters. The MLE parameter estimators can be obtained by varying the parameter values so as to find the maximum of the log-likelihood function.

The log-likelihood function of a univariate (describing one data set) two-parameter distribution is given by:

$$L = \sum_{i=1}^{n} \ln\left[f(x_i | q_1, q_2) \right] \tag{2-21}$$

where,

n = number of data points.

L = Log-likelihood function

f = Probability density function

$è_1$, $è_2$ = parameters of a two-parameter distribution

For definitions of the probability density function $f(x|\ è_1,\ è_2)$ for different parametric distributions, see Table 2-1 in Section 2.2. For some parametric probability distributions, such as the normal and lognormal distributions, analytical solutions for the maximum likelihood estimators of the parameters are available by setting the first partial derivatives of the likelihood function equal to zero. However, in many cases, an analytical solution is not readily available. In these cases, the maximum likelihood parameter estimates can be found using numerical optimization techniques. For the uniform distribution, since the density function is a constant, no MLE solution is available. Except for the uniform distribution, the estimation of the maximum

likelihood parameter values for the distributions in Table 2-1 can be formulated as the following optimization problem:

$$\text{Maximize} \qquad L = \sum_{i=1}^{n} \ln\left[f(x_i \mid \theta_1, \theta_2)\right] \tag{2-22}$$

Subject to

$\theta_1 > 0$ for beta ($\theta_1=\alpha$), gamma ($\theta_1=\alpha$)), Weibull ($\theta_1=k$)

$\theta_2 > 0$ for beta ($\theta_2=\beta$), gamma ($\theta_2=\beta$), Weibull ($\theta_2=c$)

where,

n= number of samples

The optimization problem here is a multidimensional constrained one. A variety of methods are available to solve such problems. These methods include the downhill simplex method; the direction-set method, of which Powell's method is the prototype; the penalty function method; and others (Press, *et al*., 1992). In this study, Powell's method is employed. This method is relatively easy to program, it does not require calculations of derivatives, and it typically provides good results.

Optimization solutions for the MLE parameter estimates are used in AuvTool for the gamma, Weibull, beta, and symmetric triangle distributions. In the implementation of AuvTool, for normal and lognormal distributions, analytic solutions are used. The MLE estimators for the normal and lognormal distributions are as follows (Morgan and Henrion, 1990):

MLE Parameter Estimators for the Normal Distribution

$$\mu = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{2-23}$$

$$\sigma^2 = s^2 = \frac{1}{n}\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 \tag{2-24}$$

MLE Parameter Estimators for the Lognormal Distribution

$$\mu_{\ln x} = \frac{1}{n}\sum_{i=1}^{n}(\ln x_i) \tag{2-25}$$

$$\hat{\sigma}_{\ln x} = \left(\frac{1}{n}\sum_{i=1}^{n}(\ln x_i - \hat{\mu}_{\ln x})^2\right)^{1/2} \tag{2-26}$$

2.3.3 Parameter Estimation Method Availability for Probability Distributions

Table 2-2 summarizes the availability of the MoMM and MLE parameter estimation methods for the probability distributions used in AuvTool. MoMM estimators are available for all but the Weibull distribution. For the Weibull distribution, a probability plotting method is provided in lieu of an MoMM method. MLE estimators are available for all but the uniform distribution.

2.4 Algorithms for Generating Random Samples from Probability Distributions

Computing efficiency and programming simplicity were used as the criteria for selecting methods for generating random samples from various distributions using Monte Carlo sampling. Monte Carlo simulation methods are based upon the use of a pseudo random number generator (PRNG) that produces a stream of random, independent uniformly distribution numbers. Uniformly distributed random numbers are used as the input to algorithms that generate random numbers from other types of distributions.

The most efficient and simple method for generating random variables from a particular type of probability distribution is the method of inversion (Frey and Rhodes, 1999). This method is always used when the CDF can be inverted. In many cases, however, the inverse CDF cannot be written in a closed form, and an alternative method is used. Some alternative methods are the method of composition, the method of convolution, and the acceptance-rejection method (Law and Kelton, 1991).

Table 2-2. Parameter Estimation Method Availability for Parametric Probability Distributions

Distribution Types	MoMM	MLE	Comments
Normal			Analytic solution for MLE
Lognormal			Analytic solution for MLE
Beta			Optimal Solution for MLE
Gamma			Optimal Solution for MLE
Weibull			Optimal Solution for MLE
Uniform		N/A	
Symmetric Triangle			Optimal Solution for MLE

Note: **:** The method is available for the given distribution.
: The plotting method is used instead of MOMM for Weibull distribution
N/A: The method is not available in this case

In the following subsections, the PRNG and the methods used in the AuvTool to generate random variables for the normal, lognormal, Weibull, gamma, beta, uniform, symmetric triangle and step-wise empirical distributions are described.

2.4.1 Pseudo Random Number Generator

The term pseudo-random refers to numbers which appear as if they are uniformly distributed random numbers that actually are generated in a completely deterministic manner (Barry, 1996). Pseudo random numbers are thought to be "good" when they have the following features (Rubinstein, 1981): (1) statistical uniformity, (2) statistical independence, (3) reproducibility, and (4) they can be generated quickly and economically. Another key consideration is the period length, which is the number of random values that are generated before the same sequence begins to be repeated.

There are a variety of methods for generating pseudo-random numbers (Bratley, *et al*., 1987). The most widely used method is the Linear Congruential Generator (LGC). The advantage of LGC is its speed, simplicity and portability (Barry, 1996). However, a potential problem with a LGC approach is that its period length is easily exhausted (L'Ecuyer, 1996). It is

well recognized that, for statistical reasons, the period length of a linear-type generator should be several orders of magnitude of larger than what is actually needed (L'Ecuyer, 1994; 1996).

An approach for increasing the period and improving the structure of the generator is to use combined Multiple Recursive Generators (MRGs) presented by L'Ecuyer (1996). In this method, two or more MRGs are combined. In AuvTool, a combined generator with two MRGs is used and is described as:

$$Z_n = (X_n - Y_n) \bmod m_1 \tag{2-27}$$

where the two underlying generators X_n and Y_n are:

$$X_n = (a_1 X_{n-1} + a_2 X_{n-2} + a_3 X_{n-3}) \bmod m_1 \tag{2-28}$$

and

$$Y_n = (b_1 Y_{n-1} + b_2 Y_{n-2} + b_3 Y_{n-3}) \bmod m_2 \tag{2-29}$$

with coefficients

$a_1 = 0$, $a_2 = 63308$, $a_3 = -183326$,

$b_1 = 86098$, $b_2 = 0$, $b_3 = -539608$,

and $m_1 = 2^{31} - 1 = 2147483647$ and $m_2 = 2145483479$.

The operator "mod" in the Equations (2-27~29) divides two integers and returns the remainder of the division. The period of this PRNG is 2^{205}; the six initial values for x_0, x_1, x_2 and y_0, y_1, y_2 can be any integers from 1 to $2^{31} - 1 = 2147483647$ (L'Ecuyer,1996). In AuvTool, the initial values for the x_0, x_1, x_2 and y_0, y_1, y_2 are 1973272912, 281629770, 20006270, 1280689831, 2096730329, and 1933576050, respectively. Only the value for the seed x_0 is available for users to modify.

2.4.2 Normal Distribution

Generation of random variables from a normal distribution is simplified by the fact that any normal distribution can be written in terms of the standard normal distribution, with a mean

of zero and standard deviation of one. The symbol "~" denotes "is distributed as." If $X \sim N(\mu, \sigma^2)$, and if $X' \sim N(0,1)$, which is the standard normal distribution, then

$$X = \mu + \sigma X' \tag{2-30}$$

Therefore, it is only necessary to generate random numbers from the standard normal. Standard normal random samples can be generated using an acceptance-rejection method developed by Box and Muller (1958). In this method, two uniformly distributed U(0,1) random variates, U_1 and U_2, are used to generate two N(0,1) random variates, X_1 and X_2. The Box and Muller method is used to calculate X_1 and X_2 as follows:

$$\begin{aligned} X_1 &= \sqrt{-2 \ln U_1} \cos(2\pi U_2) \\ X_2 &= \sqrt{-2 \ln U_1} \sin(2\pi U_2) \end{aligned} \tag{2-31}$$

However, a more efficient version of the Box-Muller method, called the polar method, was developed by Marsaglia and Bray (1964). The polar method is used in this study. The algorithm is presented in Law and Kelton (1991) as follows:

Step1: Generate U_1 and U_2 as independent and identically distributed (IID) uniform random samples on the interval [0,1]. Therefore, $U_1 \sim (0,1)$ and $U_2 \sim (0,1)$

Step2: Let $V_i = 2U_i - 1$ for $i = \{1, 2\}$, and let $W = V_1^2 + V_2^2$. If $W > 1$, go back to Step 1. Otherwise, let $Y = \sqrt{(-2\ln(W)/W}$, $X_1' = V_1 Y$, and $X_2' = V_2 Y$.

Step3: Then X_1' and X_2' are IID N(0,1) random variates. $X_1 = \mu + \sigma X_1'$ and $X_2 = \mu + \sigma X_2'$ so that X_1 and X_2 are IID $N(\mu, \sigma^2)$.

Since two normal random samples are generated with each call of this subroutine, in principle the procedure only needs to be implemented once for every two normal distributions that are to be simulated. If U_1 and U_2 were truly IID random variables from a uniform distribution U(0,1), then using X_1 followed by X_2 on subsequent calls to the subroutine would be

valid. It has been shown, however, that if U_1 and U_2 are sequential pseudo random numbers (as is the case in this implementation) then X_1 and X_2 will fall on a spiral in (X_1, X_2) space, rather than being truly IID. In order to ensure that all normal random variates are truly IID in this implementation, only X_1 is used and X_2 is discarded. Another option would be to generate U_1 and U_2 from separate and independent pseudo-random number streams.

2.4.3 Lognormal Distribution

Lognormal random samples are generated by using a special property of the lognormal distribution. Namely, if $Y \sim N(\mu_{lnx}, \sigma^2_{\ln x})$, then $e^Y \sim LN(\mu_{lnx}, \sigma^2_{\ln x})$. Therefore, lognormal random samples are generated by the following algorithm:

Generate $Y \sim N(\mu_{lnx}, \sigma^2_{\ln x})$,

$X = e^Y$, so that $X \sim LN(\mu_{lnx}, \sigma^2_{\ln x})$,

Note that μ_{lnx} and $\sigma^2_{\ln x}$ are the mean of lnx and standard deviation of lnx.

2.4.4 Beta Distribution

The method used in this study for generating beta random samples relies upon a special property of the beta distribution. The beta distribution can be described as a ratio comprised of gamma distributions. If $Y_1 \sim G(\alpha,1)$ and $Y_2 \sim G(\beta,1)$ and Y_1 and Y_2 are independent, then $X = Y_1/(Y_1+Y_2) \sim B(\alpha, \beta)$ (Law and Kelton, 1991). Thus, the methods described for generating random samples from a gamma distribution are used as a basis for generating random samples for the beta distribution

2.4.5 Gamma Distribution

Like the normal and lognormal distributions, the gamma distribution has no closed form solution for its CDF or inverse CDF. Therefore, the method of inversion is not feasible for generating random variables in this case. An acceptance-rejection method is used here to generate gamma random variables.

In generating G(*á*, â) random variables, it is noted that if X' ~ G(*á*,1), then X = â x' ~ G(*á* ,â). Therefore, only the G(*á*,1) distribution needs to be simulated and the results can be easily transformed to that of any G(*á*, â) distributio n. Furthermore, a gamma distribution with *á*

â. Exponential random variables are easily generated by the method of inversion as shown below (Morgan and Henrion, 1990):

$$X = -\frac{1}{\beta}\ln(U) \tag{2-32}$$

where U is a random sample from the U(0,1) distribution and *â* is the parameter of the exponential distribution.

Gamma distributions for which *á* < 1 are shaped significantly differently than gamma distributions for which *á* > 1. Therefore, two distinct acceptance-rejection algorithms are necessary.

For *á* < 1, an acceptance-rejection algorithm by Ahrens and Deiter is used in this study. A description of this method is provided in Law and Kelton (1991), where the following algorithm is also presented:

Step 1. Let $b = (e + á)/e$ (e is a constant, and e= exp(1.0) =2.718282))

Step 2. Generate U_1 ~ U(0,1), and let $P = bU_1$. If $P > 1$, go to step 4. Otherwise proceed to Step 3

Step 3. Let $Y = P^{1/á}$, and generate U_2 ~ U(0,1). If U_2 e-Y, return $X = Y$. Otherwise, go back to Step 1.

Step 4. Let Y = -ln[(b - P)/ *á*] and generate U_2 ~ U(0,1). If U_2 $Y^{á-1}$, return $X = Y$. Otherwise, go back to Step 1.

For $\acute{a} > 1$, a modified acceptance-rejection algorithm by Cheng (1977) is used to sample random samples from a Gamma distribution. A description of the method is provided in Law and Kelton (1991). Only the algorithm is presented here:

Step1. Let $a = 1/\sqrt{2\alpha - 1}$, $b = \alpha - \ln 4$, $q = \alpha + 1/a$, $\theta = 4.5$, *and* $d = 1 + \ln \theta$.

Step 2. Generate U_1 and U_2 as IID U(0,1).

Step 3. Let $V = a\ln[U_1/(1 - U_1)]$, $Y = \acute{a}\, e^{V}$, $Z = (U_1^2 U_2)$, and $W = b + qV - Y$.

Step 4. If $W + d - \acute{a}\, Z \quad 0$, return $X = Y$. Otherwise, proceed to Step 5.

Step 5. If $W \quad \ln Z$, return $X = Y$. Otherwise, go back to Step 1.

2.4.6 Weibull Distribution

The CDF for the Weibull distribution can be written as (Morgan and Henrion, 1990):

$$F(x) = 1 - \exp^{-(x/k)^c} \tag{2-33}$$

A random sample, X, from a W(k,c) can therefore be generated directly by the method of inversion using the inverse CDF:

$$X = F^{-1}(U) = k[-\ln(1 - U)]^{1/c} \tag{2-34}$$

where U is a random sample from the U(0,1) distribution.

2.4.7 Uniform distribution

The method of inversion is used in this study for generating uniform distributions with any arbitrary endpoints. The method is as follows (Morgan and Henrion, 1990):

$$X = a + (b - a)U \tag{2-35}$$

where U is a random sample from the U(0,1) distribution.

2.4.8 Symmetric Triangle Distribution

The method of inversion is used in this study for generating symmetric triangle distribution, as follows as (Morgan and Henrion, 1990):

$$X = (a - b) + b(2U)^{1/2} \qquad 0 \le U \le 0.5$$
$$X = (a + b) - b(2.0 - 2U)^{1/2} \qquad 0.5 < U \le 1.0 \tag{2-36}$$

where U is a random sample from the U(0,1) distribution.

2.4.9 Empirical Distribution

In an empirical distribution, a data set is described by a step-wise empirical cumulative distribution function, in which the probability of sampling any discrete value within the dataset is 1/n. A random re-sampled version of the original data set, of size n, is denoted by:

$$\text{`}\mathbf{X^*}=(X_1^*, X_2^*,\ldots,X_n^*) \tag{2-37}$$

The asterisks indicate that **X*** is not actual data set x, but rather a randomized or resampled version. Since the sampling is done with replacement, it is possible to have repeated values within any given random samples from an empirical distribution.

The algorithm for generating a random sample from an empirical distribution is as follows:

Step 1: Rank an original data set in an ascending order to have an ordered dataset $\mathbf{X^o}$ in which $X_m^o < X_{m+1}^o$, where, m =1,2…n.

Step 2: Generate a random number U from an U(0,1) distribution.

Step 3: Calculate an index using the following formula:

$$i = n \times U \tag{2-38}$$

where,

i is a returned smallest integer that is larger than or equal to $n \times U$ between 1 and n by rounding up the product of $n \times U$

Step 4: Retrieve the data, X_i^o, located at the ith of the ordered dataset $\mathbf{X^o}$.

2.5 Evaluation of Goodness-of-Fit of a Probability Distribution Model

There are many goodness-of-fit tests available from which to evaluate the goodness of fit of an assumed distribution model with respect to the data. Two general types of approaches for evaluating goodness of fit include probability plots and statistic tests (Cullen and Frey, 1999).

Probability plots are widely recognized to be a subjective method for determining whether or not data contradict an assumed model based upon visual inspection (Cullen and Frey, 1999). A graphical technique used in AuvTool is to compare the CDF of the fitted distribution with the original data set plotted using the Hazen plotting position method (Hazen, 1914) that was introduced in Section 2.1.

Statistical goodness-of-fit tests provide a quantitative measure of the goodness-of-fit of the assumed probability distributions, but many only apply to parametric distributions. An empirical distribution is an exact representation of the data in which each data point is assigned a probability of 1/n; therefore, a statistical goodness-of-fit test is not needed in this case. Three common goodness-of-fit tests for parametric distributions include the chi-square test, the Kolmogorov-Smirnov (K-S) test, and the Anderson-Darling (A-D) test. However, these tests may only be employed if a minimum amount of data is available (Cullen and Frey, 1999). For example, for the chi-square test, at least 25 data points should be available. The K-S test can be used with as few as five data points. The A-D test is valid if the number of samples is greater than or equal to eight.

The chi-square test involves calculating a test statistic that approximately follows a chi-square distribution only if the hypothesized model cannot be rejected as a poor fit to the data. The advantage of chi-square test is its flexibility; it can be used to test any distribution. However, a disadvantage of this method is that it has lower power than other statistical tests (Cullen and Frey, 1999). This is because the chi-square test involves binning of the data. In

binning the data, some of the information associated with individual data points is lost. Thus, the chi-square test is less discriminatory than a test that makes more sufficient use of all data points, such as the K-S test.

The K-S test involves a comparison between a stepwise empirical CDF and the CDF of a hypothesized distribution. This test is based upon evaluation of the maximum difference in the cumulative probability of the fitted distribution versus that of a data point. An attractive feature of K-S test is that it is a distribution-free test of goodness of fit. An advantage of K-S test over the chi-square test is that it can be used with smaller sample sizes. However, K-S test tends to be more sensitive to deviations of a good fit near the center of the distribution compared to at the tails (Stephens, 1974; D'Agostino and Stephens, 1986).

The A-D test is a "quadratic" test that is based upon a weighted square of the vertical distance between the empirical and fitted distributions (Cullen and Frey, 1999). The A-D test gives more weight to the tails than does the K-S test and therefore is more sensitive to deviations in the fit at the tails of a distribution (Stephens, 1974). However, the A-D test is not distribution-free test. Therefore, the critical values must be calculated specifically for each type of parametric distribution. Therefore, the A-D test is often used as a supplement to other goodness-of-fit tests.

Because the chi-square test requires at least 25 data points, and because it is not as powerful as other methods, the chi-square test was not included in AuvTool. The K-S and A-D tests are included in AuvTool.

It must be pointed out that there are some limitations with the use of statistical goodness-of-fit tests. For example, they address only one possible criterion for determining goodness-of-fit, and could imply acceptance of a fit that might be poor for reasons not addressed by the

criterion, or imply rejection of a fit that might be acceptable for reasons not addressed by the criterion. For example, it is possible that a normal distribution might not be rejected by a goodness-of-fit test. However, if the normal distribution is used to represent a quantity that must be non-negative, and if the probability of predicting negative values using a normal distribution is not negligible, then the use of a normal distribution will not make physical sense. Therefore, an uncritical application of a goodness-of-fit test can lead to an inappropriate choice of parametric distribution. Conversely, the goodness-of-fit test may imply rejection of a non-negative distribution, such as a lognormal, which might be theoretically consistent with the basis of the data. Therefore, users are strongly urged not to rely on the results of goodness-of-fit tests without inspecting the results and considering other factors that are important to the selection of an appropriate parametric distribution.

The graphical comparison of the CDF of the fitted distribution to the original data set plotted using the Hazen plotting position is more informative when confidence intervals are estimated for the fitted CDF, and when the frequency with which data are enclosed by the confidence intervals is taken into account. This approach is discussed in more detail in Section 2.6 on bootstrap simulation.

In the following subsections, methods for evaluating the adequacy of the fit of a parametric distribution with respect to the data are explained in more detail. These include the techniques used in AuvTool for: (1) visually comparing the CDF of the fitted distribution with the data; (2) using the K-S test; (3) using the A-D test; (4) and visually comparing confidence intervals for the CDF of the fitted distribution with the data. In addition, the method for automatically fitting distributions to the data in batch mode is discussed, with appropriate warnings to the user regarding limitations of the method.

2.5.1 Graphical Comparison of CDF of Fitted Distribution to the Data

The goodness-of-fit of a parametric distribution compared to the data can be visually inspected. This is accomplished by plotting the CDF of the fitted distribution versus the data. The data can be plotted using the Hazen plotting position introduced in Section 2.1.

Since analytical solutions are not available for CDFs for all of the parametric distributions used in AuvTool, the CDFs are estimated using numerical simulation. The construction of a numerically stable representation of CDF of the fitted distribution is based on statistical theory. The CDF is estimated by generating a large number of random samples from the parametric distribution and plotting them using the Hazen plotting position. With a large number of samples, the numerically simulated CDF will look as if it is a continuous smooth curve. The sample size chosen for numerical simulation of the CDF for purposes of graphical display is based upon the statement in Casella and Berger (1990) that if the sample size is large enough (e.g., >=2,000), then the sample can be assumed to be a very good representation of population distribution. Therefore, in AuvTool, 2,000 random numbers are generated for the distribution and are used to construct an empirical CDF using the Hazen plotting position. The numerically simulated CDF is considered to be a very good representation of the actual CDF of the fitted distribution, and it is plotted in the same graph with the original data set.

An example of a graphical comparison of a numerically simulated CDF for a parametric probability distribution and of the data to which the distribution was fit is shown in Figure 2-5. The data are depicted by open circles. The numerically simulated CDF is depicted by a solid line. The example shown in Figure 2-5 is for a beta distribution fit to a data set for a quantity that is bounded by zero and one. The beta distribution corresponds very closely with the data over most of the range of the observed values. Graphs similar in technical content to this example, although somewhat different in format, are produced by the AuvTool GUI.

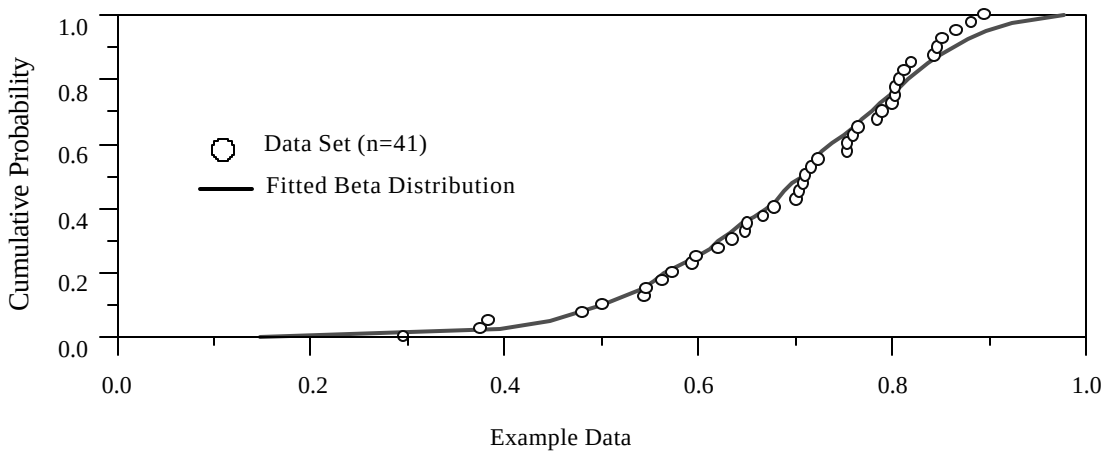


Figure 2-5. Comparison of Fitted Beta Distribution to an Example Dataset

2.5.2 Kolmogorov-Smirnov Test

As previously noted, the K-S test is based on comparison of the CDF of the fitted distribution to an empirical CDF of the data. The maximum discrepancy in the estimated cumulative probabilities for the two CDFs is identified. The maximum discrepancy is then compared to a critical value of the test statistic. If the maximum discrepancy is larger than the critical value, the hypothesized distribution is rejected (Cullen and Frey, 1999). This method is also discussed by Ang and Tang (1984), D'Agostino and Stephens (1986), and others.

The algorithm for performing the K-S test is described here:

(1) Rank the original data in an ascending order to have an ordered dataset **X** in which $X_k < X_{k+1}$, where, k =1,2…n.

(2) Develop a stepwise cumulative density function as follows:

$$S_n(x) = \begin{cases} 0 & x < x_1 \\ k/n & x_k \le x \le x_{k+1} \\ 1 & x \ge x_n \end{cases} \tag{2-39}$$

where,

$S_n(x)$ = The stepwise cumulative density function

n = The number of data points in a data set

x_k = The data

(3) Calculate the maximum difference between $S_n(x)$ and the CDF of the fitted distribution over the entire range of X. The maximum difference is denoted by:

$$D_n = \max\left|F(x) - S_n(x)\right| \tag{2-40}$$

where

D_n = The maximum difference

F(x) = The CDF of the fitted distribution

(4) Compares the calculated maximum difference from Equation (2-40) with the critical value D_n^{α} at a significance level of á.

The often-used significance level is 0.05. The critical values of D_n^{α} at a significance level of á=0.05 are tabulated in the Table 2-3.

The Table 2- . One is marked as “Specified”, another is marked as “Unspecified”. “Specified” implies that the underlying distribution type representing a data set is known, while “Unspecified” means that the information involving the underlying distribution for a data set is unknown. For example, if there is a sample for which the true values of the parameters of the population distribution are known, a “specified” critical value would be used. However, in most cases, the parameters of the distribution are estimated from the same data set for which the goodness-of-fit comparison is

Table 2-3. Citical Value of D_n^{α} the Kolmogorov-Smirnov Test

n	á=0.05 (Specified)	n	á=0.05 (Unspecified)
5	0.56	5	0.337
10	0.41	8	0.285
15	0.34	10	0.258
20	0.29	12	0.242
25	0.27	15	0.220
30	0.24	16	0.213
35	0.23	18	0.200
40	0.21	20	0.190
45	0.20	25	0.180
50	0.19	30	0.161
>50	$1.36/\sqrt{n}$	>30	$0.886/\sqrt{n}$

(Massey, 1951; Lilliefors, 1967)

made. In this latter situation, the "Unspecified" values should be used. Since this latter case is more common, the “Unspecified” critical values are used in the development of AuvTool. If the critical value of a number *n* is not listed in the Table 2-3, and when *n* is less than 30 (“Unspecified”), a linear interpolation is used to calculate the critical value for the number.

The K-S test is a distribution-free; it can be applied to normal, lognormal, beta, gamma, Weibull, uniform, and symmetric triangle distributions. However, the K-S test has several important limitations: (1) it is only valid for continuous distributions; and (2) it tends to be more sensitive near the center of the distribution than at the tails (D’Agostino and Stephens, 1986).

2.5.3 Anderson-Darling Test

The A-D test is used to test if a sample of data is from a population with a specific distribution (Stephens, 1974). It is a modification of the K-S test and gives more weight to the tails than does the K-S test. Unlike K-S test, the A-D test is not a distribution-free test. For different distributions, A-D test statistics and the corresponding critical values are different. For some distributions, relevant information for calculating the A-D test is not available in literature. These distributions include uniform, symmetric triangle and beta distributions. Therefore, in

AuvTool, the A-D test is available only for the normal, lognormal, gamma and Weibull distributions.

The A-D test statistic is defined as:

$$A^2 = -n - S \tag{2-41}$$

where,

$$S = \sum_{i-1}^{n} \frac{(2i-1)}{n}\left[\ln(F(x_i) + \ln(1.0 - F(x_{n+1-i})\right] \tag{2-42}$$

F is the cumulative distribution function of the specified distribution. X_i is the *ordered* data (Stephens, 1974; D'Agostino and Stephens, 1986).

When parameters of an assumed distribution are not known, and have to be estimated from the sample data, the A-D test statistic must be modified (D'Agostino and Stephens, 1986). For normal and lognormal distribution, the modified statistic is (D'Agostino and Stephens, 1986):

$$A^* = A^2(1.0 + 0.75/n + 2.25/n^2) \tag{2-43}$$

For the Weibull distribution, the modified statistic is (D'Agostino and Stephens, 1986):

$$A^* = A^2(1.0 + 0.2/\sqrt{n}) \tag{2-44}$$

For the gamma distribution, when both the scale and shape parameters are unknown and are estimated from the data, the A-D test statistic does not need to be modified (D'Agostino and Stephens, 1986). However, the critical value at a given significance level for the gamma distribution is dependent on the magnitude of its shape parameter (D'Agostino, Stephens, 1986).

The critical values of the A-D test for the normal, lognormal, and Weibull distributions are given in Table 2-4 and for the gamma distribution are given in Table 2-5.

Table 2-4. The Critical Values for Anderson-Darling test for Normal, Lognormal and Weibull distributions

Distribution	á=0.10	á=0.05	á=0.025	á=0.01
Normal, Lognormal	0.631	0.752	0.873	1.035
Weibull	0.637	0.757	0.877	1.038

(D'Agostino and Stephens, 1986, Table 4.7, p=123; Table 4.17, p=146)

Table 2-5. The Critical Values for Anderson-Darling test for the Gamma Distribution

Shape Parameter	Significant Level á =0.05
1	0.786
2	0.768
3	0.762
4	0.759
5	0.758
6	0.757
8	0.755
10	0.754
12	0.754
15	0.754
20	0.753
>20	0.752

(D'Agostino and Stephens, 1986, Table 4.21, p=155)

In AuvTool, linear interpolation is used to calculate the critical value of the A-D test for any given shape parameter based on the values provided in Table 2-5 for gamma distribution

2.5.4 Graphical Comparison of Confidence Intervals for CDF of Fitted Distribution to the Data

The results from bootstrap simulation can be used to help evaluate the goodness of a fit of a distribution with respect to the original data by graphically comparing confidence intervals for CDF of the fitted distribution to the data. More details on the bootstrap simulation and how the confidence intervals for CDF of the fitted distribution are estimated can be found in Sections 2.6.1, 2.6.3, and 2.6.4.

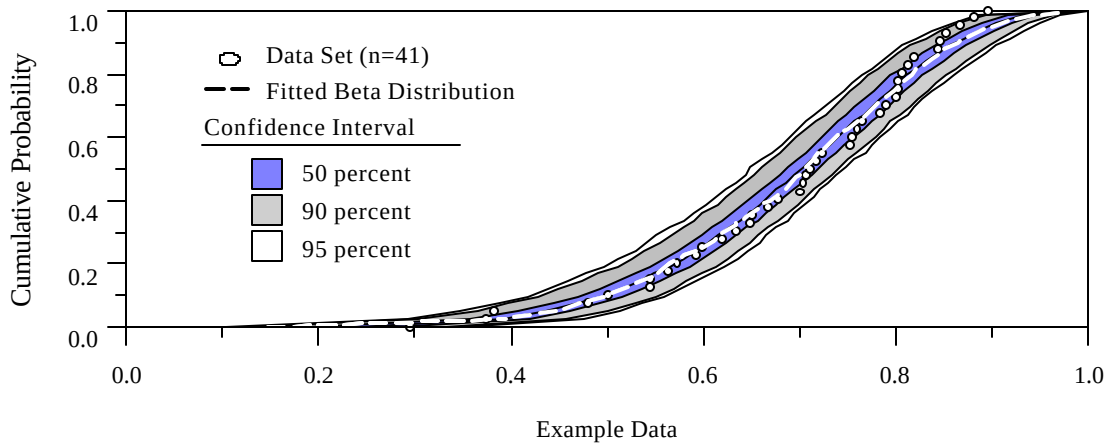


Figure 2-6. An Illustrative Example of Graphical Comparison of Confidence Intervals for CDF of Fitted Distribution to the Data

Figure 2-6 graphically shows a comparison of confidence intervals for the fitted distribution with an example data set. The results are from two-dimensional simulation with the points out of 41 are contained within the 95 percent confidence intervals. Thus, the fit in this case is a reasonably good one. On average, it is expected that 95 percent of the data will fall inside of a 95 percent confidence interval of CDF of a fitted distribution if the data are a random sample from the assumed population distribution

2.5.5 Criteria for Automatically Seeking a Best Distribution Model in Batch Mode Analysis

A technique for assisting the user in choosing a best parametric probability distribution model for a data set is included in AuvTool. The technique is based upon the of the K-S goodness-of-fit test. The technique is applicable to the normal, lognormal, beta, gamma and Weibull distribution. Although K-S test is also available for uniform and symmetric distributions, the two distributions are more often used to characterize subjective expert judgment and are not typically used when fitting distributions to data. Therefore, the uniform

and symmetric distributions are not included as options in the batch mode distribution selection technique.

A premise for using K-S test value as a criterion is that a smaller value of the K-S test statistic implies a better fit. However, it must be pointed out that there is no specific support for this premise in the literature. The evaluation and selection of fitted distributions involves many factors. A distribution model that has the smallest value of K-S test may not always be the best model for describing a variable. For example, suppose that there is a variable for which samples could be larger than 1. However, because of limited sample size, it is possible that all of the data available in a given sample have values between 0 and 1. In this case, the results from using the batch analysis feature provided in the AuvTool might suggest that a beta distribution is a better fit in terms of the K-S test values. However, the two parameter beta distribution would inappropriate as a choice for describing a random variable whose values could exceed 1. Another example is given in Cullen and Frey (1999) for a leafy vegetable producing PCB concentration data set. In the example, almost all of the many analyses pointed to the normal distribution as being a better fit to the data than the lognormal distribution. However, the use of the normal distribution in that example leads to unacceptably high probabilities of predicting negative concentrations. Therefore, the normal distribution would be an inappropriate choice, even though it provided the best fit, because the PCB concentrations cannot be negative. Therefore it must be emphasized that **uncritical application of the batch mode distribution selection procedure included in AuvTool can lead to an inappropriate selection of a parametric probability distribution model.**

The user of AuvTool is cautioned that the availability of a batch mode technique for choosing a distribution based upon the K-S test is not a substitute for the use of judgment. The

K-S test is based upon a specific criterion which may or may not be important to a particular analyst or decision maker in the context of a specific problem. The K-S test does not screen for results that may be physically implausible, such as a probability of sampling negative values for a quantity that must be non-negative. The appropriateness of selection of a distribution depends on the data quality objective of each analysis, which may differ from one situation to another. Therefore, uncritical application of the batch mode feature of AuvTool for seeking a best fit distribution is likely to lead to inappropriate selection of a probability distribution model in some cases. **It is the user's responsibility to evaluate the automatically selected parametric probability distribution for appropriateness with respect to the user's own criteria and needs**.

2.5.6 Summary of Methods for Evaluating Goodness-of-Fit

Several different techniques for evaluating goodness-of-fit of a parametric probability distribution model compared to a data set have been presented. Although it is tempting to base the selection of a parametric probability distribution model solely upon the application of a goodness-of-fit statistical test, this temptation should be strongly resisted. Instead, it is critically important to consider the following questions in making the choice of a parametric distribution:

Is the selected parametric probability distribution model consistent with the data in terms of underlying theory?

Is the selected parametric probability distribution a plausible representation of the data? For example, if the data must be non-negative, does the selected distribution also have this feature?

What characteristics of the distribution are of most concern in your specific assessment, and are these criteria the same as those for the goodness-of-fit test? If so, then the goodness-of-fit test should be treated as a useful consideration in choosing a distribution, but it should not be the only consideration. The latter is especially true if the answers to either of the first two questions are "no".

Are the criteria for the goodness-of-fit test relatively unimportant for a particular assessment? In this case, the user will find it more useful to rely upon a graphical comparison of the fitted distribution with the data, either based upon a comparison of

the CDF of the fitted distribution with the data, or based upon a comparison of the confidence intervals of the CDF of the fitted distribution with the data

In fact, both graphical comparison and statistical goodness-of-fit tests involve subjective judgment regarding what constitutes an acceptable fit (Cullen and Frey, 1999). For example, the K-S and A-D tests involve subjective judgment regarding the choice of significance levels. Many authors emphasize the subjective nature of statistical tests. Hann and Shapiro (1967) state this quite well in their excellent book:

> "One might conclude…. that a proper procedure for selecting a distribution is to consider a wide variety of possible models, evaluate each by the methods here described, and assume as correct the one that provide the best fit to the data. However, no such approach is being suggested. Where possible, **the selection of the model should be based on an understanding of the underlying physical properties**… The distributional test then provides a useful mechanism for evaluating the adequacy of the physical interpretation. Only as a last resort is the reserves procedure warranted, and then, only with much care, for, although many models might appear appropriate within the range of data, they might well be in error in the range for which predictions are desired," (pp.260-261).

2.6 Characterization of Variability and Uncertainty

The primary objective of this section is to introduce relevant methods for characterization of uncertainty in the mean, standard deviation, and parameters of a distribution. Uncertainty in a statistic attributable to random sampling error can be represented by a sampling distribution (Cullen and Frey, 1999). Sampling distributions are used to estimate confidence intervals for the parameters of a distribution. A confidence interval for a statistic is a measure of the lack of knowledge regarding the value of the statistic. There are a variety of methods for characterizing uncertainty in the mean or standard deviation, including analytical solutions and numerical simulations. Analytical solutions are available for cases in which the underlying distribution for a data set is normal or for which the variance is small enough and/or the sample size for a data set is large enough (e.g., >30). If the underlying population distribution is not normal and the sample size for a data set is small, analytical methods based upon normality may lead to

significant errors in the estimation of confidence intervals. Therefore, there is a need for a more flexible approach for estimating sampling distributions and confidence intervals. The numerical simulation method of bootstrap simulatio, may be used to estimate confidence intervals for the mean or other statistics (Efron and Tibshirani, 1993).

Bootstrap simulation, introduced by Efron in 1979, is a numerical technique originally developed for the purpose of estimating confidence intervals for statistics based upon random sampling error. This method has an advantage over analytical methods in that it can provide solutions for confidence intervals in situations where exact analytical solutions may be unavailable and in which approximate analytical solutions are inadequate. For example, in estimating uncertainty in the sample mean, bootstrap simulation does not require that the original data set be normally distributed, even for small sample sizes. This advantage over analytical methods that are based on normality assumptions makes bootstrap simulation a more versatile and robust method for estimating uncertainty in a statistic due to sampling error, especially for non-normal data sets (Cullen and Frey, 1999). Bootstrap simulation has been widely used in the prediction of confidence intervals for a variety of statistics.

The method illustrated by Frey and Rhodes (1996;1998) for using bootstrap simulation in the context of an environmental case study is the basis for the simulation technique used in AuvTool. The following subsections introduce the bootstrap method and the two major steps associated with the bootstrap method: (1) generating bootstrap samples; and (2) forming bootstrap confidence intervals. In addition, the details of the two-dimensional simulation method presented by Frey and Rhodes (1996; 1998) are described.

2.6.1 Bootstrap Method

The bootstrap method addresses uncertainty due to random sampling error by first assuming that the original data set, **x**, of sample size *n*, is a random sample from the distribution

$\hat{F}$, and then repeatedly asking the question: What if the data set had been a different set of n random values from the same distribution $\hat{F}$? This question is answered by repeatedly generating "bootstrap samples." A bootstrap sample, $\mathbf{x}^*$, is defined as a random sample of size n taken from the distribution, $\hat{F}$. Bootstrap samples may be simulated using random Monte Carlo simulation (Rhodes, 1997). A large number, B, of independent bootstrap samples ($\mathbf{x}^{*1}$, $\mathbf{x}^{*2}$, … $\mathbf{x}^{*n}$) are selected from the distribution $\hat{F}$. From each of the B bootstrap samples, a new statistic $\hat{q}^*$, is computed such that:

$$\hat{q}^{*i} = f(\mathbf{x}^{*i}) \qquad \text{for } i = 1, 2, \ldots, B \tag{2-45}$$

Each $\hat{q}^*$ is referred to as a *bootstrap replicate* of $\hat{q}$ (Rhodes, 1997; Frey and Rhodes, 1999).

The bootstrap replications ($\hat{q}^{*1}, \hat{q}^{*2}, \ldots, \hat{q}^{*B}$) are each independent realizations of an estimate of the parameter q. The dispersion of values of the bootstrap replications reflects the uncertainty in the sample estimate of the unknown parameter, q, attributable to random sampling error. The bootstrap replicate values describe an estimate of the sampling distribution of the statistic. Since a statistic is estimated from randomly drawn values, it is itself a random variable. The number of bootstrap replications necessary to reasonably approximate the true sampling distribution of the statistic depends upon the statistic being estimated. For, example, according to Efron and Tibshirani (1993), to compute the standard error of the mean (the original intent of the bootstrap technique), B = 200 is generally enough and B = 25 is often sufficient. However, for computing confidence intervals or estimating percentiles of sampling distributions, Efron and Tibshirani (1993) suggest B = 1000. In examples for computing confidence intervals given in Efron and Tibshirani (1993), the number of bootstrap replications ranges between B = 1,000 and B = 2,000.

2.6.2 Methods of Generating Bootstrap Samples

In bootstrap simulation, the sample data points, $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ are assumed to be a random sample of size n from some unknown probability distribution F. The parameter of interest, q, is a characteristic of the distribution of F, $q = f(F)$, such as the mean, variance, shape or scale parameter, or any fractile or quantile of the distribution F. An estimate of q is the statistic $\hat{\theta}$, which is determined from the data set, $\hat{\theta} = f(\mathbf{x})$.

Using the data set, $\mathbf{x}$, the distribution $\hat{F}$, is defined to be an estimate of the unknown population distribution F. The distribution may be defined as either an empirical distribution or a parametric distribution. The former is the basis for non-parametric bootstrap, and the latter is the basis for parametric bootstrap (Efron and Tibshirani, 1993). Non-parametric bootstrap is also commonly referred to as "resampling." One of the main shortcomings of resampling of a data set is that the minimum and maximum values obtained in each bootstrap sample are limited to the minimum and maximum values within the data set. When only small data sets are available, this can lead to biases in the representation of a given model input (e.g., failure to consider possible large values that are not present in the limited data set). The use of parametric distributions is one way to allow for the possibility that smaller or higher values than those observed in the data set may occur in the real system being modeled. The method of generating bootstrap samples based on an empirical distribution for non-parametric bootstrap simulation is discussed in Section 2.4.9. The algorithms for generating bootstrap samples based on parametric distributions for normal, lognormal, beta, gamma, Weibull, uniform, and symmetric triangle distributions are documented in Sections 2.4.2 through 2.4.8.

2.6.3 Methods of Forming Bootstrap Confidence Intervals

The development of good confidence intervals is an important issue in bootstrap simulation. “ Good” means that the bootstrap intervals should closely match exact confidence

intervals in those special situations where statistical theory yields an exact answer, and the interval should give dependably accurate coverage probabilities in all situations. A method that produces such a good confidence intervals should be both transformation respecting and second-order accurate (Efron and Tibshirani, 1993).

Several bootstrap confidence interval methods have been proposed in the literature (Efron and Tibshirani, 1993; Burr, 1994). These methods include the standard normal, percentile, bootstrap-t, and Efron's BC_a. The standard normal method requires the imposition of normality assumption on the bootstrap distribution and it is neither transformation respecting nor second-order accurate. Therefore, the standard normal method is not a "good technique" for forming bootstrap confidence interval. The percentile method is possibly the most frequently used in practice. Although it is only first-order accurate, the intervals obtained from this method are the simplest to use and explain (Efron and Tibshirani, 1993). The bootstrap-t and the BC_a intervals are comparable in that both have been demonstrated theoretically to be "second-order correct", but the bootstrap –t method is not transformation respecting. Burr (1994) suggests that bootstrap-t is unstable. More discussion on these methods can be found in the Efron and Tibshirani (1993), Burr (1994), and Martin (1990). Though there is no gold standard to make a definitive conclusion as to which method is the best, for simplicity and because it is the most widely used method in practice, the percentile method will be discussed and used.

When calculating a confidence interval, the intent is to develop an interval that has a (1-2α) probability of enclosing the true value of a statistic, θ. The upper and lower bounds of this confidence interval are determined by ordering the B bootstrap replicates of $\hat{\theta}^*$, ($\hat{\theta}^{*1}, \hat{\theta}^{*2}, \ldots, \hat{\theta}^{*B}$). Given these ordered statistics, the 100αth percentile (the lower bound of the confidence interval) is the $B \bullet \alpha$th largest value, $\hat{\theta}^{*B \bullet \alpha}$, and the 100(1-$\alpha$)th largest value, $\hat{\theta}^{*B \bullet (1-\alpha)}$.

For example, for B =1,000 and α = 0.05, the 90 % confidence interval for some parameter, q, is given by:

$$[\hat{q}^{*B\bullet\alpha}, \hat{q}^{*B\bullet(1-\alpha)}] = [\hat{q}^{*50}, \hat{q}^{*950}] \tag{2-46}$$

where, $\hat{q}^{*50}$ and $\hat{q}^{*950}$ are simply the 50th and 950th values in the ordered set if the bootstrap statistics.

2.6.4 Two-dimensional Simulation of Variability and Uncertainty

AuvTool features the use of the two-dimensional approach to simulation of both variability and uncertainty employed by Frey and Rhodes (1996, 1998) and that has also been implemented in other prototype software (e.g., Frey and Zheng, 2000) as described in Chapter 1.

As shown in Figure 2-7, bootstrap simulation is used to simulate the uncertainty in the parameters of a frequency distribution, $\hat{F}$, that has been fitted to a data set of sample size n. A total of B bootstrap samples of sample size n are simulated. For each bootstrap sample, a new distribution is fitted and a bootstrap replication of the distribution parameters is calculated. The bootstrap simulation produces paired parameter estimates. These multivariate sampling distributions of the parameters represent the uncertainty in the distribution parameters. In the two-dimensional simulation, a total of q different frequency distributions are simulated, where q = B in most cases presented here. Each alternative frequency distribution is based upon a different set of bootstrap replicate distribution parameters. For each alternative frequency distribution, a total of p random samples are simulated to represent one possible realization of variability within the population. For example, suppose B=500 and p = 500. Thus, a total of 250,000 samples are generated, representing 500 samples from each of 500 alternative frequency distributions. For each realization of uncertainty, the samples are sorted to represent cumulative

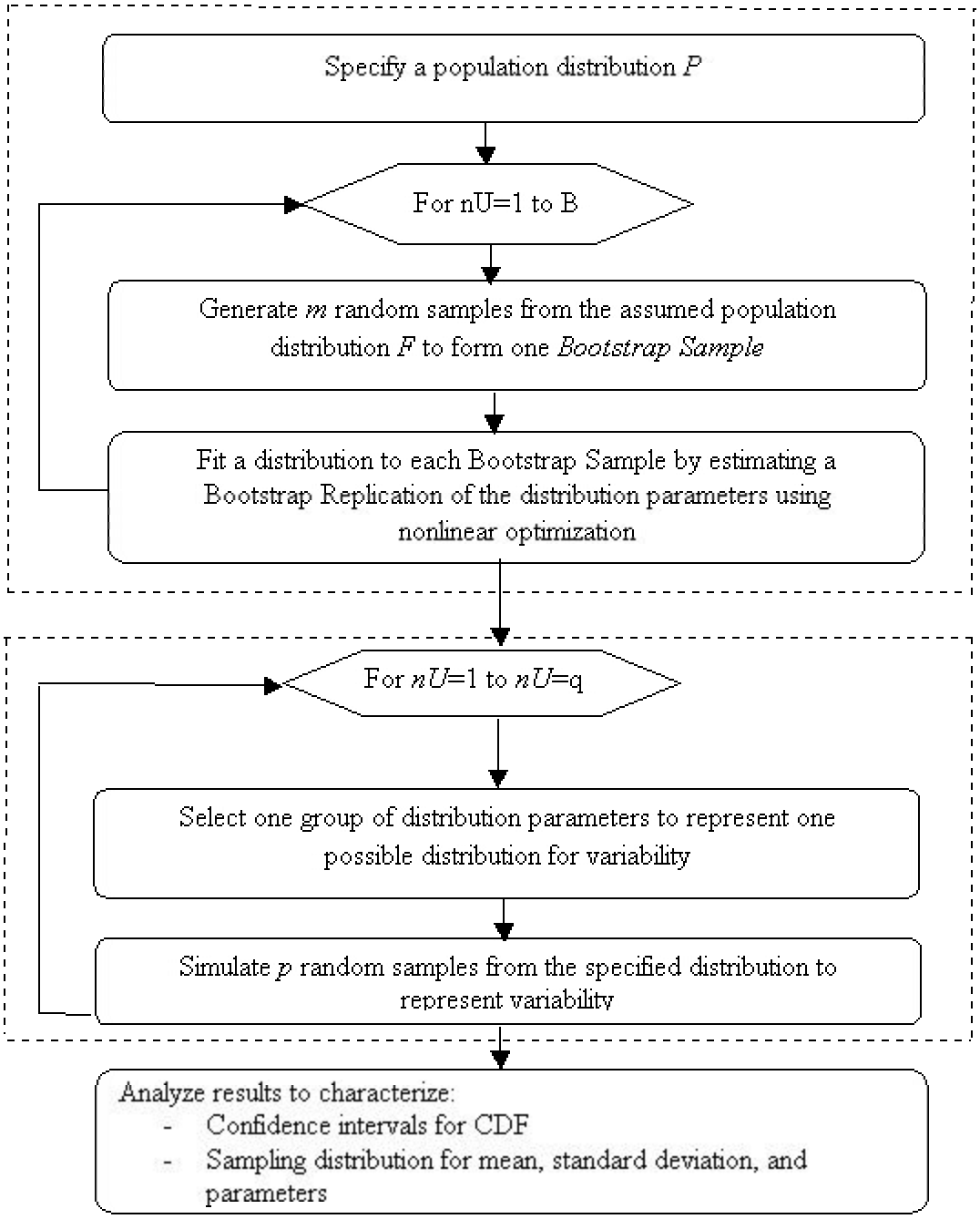


Figure 2-7. Flow Diagram For Bootstrap Simulation and Two-Dimensional Simulation of Variability and Uncertainty. (Where: B=number of Bootstrap Replications, q=Sample Size Used for Uncertainty, p=Sample Size Used of Variability) (Frey and Rhodes, 1998)

distribution functions. Thus, there are 500 values for any given statistic (e.g., mean, variance, 95th percentile of variability) which can be used to construct confidence intervals for each

statistic. An example graph of probability bands from two-dimensional simulation was shown in Figure 2-6 of Section 2.5.4.

2.7 Evaluation of Dependence or Correlation between Statistics of Interest

Possible dependence or correlation between the mean, and standard deviation, and between the parameters of distributions sometime exists. The evaluation of the dependence or correlation will guide analysts to correctly make use of the statistics (Morgan and Henrion, 1990). A sample correlation describes the strength of the linear association between variables. An association between variables means that the value of one variable can be predicted, to some extent, by the value of the other. A linear correlation is a special kind of association. A non-linear relation can be transformed into a linear one before the correlation is calculated. There are a few ways to evaluate the dependency or correlation between a set of variable pairs. These include correlation coefficients, scatter plots and regression analysis (Cullen and Frey, 1999). In the AuvTool, the sample correlation coefficient is used as an indication of an association between the sample distributions of selected statistics.

The correlation coefficient is calculated with the assumption that *both* variables are stochastic (i.e., bivariate Gaussian). It can be obtained by using the following formula (Morgan and Henrion, 1990):

$$R = \frac{\sum_{k=1}^{m}(x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum_{k=1}^{m}(x_k - \bar{x})^2 \times \sum_{k=1}^{m}(y_k - \bar{y})^2}} \tag{2-47}$$

where,

R = Correlation coefficient between two variables

x_k = Variable **x** samples

$\bar{x}$ = The mean of x_k samples

y_k = Variable **y** samples

$\bar{y}$ = The mean of y_k samples.

The magnitude of the correlation coefficient is a measure of probabilistic dependency between two uncertain variables. It varies from 0 (random relationship) to 1 (perfect linear relationship) or -1 (perfect negative linear relationship).

The use of a sample correlation as an indicator of dependence between two distributions is a potentially useful but not perfect approach. If the true dependence is non-linear, then the sample correlation coefficient may fail to give a strong indication of a potentially important dependence. Frey and Rhodes (1998) illustrate the dependence between the mean and standard deviation, and between the parameters, for selected distributions. For example, the parameters of the gamma distribution have a strong inverse nonlinear dependence. A sample correlation coefficient will not be as sensitive to this type of dependency as, for example, a rank correlation coefficient. A feature available in AuvTool is the ability to export paired data for the sampling distributions of the mean, standard deviation, and the values of both parameters for each parametric distribution. Therefore, a user can graph the paired values of the bootstrap replications of the mean and standard deviation, or of the two parameters, to identify and characterize dependences that are not fully captured by a sample correlation coefficient.

2.8 Summary

This chapter has described the technical basis for the algorithms used in AuvTool. These algorithms include the following:

- Plotting of data sets using the Hazen plotting position
- Visualization of the CDF of fitted distributions and graphical comparison of these with the data

- Estimation of parameters for parametric probability distributions using MoMM or MLE approaches
- Presentation of empirical step-wise CDFs
- Generation of random numbers from empirical step-wise CDFs or from parametric probability distribution models
- Calculation of test statistics as an aid in determined goodness-of-fit of a parametric probability distribution to a data set
- Estimation of confidence intervals of the CDF of a parametric probability distribution fitted to a dataset and graphical comparison with the data as an aid in evaluating goodness-of-fit.
- Use of bootstrap simulation to characterize sampling distributions and confidence intervals for key statistics, such as the mean, standard deviation, and parameters of parametric probability distribution models.

3.0 AUVTOOL SYSTEM DEVELOPMENT AND IMPLEMENTATION

The methodology for quantifying variability and uncertainty described in Chapter 2 was implemented in the software tool AuvTool. In this chapter, we present the design considerations, development environment and tools, structure design, and the main function modules and associated main features of AuvTool.

3.1 AuvTool Software Design Considerations

The primary goal of AuvTool is to provide a user-friendly preprocessor module for the EPA SHEDS model which incorporates appropriate algorithms for fitting distributions to model inputs and for quantifying variability and uncertainty in each input. Therefore, the main concern in the design and development of AuvTool system is to make the output of AuvTool appropriate for use as input to the SHEDS model. Because the SHEDS model involves a large number of model inputs, and because variability and uncertainty must be quantified for such inputs, a batch analysis feature was included in AuvTool. Aside from meeting the requirements of the SHEDS model, a secondary objective for AuvTool is to make it generally applicable for quantifying variability and uncertainty in other quantitative analysis fields such risk assessment and emission estimation. Thus, AuvTool was designed as a stand-alone program.

AuvTool provides output in a format of general application, but also in a format required for input to the SHEDS model. In addition, a future objective for AuvTool is to have capabilities which can allows users to specify their own models, and to propagate the variability and uncertainty from model inputs to model outputs. Therefore, the extensibility and expansion of the AuvTool was another main design concern. Based on these considerations, an object-oriented programming technique was used in the development of AuvTool system to promote modularity, extensibility, and reusability of the source code.

3.2 Development Environment and Tools

The Windows 98/ME platform was chosen as a development environment AuvTool. This choice was made to ensure compatibility with the SHEDS model. The software development tools used were Microsoft Visual C++, Graphic Server and Spread Active X controls. The reason for choosing Visual C++ lies in that it not only provides an object-oriented programming environment, which makes the software more extensible and expandable, but also facilitates the development of a user-friendly graphic interface. The Graphic Server and Spread tools can help to visualize the simulation results and organize the data input and result outputs.

3.3 Structure Design of the AuvTool System

Figure 3-1 shows the conceptual design and the relationship between modules of AuvTool system. AuvTool can currently be divided into five groups. Table 3-1 summarizes the composition of the groups and their main functions. As shown in Figure 3-1 and Table 3-1, the Data Import/Export group provides data for the Variability and Uncertainty Analysis group. The analysis results from the Variability and Uncertainty Analysis group are reported to the Variability and Uncertainty Resulting Reporting group, and to the Further Analysis group for further analysis of the sampling distribution data for the statistics of interest (e.g., mean, standard deviation, and distribution parameters). The results from the Further Analysis group are reported to the Variability and Uncertainty Resulting Reporting group for summarization. The modifications of the *Random Seed Setting* module in the Random Sampling group are passed to other analysis modules.

3.4 AuvTool Main Modules

As shown in Figure 3-1, AuvTool is composed of different function modules. The following subsections briefly describe the main functions modules and the associated features that the function modules provide.

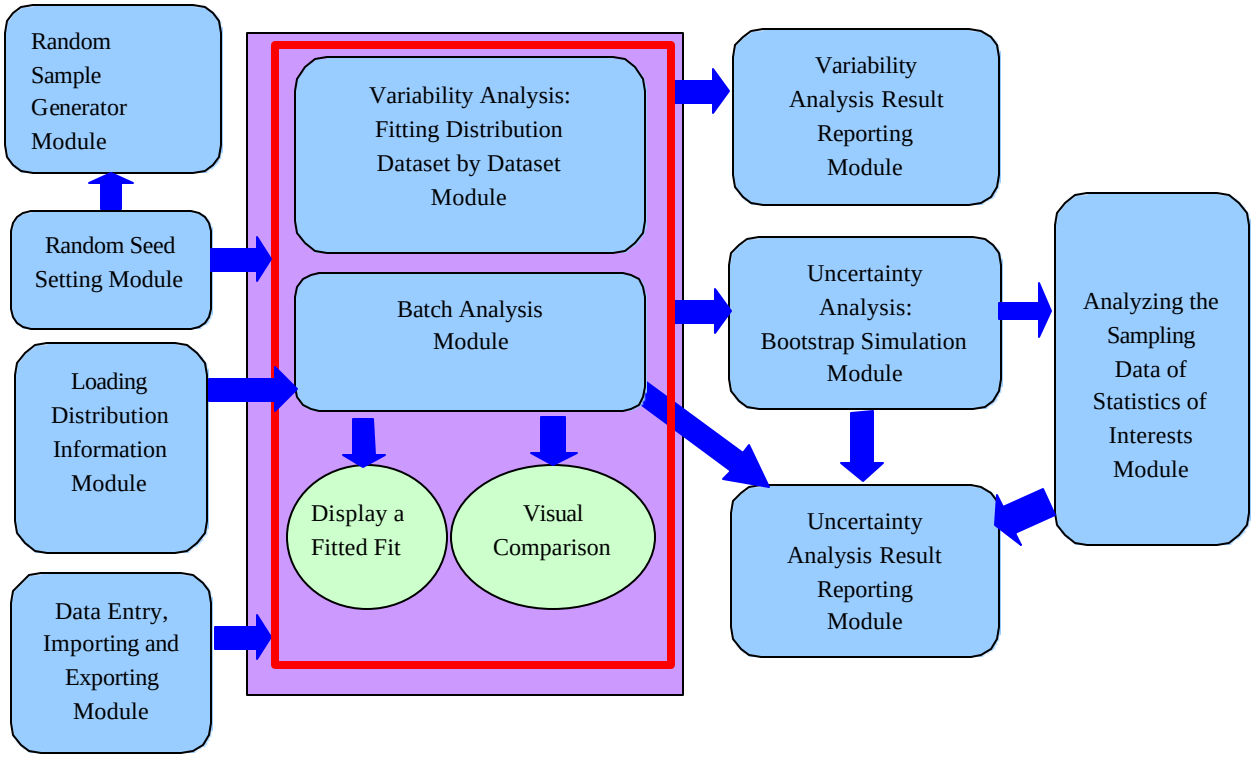


Figure 3-1. The Conceptual Structure Design of AuvTool System

Table 3-1. AuvTool Function Module Summarization Table

Group Name	Modules	Main Functions
Data Import/Export	*Data Entry, Importing and Exporting* module and *Loading Distribution Information* module	Provides the required data for variability and uncertainty analysis, and exports the input data for future analysis and other applications
Random Sampling	*Random Seed Setting* module and *Random Sample Generator* module	Sets the random seeds and generates random samples
Variability and Uncertainty Analysis	*Variability Analysis-Fitting Distribution Dataset by Dataset* module, *Batch Analysis* module and *Uncertainty Analysis* module	Implements all simulations and calculations related to variability and uncertainty analysis
Further Analysis	*Analyzing the Sampling Data of Statistics of Interest* Module	Does further analysis of the sampling data of interests of statistics from bootstrap simulation
Variability and Uncertainty Result Reporting	*Variability Analysis Result Reporting* module and *Uncertainty Analysis Result Reporting* module	Provides summarization tables for user's variability and uncertainty analysis cases

3.4.1 Data Entry, Importing and Exporting Module

The *Data Entry, Importing and Exporting* module provides a data sheet similar to a spreadsheet for users to input or output data. In this module, users can enter data from the keyboard, load an existing AuvTool format data file, and import a Microsoft Excel 97 data file or tab-delimited text files from other application programs into the main data sheet. In the data sheet, AuvTool specifies that each column represents one data set, and users can have multiple data sets by using multiple columns in the input format. Users can name each data set. The module automatically counts the number of data points in a data set and logically checks the users' inputs. For example, if there are some invalid numerical value inputs, AuvTool will prompt the user to correct their inputs before they can do variability and uncertainty analysis. This module allows the user to save their data into an AuvTool file format or to export their data to an Excel file or tab-delimited text file. The data in the module will be used in the other analysis modules as a basis of variability and uncertainty analysis.

3.4.2 Loading Distribution Information Module

It often happens that users can obtain distribution information for some variables from some other sources such as technical reports, while no original data for those variables are available. However, in this situation, it is still possible for users to do uncertainty analysis by using bootstrap simulation if they have sufficient information about the distribution describing the variable. This information includes the type of parametric distribution, the parameter values, and the sample size. The implementation of the *Loading Distribution Information* module enables users to complete uncertainty analysis for this situation. This module allows users to provide the distribution information from the keyboard, from an existing AuvTool disk file, or from other file formats such as Excel. The information is passed to the batch analysis module to do uncertainty analysis. Currently, the module allows users to provide common single

component parametric distributions. The distribution models include normal, lognormal, gamma, beta, Weibull, uniform, and symmetric triangle distributions.

3.4.3 Random Seed Setting and Random Sample Generator Modules

By default, any analysis modules will use the default random seed provided by AuvTool. However, in some situations, users want to change the random seed for their needs. For example, they want to check the repeatability of simulation results for different random sample series. Keeping the same seed will help users to duplicate the simulation results. The *Random seed setting* module implemented in AuvTool provides options for users to keep or modify the default random seed. The choice of random setting in this module is passed to all other modules. AuvTool also provides a *random sample generator* module, in which users can generate random samples by specifying the corresponding distribution information and the number of random samples they want to generate. This module can generate random samples based on an empirical distribution. The results generated in the module can be easily copied or exported to other application programs, for example, Excel or Notepad.

3.4.4 Variability Analysis-Fitting Distribution Dataset by Dataset Module

The *variability analysis-fitting distribution dataset by dataset* module automatically lists the data sets needing to be analyzed based on the data that users provide in the *data entry, importing and exporting* module. In the module, users are able to perform variability analysis data set by data set. This modules provides seven distribution types which include normal, lognormal, beta, gamma, Weibull, uniform and symmetric triangle distributions that can be fit to a data set, and (in most cases) a choice of two parameter estimation methods, including method of matching moments (MoMM) and maximum likelihood estimation (MLE).

The user can choose the K-S and A-D statistical goodness-of-fit tests, where applicable, to help in choosing a best fitting parametric distribution for a particular dataset. When users

select a data set to analyze, the module allows users to choose the parameter estimation method and the preferred distribution type. The data set and fitted distribution will be graphically and instantly visualized, which will help users to judge if the distribution they chose is a good representation of the data set or not. The K-S test and A-D statistical test results are presented on the right side of the user interface, which shows the value of the calculated test statistic; the critical value of the test statistic and whether or not the test was passed. If the users find that no parametric distribution offers a good enough fit to represent a data set, they can choose an empirical distribution. The decisions made via the module provide a basis for uncertainty analysis as described in Section 3.4.6. The variability analysis results in the module are reported to the *variability analysis reporting* module.

3.4.5 Batch Analysis Module

The *batch analysis* module is a core one in the AuvTool. Based on data provided in the *data entry, importing and exporting* module and the distribution information in the *loading distribution information* module, the *batch analysis* module automatically generates the control options for each data set or variable being analyzed. In the sheet inside the module, each row represents a data set or a variable; any choices and actions made on the selected row will only be effective for the data set or variable on the row.

For any data sets or variables with original data, the program will set "Auto" as the default option in the column of Distribution Choice. The user can modify the default option to one of the specific distribution types listed in the Distribution Choice combo box. "Auto" is not a distribution type, but an option, in which the user lets the program automatically choose a good fit for the selected data set.

For those cases that do not have original data, there is no "Auto" option available, and distribution information is from the data provided in the *loading distribution information* module. Users cannot modify the distribution type in these cases.

The module also allows users to choose parameter estimation methods. By default, the program will choose MoMM for cases with original data. For those cases without original data, and if no information is available for the parameter estimation methods, the program will mark "NA" on the row of the dataset. However, in uncertainty analysis, the program will by default assign MoMM to these cases. The module provides a feature to graphically display the fitted distribution and the data set. Another main feature of the module is that it allows users to visually compare different distributions fitted to a data set by graphically showing all reasonable fitted distributions in the same window, which will help users to choose a good fit.

The main advantage of this module is that not only it covers all features implemented in the *variability analysis-fitting distribution dataset by dataset* module, but also it provides features of automatic batch variability and uncertainty analysis, visual comparisons of different distribution types fitted to a data set, and uncertainty analysis for the variables without original data. In the module, if users prefer to use the default settings for all data sets analyzed, they do not need to make any choice or to go to any other analysis modules, but they still can complete their variability and uncertainty analyses.

The program will automatically help users to choose best fits and to do uncertainty analysis. This feature will be very helpful if users have a large number of data sets to be analyzed simultaneously. It must be pointed out that automatically choosing a best fit is based on a specified criterion. The criterion used in the AuvTool 1.0 is the minimum K-S test value. However, it must also be mentioned that a best fit in terms of the minimum K-S test statistic

value does not mean that the fit is the most reasonable one. In fact, users are cautioned that blind application of the K-S criterion to choosing a best fit may lead to selections of parametric distributions that are less than ideal fits in ways not captured by the K-S statistic or that may not have the most relevant theoretical underpinnings.

As mentioned above, the *batch analysis* module allows users to do uncertainty analysis based on the users' own judgments or selections. Any choices made via the module will be passed to the *uncertainty analysis-bootstrap simulation* module to do bootstrap simulations, and will be reported to the *variability analysis-reporting* module.

3.4.6 Uncertainty Analysis-Bootstrap Simulation Module

The *uncertainty analysis-bootstrap simulation* module features the use of bootstrap simulation and two-dimensional Monte Carlo simulation for simultaneously quantifying variability and uncertainty. The simulations are based on the choices of distribution types and parameter estimation results from the *variability analysis-fitting distribution dataset by dataset* module or *batch analysis* module.

The module allows users to modify the parameters for bootstrap simulations. For example, users can specify the number of bootstrap replications, and the sample size for variability. The program will by default show the probability band graph for the selected variable or data set when the bootstrap simulation is done. An example of band graph is shown in Figure 2-7 in the Chapter 2. The probability band depicts a plausible range which may enclose the "true" but unknown distribution. For example, the 95 percent probability band may be thought of as a 95 confidence interval. This interval has a 95 percent probability of enclosing the true but unknown distribution. The probability bands tend to be wider with very small datasets and/or in situations with large variation within the available sample of data. From the probability bands users can obtain a confidence interval for any percentile of the distribution.

This module also can graphically display the sampling distributions of the statistics of interest for the selected variable. The sampling distributions are the basis for constructing confidence intervals for the statistics. These statistics include the mean, standard deviation and distribution parameters. Because there are no parameters for an empirical distribution, the statistics for which sampling distributions are reported include only the mean and standard deviation. The module also provides a data sheet to hold the simulation data for the current variable in the data page of the module where users can export the results to other application programs. The simulation results from the module will be passed to the *analyzing the sampling data of statistics of interest* module.

3.4.7 Analyzing the Sampling Data of Statistics of Interest Module

The sampling distribution data from bootstrap simulation, which describe uncertainty for the selected statistics, are often described using an empirical distribution. The advantage of using empirical distributions is that they do not need any parametric distribution assumptions. However, a potential problem is that there is a large data storage requirement to save all of the replicate values of each statistic. A parametric probability distribution can also be used to represent the sampling distribution for the statistics in a more compact form. For example, in classical statistical theory, the confidence interval for the mean is often described using a normal distribution if the sample data are from a normal distribution or if the sample size is large enough. The use of bootstrap simulation makes the sampling data for statistics available for all other parametric population distribution and eliminates the often restrictive or incorrect normality assumption imposed upon the sampling distribution of the mean in case with small sample size and skewed data. Therefore, it is often the case that other parametric distributions besides the normal distribution should be used to represent the sampling distribution data for statistics such as the mean. The role of the *analyzing the sampling data of statistics of interest*

module in the AuvTool is to implement the further analysis of the sampling data from the bootstrap simulations feature. The batch analysis feature and the further analysis feature in the module embody the advantage of the AuvTool over the other commercial software packages.

This module is very similar to the *variability analysis-fitting distribution module*. The main difference is that the former analyzes the sampling data of statistics from bootstrap simulation for a chosen variable, and uses a parametric distribution model to represent the uncertainty for a statistic, while the later focus on characterizing the variability of a variable based on an original data set using a distribution model. Another difference is that this module also has a feature that can automatically help users to choose a best fit to the sampling distribution data of a statistic; while the *variability analysis-fitting distribution* module does not. Like the *variability analysis-fitting distribution* module, the module also allows users to choose different distribution types and different parameter estimation methods when they analyze a statistic for a selected variable or data set. The choices made via the module will be used to construct the uncertainty analysis summary table in *the uncertainty analysis result- reporting* module.

3.4.8 Variability and Uncertainty Reporting Analysis Modules

The purposes of the *Variability and Uncertainty Reporting Analysis* modules are to report the variability and uncertainty analysis results in a tabular form and to facilitate export of the results to other application programs such as Microsoft Excel. The *variability analysis result-reporting* module summarizes the variability analysis results from the *variability analysis-fitting distribution* module or *batch analysis* modules. These results include the summarization of the variable or data set names analyzed, the number of data points for each variable or data set, the distribution types representing variability, the corresponding distribution parameters, the parameter estimation methods, and the K-S and A-D test results. For the beta, uniform and

symmetric triangle distributions, the A-D test is not available, and the corresponding cells will be marked “NA”.

The *uncertainty analysis result-reporting* module summarizes the 95 percent confidence intervals for the mean, standard error, and the variable or data set names analyzed, the number of bootstrap replication for each variable or data set, the distribution types fitted to the sampling distributions of the statistics of mean, standard error and distribution parameters, and the K-S and A-D statistical test results for those distributions. The module also reports all pair-wise sampling data combinations of all possible statistics and the correlation coefficients between all statistics.

4.0 VERIFCATION OF AUVTOOL

A key task of this project was a comprehensive evaluation and refinement of AuvTool. The evaluation was based upon extensive testing of the software by several persons who were not involved in the development of the software. These persons constituted a verification test team. The team was comprised of three people. The testing was done according to a predefined testing plan.

The scope of the testing included: (1) evaluation of the Graphical User Interface, including identification of any instabilities or errors associated with the interface; (2) verification of methods for input and output of data worked correctly; (3) verification of algorithms for generating random numbers from each type of distribution (i.e. normal, lognormal, gamma, Weibull, beta, uniform, symmetric triangle, uniform, and empirical); (4) verification of algorithms for estimating the parameters of the parametric distribution, including the maximum likelihood estimators and the method of matching moment estimators; (5) verification of the results of bootstrap simulations for confidence intervals of the mean and, in selected cases, confidence intervals for the standard deviation; (6) verification of algorithms for the goodness-of-fit tests, with a focus on the K-S test; and (7) verification of the numerical stability of the bootstrap simulation results.

Any problems that were identified by the verification test team were reported to the individual responsible for software development, and new versions of the program were provided that corrected the identified problems. The program was found to perform well in terms of the graphical user interface and regarding the input and output of data.

This chapter focuses on documentation of items (3) through (7) in the list above. These items pertain to random number generation, parameter estimation, evaluation of bootstrap

simulation results for selected confidence intervals for selected statistics, goodness-of-fit testing, and the numerical stability of bootstrap results. The details of the specific testing approach for each of these five items are described in the following sections.

4.1 Verification of Random Number Generation for Probability Distribution Models

AuvTool includes algorithms for generating random numbers from specified probability distribution models, including the normal, lognormal, gamma, Weibull, beta, uniform, symmetric triangle, and empirical distributions. The objective of the first test applied to AuvTool was to verify that the algorithms for generating random numbers from a specified probability distribution perform correctly. The random number generation algorithms are a key component of AuvTool. For example, they are the basis for performing bootstrap simulation.

The evaluation of the performance of the random number generation algorithms was based upon the following approach:

1. Specify an assumed population distribution. In the case of the parametric distribution, this required specification of parameter values. In the case of the empirical distribution, this was done by specifying a data set.
2. Generate 1,000 random numbers from each distribution.
3. Graphically compare the 1,000 random numbers from Step 2 with the respective specified distribution from Step 1.
4. Perform a quantitative statistical goodness-of-fit test in which the random samples are compared to the specified distribution.
5. Make a finding as to whether the randomly generated numbers are an acceptable sample from the specified distribution.

The results of each of these steps are described in the following subsections.

4.1.1 Specifying Parameters and Generating Random Numbers

For each of the seven parametric probability distribution models, two parameters have to be specified. In testing the random number generators for each of the parametric distributions, it was desired to include test cases with different ranges of variability in order to evaluate the

robustness of the random number generator with respect to relative variation. Relative variation was quantified based upon the coefficient of variation (CV), which is the standard deviation divided by the mean.

For each of the seven parametric distributions, parameters were specified for three coefficients of variation, as summarized in Table 4-1. In most cases, the coefficients of variation used were 0.5, 1.0, and 2.0. These three values of CV were used for the normal, lognormal, gamma, Weibull, uniform, and symmetric triangle distributions. For these same six distributions, the assumed mean in each test case was 1.0. The parameters of the distributions associated with a mean of 1.0 and the assigned CV were calculated using MoMM. The MoMM parameter estimation method is described in detail in Chapter 2. The calculated parameter values are given in Table 4-1. In an independent check by members of the study team who were not involved in writing Chapter 2 or in coding the AuvTool software, the definitions of the probability distribution models and of the parameter estimation equations were verified by comparison to published information in sources such as Morgan and Henrion (1990).

The values of CV used for the beta distribution were 0.2, 0.5, and 0.8. These values were chosen because the two-parameter beta distribution is defined only for a finite range of values from 0 to 1. Therefore, the standard deviation for a beta distribution is not unbounded. A mean value of 0.5 was selected. A judgment was made that the highest value of CV to use for test purposes should be 0.8.

For the empirical distribution, one test case was evaluated based upon an arbitrarily generated data set. There were three test cases for each of seven parametric distributions, resulting in a total of 21 test cases for the parametric distributions and 22 test cases for all distributions, including the empirical distribution test case.

Table 4-1. Parameters of the Tested Distributions.

Distribution[a]	CV[b]	Parameter1[c]	Parameter2[c]
Normal	0.5	1	0.5
	1	1	1
	2	1	2
Lognormal	0.5	-0.112	0.472
	1	-0.347	0.833
	2	-0.804	1.269
Gamma	0.5	4	0.25
	1	1	1
	2	0.25	4
Weibull	0.5	2.101	1.129
	1	1	1
	2	0.543	0.575
Beta	0.2	12	12
	0.5	1.5	1.5
	0.8	0.281	0.281
Uniform	0.5	-0.5	2.5
	1	-2	4
	2	-5	7
Symmetric Triangle	0.5	1	1.225
	1	1	2.450
	2	1	4.899

[a] For normal, lognormal, gamma, Weibull, , uniform, symmetric triangle distributions, mean equals 1. For beta distribution mean equals 0.5.
[b] CV = coefficient of variation, which is the standard deviation divided by the mean.
[c] Parameter definition are given in Chapter 2, Section 2.2.

For each test case, 1,000 random numbers were generated using AuvTool. This was done by using the Random Sample Generator feature of AuvTool described in Chapter 7 of the User's Guide (Zheng and Frey, 2002). In the case of parametric distributions, random number generation was done by specifying the type of parametric distribution and the numerical values of the parameters, obtained from Table 4-1, and executing AuvTool to generate 1,000 random values in each case. For the empirical distribution, a data set was entered to specify the distribution. AuvTool was used to generate 1,000 random values from the specified empirical distribution.

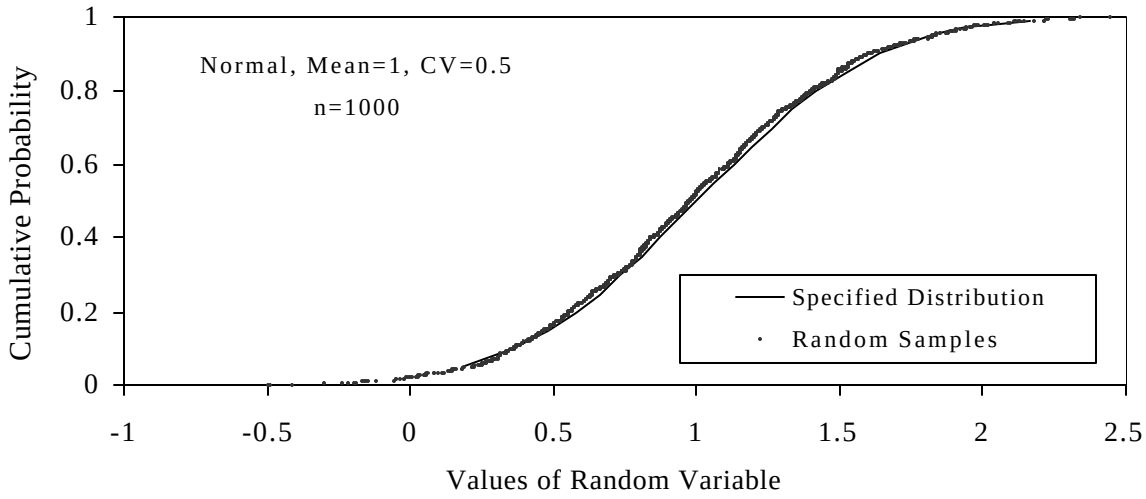


Figure 4-1. Comparison of Random Samples and Specified Normal Distribution, μ=1, σ^2=0.25

4.1.2 Graphical Comparison of Random Numbers with Specified Distribution

The objective of graphical comparison is to visualize the degree of match between the random samples and the corresponding specified distribution. An example of a graphical comparison is given in Figure 4-1 for the case of a normal distribution with a mean of one and a CV of 0.5. The 1,000 random numbers generated from AuvTool are shown as dots. These 1,000 random numbers were plotted using the Hazen plotting position described in Chapter 2. Because there are so many data points, they are very close together and appear as if they are a continuous line except at the ends of the lower and upper tails. The CDF of the normal distribution is shown as a solid line. The CDF was plotted using a function for the normal distribution available in Microsoft Excel™. Some discrepancy between the random samples and the CDF of the specified distribution is expected because of random variation within a data set of a finite sample size. The graphical comparison implies that there is good agreement between the sample of 1,000 random numbers and the specified distribution in this case.

Graphical comparisons for all of the other test cases are given in Appendix A. For all of the distributions, including normal, lognormal, gamma, Weibull, beta, uniform, symmetric triangle, and empirical, there is good agreement between the random sample of 1,000 values and the CDF of the specified distribution. Thus, the graphical comparison provides a qualitative indication that the random number generators are performing properly.

4.1.3 Goodness-of-Fit Tests for Random Numbers

To augment the graphical comparison of random numbers generated from AuvTool and the specified distributions, quantitative goodness-of-fit tests were conducted. The K-S test is applicable to evaluating goodness-of-fit for continuous distributions. Therefore, it was used to assess the correspondence between the 1,000 random numbers and the specified distribution for each case pertaining to the normal, lognormal, gamma, Weibull, beta, uniform, and symmetric triangle distributions. The chi-squared test was used to evaluate the goodness of fit of the generated random numbers compared to the empirical distribution. The chi-squared test was used instead of the K-S test in the case of the empirical distribution because the empirical distribution is discrete. Therefore, strictly speaking, the K-S test should not be used to evaluate goodness-of-fit for the empirical distribution. The details of the K-S test are described in Chapter 2. The verification team reviewed the description of the K-S test in Chapter 2 and verified that it was consistent with information available in the literature.

The results for evaluation of the goodness-of-fit of the specified distributions and the randomly generated numbers are summarized in Table 4-2 for all of the parametric distributions. The critical value of the K-S test at the α level for a sample size of n is expressed as D_n^a . In this work, the critical level α is set to be 0.05 and the sample size is 1,000. Therefore, the critical value used was $D_{1000}^{0.05}$ = 0.0430. The test is passed if the calculated value of the test statistic is

Table 4-2. Summary of the Results of K-S Test with critical value of $D_{1000}^{0.05}$ = 0.0430.

Distribution	Mean	CV	D[a]
Normal	1	0.5	0.0286
		1	0.0256
		2	0.0317
Lognormal	1	0.5	0.0286
		1	0.0273
		2	0.0337
Gamma	1	0.5	0.0238
		1	0.0254
		2	0.0254
Weibull	1	0.5	0.0320
		1	0.0250
		2	0.0292
Beta	0.5	0.2	0.0314
		0.5	0.0260
		0.8	0.0220
Uniform	1	0.5	0.0207
		1	0.0375
		2	0.0307
Symmetric Triangle	1	0.5	0.0320
		1	0.0320
		2	0.0320

[a] All tests were passed for n=1000, α=0.05, compared to a critical value of $D_{1000}^{0.05}$ = 0.0430.

less than the critical value. As shown in Table 4-2, the calculated values of the test statistics are less than 0.0430 in all cases. The largest test statistic value was 0.0375 in the case of the uniform distribution with CV=1. In most cases, the value of the test statistic was less than 0.03. Based upon the results of the goodness-of-fit tests, the hypothesis that the data are a random sample from the specified distribution cannot be rejected. Therefore, the random number generator is found to perform properly for the parametric distributions.

Although the details of the K-S test are mentioned in Chapter 2, the chi-squared test is only briefly introduced. Therefore, more detail regarding the chi-squared test is provided here. The chi-squared test was used for the empirical distribution. The chi-squared test involves calculating a test statistic that approximately follows a chi-square distribution only if the

hypothesized model cannot be rejected as a poor fit to the data. The chi-squared test includes grouping the values in to cells in which each cell has at least five data points. The probability of obtaining values within the range of each cell is calculated based on the hypothesized distribution. Then a test statistic is calculated and evaluated. The number of cells that should be used in a matter of judgment. The number of cells that should be used can be estimated as follows:

$$\text{For} \qquad n <\sim 200; \qquad k \le \frac{n}{5} \tag{4-1}$$

$$\text{For} \qquad n >\sim 200; \qquad k = \text{integer}\{4 \times [0.75 \times (n-1)^2]^{0.2}\} \tag{4-2}$$

where: k is the number of cells. A test statistic is computed as follows:

$$X^2 = \sum_{i-1}^{k} \frac{(M_i - E_i)^2}{E_i} \tag{4-3}$$

where: M_i is the number of data values in each cell: and E_i is the expected number of data values in each cell. The test is passed if the following condition is met (Ang and Tang, 1975):

$$\sum_{i-1}^{k} \frac{(M_i - E_i)^2}{E_i} < X^2_{1-\alpha,k-1} \tag{4-4}$$

where: $X^2_{1-\alpha,k-1}$ is the value the chi square distribution, X^2_{k-1}, at a cumulative probability (1- α).

For the test of the random number generator for the empirical distribution, a data set of n=1,000 that was simulated from a uniform distribution with a minimum of one and a maximum of 10 was used to specify an empirical distribution in AuvTool. AuvTool was used to generate 1,000 random numbers from the specified empirical distribution. The chi-square test was used to test if the simulated samples obtained from AuvTool and the original samples that were input to AuvTool were from the same distribution.

Table 4-3. Procedure of the Chi-Squared test

C^a	Range of Data		E_i^b	M_i^c	D_i^d	C^a	Range of Data		E_i^b	M_i^c	D_i^d
	L	U					L	U			
1	1	1.15	14	11	0.643	31	5.5	5.65	23	26	0.391
2	1.15	1.3	24	21	0.375	32	5.65	5.8	16	16	0.000
3	1.3	1.45	8	8	0.000	33	5.8	5.95	15	18	0.600
4	1.45	1.6	22	15	2.227	34	5.95	6.1	20	20	0.000
5	1.6	1.75	17	17	0.000	35	6.1	6.25	12	16	1.333
6	1.75	1.9	12	11	0.083	36	6.25	6.4	16	16	0.000
7	1.9	2.05	16	21	1.563	37	6.4	6.55	13	15	0.308
8	2.05	2.2	22	20	0.182	38	6.55	6.7	9	10	0.111
9	2.2	2.35	17	19	0.235	39	6.7	6.85	17	17	0.000
10	2.35	2.5	16	12	1.000	40	6.85	7	13	11	0.308
11	2.5	2.65	14	12	0.286	41	7	7.15	19	26	2.579
12	2.65	2.8	22	20	0.182	42	7.15	7.3	17	8	4.765
13	2.8	2.95	21	20	0.048	43	7.3	7.45	15	19	1.067
14	2.95	3.1	25	18	1.960	44	7.45	7.6	15	17	0.267
15	3.1	3.25	13	14	0.077	45	7.6	7.75	19	11	3.368
16	3.25	3.4	14	14	0.000	46	7.75	7.9	15	16	0.067
17	3.4	3.55	22	24	0.182	47	7.9	8.05	14	8	2.571
18	3.55	3.7	22	18	0.727	48	8.05	8.2	7	6	0.143
19	3.7	3.85	17	21	0.941	49	8.2	8.35	20	21	0.050
20	3.85	4	18	22	0.889	50	8.35	8.5	17	17	0.000
21	4	4.15	15	15	0.000	51	8.5	8.65	13	14	0.077
22	4.15	4.3	23	26	0.391	52	8.65	8.8	15	22	3.267
23	4.3	4.45	13	8	1.923	53	8.8	8.95	21	21	0.000
24	4.45	4.6	18	21	0.500	54	8.95	9.1	14	15	0.071
25	4.6	4.75	15	11	1.067	55	9.1	9.25	8	10	0.500
26	4.75	4.9	10	6	1.600	56	9.25	9.4	22	26	0.727
27	4.9	5.05	25	23	0.160	57	9.4	9.55	24	22	0.167
28	5.05	5.2	16	20	1.000	58	9.55	9.7	14	19	1.786
29	5.2	5.35	15	12	0.600	59	9.7	9.85	21	20	0.048
30	5.35	5.5	16	14	0.250	60	9.85	10	14	23	5.786
Total Chi-Squared test value = sum of D_i = 49.45											

[a] Cell number.
[b] Number of original data sets in each cell.
[c] Number of simulated data sets in each cell.
[d] $D_i = \frac{(M_i - E_i)^2}{E_i}$

In setting up the chi-square test, the number of cells was calculated to be 60 based upon Equation (4-2). Both the original and simulated data sets were divided into 60 cells, as shown in

Table 4-4. Chi-Squared Test for Random Samples and Specified Empirical Distribution

Hypothesis H_0:	Simulated sample comes from the original empirical distribution
Test statistic:	49.45
Degrees of freedom:	59
Significance level:	0.05
Critical value[a]:	77.93
Conclusions:	Accept H_0

[a.] Calculated by MATLAB function chi2inv (0.95, 59).

Table 4-3. The endpoints of each cell were determined based upon dividing the domain of the distribution into equal intervals. Although the width of each cell was the same, there was random fluctuation regarding the number of data points in each cell both for the original and for the simulated data. The test statistic was calculated by summing the values obtained for each cell for the quantity D_i, as shown in Table 4-3. The test statistic was found to be 49.45. In contrast, the critical value for the test statistic, based upon 59 degrees of freedom and a significance level of 0.05, was found to be 77.93. The critical value was estimated using the chi2inv function in MATLAB. A summary of the test results is given in Table 4-4. Because the value of the test statistic is less than the critical value, the empirical distribution specified in AuvTool cannot be rejected as an inappropriate fit to the data generated by AuvTool. Therefore, the random number generator for the empirical distribution was found to perform properly.

4.1.4 Summary of Results for Verification of the Random Number Generator

Based upon graphical comparisons and quantitative statistical goodness-of-fit tests, the random number generators for the normal, lognormal, gamma, Weibull, beta, uniform, symmetric triangle, and empirical distributions were verified to perform properly. Therefore, AuvTool was demonstrated to correctly simulate random numbers for these distributions.

4.2 Verification of Parameter Estimation Algorithms

The objective of the work related to verification of parameter estimation algorithms was to verify that the algorithms for parameter estimation incorporated into AuvTool work correctly.

In order to verify the parameter estimation algorithms, it is necessary to use both AuvTool and independent software and/or calculations to estimate parameters based upon the same data sets. The parameter estimates obtained from AuvTool and the independent methods were then compared. If the results were the same or sufficiently similar, then the results of AuvTool were deemed to be correct. A key assumption in doing such comparisons is that the correct algorithms are used in the independent software and/or calculations. A key limitation in identifying software to use for comparison purposes is that the parameter estimation algorithms are often not well-documented. For example, two popular software tools for probabilistic analysis are Crystal Ball and @Risk. However, in many cases, the definition of the PDF and the parameters is not given, and/or information regarding the algorithm used for parameter estimation is not given. Therefore, it was not possible to rigorously compare parameter estimation results from Crystal Ball and @Risk with those from AuvTool.

4.2.1 Method for Verification and Comparison of Parameter Estimates

As documented in Chapter 2, AuvTool incorporates both MoMM and MLE parameter estimation methods for the parametric probability distribution models, with a few exceptions. The exceptions are that the MLE method is not available for the uniform distribution, and MoMM is not included for the Weibull distribution. With these exceptions, the verification task included development of alternative, independent methods for calculating both MoMM and MLE parameter estimates for selected data sets.

In order to test and verify the implementation of the parameter estimation algorithms given in Chapter 2, several steps were followed. The first was to verify the correctness of the equations given in Chapter 2 by reviewing the literature. This work was done by the independent verification team, and not by the persons who implemented the software or who wrote Chapter 2. The equations reported in Chapter 2 were verified to be accurate. A second

step was to develop an independent method for calculating the parameters of distributions based upon the algorithms given in Chapter 2. The approach chosen for the independent calculations was to perform parameter estimation using Excel spreadsheets.

In the case of MoMM, the MoMM estimators were programmed into an Excel spreadsheet and parameters were estimated for several sets of test data. In the case of MLE, two different types of calculations were done. For the normal and lognormal distributions, analytical solutions for the MLE estimator are given in Chapter 2 and these were implemented into Excel. For the gamma, Weibull, beta, and symmetric triangle distributions, the log-likelihood function was entered into Excel. The MLE parameter estimates were obtained by optimization. The "Solver" in Excel was used in finding the optimal solution for the parameter estimates. Specifically, the maximum value of the log-likelihood function was found by varying the two parameters using the Solver. The log-likelihood functions for the gamma, Weibull, beta, and symmetric triangle distributions are given here:

<u>Log-likelihood function for the gamma distribution</u>:

$$J(\alpha,\beta) = -n\{\alpha \ln(\beta) + \ln[\Gamma(\alpha)]\} + \sum_{i=1}^{n}\left\{(\alpha-1)\ln(x_i) - \frac{x_i}{\beta}\right\} \tag{4-5}$$

where:

n	= the number of data points
J	= Log-likelihood function
G(a)	= the *gamma function* of *a*,
a, b	= the parameters in the gamma distribution.

<u>Log-likelihood function for the Weibull distribution</u>:

$$J(\alpha,\beta) = -n\ln(\frac{\alpha}{\beta}) + \sum_{i=1}^{n}\left\{(\beta-1)\ln(\frac{x_i}{\alpha}) - (\frac{x_i}{\alpha})^{\beta}\right\} \tag{4-6}$$

where:

a, b = the parameters in the Weibull distribution.

<u>Log-likelihood function for the beta distribution</u>:

$$J(\alpha,\beta) = n \ln \frac{1}{B(\alpha,\beta)} + \sum_{i=1}^{n} [(\alpha - 1)\ln x_i + (\beta - 1)\ln(1 - x_i)] \tag{4-7}$$

where:

B = Beta function,

a, b = the parameters in the beta distribution.

<u>Log-likelihood function for the symmetric triangle distribution</u>:

$$L(a,b) = \prod_{i=1}^{n} \frac{b - |x_i - a|}{b^2} \tag{4-8}$$

where:

a, b = the parameters in the symmetric triangle distribution.

The calculations that were performed in Excel are referred to as "Manual" calculations because it was necessary for the verification team to enter the data and the estimation algorithms into Excel, and to manually execute the Solver in the cases of the optimization solution.

In addition to the use of Excel as a platform for performing independent calculations of parameter values, an attempt was made to use both Crystal Ball and @Risk. Both programs have a capability to estimate parameter values from data sets. However, as previously noted, these two programs are not sufficiently documented with regard to the definition of the PDF or of the parameter estimation methods employed for a particular distribution. Although AuvTool uses definitions of the PDF that are found in the literature, there are alternative ways to define many distributions that in turn lead to different parameter definitions and different numerical values of the parameters. For example, the parameters of the lognormal distribution can be defined as the arithmetic mean and standard deviation of log-transformed data, as is done in AuvTool. Alternatively, one could define the lognormal distribution based upon the geometric mean and geometric standard deviation, or based upon the arithmetic mean and arithmetic standard

deviation of untransformed data, or other approaches (e.g., Small, 1990). Similarly, the parameters of the gamma distribution can be defined in a variety of ways. In the absence of knowledge of the actual definitions used in Crystal Ball and @risk, it is quite likely that differences in parameter estimates obtained with these two software in comparison to AuvTool could be because of different definition.

Even if the definition of the PDF and the parameters of a distribution is the same in different software tools, the numerical values obtained for the parameters can be different if different parameter estimation methods are used. Although AuvTool contains two of the most common parameter estimation algorithms, there are other methods that may have been used in software such as Crystal Ball and @Risk for parameter estimation. For example, probability plotting methods are sometimes used by practitioners to estimate parameters, although this approach is not recommended (e.g., Cullen and Frey, 1999). Thus, without knowledge of the specific parameter estimation used in Crystal Ball and @Risk, it is possible that any differences compared to AuvTool could be because of different parameter estimation methods, and not because of an error in AuvTool. Therefore, Crystal Ball and @Risk could not be used to verify AuvTool.

Even though Crystal Ball and @Risk are poorly documented with regard to the definitions of the PDF, the definitions of the parameters, and/or the specifics of the parameter estimation algorithms used for each of the parametric distributions tested, a choice was made to include comparisons of Crystal Ball and @Risk with AuvTool were possible. The reason for doing so is that many practitioners commonly use Crystal Ball and @Risk. Therefore, if a practitioner were to estimate parameters for a given data set with either of these two tools and also with AuvTool, there may be situations in which different results would be obtained because

of undocumented differences in definitions and algorithms. Thus, it was deemed useful to identify situations in which such differences occurred even though the differences could not be explained at this time. It is recommended that developers of commercial software should properly document the definitions of PDFs, parameters, and algorithms used for parameter estimation.

Because the random number generator was verified in Section 4.1, the random number generator of AuvTool was used to generate data sets of sample sizes 10, 20, and 50 for use in the parameter estimation algorithm testing. These data sets are given in Appendix B.

4.2.2 Results of Parameter Calculations

The results of calculations for parameter estimates from AuvTool, manual calculations performed in Excel, and @Risk and Crystal Ball (where applicable) are given in Tables 4-5 through 4-11 for the normal, lognormal, gamma, Weibull, beta, uniform, and symmetric triangle distributions, respectively. Each table shows the results obtained with three different test data sets of sample sizes 10, 20, and 50.

For Crystal Ball, definitions of the PDF and of the parameters are not given in either the user manual or help files. For @Risk, some parameter definitions are the same as those used in AuvTool, while the others are different from AuvTool. The comparisons of the PDF definitions in @Risk are discussed case by case in the following paragraphs. However, for all cases, there is no documentation of the parameter estimation methods used in @Risk or Crystal Ball. Therefore, for all cases, it is not known whether @Risk and Crystal Ball used MoMM or MLE estimates, or whether other methods are used instead (e.g., probability plotting, approximate solutions for MLE, etc.).

The results for the normal distribution in Table 4-5 include parameters estimated using MoMM in AuvTool and manually using Excel as the calculation platform. The results of these

Table 4-5. Estimated Normal Parameters

Method	Software	Sample size n=10		Sample size n=20		Sample size n=50	
		μ	σ	μ	σ	μ	σ
MoMM[a]	AuvTool	0.241	0.683	1.067	0.522	1.036	2.007
	Excel	0.241	0.683	1.067	0.522	1.036	2.007
MLE[b]	AuvTool	0.241	0.648	1.067	0.509	1.036	1.987
	Excel	0.241	0.648	1.067	0.509	1.036	1.987
@Risk		0.241	0.683	1.067	0.522	1.036	2.007
Crystal Ball		0.83	0.25	1.20	1.44	1.10	2.06

[a] Method of Matching Moment.
[b] Maximum Likelihood Estimates.

two sets of parameter estimates are identical. Therefore, it is verified that the parameter estimation equations for the mean and standard deviation are implemented correctly in AuvTool and that they provide the correct MoMM estimates. Similarly, the same parameter values were obtained for MLE using AuvTool and the manual calculation method implemented in Excel.

Because MoMM and MLE are based upon different approaches to parameter estimation, the values for the standard deviation parameter for a given data set are not the same when comparing the two methods. It is expected that MLE will give a different estimate for the standard deviation than does MoMM. In the case of MoMM, the standard deviation of the distribution will be the same as that of the data. In the case of MLE, the standard deviation of the distribution is associated with the best fitting distribution that maximizes the likelihood function based upon the observed data values. Thus, the distribution estimated using MLE will have central moments different from that of the original data.

For comparison purposes, parameter estimates obtained from @Risk are shown. @Risk has the same PDF definition as the one used in AuvTool and also reported the same result as the AuvTool MoMM estimates. Therefore, although not documented, it is likely that @Risk is using the same MoMM parameter estimation method as AuvTool. Parameter values obtained from

Table 4-6. Estimated Lognormal Parameters

Method	Software	Sample size n=10		Sample size n=20		Sample size n=50	
		μ_{lnx}	σ_{lnx}	μ_{lnx}	σ_{lnx}	μ_{lnx}	σ_{lnx}
MoMM	AuvTool	-0.233	0.296	-0.270	0.947	-0.664	1.229
	Excel	-0.233	0.296	-0.270	0.947	-0.664	1.229
MLE	AuvTool	-0.233	0.306	-0.235	0.847	-0.782	1.261
	Excel	-0.233	0.306	-0.235	0.847	-0.782	1.261
@Risk		Reject[a]	Reject[a]	-0.401	0.981	-0.505	1.106
Crystal Ball[b]		0.83	0.26	0.66	0.38	1.01	2.00

[a] Rejected the distribution as a candidate fit.
[b] The definition of the parameters of the lognormal distribution in Crystal Ball is not available.

Crystal Ball are based upon unknown parameter definitions and an unknown parameter estimation algorithm, and they do not correspond to any of the other values shown in Table 4-5.

For the lognormal distribution, manual calculations performed in Excel verified that AuvTool correctly calculated the two parameters using both MoMM and MLE. Parameter values calculated in @Risk and Crystal Ball were different than those obtained from AuvTool. Since the parameter definitions and algorithms used in these two softwares were not sufficiently documented, the basis for the difference could not be determined. @Risk would not estimate parameters for the smallest data set. @Risk has a feature to automatically choose a "best" fit distribution, and it does not provide an option to allow a user to arbitrarily fit a distribution to a dataset. Therefore, it was not possible to estimate the parameters for a lognormal distribution for the data set with n=10.

For the gamma distribution, the results in Table 4-7 demonstrate exact agreement between AuvTool and the manual calculations performed using Excel for the MoMM parameter estimates. For the MLE parameter estimates, the results from AuvTool and from the calculations in Excel were similar in all cases and were identical in some cases. The reason for the small differences (e.g., AuvTool estimate of 0.695 for the second parameter versus a manually calculated value of 0.694) can be attributed to the fact that a numerical optimization method was

Table 4-7. Estimated Gamma Parameters

Method	Software	Sample size n=10		Sample size n=20		Sample size n=50	
		á	â	á	â	á	â
MoMM	AuvTool	0.580	1.535	0.956	1.16	1.11	0.763
	Excel	0.580	1.535	0.956	1.16	1.11	0.763
MLE	AuvTool	0.852	1.05	0.948	1.17	1.22	0.695
	Excel	0.886	1.01	0.948	1.17	1.22	0.694
@Risk		Reject	Reject	Reject	Reject	1.141	0.738
Crystal Ball[a]		0.05	1.98	0.07	2.06	0.01	0.93

[a] The definition of the parameters of the gamma distribution in Crystal Ball is not available.

used in both AuvTool and in Excel. The numerical optimization method has a tolerance within which convergence is assumed. The tolerance can lead to some differences in the value of the log-likelihood function, and in the parameter values, at which the numerical method converges on a solution. The differences in the results for AuvTool and the manual calculation results are relatively small. Therefore, it is concluded that the AuvTool MLE parameter estimates were verified.

For comparison purposes, gamma distribution parameter estimates obtained from @Risk and from Crystal Ball are shown. @Risk did not select the gamma distribution as the "best" fit for the data sets of sample sizes 10 and 20. Therefore, parameters for a gamma distribution could not be estimated for these cases in @Risk. The parameter values obtained from @Risk are similar to, but not the same as, those obtained by AuvTool and the manual calculation method in Excel. @Risk has the same parameter definitions as AuvTool, but the parameter estimation method is undocumented. It is likely that @Risk uses a different parameter estimation method than the ones available in AuvTool. The parameter estimates from Crystal Ball are substantially different. It is likely that Crystal Ball is using a different definition of the PDF, different definitions of the parameters, and/or a different parameter estimation method.

Table 4-8. Estimated Weibull Parameters

Method	Software	Sample size n=10		Sample size n=20		Sample size n=50	
		á	â	á	â	á	Â
Regression	AuvTool	0.824	0.942	1.10	0.938	0.903	1.09
MLE	AuvTool	0.835	0.889	1.09	0.964	0.885	1.12
	Excel	0.835	0.889	1.09	0.964	0.885	1.12
@Risk		Reject	Reject	Reject	Reject	Reject	Reject
Crystal Ball[a]		0.01	0.76	0.02	1.07	-0.01	0.88

[a] The definition of the parameters of the Weibull distribution in Crystal Ball is not available

For the Weibull distribution, AuvTool uses a regression method based upon probability plotting and MLE as the parameter estimation methods. The MLE results from AuvTool were reproduced exactly in the manual calculations performed using Excel. The MLE results are similar to, but not the same as, those obtained from the regression method. The regression method is not expected to produce the same results, and the differences in results between the regression and MLE methods are considered to be within the range of expected differences. Therefore, the parameter estimation methods for the Weibull distribution are deemed to have been verified. For comparison purposes, results obtained with @Risk and Crystal Ball are shown. @Risk did not choose the Weibull distribution as the "best" fit to the three data sets used; therefore, it was not possible to obtain parameter estimates for the Weibull distribution using @Risk. Crystal Ball produced parameter estimates that were very different from those of AuvTool and the manual calculation. Because documentation of how Crystal Ball calculates the parameter estimates, and regarding the definition of the PDF and the parameters, was not available, it is not possible to explain the differences.

The MoMM and MLE parameter estimates for the beta distribution obtained from AuvTool were verified by comparison to values calculated manually using Excel. The results agreed identically for the MoMM estimates. The results agreed within the tolerance of the numerical optimization methods in the case of the MLE estimates. @Risk provided estimates for

Table 4-9. Estimated Beta Parameters

Method	Software	Sample size n=10		Sample size n=20		Sample size n=50	
		á	â	á	â	á	â
MoMM	AuvTool	2.79	2.09	1.20	1.65	0.98	1.07
	Excel	2.79	2.09	1.20	1.65	0.98	1.07
MLE	AuvTool	2.79	2.27	1.10	1.54	1.12	1.24
	Excel	2.79	2.27	1.09	1.53	1.12	1.24
@Risk[a]		Diff	Diff	Diff	Diff	Diff	Diff
Crystal Ball[b]		1.24	0.24	1.16	1.53	0.86	0.84

[a] The beta distribution used in @Risk is defined based upon four parameters and is not directly comparable to the two-parameter beta distribution used in AuvTool.
[b] The definition of the parameters of the beta distribution in Crystal Ball is not available.

a four parameter beta distribution, instead of for the two parameter beta distribution used in AuvTool. Therefore, a comparison of the results from @Risk with those from AuvTool is not meaningful. Crystal Ball produced different values of the parameters than did AuvTool, although the discrepancy appears to decrease as the same size increases. However, since there is insufficient documentation in Crystal Ball regarding the algorithms used, it is not possible to explain the difference.

For the uniform distribution, only MoMM is used as a parameter estimation method in AuvTool. The results obtained in AuvTool were verified exactly by manual calculations performed in Excel. Results are also shown for parameter values obtained from @Risk and Crystal Ball for the same data sets. While the results are not identical, they appear to become more similar as the sample size increases. The PDF used in @Risk is reported and is the same as that used in AuvTool. The PDF used in Crystal Ball is not reported. Because the parameter estimation methods used in @Risk and Crystal Ball are not documented, no specific explanation can be offered with regard to differences.

For the symmetric triangle distribution, the MoMM and MLE parameter estimates obtained in AuvTool were verified exactly in the manual calculations performed in Excel.

Table 4-10. Estimated Uniform Parameters

Method	Software	Sample size n=10		Sample size n=20		Sample size n=50	
		a (min)	b (max)	a (min)	b (max)	a (min)	b (max)
MoMM	AuvTool	0.219	0.925	-0.0154	0.857	-0.0179	0.974
	Excel	0.219	0.925	-0.0154	0.857	-0.0179	0.974
@Risk		0.00866	0.841	-0.0439	0.992	-0.00348	0.959
Crystal Ball[a]		0.01	0.84	-0.04	0.99	0.00	0.96

[a] The definition of the parameters of the uniform distribution in Crystal Ball is not available.

Table 4-11. Estimated Symmetric Triangle Parameters

Method	Software	Sample size n=10		Sample size n=20		Sample size n=50	
		a	b	a	b	a	b
MoMM	AuvTool	0.572	0.500	0.421	0.617	0.478	0.701
	Excel	0.572	0.500	0.421	0.617	0.478	0.701
MLE	AuvTool	0.528	0.570	0.448	0.574	0.480	0.607
	Excel	0.521	0.504	0.448	0.574	0.480	0.607

Because neither @Risk nor Crystal Ball have a capability to estimate parameters for a symmetric triangle distribution, no comparison with these two programs is shown.

4.2.3 Summary of Parameter Estimation Verification

The parameter estimation methods used in AuvTool for the normal, lognormal, gamma, Weibull, beta, uniform, and symmetric triangle distributions were verified in call cases for which MoMM and/or MLE solutions are included in AuvTool. The primary method for verification was based upon verifying the definitions of the PDFs and parameter estimation functions with respect to published equations in the literature, and performing manual calculations to solve for parameter values for specific test cases. The manual calculations were implemented in Excel. In the case of MoMM parameter estimates, exact agreement was observed between the AuvTool parameter estimates and the manually calculated estimates. In the case of MLE parameter estimates, the agreement with the manually calculated estimates was exact in most cases, and was within the precision of the numerical optimization method in all cases.

The results from AuvTool were compared with popular commercially available software packages, including @Risk and Crystal Ball. The PDFs and parameter definitions in @Risk for the normal, lognormal, gamma, Weibull, and uniform distributions are the same as those used in AuvTool. However, the parameter estimation algorithms used in @Risk are not documented. The same results were obtained from @Risk as from AuvTool using MoMM for the normal distribution. Therefore, in this case, it is likely that @Risk used the same MoMM solution for the parameter estimates as did AuvTool. However, the parameter estimates obtained from @Risk for the lognormal, gamma, and uniform distributions differed from those obtained from AuvTool, although the parameter estimates were often similar in magnitude. These results suggest that @Risk employs parameter estimation methods different than those used in AuvTool. For the Weibull distribution, a direct comparison was not possible because @Risk did not select the Weibull distribution as a best fit to the test data. @Risk does not allow a user to over-ride its choice of a best fit distribution.

For Crystal Ball, there was no documentation available in the user's manual or on-line regarding the definitions used for the PDFs or the parameters, or regarding the parameter estimation algorithms employed. Crystal Ball typically provided parameter estimates that were different from those of both AuvTool and @Risk. The lack of documentation of the technical basis of Crystal Ball should be taken into account by serious practitioners when evaluating alternative software packages to be used for scientific and/or engineering applications.

Overall, the parameter estimation algorithms in AuvTool were verified to perform as intended, and they provided correct parameter estimates for all test cases. The test cases included seven parametric distributions, data sets of three samples sizes, and evaluation of both MoMM and MLE parameter estimation methods. Thus, the testing performed was

comprehensive and thorough. A user can expect that the parameter estimates obtained with @Risk will typically be similar to but not the same as that of AuvTool. A user of Crystal Ball is cautioned that insufficient documentation was available to evaluate the parameter definitions or estimates used, and that results obtained from Crystal Ball were generally not comparable to those obtained from the verified AuvTool software.

4.3 Verification of Confidence Intervals

The objective of this test was to verify the accuracy of confidence intervals estimates by AuvTool using bootstrap simulation for selected statistics, including the mean and standard deviation. The method employed was to compare the AuvTool results with analytical solutions, where available. Bootstrap simulation was used in AuvTool to estimate the 95 percent confidence interval for the mean for each of the following eight distributions: normal, lognormal, gamma, Weibull, beta, uniform, symmetric triangle, and empirical.

For the normal and lognormal distributions, an exact analytical solution for the 95 percent confidence interval of the mean can be calculated. For the other distributions, the 95 percent confidence interval of the mean should asymptotically converge to the exact analytical solution as the standard error of the mean becomes small compared to the mean value. The asymptotic properties of the 95 percent confidence interval for the mean are based upon the central limit theorem. The standard error of the mean is influenced by the variability in the assumed population distribution, or in the data set, and by the sample size. As the variability decreases, and/or as the sample size increases, the 95 percent confidence interval for the mean estimated from bootstrap simulation should converge to the analytical solution.

In addition to evaluating the confidence interval solutions for the mean of all distributions included in AuvTool, an evaluation was made of the solution for the 95 percent confidence

interval for the standard deviation of the normal distribution, for which an analytical solution is available.

The key steps performed in the test cases included the following:

1. Develop data sets to use as test cases. This was done by specifying a population distribution and generating random samples of different sample sizes (e.g., n=10, 20, 50, 1,000) from the specified distribution using AuvTool. The population distributions for all seven parametric distributions were specified for three different coefficients of variation (e.g., CV = 0.5, 1, 2) as given in Table 4-2. In addition, a test case was developed for the empirical distribution. The data sets are documented in Appendix B. An explanation of the special considerations for setting up test cases of the beta distribution is given in Section 4.1.

2. Use AuvTool to fit distributions to the test data sets and to perform bootstrap simulation. Where applicable, obtain results based upon both the MoMM and MLE parameter estimation methods. The key results obtained were the bootstrap means, the 2.5^{th} percentile of the sampling distribution of the mean, and 97.5^{th} percentile of the sampling distribution of the mean.

3. Calculate the analytical solution for the 95% confidence interval of the mean assuming normality.

4. Compare the bootstrap results with the analytical solutions

5. For the test of the confidence interval of the standard deviation for the normal distribution, the steps are the same as above with the exception that results from AuvTool were recorded for the sampling distribution of the standard deviation and were compared with the analytical solution of the confidence interval of the standard deviation.

The results of the comparison of the confidence intervals for the mean are given in Section 4.3.1. Results for the comparison of the confidence intervals for the standard deviation of the normal distribution are given in Section 4.3.2. The key findings of the test of the confidence intervals obtained from AuvTool are given in Section 4.3.3.

4.3.1 Verification of the Confidence Interval for Mean

The two key components of the verification of the confidence intervals for the mean obtained from AuvTool include calculation of the analytical solution and simulation of the confidence interval in AuvTool using bootstrap simulation. The bootstrap simulation method is

described in detail in Chapter 2. The method for calculating the analytical solution of the confidence interval when normality conditions are satisfied and when the distribution parameters are known is briefly summarized here. More information regarding the calculation of confidence intervals can be found in Ang and Tang (1984), Hahn and Shapiro (1967), and Cullen and Frey (1999).

Let x_1, x_2, ….x_n be a random sample of sample size *n* from a normal distribution, N(μ, σ^2). The analytical solution of the confidence interval of the mean can be calculated with the following equation (Casella and Berger, 1990):

$$\mu_{1-\alpha} = (\overline{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}, \overline{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}) \tag{4-9}$$

where:

$1-\alpha$ = the specified confidence level

$t_{\alpha/2}$ = the percentile value of t-distribution with (n-1) degrees of freedom.

$\overline{x}$ = mean of the original samples.

S = standard deviation of the original samples

For a lognormal distribution, we can take ln(x_1), ln(x_2), ….ln(x_n), as samples from a normal distribution, N(μ, σ^2), and get the analytical solution of confidence interval of mean for lognormal distribution as follows:

$$\mu_{1-\alpha} = (\frac{1}{n}\sum_{i=1}^{n}\ln(x_i) - t_{\alpha/2,1-n}\frac{S_{\ln x}}{\sqrt{n}}, \frac{1}{n}\sum_{i=1}^{n}\ln(x_i) + t_{\alpha/2,1-n}\frac{S_{\ln x}}{\sqrt{n}}) \tag{4-10}$$

where

x_i = the original samples, i = 1 to n.

$S_{\ln x}$ = the standard deviation of the variable ln(x)

For other distributions, when the sample size is large enough and/or when the variability in the distribution is small enough, the sampling distribution of the mean asymptotically

approaches a normal distribution based upon the central limit theorem. Thus, it is not expected that the confidence intervals for the mean will agree with the analytical solutions for small sample sizes and/or large variability in the cases of the gamma, Weibull, beta, uniform, symmetric triangle, and empirical distributions. However, it is expected that the solutions for the 95 percent confidence interval for the mean of these distributions will approach the analytical solution as the sample size becomes large and/or as variability in the distribution is reduced. Thus, it is useful to compare the numerical simulation results from AuvTool with the analytical solutions to evaluate whether the asymptotic trend that is expected is actually observed in practice.

For the normal and lognormal distributions, samples with sample sizes of 10, 20, 50 and 1000 were generated for each of the three values of CV that were tested, including CV = 0.5, 1, and 2. It should be noted that the mean and standard deviation of the population distribution from which random samples were generated were specified as per the values given in Table 4-2. However, it is not the case that each random sample has a mean and standard deviation equal to that of the assumed population distribution. Because of the finite sample sizes and because of random fluctuation in the sample values, the mean and standard deviation of each sample will be different than that of the specified population distribution. For the distributions other than the normal and lognormal, samples with sample size of 50 and 1000 were generated. Sample sizes of 10 and 20 were not used for the other distributions because the numerical solutions of the confidence intervals in those cases were not expected to be similar to the normality assumption of the analytical solution.

For the randomly generated data sets used for each test case, AuvTool was used to calculate the 95% confidence interval of the mean based upon bootstrap simulation. Where

Table 4-12. Absolute 95% Confidence Interval of Mean for Normal Distribution

Case Study	n[a]	Mean of Data	Bootstrap Means		2.5th percentile of Mean			97.5th percentile of Mean		
			MoMM	MLE	MoMM	MLE	A[b]	MoMM	MLE	A[b]
μ=1, CV=0.5	10	0.87	0.87	0.87	0.68	0.65	0.66	1.12	1.13	1.08
	20	1.07	1.08	1.05	0.89	0.80	0.84	1.28	1.31	1.30
	50	1.01	1.01	1.01	0.83	0.88	0.87	1.16	1.16	1.15
	10^3	0.98	0.98	0.98	0.96	0.95	0.95	1.01	1.01	1.01
μ=1, CV=1	10	0.74	0.73	0.75	0.35	0.32	0.32	1.14	1.18	1.17
	20	1.13	1.14	1.12	0.69	0.63	0.68	1.57	1.56	1.59
	50	1.02	1.00	1.02	0.70	0.75	0.74	1.30	1.30	1.30
	10^3	0.96	0.97	0.96	0.90	0.90	0.90	1.03	1.03	1.03
μ=1, CV=2	10	0.48	0.45	0.45	-0.43	-0.42	-0.36	1.42	1.28	1.33
	20	1.27	1.32	1.29	0.31	0.39	0.35	2.37	2.23	2.18
	50	1.04	1.04	1.02	0.54	0.45	0.48	1.57	1.58	1.59
	10^3	1.07	1.07	1.07	0.95	0.93	0.94	1.20	1.21	1.19

[a] Sample size.
[b] Analytical solution.

applicable, both MoMM and MLE were used as the basis for the bootstrap simulations. Equation (4-9) was used to calculate the analytical solutions for comparison to the normal, gamma, Weibull, beta, uniform, symmetric triangle, and empirical distributions. Equation (4-10) was used to calculate the analytical solution for comparison to the lognormal distribution. The results from AuvTool and analytical calculations were compared. The results are shown in Tables 4-12 through Table 4-19 for the normal, lognormal, gamma, Weibull, beta, uniform, symmetric triangle and empirical distributions. In the bootstrap simulation process, the number of replications used was B=1,000.

The results for the normal distribution in Table 4-12 reveal generally very good agreement between the boostrap results obtained from AuvTool and the analytical results for the 95 percent confidence interval of the mean for all sample sizes and all values of CV that were tested. For example, in the case of CV=0.5 with n=1,000, the mean values obtained from

Table 4-13. Absolute 95% Confidence Interval of Mean for Lognormal Distribution

Case Study	n[a]	Mean of Data	Bootstrap Means		2.5th percentile of Mean			97.5th percentile of Mean		
			MoMM	MLE	MoMM	MLE	A[b]	MoMM	MLE	A[b]
μ=1, CV=0.5	10	0.83	0.83	0.83	0.67	0.67	0.67	1.00	1.00	0.98
	20	1.08	1.09	1.07	0.84	0.85	0.80	1.39	1.32	1.36
	50	1.01	1.02	1.00	0.87	0.88	0.86	1.19	1.15	1.16
	10^3	0.98	0.98	0.98	0.95	0.95	0.95	1.01	1.01	1.01
μ=1, CV=1	10	0.65	0.65	0.66	0.46	0.45	0.45	0.89	0.94	0.86
	20	1.20	1.20	1.12	0.71	0.75	0.57	1.98	1.72	1.83
	50	1.04	1.04	1.00	0.78	0.77	0.73	1.40	1.30	1.35
	10^3	0.97	0.97	0.96	0.91	0.90	0.91	1.04	1.02	1.03
μ=1, CV=2	10	0.43	0.43	0.45	0.27	0.24	0.24	0.65	0.77	0.63
	20	1.43	1.45	1.21	0.66	0.53	0.15	3.01	2.77	2.70
	50	1.10	1.10	1.02	0.68	0.61	0.53	1.74	1.73	1.67
	10^3	0.95	0.95	0.94	0.85	0.83	0.85	1.06	1.06	1.06

[a] Sample size.
[b] Analytical solution.

Table 4-14. Absolute 95% Confidence Interval of Mean for Gamma Distribution

Case Study	n[a]	Mean of Data	Bootstrap Means		2.5th percentile of Mean			97.5th percentile of Mean		
			MoMM	MLE	MoMM	MLE	A[b]	MoMM	MLE	A[b]
μ=1, CV=0.5	50	1.00	1.00	1.00	0.87	0.87	0.86	1.15	1.14	1.14
	10^3	0.99	0.99	0.99	0.96	0.96	0.96	1.02	1.02	1.02
μ=1, CV=1	50	1.15	1.16	1.15	0.88	0.90	0.85	1.49	1.45	1.45
	10^3	0.98	0.98	0.98	0.92	0.93	0.92	1.04	1.04	1.04
μ=1, CV=2	50	0.99	0.99	0.97	0.27	0.24	0.24	1.62	1.56	1.54
	10^3	0.98	0.98	0.98	0.86	0.86	0.86	1.11	1.10	1.10

[a] Sample size.
[b] Analytical solution.

bootstrap simulation using either MoMM or MLE were the same as the mean value of the data, and the 95 percent confidence interval for the mean was from approximately 0.95 to 1.01 as estimated by all approaches. In general, the most agreement is expected for the smallest CV and the largest sample size, and the least agreement is expected for the largest CV and the smallest sample size. Thus, the results for CV=2 and n=10 are expected to be the worst among all of the cases, on average. In general, the results agreed well for all values of CV and all values of n.

Table 4-15. Absolute 95% Confidence Interval of Mean for Weibull Distribution

Case Study	n[a]	Mean of Data	Bootstrap Means		2.5th percentile of Mean			97.5th percentile of Mean		
			Reg [b]	MLE	Reg [b]	MLE	A[c]	Reg [b]	MLE	A[c]
μ=1, CV=0.5	50	1.00	1.02	1.00	0.87	0.86	0.87	1.18	1.13	1.13
	10^3	1.00	1.00	1.00	0.97	0.97	0.97	1.04	1.03	1.03
μ=1, CV=1	50	0.91	0.90	0.91	0.66	0.67	0.66	1.17	1.19	1.15
	10^3	1.00	1.00	0.99	0.94	0.94	0.93	1.06	1.06	1.06
μ=1, CV=2	50	1.66	1.73	1.66	1.30	1.27	1.30	2.22	2.09	2.03
	10^3	2.06	2.07	2.05	1.96	1.95	1.95	2.18	2.16	2.16

[a] Sample size.
[b] Reg = Regression, and refers to the use of regression analysis as a basis for estimating distribution parameters.
[c] Analytical solution.

Table 4-16. Absolute 95% Confidence Interval of Mean for Beta Distribution

Case Study	n[a]	Mean of Data	Bootstrap Means		2.5th percentile of Mean			97.5th percentile of Mean		
			MoMM	MLE	MoMM	MLE	A[b]	MoMM	MLE	A[b]
μ=0.5, CV=0.2	50	0.51	0.51	0.51	0.48	0.48	0.48	0.54	0.54	0.54
	10^3	0.50	0.50	0.50	0.49	0.49	0.49	0.50	0.50	0.50
μ=0.5, CV=0.5	50	0.54	0.54	0.54	0.48	0.48	0.48	0.61	0.61	0.61
	10^3	0.50	0.50	0.50	0.48	0.48	0.48	0.51	0.51	0.51
μ=0.5, CV=0.8	50	0.51	0.52	0.51	0.41	0.40	0.40	0.68	0.64	0.62
	10^3	0.50	0.50	0.50	0.47	0.47	0.47	0.52	0.52	0.52

[a] Sample size.
[b] Analytical solution.

Table 4-17. Absolute 95% Confidence Interval of Mean for Uniform Distribution

Case Study	n[a]	Mean of Data	Bootstrap Means		2.5th percentile of Mean			97.5th percentile of Mean		
			MoMM	MLE	MoMM	MLE	A[b]	MoMM	MLE	A[b]
μ=1, CV=0.5	50	1.13	1.13	N/A	0.89	N/A	0.89	1.38	N/A	1.37
	10^3	0.99	0.99	N/A	0.94	N/A	0.93	1.04	N/A	1.04
μ=1, CV=1	50	1.30	1.30	N/A	0.76	N/A	0.81	1.80	N/A	1.79
	10^3	1.02	1.02	N/A	0.92	N/A	0.91	1.13	N/A	1.13
μ=1, CV=2	50	0.60	0.59	N/A	-0.28	N/A	-0.28	1.50	N/A	1.49
	10^3	0.88	0.87	N/A	0.66	N/A	0.66	1.10	N/A	1.10

[a] Sample size.
[b] Analytical solution.

Table 4-18. Absolute 95% Confidence Interval of Mean for Symmetric Triangle Distribution

Case Study	n[a]	Mean of Data	Bootstrap Means		2.5th percentile of Mean			97.5th percentile of Mean		
			MoMM	MLE	MoMM	MLE	A[b]	MoMM	MLE	A[b]
μ=1, CV=0.5	50	1.09	1.09	1.10	0.95	0.98	0.95	1.22	1.24	1.22
	10^3	0.98	0.99	0.99	0.95	0.96	0.95	1.01	1.02	1.01
μ=1, CV=1	50	0.92	0.93	0.86	0.69	0.62	0.70	1.15	1.07	1.14
	10^3	1.00	1.00	0.98	0.94	0.92	0.94	1.06	1.04	1.06
μ=1, CV=2	50	1.38	1.38	1.50	0.88	1.02	0.89	1.87	2.07	1.87
	10^3	1.00	1.00	0.97	0.87	0.86	0.87	1.13	1.12	1.12

[a] Sample size.
[b] Analytical solution.

Table 4-19. Absolute 95% Confidence Interval of Mean for Empirical Distribution[a]

Case Study	n[b]	Mean of Data	Bootstrap Means of AuvTool	2.5th percentile of Mean		97.5th percentile of Mean	
				AuvTool	Analytical	AuvTool	Analytical
μ=1, CV=1	50	1.15	1.13	0.86	0.85	1.43	1.45
	10^3	0.98	0.98	0.92	0.92	1.04	1.04

[a] For empirical distribution, because there is no specified parametric distribution, we can choose any random numbers. Here, the same samples with sample size 50 and 1000 are used as gamma distribution for coefficient of variation equals to 1. No parameter estimation is used here.
[b] Sample size.

For example, for all three values of CV, the AuvTool results for n=1,000 agreed well with the analytical solutions.

In general, excellent agreement was found for the results of AuvTool versus the analytical solutions in the case of the lognormal distribution. For n=1,000, the upper and lower bounds of the confidence intervals agreed with the analytical solution typically within two decimal places. For smaller sample sizes and larger variability the results tend to vary from the analytical solution to some extent. For example, for CV=2 with n = 10, the 95 percent confidence interval for the mean was estimated to be 0.27 to 0.65 based upon bootstrap simulation using MoMM, 0.24 to 0.77 based upon bootstrap simulation using MLE, and 0.24 to 0.63 based upon the analytical solution. Thus, the results were similar for the lower bound of the

confidence interval, but differed somewhat for the upper bound. In contrast, as the sample size increased, the concordance between the numerical simulation and the analytical solution improved. For example, for CV=2 and n=1,000, the 95 percent confidence interval for the mean was approximately 0.85 to 1.06 based upon all methods.

The results from AuvTool for the 95 percent confidence interval of the mean agreed well with the analytical solutions for sample sizes of n=50 and n=1,000 for the gamma, Weibull, beta, uniform, symmetric triangle, and empirical distributions. In particular, for n=1,000, the results were typically exactly the same or very close for all distributions and for all values of CV that were tested. Thus, the asymptotic performance of the bootstrap solution for the 95 percent confidence interval for the mean was verified.

4.3.2 Verification of the Confidence Interval for the Standard Deviation of the Normal Distribution

The analytical solution for the confidence interval of the standard deviation of a normal distribution is based upon the chi square distribution. Let x_1, x_2, ….x_n be a random sample from a $N(\mu, \sigma^2)$ distribution. According to Casella and Berger (1990), the confidence interval of standard deviation is given by:

$$\sigma_{1-\alpha} = \left(\sqrt{\frac{(n-1)s^2}{c^2_{1-a/2,n-1}}}, \sqrt{\frac{(n-1)s^2}{c^2_{a/2,n-1}}}\right) \qquad (4\text{-}11)$$

where:

$1-\alpha$ = the specified confidence level,

$c^2_{a/2,n-1}$ = the percentile value of c^2 distribution with (n-1) degrees of freedom at the level of $\alpha/2$.

Using the same random samples for the normal distribution that were used in the test of the confidence interval for mean, the confidence interval for standard deviation was also tested. AuvTool was used to calculate the 95% confidence interval of the standard deviation, using

Table 4-20. Absolute 95% Confidence Interval of Standard Deviation for Normal Distribution

Case Study	N[a]	Std. Dev. of Data	Std. Dev. of Bootstrap Samples		2.5th percentile of Std. Dev.			97.5th percentile of Std. Dev.		
			MoMM	MLE	MoMM	MLE	A[b]	MoMM	MLE	A[b]
μ=1, CV=0.5	10	0.34	0.33	0.30	0.20	0.17	0.24	0.50	0.43	0.62
	20	0.52	0.51	0.49	0.37	0.33	0.40	0.68	0.67	0.76
	50	0.50	0.50	0.49	0.41	0.38	0.42	0.60	0.60	0.63
	10^3	0.49	0.49	0.49	0.47	0.47	0.47	0.51	0.52	0.52
μ=1, CV=1	10	0.68	0.67	0.61	0.35	0.35	0.47	1.01	0.90	1.25
	20	1.04	1.04	0.98	0.74	0.71	0.79	1.38	1.37	1.52
	50	1.00	0.99	0.97	0.81	0.74	0.84	1.18	1.14	1.25
	10^3	0.99	0.99	0.99	0.94	0.94	0.95	1.03	1.03	1.03
μ=1, CV=2	10	1.37	1.31	1.20	0.71	0.76	0.94	1.93	1.73	2.50
	20	2.09	2.09	1.96	1.45	1.46	1.59	2.73	2.68	3.05
	50	2.01	2.02	1.96	1.66	1.63	1.68	2.32	2.38	2.50
	10^3	1.99	1.99	1.99	1.90	1.90	1.91	2.07	2.09	2.09

[a] Sample size.
[b] Analytical solution.

bootstrap simulation with MoMM and with MLE. Equation 4-11 was used to calculate the analytical solution for the 95% confidence interval of standard deviation. The results from AuvTool and the analytical calculation are compared in Table 4-20.

For all cases in which n=1,000, the results from AuvTool agree almost exactly with the analytical solution. For smaller sample sizes, such as n=10, 20, and 50, MoMM typically gives somewhat wider estimates of the confidence interval than does MLE. The lower bound and the upper bound of the confidence interval was generally underestimated to some extent for the smaller sample sizes. These results suggest that the MoMM method gives wider coverage of the confidence interval compared to MLE, which is expected. The results also suggest that the numerical simulation results with the boostrap percentile method may tend to underestimate the analytical results in the case of the confidence interval for the standard deviation when the sample size is relatively small.

4.3.3 Summary of Results for Verification of Confidence Intervals

The verification case studies demonstrate that AuvTool performs well in estimating confidence intervals for the mean. The results agree well with the analytical solution for the normal and lognormal distributions, and results for other distributions asymptotically approach the analytical solution as expected. AuvTool performed well in estimating confidence intervals for the standard deviation of the normal distribution for large sample sizes. For small sample sizes, the MoMM parameter estimation method resulted in more coverage of the actual confidence interval than did the MLE method.

4.4 Verification of Algorithms for Goodness-of-Fit Tests

The objective of this section is to document the verification of the K-S goodness of fit test algorithm that is employed in AuvTool. The procedure used in this task was as follows:

1. Specify a test data set. The test data sets are documented in Appendix B. Three sample sizes of n=10, 20, and 50 were used to evaluate the robustness of results to different sample sizes.
2. For each test data set, estimate the parameters of the distribution using both MoMM and MLE.
3. For each distribution fitted to a given data set, calculate the K-S test statistic in AuvTool.
4. For each distribution fitted to a given data set, calculate the K-S statistic manually.
5. Compare the values of the K-S statistic calculated in AuvTool with those calculated manually.
6. Manually calculate the critical value of the test statistic for a given sample size.
7. Compare the calculated K-S statistic values to the critical value of the K-S statistic.

The algorithm used for the K-S test as documented in Chapter 2 was verified by comparison to the literature by the verification study team. The K-S test is applicable to continuous distributions. Therefore, verification was done for the normal, lognormal, gamma,

Table 4-21. Verification of K-S Test Results for Normal Distribution

Method	C[a]	Sample size n=10			Sample size n=20			Sample size n=50		
		D^b	$D_{10}^{0.05\ c}$	R^d	D^b	$D_{20}^{0.05\ c}$	R^d	D^b	$D_{50}^{0.05\ c}$	R^d
MoMM	AuvTool	0.163	0.258	P	0.148	0.190	P	0.0889	0.125	P
	Manual	0.163		P	0.147		P	0.0889		P
MLE	AuvTool	0.161	0.258	P	0.146	0.190	P	0.0872	0.125	P
	Manual	0.161		P	0.146		P	0.0872		P

[a] Calculation method.
[b] Test statistics.
[c] Critical value at α=0.05 significance level.
[d] Results: P=pass, F=fail.

Table 4-22. Verification of K-S Test Results for Logormal Distribution

Method	C[a]	Sample size n=10			Sample size n=20			Sample size n=50		
		D^b	$D_{10}^{0.05\ c}$	R^d	D^b	$D_{20}^{0.05\ c}$	R^d	D^b	$D_{50}^{0.05\ c}$	R^d
MoMM	AuvTool	0.158	0.258	P	0.147	0.190	P	0.118	0.125	P
	Manual	0.158		P	0.147		P	0.118		P
MLE	AuvTool	0.161	0.258	P	0.146	0.190	P	0.0873	0.125	P
	Manual	0.161		P	0.146		P	0.0873		P

[a] Calculation method.
[b] Test statistics.
[c] Critical value at α=0.05 significance level.
[d] Results: P=pass, F=fail.

Table 4-23. Verification of K-S Test Results for Gamma Distribution

Method	C[a]	Sample size n=10			Sample size n=20			Sample size n=50		
		D^b	$D_{10}^{0.05\ c}$	R^d	D^b	$D_{20}^{0.05\ c}$	R^d	D^b	$D_{50}^{0.05\ c}$	R^d
MoMM	AuvTool	0.154	0.258	P	0.112	0.190	P	0.0949	0.125	P
	Manual	0.154		P	0.112		P	0.0949		P
MLE	AuvTool	0.193	0.258	P	0.112	0.190	P	0.0749	0.125	P
	Manual	0.193		P	0.112		P	0.0749		P

[a] Calculation method.
[b] Test statistics.
[c] Critical value at α=0.05 significance level.
[d] Results: P=pass, F=fail.

Weibull, beta, uniform, and symmetric triangle distributions. The results are shown in Tables 4-21 through 4-27, respectively, for these seven continuous parametric distributions.

Table 4-24. Verification of K-S Test Results for Weibull Distribution

Method	C[a]	Sample size n=10			Sample size n=20			Sample size n=50		
		D[b]	$D_{10}^{0.05}$ [c]	R[d]	D[b]	$D_{20}^{0.05}$ [c]	R[d]	D[b]	$D_{50}^{0.05}$ [c]	R[d]
MoMM	AuvTool	0.174	0.258	P	0.110	0.190	P	0.0818	0.125	P
	Manual	0.174		P	0.110		P	0.0818		P
MLE	AuvTool	0.178	0.258	P	0.113	0.190	P	0.0785	0.125	P
	Manual	0.178		P	0.113		P	0.0785		P

[a] Calculation method.
[b] Test statistics.
[c] Critical value at α=0.05 significance level.
[d] Results: P=pass, F=fail.

Table 4-25. Verification of K-S Test Results for Beta Distribution

Method	C[a]	Sample size n=10			Sample size n=20			Sample size n=50		
		D[b]	$D_{10}^{0.05}$ [c]	R[d]	D[b]	$D_{20}^{0.05}$ [c]	R[d]	D[b]	$D_{50}^{0.05}$ [c]	R[d]
MoMM	AuvTool	0.212	0.258	P	0.124	0.190	P	0.0833	0.125	P
	Manual	0.212		P	0.124		P	0.0833		P
MLE	AuvTool	0.246	0.258	P	0.134	0.190	P	0.0958	0.125	P
	Manual	0.246		P	0.134		P	0.0958		P

[a] Calculation method.
[b] Test statistics.
[c] Critical value at α=0.05 significance level.
[d] Results: P=pass, F=fail.

Table 4-26. Verification of K-S Test Results for Uniform Distribution

Method	C[a]	Sample size n=10			Sample size n=20			Sample size n=50		
		D[b]	$D_{10}^{0.05}$ [c]	R[d]	D[b]	$D_{20}^{0.05}$ [c]	R[d]	D[b]	$D_{50}^{0.05}$ [c]	R[d]
MoMM	AuvTool	1	0.258	F	0.155	0.190	P	0.076	0.125	P
	Manual	0.290	0.258	F	0.161	0.190	P	0.0877	0.125	P

[a] Calculation method.
[b] Test statistics.
[c] Critical value at α=0.05 significance level.
[d] Results: P=pass, F=fail.

For the normal, lognormal, gamma, Weibull, beta, and symmetric triangle distributions, the values of the K-S statistic calculated manually agreed exactly with the values reported by AuvTool. Both AuvTool and the manual calculations resulted in the same decision as to whether the distribution was rejected or not. The values of the K-S statistic typically differed depending on which parameter estimation method was used. This is because the distribution fitted using

Table 4-27. Verification of K-S Test Results for Symmetric Triangle Distribution

Method	C[a]	Sample size n=10			Sample size n=20			Sample size n=50		
		D[b]	$D_{10}^{0.05}$ [c]	R[d]	D[b]	$D_{20}^{0.05}$ [c]	R[d]	D[b]	$D_{50}^{0.05}$ [c]	R[d]
MoMM	AuvTool	0.210	0.258	P	0.125	0.190	P	0.130	0.125	P
	Manual	0.210		P	0.125		P	0.130		P
MLE	Auvtool	0.298	0.258	P	0.159	0.190	P	0.162	0.125	P
	Manual	0.298		P	0.159		P	0.162		P

[a] Calculation method.
[b] Test statistics.
[c] Critical value at α=0.05 significance level.
[d] Results: P=pass, F=fail.

MoMM is not the same as that fitted using MLE. Neither method was consistently better at fitting a distribution to the data, per the results of the K-S test statistic.

The manually calculated results for the K-S statistic for the uniform distribution were different than those obtained from AuvTool, as indicated in Table 4-26. The magnitude of the difference in results appears to decrease as the sample size increases. The exact reason for the difference is not yet known, although it is possible that there may be an implementation error in AuvTool regarding the K-S test statistic for the uniform distribution.

The K-S test procedure used in AuvTool was verified by comparison to an independent software tool. In Crystal Ball, no clear definition was given with regard to how the K-S test is implemented. Therefore, it was decided not to use Crystal Ball for comparison purposes. Instead, for the standard normal distribution, the result of the K-S test statistic from AuvTool was verified by comparison to the "kstest" function in MATLAB. MATLAB only reports a K-S test for the standard normal distribution with specified parameters $m = 0, s = 1$. In order to test the K-S test procedure in AuvTool, the code of the K-S test with a specified standard normal distribution was added and the results were compared. Both programs gave exactly the same results.

Overall, the key findings are that AuvTool correctly implements the K-S test for all of the continuous parametric distributions with the possible exception of the uniform distribution. Future work should be aimed at identifying and correcting the apparent problem in the implementation of the K-S test for the uniform distribution.

4.5 Evaluation of the Stability of Bootstrap Simulation Results

Bootstrap simulation is a numerical method based upon random sample for estimating confidence intervals for selected statistics. Because bootstrap simulation is based upon a finite number of bootstrap samples, there is some random sampling error inherent in any results obtained from this method. Therefore, the objective of this task was to perform a test case in which bootstrap simulation was repeated several times for the same data set and the same number of bootstrap replications. The results of the multiple bootstrap simulations were compared to evaluate the variability in results obtained from one simulation to another. The comparison provides insight into the robustness of the bootstrap results.

The method for performing the evaluation was based upon specifying a random data set as the basis for bootstrap simulation. For this purpose, 20 random values were generated from a gamma distribution with both scale and shape parameters equal to one. The choice of parametric distribution and parameter values was arbitrary. The 20 random values were all copied into each of 10 data columns in the input sheet for AuvTool. Therefore, there were 10 data sets, each with n=20 and each with the same numerical values of the 20 samples. AuvTool was executed to perform one bootstrap simulation, with B=1,000, for each of the 10 identical data sets. The results from each of the 10 bootstrap simulations are summarized in Table 4-28 for the mean value and 95 percent confidence interval of four statistics: mean, standard deviation, first parameter, and second parameter. Each row labeled "1" through "10" in the table represents one of the 10 bootstrap simulation results. In the next to last row of the table, the mean of the 10

Table 4-28. Stability of Bootstrap Simulation Results for the Mean and 95 Percent Confidence Intervals of the Mean, Standard Deviation, and Distribution Parameters[a]

Test No.	Mean			Std. Dev.			1st parameter			2nd parameter		
	M[b]	L[c]	U[d]	M[b]	L[c]	U[d]	M[b]	L[c]	U[d]	M[b]	L[c]	U[d]
1	1.14	0.667	1.80	1.16	0.584	2.09	1.17	0.497	2.30	1.11	0.403	2.54
2	1.12	0.677	1.73	1.15	0.578	2.07	1.16	0.494	2.14	1.09	0.439	2.42
3	1.11	0.652	1.71	1.13	0.587	2.11	1.17	0.510	2.16	1.08	0.436	2.32
4	1.13	0.673	1.71	1.14	0.599	2.01	1.17	0.512	2.31	1.09	0.406	2.38
5	1.10	0.678	1.71	1.12	0.557	2.06	1.17	0.494	2.29	1.08	0.402	2.45
6	1.14	0.664	1.74	1.15	0.584	1.95	1.16	0.493	2.27	1.10	0.419	2.43
7	1.14	0.662	1.71	1.15	0.579	2.05	1.19	0.501	2.26	1.07	0.422	2.39
8	1.12	0.657	1.71	1.13	0.598	1.96	1.19	0.530	2.25	1.07	0.398	2.31
9	1.11	0.691	1.73	1.12	0.550	1.94	1.14	0.513	2.26	1.12	0.410	2.32
10	1.13	0.665	1.72	1.13	0.588	1.95	1.16	0.503	2.25	1.11	0.431	2.57
Average	1.12	0.664	1.73	1.14	0.580	2.02	1.17	0.505	2.25	1.09	0.417	2.41
CV[e]	.013	.017	.016	.012	.027	.032	.013	.023	.025	.016	.034	.037

[a] Results are based upon 10 bootstrap simulations of B=1,000 for a data set of n=20 obtained from a gamma distribution with scale parameter = 1 and shape parameter = 1.
[b] M = the mean value of the statistic.
[c] L = the lower limit of the 95 percent confidence interval, which is the 2.5th percentile of the statistic.
[d] U = the upper limit of the 95 percent confidence interval, which is the 97.5th percentile of the statistic.
[e] the standard deviation of the 10 test runs divided by the average of the 10 test runs.

values given above in the same column is calculated. For example, the average of the data in the second column represents the average of the 10 bootstrap simulation means. The last row of the table displays the coefficient of variation for data in the column above. For example, the coefficient of variation of 0.013 for the mean implies that the standard deviation of the 10 values of the mean estimate of the mean statistic is 1.3 percent the value of the mean estimate.

The coefficients of variation for each statistic estimated from the 10 bootstrap simulation results were less than 0.04 in all cases, and were less than 0.02 in six of the 12 cases given in the table. The 12 cases include the mean, lower confidence bound, and upper confidence bound for

each of four statistics. The mean value of the sampling distribution for the mean varied from 1.10 to 1.14 among the 10 bootstrap simulations. These results imply that some variability in results can be expected from one bootstrap simulation to another. However, the relatively small values of the CV imply that the bootstrap simulation results are precise to within approximately two significant figures in these cases. The precision of the results can be improved by choosing larger values for the number of bootstrap samples (B). Overall, the results among the 10 bootstrap simulations are found to be reasonably stable.

4.6 Overall Findings Regarding Verification of AuvTool

AuvTool was rigorously evaluated with respect to the generation of random numbers, estimation of parameters of distributions using both MoMM and MLE, confidence intervals estimated based upon bootstrap simulation, estimation of the K-S statistic, and the stability of bootstrap simulation results. Factors such as variation in sample sizes and CV were considered in the evaluation of AuvTool. Overall, AuvTool was verified to perform calculations correctly, and the results of AuvTool were found to be robust to different sample sizes and to different CVs. The random sample generation and parameter estimation of AuvTool is correct for all cases. AuvTool provides good estimates of the 95 percent confidence interval of the mean for the normal and lognormal distribution in all cases, and performs as expected for other distributions with regard to asymptotic convergence of the numerical solution to the analytical solution as the sample size increases. The K-S goodness of fit test is implemented correctly and provides correct results in all cases except for the uniform distribution. The bootstrap simulation results were found to be stable for an illustrative test case. The technical basis of AuvTool is more thoroughly documented than many other software packages, such as @Risk and Crystal Ball. Because of the lack of documentation of other software, it was difficult to make meaningful comparisons of results from AuvTool to those of other programs.

5.0 AN ILLUSTRATIVE CASE STUDY USING AUVTOOL

In this chapter, a case study is presented to help illustrate the use of AuvTool. In this case study, the batch analysis feature of AuvTool is demonstrated. Results of analysis of variability and uncertainty are also presented.

5.1 Introduction to Case Study

There are five datasets with original data, which are named as "Dataset 1", "Dataset 2", "Dataset 3", "Dataset 4" and "Dataset 5", and the sample sizes for the data sets are 10, 20, 50, 1000, and 10, respectively. These data represent five different variables. These data are entered into the main sheet of the *data entry, importing and exporting* module. There are also three variables without original data for which specified distribution information is available. The three variables are named as "NoData Name 1", " NoData Name 2" and "No DataName 3", respectively. The corresponding distribution information is summarized in the Table 5-1. The information is provided in the *loading distribution information* module.

5.2 AuvTool Analysis Results

In the example case study, the batch analysis feature in the AuvTool is used. All default settings in AuvTool are kept except that the bootstrap replication numbers for "NoDataset 2", "NoDataset 4" and "NoData Name 2" were modified to 500, 1000, 500, respectively (the default number is 200); and the parameter estimation method was modified to MLE (MoMM is the default method) for "NoDataset 1" and "NoDataset 3". The program automatically chose best fits for the variables with original data and for the sampling distribution data for the statistics of interest. By invoking the *batch analysis* module and modifying the corresponding settings as introduced above, it took a few minutes for the program to report the variability and uncertainty analysis results to the *variability and uncertainty analysis result-reporting* module.

Table 5- 1. Summarization Table of Input Distribution Information for the Variables without Original Data

Variable Name	Sample Size	Distribution Type	First Parameter [a]	Second Parameter [b]	Estimation Method
NoDataName 1	15	Normal	10	5	Moment
NoDataName 2	20	Lognormal	0.5	0.25	MLE
NoDataName 3	25	Gamma	20.5	10	NA

[a.] First Parameter: mean for normal, mean of *lnx* for lognormal, scale parameter for gamma and Weibull, shape parameter for Beta distributions

[b]. Second Parameter: standard deviation for normal, standard deviation of *lnx* for lognormal, shape parameter for gamma, Weibull and beta distributions

Table 5-2. AuvTool Variability Analysis Result Summarization Table

Dataset or Variable Name	No. Of Data Points	Dist. Type	Estimation Method	First [a] Para.	Second [b] Para.	KS Test	AD Test	KS Test Passed	AD test Passed
Dataset 1	10	Weibull	MLE	0.916	3.932	0.150	0.182	Passed	Passed
Dataset 2	20	Lognormal	Moment	-0.072	0.547	0.148	0.465	Passed	Passed
Dataset 3	50	Lognormal	MLE	-0.103	0.469	0.087	0.285	Passed	Passed
Dataset 4	1000	Lognormal	Moment	-0.130	0.473	0.016	0.306	Passed	Passed
Dataset 5	10	Normal	Moment	0.651	0.330	0.137	0.206	Passed	Passed
NoData Name 1	15	Normal	Moment	10	5	NA	NA	NA	NA
NoData Name 2	20	Lognormal	MLE	0.5	0.25	NA	NA	NA	NA
NoData Name 3	25	Gamma	NA	20.5	10	NA	NA	NA	NA

[a.] First Parameter: mean for normal, mean of *lnx* for lognormal, scale parameter for gamma and Weibull, shape parameter for Beta distributions

[b]. Second Parameter: standard deviation for normal, standard deviation of *lnx* for lognormal, shape parameter for gamma, Weibull and beta distributions

Table 5-2 lists the variability analysis results from AuvTool for the eight variables. The table includes the distribution information from the variables without original data. For the latter, since there are no original data available and no fitting was necessary to be done, no goodness-of-fit statistical tests are needed. Therefore, the relevant cells were marked as “NA” in those cases.

Table 5-3 shows the confidence intervals of mean and standard deviation for all of the variables analyzed. The program can also summarize the confidence intervals for

Table 5-3. AuvTool Summarization Table for Confidence Intervals of Mean and Standard Deviation

Dataset Name	Mean 2.5% Percentile	Mean Mean	Mean 97.5% Percentile	Std. Deviation 2.5% Percentile	Std. Deviation Mean	Std. Deviation 97.5% Percentile
Dataset 1	0.665	0.829	0.983	0.121	0.222	0.323
Dataset 2	0.830	1.06	1.33	0.355	0.604	0.994
Dataset 3	0.886	1.01	1.18	0.350	0.491	0.691
Dataset 4	0.949	0.982	1.01	0.454	0.490	0.532
Dataset 5	0.424	0.650	0.855	0.181	0.316	0.484
NoData Name 1	7.87	10.1	12.5	3.25	4.88	7.06
NoData Name 2	1.51	1.70	1.90	0.286	0.421	0.591
NoData Name 3	186	205	223	29.1	44.1	60.8

Table 5-4. AuvTool Summarization Table for the Fitted Distribution to the Sampling Data of Mean Statistic

Variable Name	No.Of B. R. [a]	Dist. Type	Est. Method	First Para.[b]	Second Para. [c]	KS Value	KS Passed	AD Value	AD Passed
Dataset 1	200	Normal	Moment	0.829	0.078	0.039	Passed	0.187	Passed
Dataset 2	500	Lognormal	Moment	0.062	0.125	0.0252	Passed	0.438	Passed
Dataset 3	200	Gamma	Moment	193.	0.005	0.0254	Passed	0.133	Passed
Dataset 4	1000	Lognormal	Moment	-0.019	0.016	0.0203	Passed	0.341	Passed
Dataset 5	200	Beta	Moment	17.3	9.35	0.037	Passed	NA	NA
NoData Name 1	200	Lognormal	Moment	2.31	0.126	0.0441	Passed	0.399	Passed
NoData Name 2	500	Normal	Moment	1.70	0.093	0.0245	Passed	0.253	Passed
NoData Name 3	200	Normal	Moment	205	9.15	0.0435	Passed	0.278	Passed

B.R.[a]. The bootstrap replication number

[b]. First Parameter: mean for normal, mean of *lnx* for lognormal, scale parameter for gamma and Weibull, shape parameter for Beta distributions

[c]. Second Parameter: standard deviation for normal, standard deviation of *lnx* for lognormal, shape parameter for gamma, Weibull and beta distributions

distribution parameters; those results are not shown here. Uncertainty analysis results are available for the variables without original data.

Table 5-4 shows the summarization of results of fitting a parametric distribution to the sampling data for the mean. Similar analysis results are also available for the statistics of standard deviation, distribution parameters, but are not shown here. These results are automatically calculated by AuvTool. As previously mentioned, the normal distribution is often

used to represent the uncertainty in the mean; however, the assumption is only valid if the sampling distribution is from normal distribution and/or the sample size is large enough. The use of bootstrap simulation enables the use of other distribution types to represent the uncertainty in the mean as shown in Table 5-4. The statistical test results show that the distribution types describing the mean for all variables in the Table 5-4 are good representatives of the corresponding means. The compact forms of representing mean will be convenient in propagating the uncertainty from model inputs to model output in other software, such as SHEDS.

It must be pointed out that the result formats shown above are not the ones that the SHEDS model requires. A special variability and uncertainty analysis result output format for the EPA SHEDS model is designed inside AuvTool but not shown here.

6.0 CONCLUSIONS AND RECOMMENDATIONS

This chapter presents key conclusions for the project and offers recommendations for future work.

6.1 Conclusions

The objective of this project was to develop a stand-alone software tool that can conduct statistical analysis of variability and uncertainty associated with fitting probability distributions to data sets for use with the SHEDS modeling framework. Secondary objectives were to develop a tool that would be user-friendly, to develop a tool so that it could be used for general purpose applications, and to verify the new software through extensive testing of its algorithms.

The project succeeded in meeting the objectives that were set forth at the beginning of the project. Specifically, a stand-alone software tool was developed that can conduct statistical analysis of variability and uncertainty associated with fitting probability distributions to data sets. The new software tool, AuvTool, is capable of producing output in a format required by the EPA SHEDS modeling framework. The new software is user friendly. It includes a graphical user interface. A separate user's manual was prepared as part of this project and it is included as on-line help in AuvTool. Because the software is modular and was written as a stand-alone program, it can be used for any applications in which there is a need to fit distributions to data and/or to characterize variability and uncertainty associated with the fitted distribution. The technical basis of the software is documented in detail in this report. As reported in Chapter 4, the software was extensively tested to verify the technical correctness of the algorithms and to prove the stability and utility of the program.

AuvTool was rigorously evaluated with respect to the generation of random numbers, estimation of parameters of distributions using both MoMM and MLE, confidence intervals estimated based upon bootstrap simulation, estimation of the K-S statistic, and the stability of bootstrap simulation results. Factors such as variation in sample sizes and CV were considered in the evaluation of AuvTool. Overall, AuvTool was verified to perform calculations correctly, and the results of AuvTool were found to be robust to different sample sizes and to different CVs. The random sample generation and parameter estimation of AuvTool is correct for all cases. AuvTool provides good estimates of the 95 percent confidence interval of the mean for the normal and lognormal distribution in all cases, and performs as expected for other distributions with regard to asymptotic convergence of the numerical solution to the analytical solution as the sample size increases. The K-S goodness of fit test is implemented correctly and provides correct results in all cases except for the uniform distribution. The bootstrap simulation results were found to be stable for an illustrative test case. The technical basis of AuvTool is more thoroughly documented than many other software packages, such as @Risk and Crystal Ball. Because of the lack of documentation of other software, it was difficult to make meaningful comparisons of results from AuvTool to those of other programs.

6.2 Recommendations

As a result of this project, several recommendations were developed. These recommendations pertain to methods for documenting new software, the appropriate uses of AuvTool, and needs for continued development of AuvTool.

When developing new software that is intended to be used for policy analysis or policy making purposes, a thorough approach to software development, documentation,

and testing should be taken, as demonstrated by this project. Specifically, in addition to developing the program itself, this project resulted in this technical document and a companion user's guide. Furthermore, the technical document includes documentation of extensive testing and verification of the software. The software and the documents should be made publicly available to facilitate review of any policy analyses performed with the aid of the software.

The results of AuvTool were compared to two commercially available programs that include a capability for fitting distributions to data. The commercially available programs did not include sufficient documentation to enable meaningful comparisons. Therefore, a recommendation is that the vendors of such programs should provide more thorough and complete technical documentation of the definitions of parametric distributions and of the parameter estimation methods used.

AuvTool should be used for its intended purpose, taking into account the limitations of the software and its technical basis. AuvTool was designed to fit parametric probability distributions to data and to characterize uncertainty in key statistics based upon random sampling error. Other sources of uncertainty, such as measurement error, are not addressed by the current version of AuvTool.

The user is strongly cautioned to be very careful in using the batch analysis feature of AuvTool. As mentioned both in this technical document and in the user's guide, the batch method for selecting a "best" fitting distribution is based upon only one criterion and uncritically application of this criterion can lead to potentially incorrect results. Because the data quality objective of an analysis differs from one case to another, it is not likely that any one criterion will be adequate for all applications. It is

the user's responsibility to review the output from AuvTool to make sure that the distributions selected by AuvTool for a given case are appropriate to the user's needs.

AuvTool should be continually reviewed and revised to ensure that it meets the ongoing needs of key users, such as analysts working with the SHEDS model and other user communities. Examples of key areas in which AuvTool should be augmented include the following:

- Incorporation of other parametric distributions, as appropriate.
- Evaluation of whether mixed empirical-parametric (MEP) distributions should be included in AuvTool.
- Capability to fit mixture distributions to data and to quantify uncertainty in the statistics and CDFs of mixture distributions using bootstrap simulation.
- Calculation of other statistics (e.g., skewness, kurtosis).
- Incorporation of additional methods for assisting the user in choosing parametric distributions for fitting to data sets (e.g., moment planes, probability plots, other goodness-of-fit tests).
- Incorporation of other bootstrap simulation methods for characterizing confidence intervals for statistics (e.g., BC_a method).
- Capability to handle measurement error in addition to random sampling error.
- Capability to interface directly with other models for purposes of controlling a probabilistic simulation of variability and uncertainty in model inputs and outputs.

The basis of these recommendations is briefly described.

AuvTool includes a limited set of seven parametric distributions. There may be other parametric distributions that the user community has found to be convenient and that should be added to AuvTool. Although AuvTool includes capabilities to work with empirical distributions and with selected parametric distributions, methods for dealing with "mixed empirical parametric" (MEP) distributions should also be considered for inclusion in AuvTool. These methods typically involve estimating an upper and lower

tail that extends beyond the minimum and maximum values, respectively, of an empirical distribution.

A key need for probabilistic analysis in the context of exposure and risk assessment is to be able to handle mixtures of two or more distributions. For example, many data sets are based upon observations of two or more subgroups. Such data can be multi-modal and would more accurately be described based upon a weighted combination of two or more parametric distributions (e.g., two lognormal distributions). The current version of AuvTool does not support the use of mixture distributions. A capability to fit mixture distributions to data and to estimate uncertainty in the statistics of mixture distributions should be included in future versions of AuvTool.

AuvTool currently supports calculation of point estimates, sampling distributions, and confidence intervals for selected statistics, including the arithmetic mean, arithmetic standard deviation, and distribution parameters. There may be other statistics that would be useful to the user community. For example, the skewness and kurtosis are measures of the shape of a distribution.

AuvTool currently includes a limited number of techniques for assessing the goodness-of-fit of a parametric distribution compared to an data set. These techniques include the K-S test and the A-D test. Other statistical goodness-of-fit tests could be included as options. In addition, graphical methods for evaluating goodness of fit, based upon probability plots and moment planes, could be incorporated. There are a variety of probability plotting methods that enable the analyst to evaluate different aspects of goodness-of-fit. Some methods provide more sensitivity with respect to goodness-of-fit of the tails while others provide more sensitivity with respect to the central portions of

the distribution. Moment planes are a graphical technique for comparing the shape of the empirical distribution of the data with the shape of possible candidate parametric distribution. Moment planes can be developed based upon the skewness and kurtosis.

The bootstrap simulation method used in AuvTool is the percentile method, which is versatile. However, there are other bootstrap methods, such as BCa, that may give more accurate confidence interval estimates in some cases. Such methods are typically not as versatile as the percentile methods, but they should be evaluated for incorporation into future versions of AuvTool.

AuvTool provides a method for quantifying uncertainty in the statistics of a distribution based upon random sampling error. While random sampling error may often be a dominate source of uncertainty, it is also possible that other sources of uncertainty may be equally or more important in some cases. For example, measurement errors in each individual data value may contribute substantially to uncertainty in key statistics or regarding the fit of a parametric distribution to the data. Methods for dealing with measurement error as a source of uncertainty should be evaluated for possible incorporation into AuvTool.

Because AuvTool is modular and based upon an object-oriented programming approach, it is possible to extend AuvTool as a basis for quantifying variability and uncertainty in model inputs, propagating variability and uncertainty to model outputs, and analysis of results. Approaches for extending the capabilities of AuvTool to include an ability to interface with other models and to analyze probabilistic analysis results should be explored.

7.0 REFERENCES

Ang A. H-S, and Tang, W.H., 1984, *Probability Concepts in Engineering Planning and Design, Volume 2*, John Wiley and Sons: New York.

Abdel-Aziz, A., and H.C. Frey, "Quantification of Variability and Uncertainty in Hourly NOx Emissions from Coal-Fired Power Plants," *Proceedings, Annual Meeting of the Air & Waste Management Association*, Pittsburgh, PA, June 2002 (in press).

Barry, T., 1996, "Recommendations on the Testing and Use of Pseudo-Random Number Generators Used in Monte Carlo Analysis for Risk Assessment," *Risk Analysis*, 16(1): 93-105.

Bogen, K.T., and Spear, R.C., 1987, "Integrating Uncertainty and Interindividual Vaiability in Environmental Risk Assessment," *Risk Analysis*, 7(4): 427-436.

Bogen, K.T., 1990, *Uncertainty in Environmental Health Risk Assessment*, Garland Publishing Inc., New York.

Bogen, K.T., 1992, *RiskQ: An Interactive Approach to Probability, Uncertainty, and Statistics for Use with Mathematica (Reference Manual*), Lawrence Livermore National Laboratory, Livermore, CA.

Box, G.E.P., M.E. Muller, 1958, "A Note on the Generation of Random Normal Deviates," *Annals of Mathematical Statistics*, 29: 610-611.

Bratley, P., B.L. Fox, and L.E. Schrage, 1987, *A Guide to Simulation*, 2nd ed. Springer-Verlag: New York.

Burmaster, D.E., R.H. Harris, 1994, "The Magnitude of Compounding Conservatisms in Superfund Risk Assessments", *Risk Analysis*, 13(2):131-143.

Burr, D., 1994, "A Comparison of Certain Confidence Intervals in the Cox model," *J. of the American Statistical Association*, 89, 1290-1302.

Casella, G., R. L. Berger, 1990, *Statistical Inference*, Duxbury Press: Belmont, CA

Cheng, R. C. H., 1977, "The Generation of Gamma Variables with Non-Integral Shape Parameter," *Applied Statistics*, 26:71-75.

Cohen, A.C., B. Whitten, 1988, *Parameter Estimation in Reliability and Life Span Models*, M. Dekker: New York.

Cohen, J.T., M.A., Lampson, and S. Bowers, 1996, "The Use of Two-Stage Monte Carlo Simulation Techniques to Characterize Variability and Uncertainty in Risk Analysis," *Human and Ecological Risk Assessment*, 2(4): 939-971.

Cooke, R.M., 1991, *Experts in Uncertainty: Opinion and Subjective Probability in Science*, Oxford Press.

Cullen, A.C., 1994, "Measures of Conservatism in Probabilistic Risk Assessment," *Risk Analysis*, 14: 389-392.

Cullen, A.C., H.C. Frey, 1999, *Use of Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs*, Plenum Press: New York.

D'Agostino, R.B., M.A. Stephens, 1986, *Goodness-of-Fit Techniques*, M. Dekker: New York.

Decisioneering, http://www.decisioneering.com (Accessed 01/20/2001).

Efron, B., 1979, "Bootstrap Method: Another Look at the Jackknife," *The Analysis of Statistics*, 7(1): 1-26.

Efron, B, R.J. Tibshirani, 1993, *An Introduction to the Bootstrap*, Chapman & Hall: London, UK.

EPA, 1992, "Guidelines for Exposure Assessment," Exposure Assessment Group, Office of Health and Environmental Assessment, U.S. Environmental Protection Agency, Washington, DC.

EPA, 1995, *Compilation of Air Pollutant Emission Factors* 5th Ed., AP-42 and Supplements, Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC.

EPA, 1996, *Summary Report for the Workshop on Monte Carlo Analysis*, EPA/630/R-96/010, Risk Assessment Forum, Office of Research and Development, U.S. Environmental Protection Agency, Washington, DC.

EPA, 1997, *Guiding Principles for Monte Carlo Analysis*, EPA/630/R-97/001, U.S. Environmental Protection Agency, Washington, DC.

EPA,1999a, *Report of the Workshop on Selecting Input Distributions for Probabilistic Assessment*, EPA/630/R-98/004, U.S. Environmental Protection Agency, Washington, DC.

EPA, 1999b, RAGS 3A - *Process for Conducting Probabilistic Risk Assessment*, Draft, U.S. Environmental Protection Agency, Washington, DC.

Evans, J.S., Graham, J.D., Gray, G.M., and Sielken, R.L., 1994a, "A Distributional Approach to Characterizing Low-Dose Cancer Risk," *Risk Analysis*, 14(1):25-34.

Finkel, A.M., 1990, *Confronting Uncertainty in Risk Assessment: A Guide for Decision Makers*, Center for Risk Management, Resources for the Future, Washington, DC.

Frey, H.C., 1992, "Quantitative Analysis of Uncertainty and Variability in Environmental Policy Making," Directorate for Science and Policy Programs, American Association for the Advancement of Science, Washington, DC.

Frey, H.C., 1997, "Variability and Uncertainty in Highway Vehicle Emission Factors," in Emission Inventory: Planning for the Future (held October 28-30 in Research Triangle Park, NC), Air and Waste Management Association, Pittsburgh, Pennsylvania, pp. 208-219.

Frey, H.C., and S. Bammi, 2002a, "Quantification of Variability and Uncertainty in Lawn and Garden Equipment NO_x and Total Hydrocarbon Emission Factors," *Journal of the Air & Waste Management Association*, accepted January 2002 for publication.

Frey, H.C., and S. Bammi, 2002b, "Probabilistic Nonroad Mobile Source Emission Factors," *Journal of Environmental Engineering*, tentatively accepted pending revisions.

Frey, H.C., and R. Bharvirkar, 2002, "Quantification of Variability and Uncertainty: A Case Study of Power Plant Hazardous Air Pollutant Emissions," Chapter 10 in *Human and Ecological Risk Analysis*, D. Paustenbach, Ed., John Wiley and Sons: New York. In press.

Frey, H.C., R. Bharvirkar, J. Zheng, 1999, "Quantitative Analysis of Variability and Uncertainty in Emissions Estimation," Final Report, Prepared by North Carolina State University for Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Frey, H.C., and D.A. Eichenberger, *Remote Sensing of Mobile Source Air Pollutant Emissions: Variability and Uncertainty in On-Road Emissions Estimates of Carbon Monoxide and Hydrocarbons for School and Transit Buses*, FHWY/NC/97-005, Prepared by North Carolina State University for North Carolina Department of Transportation, Raleigh, June 1997.

Frey, H.C., N.M. Rouphail, A. Unal, and J.D. Colyar, 2001, Emissions Reduction Through Better Traffic Management: An Empirical Evaluation Based Upon On-Road Measurements, FHWY/NC/2002-001, Prepared by North Carolina State University for North Carolina Department of Transportation, December 2001. 323 pp.

Frey, H.C., D.S. Rhodes, 1996, "Characterizing, Simulating, and Analyzing Variability and Uncertainty: An Illustration of Methods Using an Air Toxics Emissions Example," *Human and Ecological Risk Assessment*, 2(4):762-797.

Frey, H.C., D.S. Rhodes, 1998, "Characterization and simulation of uncertain frequency distributions: Effects of Distribution Choice, Variability, Uncertainty, and Parameter Dependence," *Human and Ecological Risk Assessment*, 4(2):423-468.

Frey, H.C., D.S. Rhodes, 1999, "Quantitative Analysis of Variability and Uncertainty in Environmental Data and Models: Volume 1. Theory and Methodology Based Upon Bootstrap Simulation," Report No. DOE/ER/30250--Vol. 1, Prepared by North Carolina State University for the U.S. Department of Energy, Germantown, MD.

Frey, H.C., A. Unal, and J. Chen, 2002, Recommended Strategy for On-Board Emission Data Analysis and Collection for the New Generation Model, Prepared by North Carolina State University for the Office of Transportation and Air Quality, U.S. Environmental Protection Agency, Ann Arbor, MI. February 2002.

Frey, H.C., J. Zheng, 2000, "Methods and Example Case Study for Analysis of Variability and Uncertainty in Emissions Estimation (AUVEE)," Prepared by North North Carolina State University for Office of Air Quality Planning and Standards, U.S. Environmental Protection Agency, Research Triangle Park, NC.

Frey, H.C., J. Zheng, 2002, "Quantification of Variability and Uncertainty in Utility NO_x Emission Inventories," *J. of Air & Waste Manage. Assoc.*, tentatively accepted pending revisions.

Hoffman, F.O., J.S. Hammonds, 1994, "Propagation of Uncertainty in Risk Assessments: The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability," *Risk Analysis*, 14(5):707-712.

Hahn, G.J., S.S. Shapiro, 1967, *Statistical Models in Engineering*, John Wiley and Sons: New York.

Hanna, S.R., Z. Lu, H.C. Frey, N. Wheeler, J. Vukovich, S. Arunachalam, M. Fernau, and D.A. Hansen, 2001, "Uncertainties in Predicted Ozone Concentrations due to Input Uncertainties for the UAM-V Photochemical Grid Model Applied to the July 1995 OTAG Domain," *Atmospheric Environment*, 35(5):891-903 (2001).

Harter, L.H., 1984, "Another Look at Plotting Positions," *Communications in Statistical-Theoretical Methods*, 13(13): 1613-1633.

Hattis, D., Burmaster, D.E., 1994, "Assessment of Variability and Uncertainty Distributions for Practical Risk Analyses," *Risk Analysis*, 14(5):713:729.

Hazen, A., 1914, "Storage to be Provided in Impounding Reservoirs for Municipal Water *Transactions of the American Society of Civil Engineers*, 77: 1539-1640.

Helton, J.,1996, "Probability, Conditional Probability and Complementary Cumulative Distribution Functions in Performance Assessment for Radioactive Waste Disposal," Sandia National Laboratories, Albuquergue, NM.

Helton et al., 1996, "Computational Implementation of a System Prioritization Methodology for the Waste Isolation Pilot Plant: A Preliminary Example," Sandia National Laboratories, Albuquergue, NM.

Henrion, M., and Fischoff, B., 1986, "Assessing Uncertainty in Physical Constants," *Am J Phys*, 54:791-797.

Hodges, J.S. (1987), "Uncertainty, Policy Analysis and Statistics," *Statistical Science*, 2, 259-291.

Holland, D.M., T. Fitz-Simmons, 1982, " Fitting Statistical Distributions to Air Quality Data by Maximum Likelihood Method," *Atmospheric Environment*, 16(5): 1071-1076.

IAEA 1989, "Evaluating the Reliability of Predictions Made Using Environmental Transfer Models," Safety Series, No. 100, *International Atomic Energy Agency* Vienna, Austria.

IPCC, 2000, *Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories*. National Greenhouse Gas Inventories Program, Intergovernmental Panel on Climate Change (IPCC), 2000.

Kahneman, D., Slovic, P., and Tversky, A. (eds.), 1982, *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, New York.

Kini, M.D., and H.C. Frey, 1997, *Probabilistic Evaluation of Mobile Source Air Pollution: Volume 1, Probabilistic Modeling of Exhaust Emissions from Light Duty Gasoline Vehicles*, Prepared by North Carolina State University for Center for Transportation and the Environment, Raleigh, December.

Law, A.M., W.D. Kelton, 1991, *Simulation Modeling and Analysis* 2d ed., McGraw-Hill: New York.

L'Ecuyer, P, 1994, "Uniform Random Number Generation," *Annals of Operation Research*, 53, 77-120.

L'Ecuyer, P., 1996, "Multiple Recursive Random Number Generators," *Operations Research*, 44:816-822.

Li, S., and H.C. Frey, 2002, "Methods and Example for Development of a Probabilistic Per-Capita Emission Factor for VOC Emissions from Consumer/Commercial Product Use", *Proceedings, Annual Meeting of the Air & Waste Management Association*, Pittsburgh, PA, June 2002 (in press).

Lilliefors, H.W., 1967, "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," *Journal of the American Statistical Association*, 62, 399-402.

Lumina, 1996, *Analytica User Guide*, Lumina Decision Systems, Los Altos, CA.

Mandel, J., 1969, *The Statistical Analysis of Experimental Data*, John Wiley and Sons, New York.

Mage, D. T., 1988. Correspondence: Comment on "Parametric Distributions of Regional Lake Chemistry: Fitted and Derived," *Environ. Sci. Technol*. 22, 1367-1368

Marsaglia, G., T.A. Bray, 1964, "A Convenient Method for Generating Normal *SIAM Review*, 6:260-264.

Martin, M.A., 1990, "On Bootstrap Iteration for Converge Correction in Confidence Intervals," *Journal of the American Statistical Association*, 85, 1105-1108.

Massey, F.J., 1951, "The Kolmogorov-Smirnov Test for Goodness of Fit," *Journal of the American Statistical Association*, 46, 68-78.

McKay, M.D., R.J. Beckman, W.J. Conover, 1979, " A Comparison of Three Method for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics* 21(2):239-245.

Morgan, M.G., and M. Henrion, 1990, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press: New York.

Murray, D.M., D.E. Burmaster, 1993, "Review of RiskQ: An Interactive Approach to Probability, Uncertainty, and Statistics for Use with Mathematica", *Risk Analysis*, 13(4): 479-482.

NCRP, 1996, *A Guide for Uncertainty Analysis in Dose and Risk Assessments Related to Environmental Contamination*, NCRP Commentary, Number 14, National Council on Radiation Protection and Measurement, Bethesda, MD, May 1996.

NRC, 1994, *Science and Judgment in Risk Assessment*, National Research Council, National Academy Press, Washington, D.C.

NRC, 2000, *Modeling Mobile Source Emissions*, National Research Council, National Academy Press, Washington, D.C.

Palisades, http://www.palisade.com (Accessed 01/20/2001).

Pollack, A.K., P. Bhave, J. Heiken, K. Lee, S. Shepard, C. Tran, G. Yarwood, R.F. Sawyer, and B.A. Joy, 1999, *Investigation of Emission Factors in the California EMFAC7G Model*, PB99-149718INZ, Prepared by ENVIRON International Corp, Novato, CA, for Coordinating Research Council, Atlanta, GA.

Press, W.H., S.A Teukolsky, W.T. Vetterling, and B.P. Flannery, 1992, *Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd ed.*, Cambridge University Press: New York.

Radian, 1996, *Evaluating the Uncertainty of Emission Estimates, Final Report*, Prepared by Radian Corporation for the Emission Inventory Improvement Program, State and Territorial Air Pollution Control Officers' Association, Association of Local Air Pollution Control Officers, and U.S. Environmental Protection Agency, Research Triangle Park, NC.

Rhodes, D.S., 1997, *Quantitative Analysis of Variability and Uncertainty in Environmental and Risk Assessment*, Masters Thesis, Department of Civil Engineering, North Carolina State University, Raleigh, NC.

Rubinstein, R. Y., 1981, *Simulation and the Monte Carlo Method*, John Wiley & Sons: New York.

Seinfeld, J.H., 1986, *Atmospheric Chemistry and Physics of Air Pollution*, John Wiley and Sons: New York.

Seiler, F.A., J.L. Alvarez, 1996, "On the Selection of Distributions for Stochastic *Risk Analysis*, 16(1):5-1.

Small, M.J., 1990. "Probability Distributions and Statistical Estimation," Chapter 5 in *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Morgan, M.G., and Henrion, M., Cambridge University Press: New York.

Stephens, M.A., 1974, "EDF Statistics for Goodness-of-Fit and Some Comparisons," *Journal of the American Statistical Association*, 69, 730-737.

Thompson, K.M., J.D. Graham, 1996, "Going Beyond the Single Number: Using Probabilistic Risk Assessment to Improve Risk Management", *Human and Ecological Risk Assessment*, 2(4):1008-1034.

Walpole, R.E., R.H. Myers, 1993. *Probability and Statistics for Engineers and Scientists*, Macmillan Publishing Company: New York.

Zheng, J., H.C. Frey, 2002, "AuvTool User's Guide," Prepared by North Carolina State University for Office of Research and Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, February, 2002

APPENDIX A. GRAPHICAL COMPARISON FOR RANDOM SAMPLE GENERATION

1. Normal Distribution

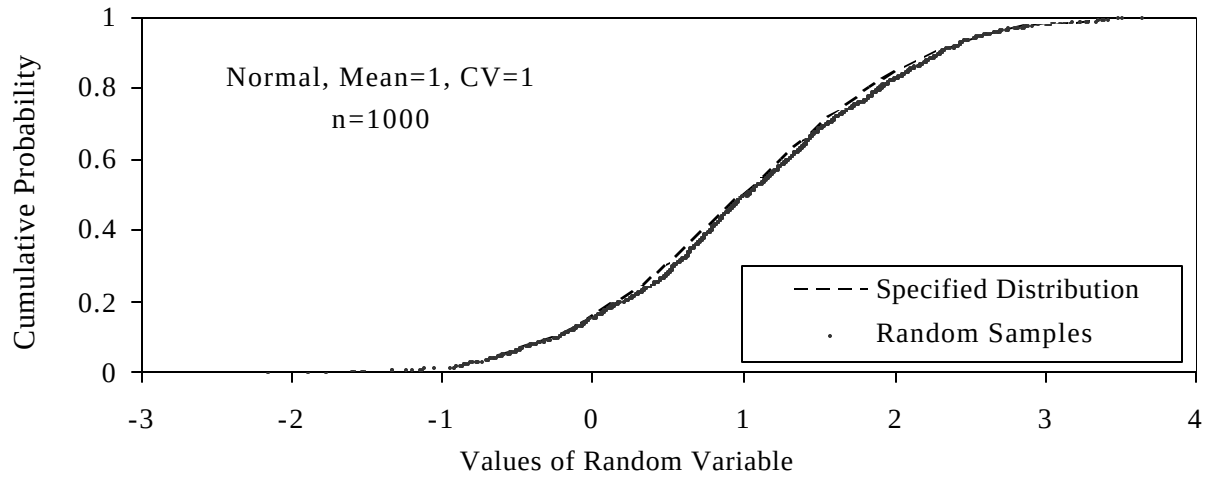


Figure A-1. Comparison of Random Samples and Specified Normal Distribution, $\mu=1, \sigma^2=1$.

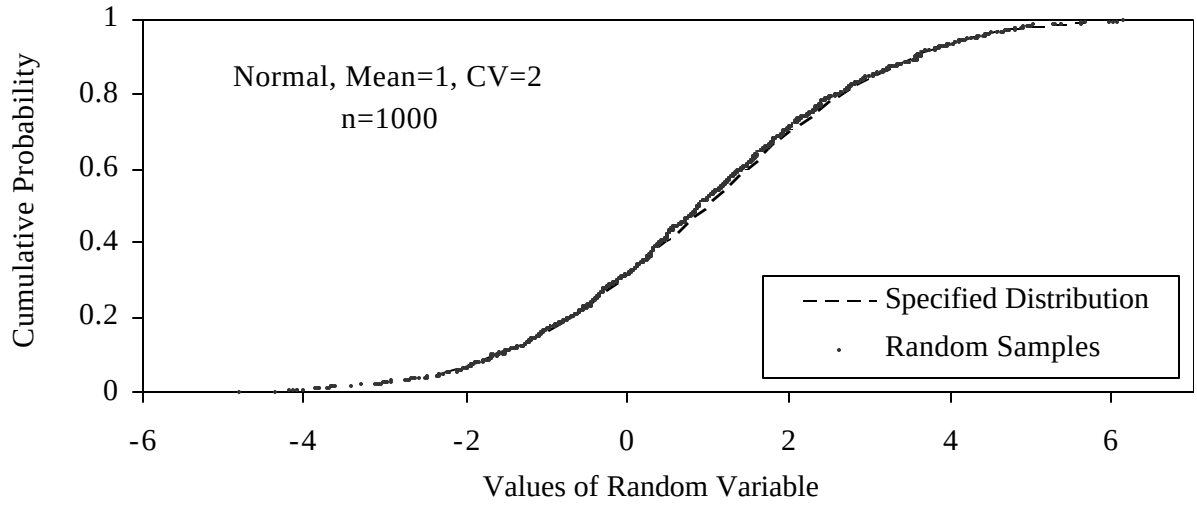


Figure A-2. Comparison of Random Samples and Specified Normal Distribution, $\mu=1, \sigma^2=4$.

2. Lognormal Distribution

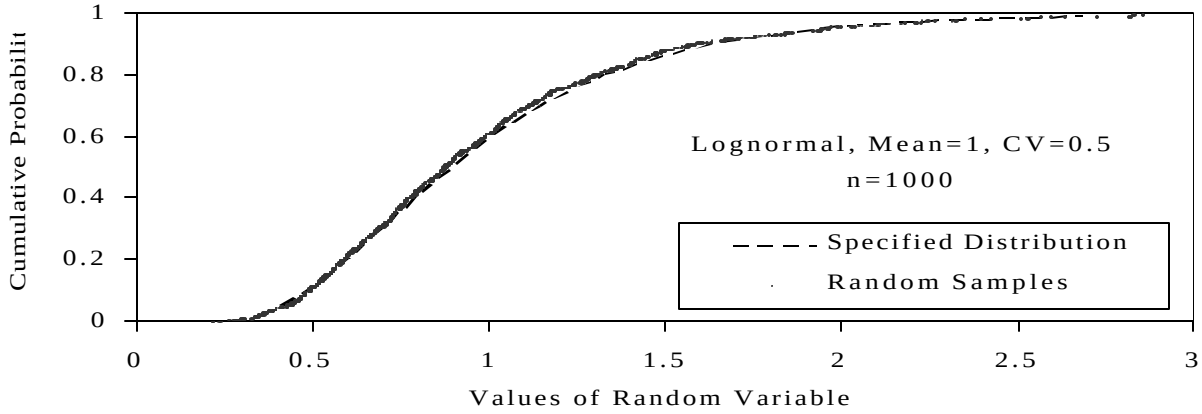


Figure A-3. Comparison of Rand. Samp. and Specified Lognormal Dist., μ_{lnx}=-0.112, σ_{lnx}=0.472.

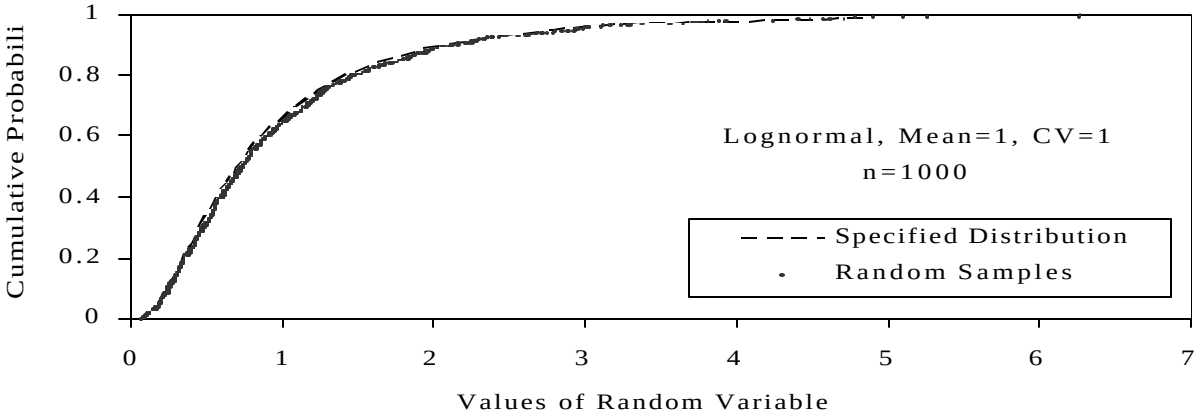
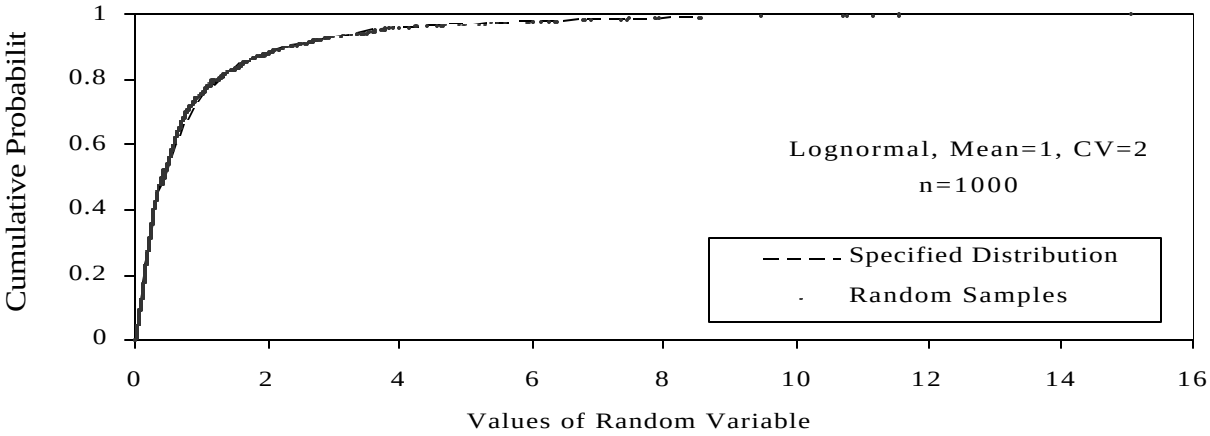


Figure A-4. Comparison of Rand. Samp. and Specified Lognormal Dist., μ_{lnx}=-0.347, σ_{lnx}=0.833.



igure A-5. Comparison of Rand. Samp. and Specified Lognormal Dist., μ_{lnx}=-0.804, σ_{lnx}=1.269.

3. Gamma Distribution

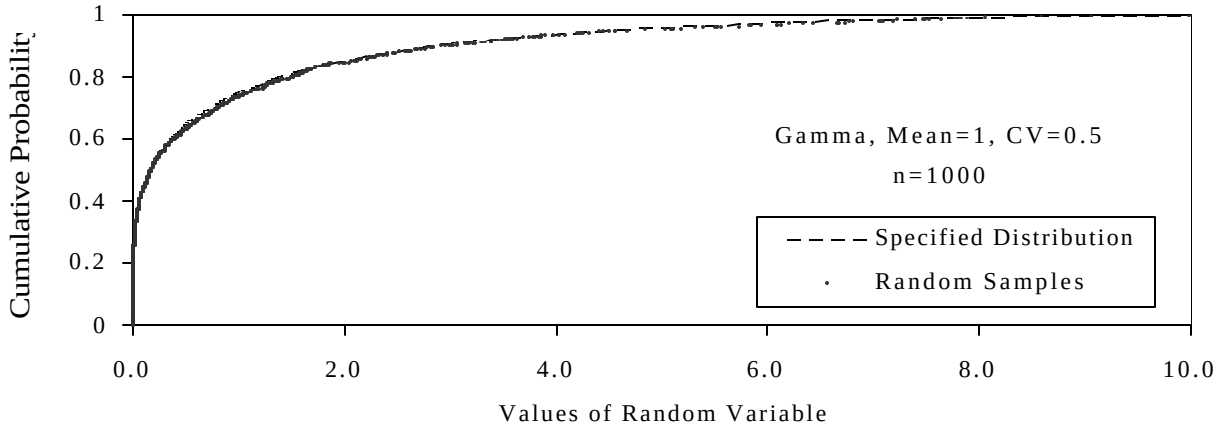


Figure A-6. Comparison of Random Samples and Specified Gamma Distribution, α=4, β=0.25.

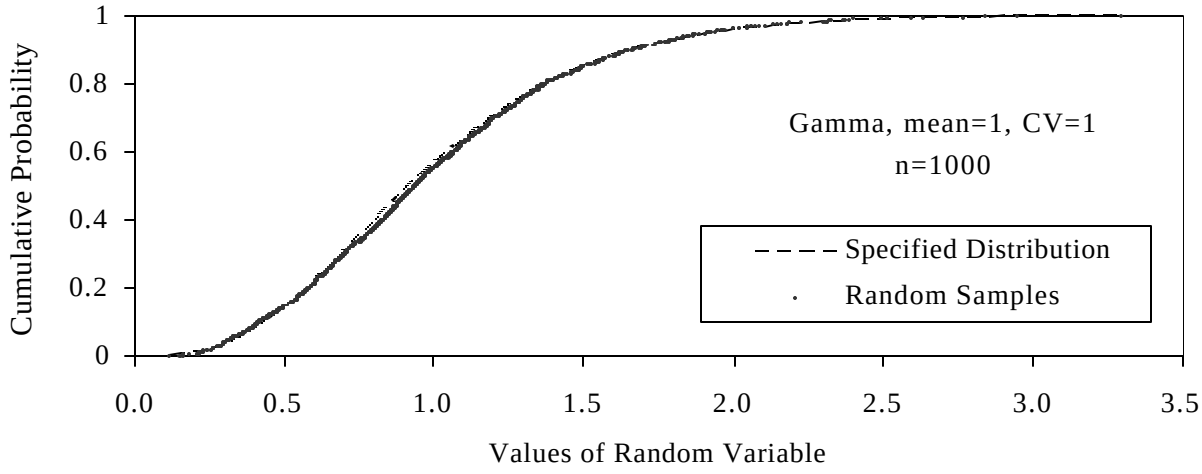


Figure A-7. Comparison of Random Samples and Specified Gamma Distribution, α=1, β=1.

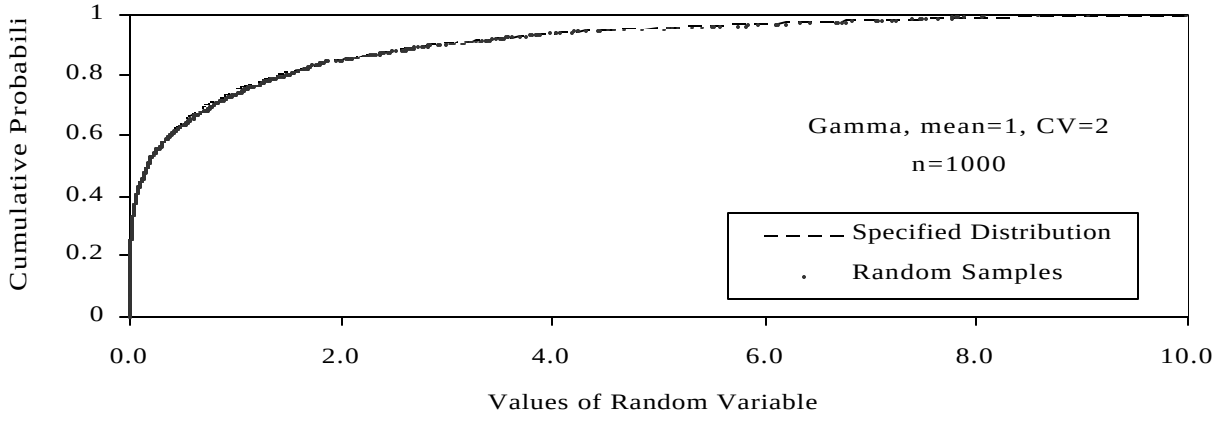


Figure A-8. Comparison of Random Samples and Specified Gamma Distribution, α=0.25, β=4.

4. Weibull Distribution

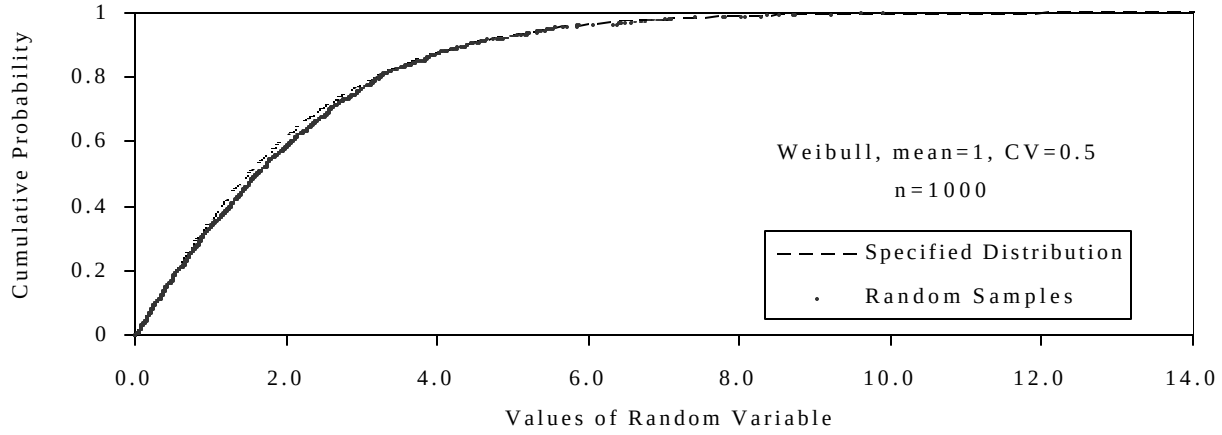


Figure A-9. Comparison of Random Samples and Specified Weibull Dist., α=2.101, β=1.129.

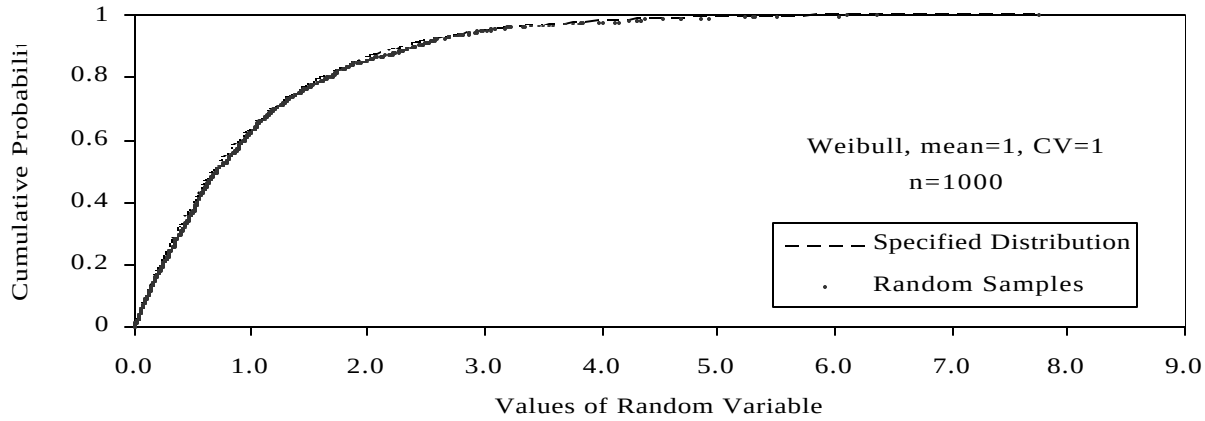


Figure A-10. Comparison of Random Samples and Specified Weibull Distribution, α=1, β=1.

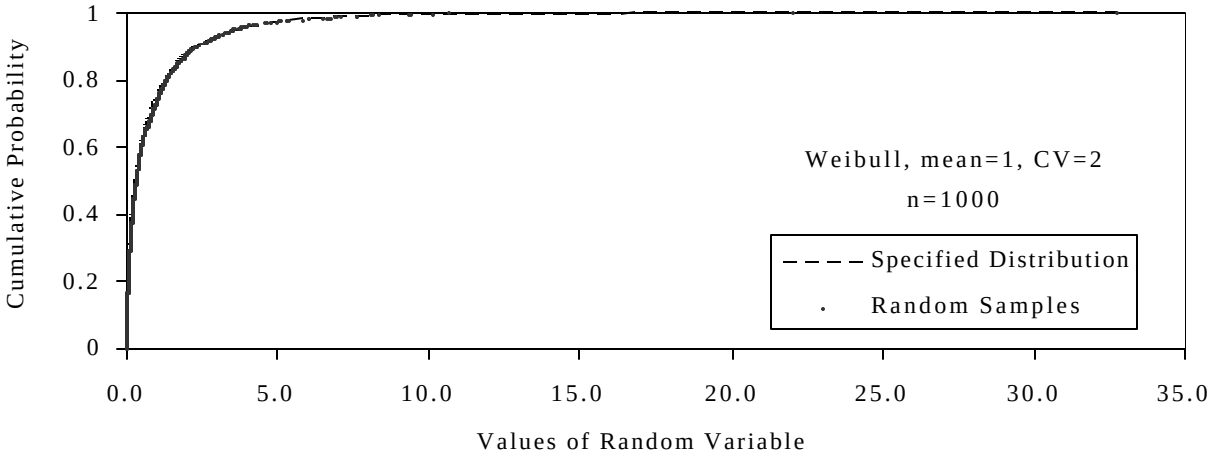


Figure A-11. Comparison of Random Samples and Specified Weibull Dist., α=0.543, β=0.575.

5. Beta Distribution

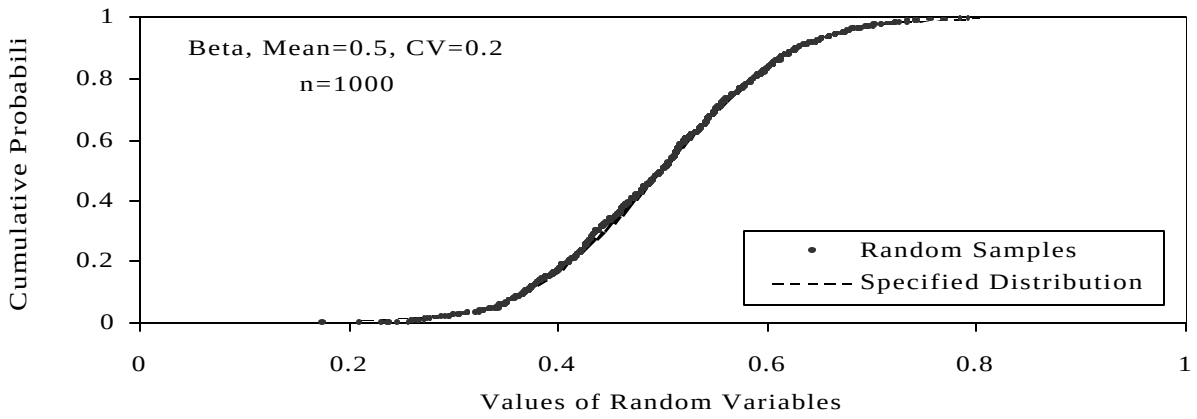


Figure A-12. Comparison of Random Samples and Specified Beta Distribution, α=12, β=12.

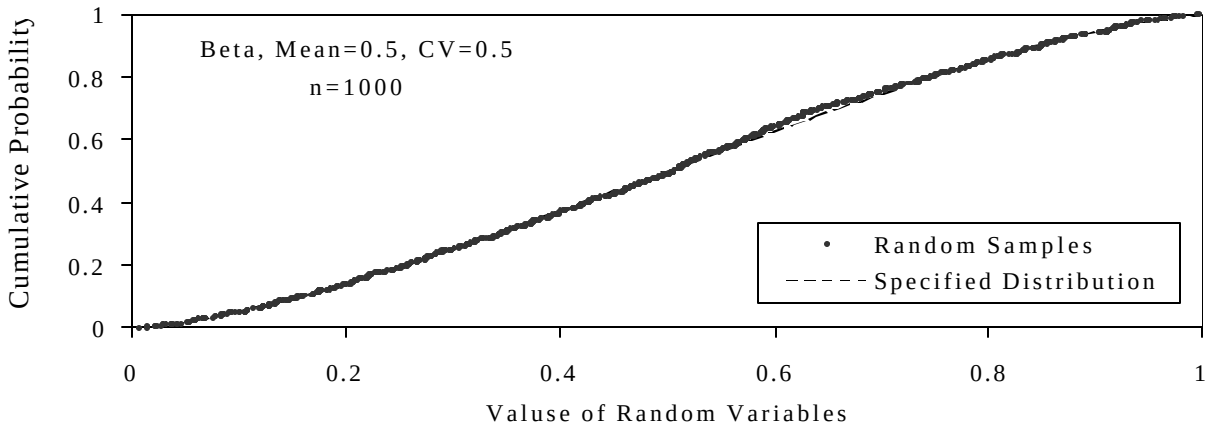


Figure A-13. Comparison of Random Samples and Specified Beta Distribution, α=1.5, β=1.5.

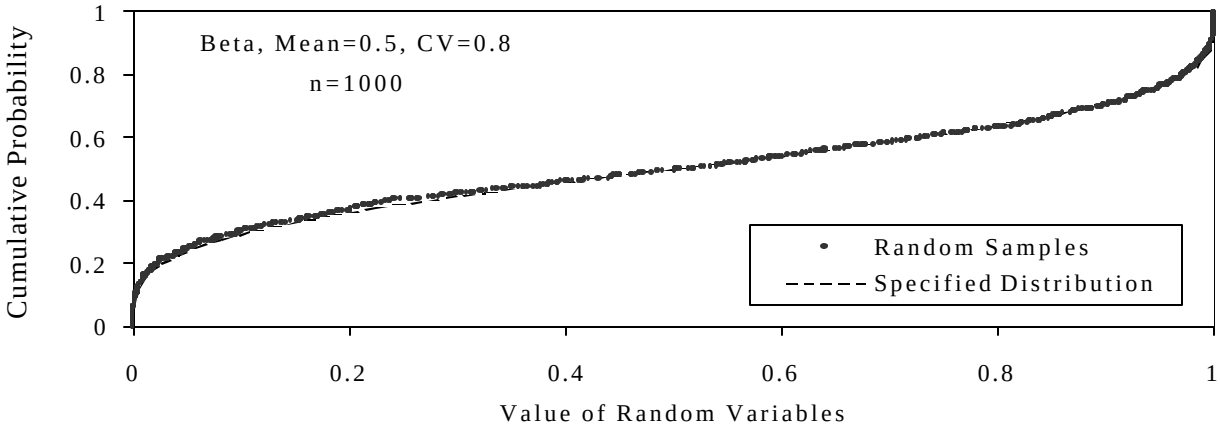


Figure A-14. Comparison of Random Samples and Specified Beta Dist., α=0.281, β=0.281.

6. Uniform Distribution

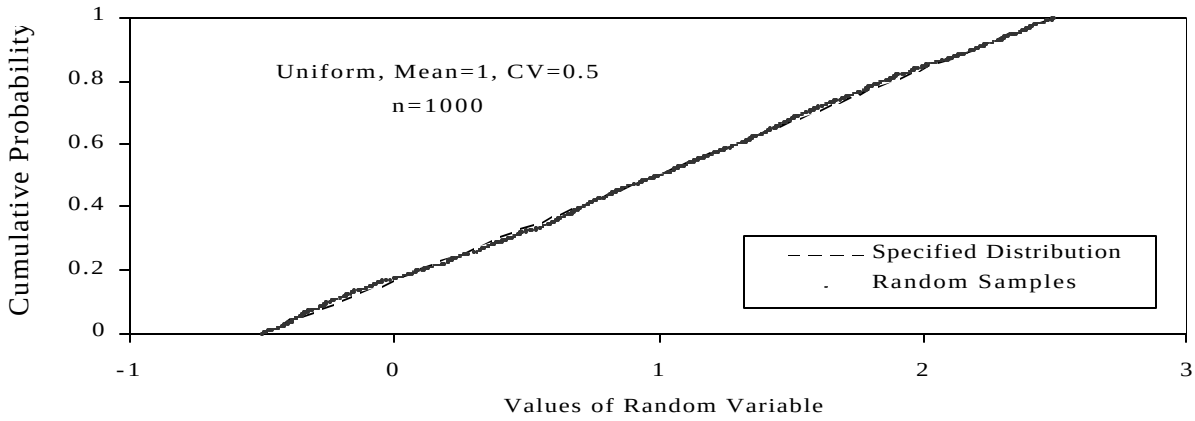


Figure A-15. Comparison of Random Samples and Specified Uniform Dist., a=-0.5, b=2.5.

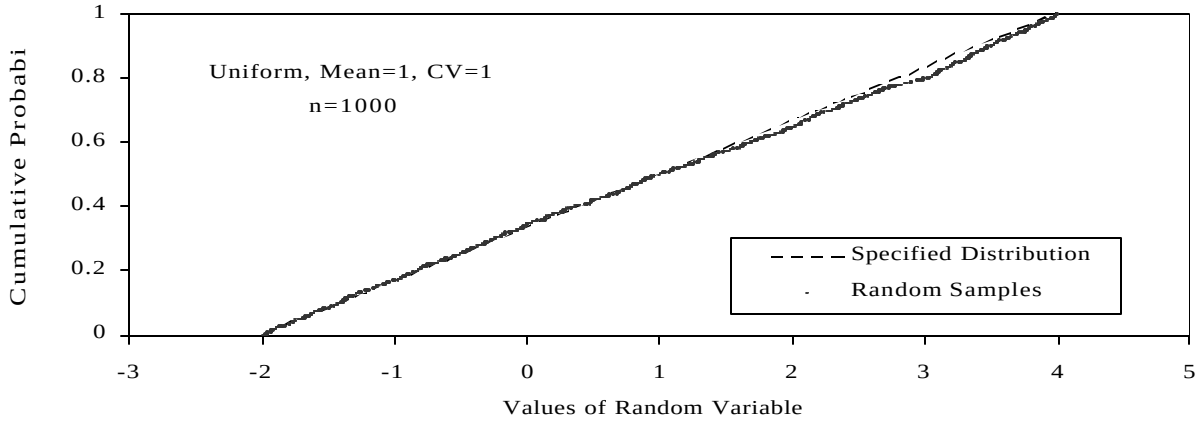


Figure A-16. Comparison of Random Samples and Specified Uniform Distribution, a=-2, b=4.

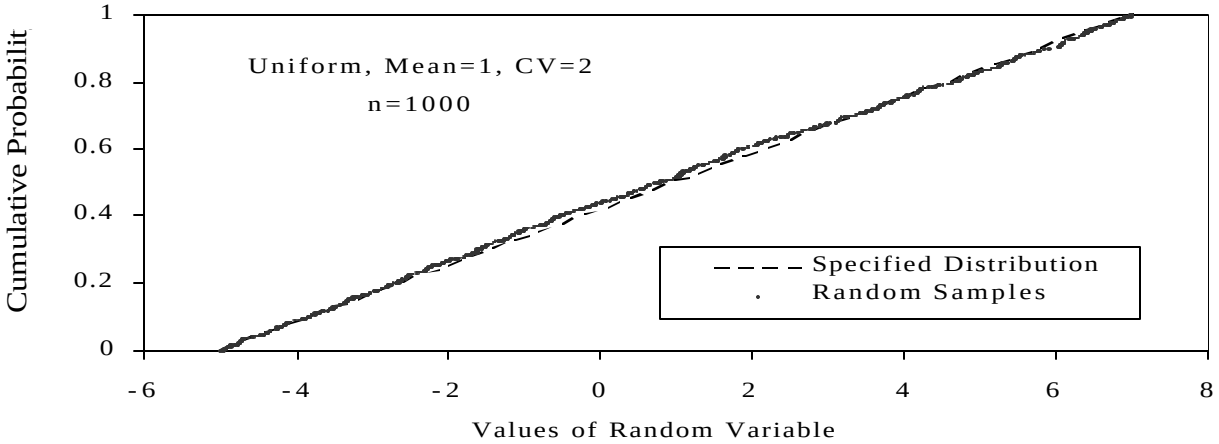


Figure A-17. Comparison of Random Samples and Specified Uniform Distribution, a=-5, b=7.

7. Symmetric Triangle Distribution

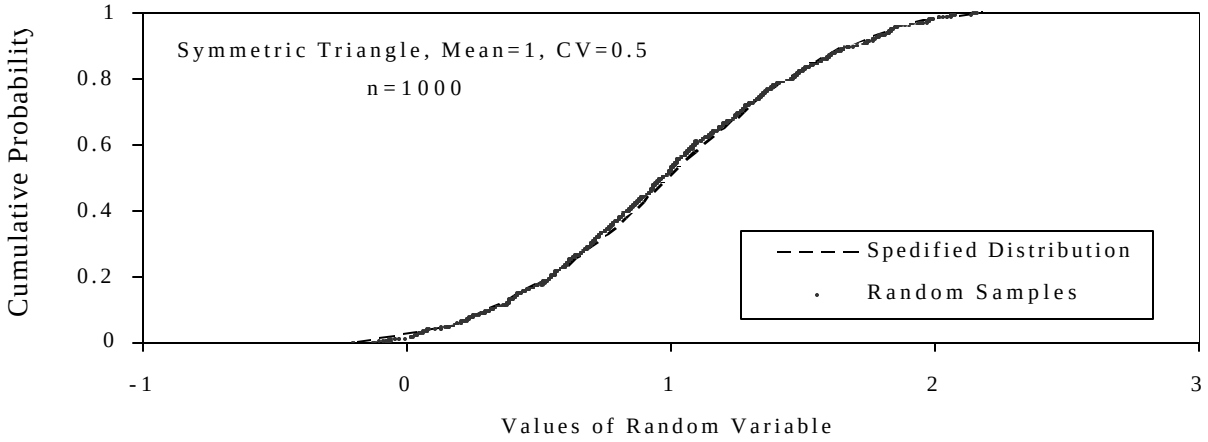


Figure A-18. Comparison of Rand. Samp. and Specified Symmtr. Triangle Dist., a=1, b=1.225.

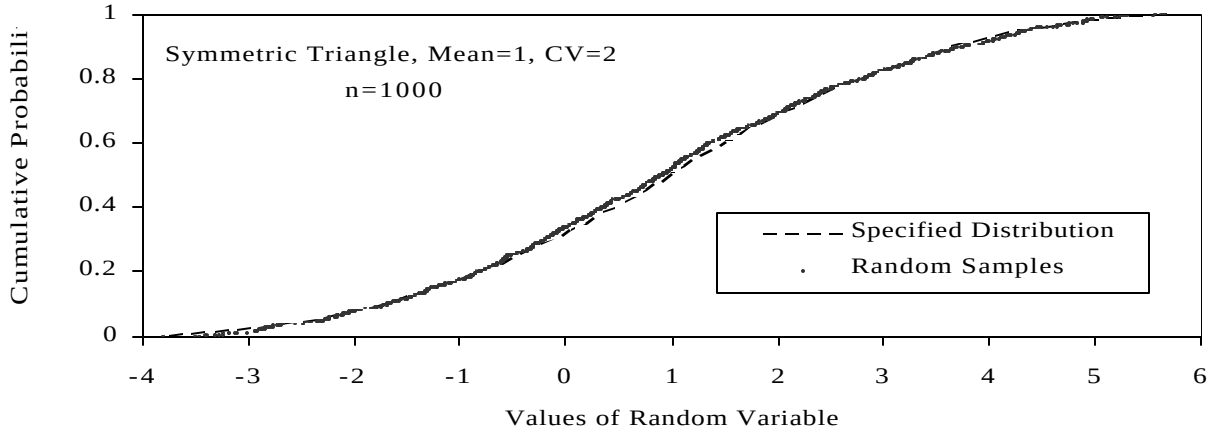


Figure A-19. Comparison of Rand. Samp. and Specified Symmtr. Triangle Dist., a=1. b=2.450.

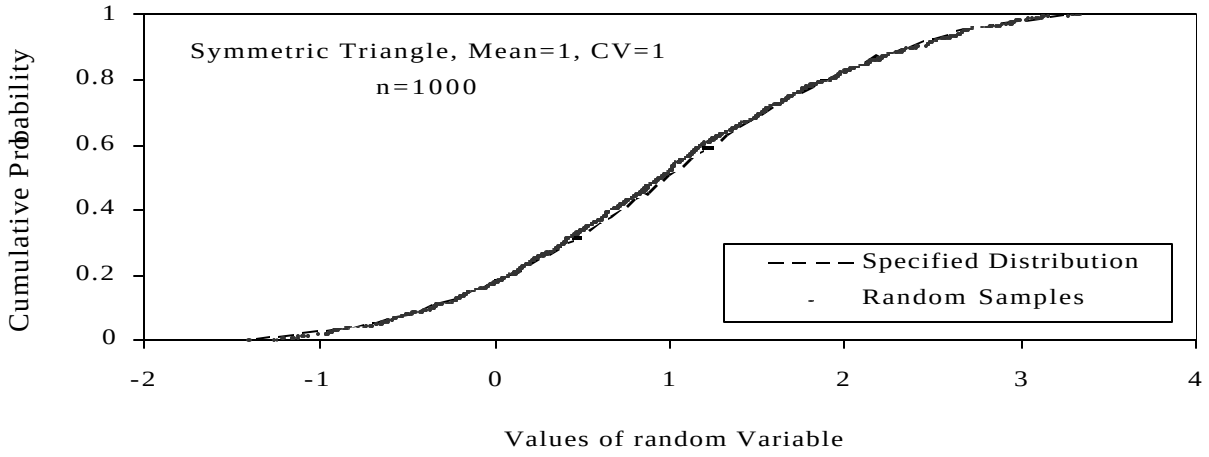


Figure A-20. Comparison of Rand. Samp. and Specified Symmtr. Triangle Dist., a=1, b=4.899.

8. Empirical Distribution

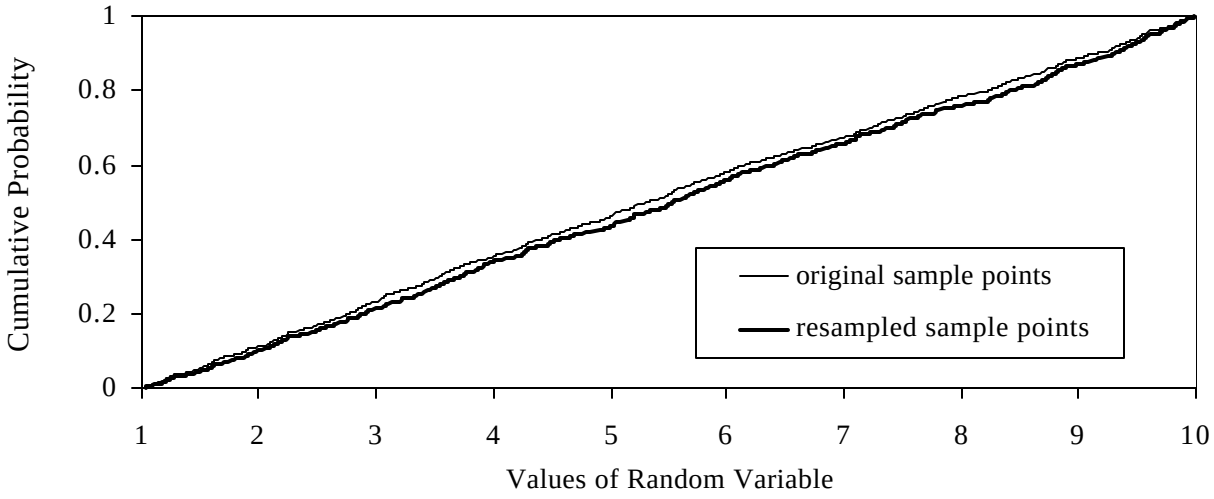


Figure A-21. Comparison of Empirical Distribution of Randomly Generated Data and Analytical CDF of Assumed Population Distribution for the Empirical Distribution.

APPENDIX B: DATA USED IN VERIFICATION TESTS

Table B-1. Data Sets for Normal Distribution for Parameter Estimation and K-S Test

n=10	n=20	n=50
-0.9892	0.2554	-2.7846
-0.4436	0.3347	-2.5341
-0.4265	0.5029	-2.2542
0.1772	0.5282	-1.9784
0.2526	0.5368	-1.6614
0.4170	0.8386	-1.3167
0.5368	0.8763	-1.0314
0.7948	0.9585	-0.9882
0.9033	0.9735	-0.8873
1.1912	1.0184	-0.8530
	1.1430	-0.5252
	1.1474	-0.4655
	1.1664	-0.2425
	1.1841	-0.1543
	1.2017	0.0552
	1.3456	0.0646
	1.3495	0.1266
	1.7138	0.2434
	1.9437	0.2570
	2.3228	0.3544
		0.5052
		0.5060
		0.8341
		0.8940
		0.9776
		1.0049
		1.0631
		1.0735
		1.1502
		1.5185
		1.5718
		1.5895
		1.6526
		1.6654
		1.7366
		1.8067
		1.9102
		1.9256

Continued on next page.

Table B-1. Continued.

		2.1029
		2.3824
		2.3980
		3.0385
		3.3135
		3.6305
		3.8550
		4.0295
		4.3277
		4.7749
		4.8574
		6.2912

Table B-2. Data Sets of Normal Distribution for Confidence Interval Test

μ=1, CV=0.5			μ=1, CV=1			μ=1, CV=2		
n=10	n=20	n=50	n=10	n=20	n=50	n=10	n=20	n=50
0.9585	0.9585	0.9585	0.9170	0.9170	0.9170	0.8341	0.8341	0.8341
0.8763	0.8763	0.8763	0.7526	0.7526	0.7526	0.5052	0.5052	0.5052
0.5282	0.5282	0.5282	0.0564	0.0564	0.0564	-0.8873	-0.8873	-0.8873
1.0184	1.0184	1.0184	1.0368	1.0368	1.0368	1.0735	1.0735	1.0735
1.3456	1.3456	1.3456	1.6912	1.6912	1.6912	2.3824	2.3824	2.3824
1.1474	1.1474	1.1474	1.2948	1.2948	1.2948	1.5895	1.5895	1.5895
1.2017	1.2017	1.2017	1.4033	1.4033	1.4033	1.8067	1.8067	1.8067
0.8386	0.8386	0.8386	0.6772	0.6772	0.6772	0.3544	0.3544	0.3544
0.5368	0.5368	0.5368	0.0735	0.0735	0.0735	-0.8530	-0.8530	-0.8530
0.2554	0.2554	0.2554	-0.4892	-0.4892	-0.4892	-1.9784	-1.9784	-1.9784
	1.1841	1.1841		1.3683	1.3683		1.7366	1.7366
	0.5029	0.5029		0.0059	0.0059		-0.9882	-0.9882
	0.3347	0.3347		-0.3307	-0.3307		-1.6614	-1.6614
	0.9735	0.9735		0.9470	0.9470		0.8940	0.8940
	2.3228	2.3228		3.6456	3.6456		6.2912	6.2912
	1.1664	1.1664		1.3327	1.3327		1.6654	1.6654
	1.3495	1.3495		1.6990	1.6990		2.3980	2.3980
	1.1430	1.1430		1.2859	1.2859		1.5718	1.5718
	1.7138	1.7138		2.4275	2.4275		3.8550	3.8550
	1.9437	1.9437		2.8874	2.8874		4.7749	4.7749
		0.8109			0.6217			0.2434
		0.6894			0.3788			-0.2425
		1.2276			1.4551			1.9102
		0.6187			0.2374			-0.5252
		1.8319			2.6639			4.3277
		1.1632			1.3263			1.6526
		1.7574			2.5148			4.0295
		1.6576			2.3152			3.6305
		0.7817			0.5633			0.1266
		1.2757			1.5514			2.1029
		0.8765			0.7530			0.5060
		1.0012			1.0024			1.0049
		1.9644			2.9287			4.8574
		1.0376			1.0751			1.1502
		1.2314			1.4628			1.9256
		0.4922			-0.0157			-1.0314
		1.0158			1.0315			1.0631
		1.5784			2.1567			3.3135
		0.1165			-0.7671			-2.5341

Continued on next page.

Table B-2. Continued.

		1.5096			2.0193			3.0385
		0.9944			0.9888			0.9776
		0.7638			0.5276			0.0552
		0.8143			0.6285			0.2570
		1.1296			1.2593			1.5185
		0.0538			-0.8923			-2.7846
		0.7662			0.5323			0.0646
		0.1864			-0.6271			-2.2542
		0.6336			0.2672			-0.4655
		0.7114			0.4229			-0.1543
		0.4208			-0.1584			-1.3167

Table B-3. Data Sets for Lognormal Distribution for Parameter Estimation and K-S Test

n=10	n=20	n=50
0.8601	0.6599	0.4025
0.7958	0.5755	0.3268
0.5727	0.3223	0.1351
0.9101	0.7291	0.4686
1.2398	1.2572	1.0748
1.0280	0.9038	0.6500
1.0822	0.9893	0.7460
0.7679	0.5405	0.2969
0.5774	0.3270	0.1381
0.4426	0.2047	0.0676
	0.9608	0.7135
	0.3091	0.1267
	0.2335	0.0827
	0.6766	0.4181
	6.3983	12.8273
	0.9328	0.6821
	1.2654	1.0855
	0.8971	0.6427
	2.3208	2.7354
	3.4035	4.9025
		0.2768
		0.2033
		0.7967
		0.1700
		3.6918
		0.6765
		3.0556
		2.3723
		0.2570
		0.9002
		0.3269
		0.4486
		5.1661
		0.4919
		0.8044
		0.1233
		0.4655
		1.9402
		0.0475
		1.6296

Continued on next page.

Table B-3. Continued.

		0.4409
		0.2456
		0.2792
		0.6214
		0.0405
		0.2471
		0.0568
		0.1765
		0.2150
		0.1029

Table B-4. Data Sets of Lognormal Distribution for Confidence Interval Test

μ=1, CV=0.5			μ=1, CV=1			μ=1, CV=2		
n=10	n=20	n=50	n=10	n=20	n=50	n=10	n=20	n=50
0.8601	0.8601	0.8601	0.6599	0.6599	0.6599	0.4025	0.4025	0.4025
0.7958	0.7958	0.7958	0.5755	0.5755	0.5755	0.3268	0.3268	0.3268
0.5727	0.5727	0.5727	0.3223	0.3223	0.3223	0.1351	0.1351	0.1351
0.9101	0.9101	0.9101	0.7291	0.7291	0.7291	0.4686	0.4686	0.4686
1.2398	1.2398	1.2398	1.2572	1.2572	1.2572	1.0748	1.0748	1.0748
1.0280	1.0280	1.0280	0.9038	0.9038	0.9038	0.6500	0.6500	0.6500
1.0822	1.0822	1.0822	0.9893	0.9893	0.9893	0.7460	0.7460	0.7460
0.7679	0.7679	0.7679	0.5405	0.5405	0.5405	0.2969	0.2969	0.2969
0.5774	0.5774	0.5774	0.3270	0.3270	0.3270	0.1381	0.1381	0.1381
0.4426	0.4426	0.4426	0.2047	0.2047	0.2047	0.0676	0.0676	0.0676
	1.0644	1.0644		0.9608	0.9608		0.7135	0.7135
	0.5592	0.5592		0.3091	0.3091		0.1267	0.1267
	0.4770	0.4770		0.2335	0.2335		0.0827	0.0827
	0.8723	0.8723		0.6766	0.6766		0.4181	0.4181
	3.1210	3.1210		6.3983	6.3983		12.8273	12.8273
	1.0467	1.0467		0.9328	0.9328		0.6821	0.6821
	1.2444	1.2444		1.2654	1.2654		1.0855	1.0855
	1.0238	1.0238		0.8971	0.8971		0.6427	0.6427
	1.7555	1.7555		2.3208	2.3208		2.7354	2.7354
	2.1815	2.1815		3.4035	3.4035		4.9025	4.9025
		0.7481			0.5161			0.2768
		0.6670			0.4216			0.2033
		1.1090			1.0329			0.7967
		0.6239			0.3748			0.1700
		1.9629			2.8255			3.6918
		1.0435			0.9278			0.6765
		1.8294			2.4956			3.0556
		1.6648			2.1137			2.3723
		0.7277			0.4916			0.2570
		1.1606			1.1191			0.9002
		0.7959			0.5757			0.3269
		0.8955			0.7085			0.4486
		2.2245			3.5225			5.1661
		0.9267			0.7527			0.4919
		1.1130			1.0395			0.8044
		0.5536			0.3036			0.1233
		0.9078			0.7259			0.4655
		1.5447			1.8524			1.9402
		0.3882			0.1624			0.0475

Continued on next page.

Table B-4. Continued.

		1.4476			1.6521			1.6296
		0.8897			0.7005			0.4409
		0.7155			0.4772			0.2456
		0.7505			0.5190			0.2792
		1.0110			0.8775			0.6214
		0.3659			0.1463			0.0405
		0.7171			0.4791			0.2471
		0.4147			0.1825			0.0568
		0.6327			0.3842			0.1765
		0.6810			0.4373			0.2150
		0.5175			0.2696			0.1029

Table B-5. Data Sets for Beta, Uniform and Symmetric Triangle Distribution for Parameter Estimation and K-S Test

n=10	n=20	n=50
0.666686	0.056167	0.479767
0.520708	0.204888	0.847666
0.084358	0.416095	0.1422
0.688581	0.681037	0.75929
0.765656	0.674789	0.202192
0.719637	0.314731	0.940327
0.761868	0.187021	0.161397
0.518099	0.942398	0.291832
0.465871	0.50316	0.290142
0.527589	0.74777	0.69741
	0.465847	0.795787
	0.293581	0.640029
	0.473296	0.932238
	0.448283	0.717781
	0.797148	0.540517
	0.005394	0.633189
	0.182474	0.762339
	0.288185	0.142484
	0.475007	0.549227
	0.255269	0.170821
		0.790685
		0.528421
		0.834544
		0.268829
		0.384195
		0.391376
		0.912772
		0.543857
		0.434066
		0.16911
		0.037326
		0.788454
		0.615929
		0.24517
		0.027483
		0.119344
		0.289228
		0.715284
		0.015397

Continued on next page.

Table B-5. Continued.

		0.422466
		0.24964
		0.925673
		0.241992
		0.811826
		0.158428
		0.129122
		0.316524
		0.788725
		0.281392
		0.758638

Table B-6. Data Sets of Beta Distribution for Confidence Interval Test

μ=0.5, CV=0.2, n=50	μ=0.5, CV=0.5, n=50	μ=0.5, CV=0.8, n=50
0.024467	0.002633	1.00E-10
0.036959	0.002684	1.00E-10
0.051817	0.004858	1.00E-10
0.051899	0.004995	1.00E-10
0.054774	0.006383	1.00E-10
0.055357	0.007085	1.00E-10
0.055589	0.007203	8E-07
0.055947	0.011039	1.6E-06
0.057648	0.012053	0.000002
0.059847	0.0219	2.6E-06
0.06246	0.022339	4.8E-06
0.064651	0.024703	1.06E-05
0.065029	0.026111	1.71E-05
0.065164	0.026281	2.27E-05
0.069831	0.026447	4.69E-05
0.071273	0.034833	0.00013
0.074798	0.035238	0.000141
0.077348	0.035755	0.000158
0.077413	0.035855	0.000168
0.077604	0.039402	0.000187
0.07854	0.04344	0.000221
0.086017	0.045545	0.00026
0.088675	0.054316	0.000354
0.089772	0.061863	0.000884
0.089893	0.063722	0.001193
0.092192	0.07151	0.001745
0.097064	0.072251	0.001804
0.09716	0.077617	0.002203
0.098054	0.079543	0.002609
0.102059	0.081759	0.003233
0.104372	0.085027	0.003602
0.108084	0.089248	0.003891
0.110239	0.0996	0.005033
0.111961	0.113736	0.012219
0.116976	0.116015	0.012529
0.117631	0.116034	0.015099
0.118047	0.128622	0.015686
0.119931	0.140123	0.024592
0.122798	0.148648	0.028731
0.125703	0.166565	0.039662

Continued on next page.

Table B-6. Continued.

0.138469	0.187798	0.050859
0.151853	0.203531	0.053434
0.15532	0.242457	0.080479
0.166411	0.245294	0.152349
0.167765	0.263492	0.193202
0.17457	0.323808	0.206504
0.178525	0.324956	0.208986
0.196136	0.338322	0.2738
0.197207	0.403492	0.406509
0.25337	0.407709	0.412991

Table B-7. Data Sets of Uniform Distribution for Confidence Interval Test

μ=1, CV=0.5, n=50	μ=1, CV=1, n=50	μ=1, CV=2, n=50
-0.32498	1.221523	6.07E+00
0.30388	0.735288	6.39E+00
0.583567	-1.25886	8.96E-01
2.225166	3.913587	-4.49E+00
2.46808	-1.3389	-1.27E+00
1.529386	3.709192	-3.13E+00
0.519346	-1.61759	-0.1291269
0.355154	-0.18521	-2.7758997
2.343876	0.195226	3.6908817
-0.20759	-0.63884	5.8900257
0.068129	-1.70059	6.1600923
2.22944	3.43418	4.2930695
1.01877	3.265578	0.3838651
0.194764	-0.60356	0.3473017
1.489041	-0.60929	-0.4435245
1.817975	-0.57964	-2.696858
1.510918	2.498651	-2.0761322
-0.19813	0.939452	0.7750195
1.583366	2.895536	-4.6393945
2.100992	3.097996	3.3696667
0.800552	2.275948	-4.2462408
0.131175	1.671725	0.9784055
0.642738	2.936652	-1.3917915
1.58287	1.144557	4.7880763
0.142016	-0.4461	-2.6372971
1.05426	3.134434	-0.6632113
1.109989	3.011648	2.2567119
1.570975	2.699634	3.3643972
0.714458	-1.82872	0.2817935
0.510443	1.839341	2.4258328
0.183517	1.56536	6.6566476
0.650184	3.94612	3.4202417
-0.04149	2.158811	-0.2463783
2.358758	1.141405	-1.4996307
0.889038	1.860129	-4.3053673
2.346991	-1.82247	-4.7690753
1.173009	2.3823	-1.8921308
1.074553	-1.09907	3.1037538
2.440008	-0.09205	-1.4702646
0.771399	2.246039	-1.8078144

Continued on next page.

Table B-7. Continued.

2.190604	3.630662	2.3860321
1.515629	3.77217	0.7263638
0.119436	1.044941	0.0872605
2.363081	2.912202	1.6929773
1.830178	2.41365	-2.0452632
-0.14426	2.069731	0.2477471
1.263888	-0.35185	1.2236824
1.324095	0.846339	-1.8524161
2.227001	2.679161	4.6425357
2.0533	-0.16315	4.0404926

Table B-8. Data Sets of Symmetric Triangle Distribution for Confidence Interval Test

μ=1, CV=0.5, n=50	μ=1, CV=1, n=50	μ=1, CV=2, n=50
0.193447	1.225817	-1.10E+00
0.67178	1.93194	1.62E+00
0.816162	1.181863	1.01E+00
1.700645	0.295435	2.21E+00
2.046302	0.878902	-3.63E-01
1.239597	1.149237	4.27E+00
0.784837	1.358888	0.4972708
0.699938	0.989834	0.2721264
1.829793	0.813274	1.9241575
0.315865	-1.18277	-1.152917
0.528901	0.654205	0.5490978
1.704738	-0.67189	-2.056923
1.007688	1.128859	2.2209788
0.608698	2.720267	4.0099247
1.219326	1.268958	0.999532
1.39898	0.867484	1.1666594
1.230267	0.966704	-0.987974
0.324538	-0.58699	3.041046
1.267391	0.774517	1.5613391
1.593197	0.379216	0.3800469
0.915655	0.577578	3.7338996
0.569631	0.439358	0.3809431
0.844212	1.178303	4.433354
1.267131	1.946709	-0.865622
0.576426	0.98284	0.5462138
1.02236	0.87753	3.1453094
1.045767	0.512284	0.9832992
1.26094	2.249964	4.5257123
0.877254	-0.11401	3.8048765
0.780417	2.156753	1.7004283
0.601923	2.177259	-1.360886
0.84769	0.507348	-0.16514
0.452278	0.764985	1.6787819
1.849099	1.168802	0.6397555
0.95382	1.813824	1.7637404
1.833754	1.848142	-1.15271
1.072809	0.905758	1.2388257
1.03083	0.157541	0.8716614
1.980016	-0.09441	2.0387728
0.902798	0.584578	5.4352119

Continued on next page.

Table B-8. Continued.

1.66865	0.735153	1.423675
1.232639	0.717552	0.5396904
0.562207	1.0998	4.016813
1.854897	0.822438	0.9173806
1.406403	0.951848	2.9746332
0.371561	1.101905	0.5656432
1.112963	0.905231	0.9745913
1.140383	2.466092	-1.047837
1.702398	1.113093	2.5144999
1.556505	-0.51151	2.6021935

Table B-9. Data Sets for Gamma and Weibull Distributions for Parameter Estimation and K-S Test

n=10	n=20	n=50
0.0512	0.4854	0.5399
1.4647	0.2333	0.8592
0.4995	0.0814	0.6626
0.7216	0.3035	1.0968
0.1151	1.7358	0.8372
0.2717	0.9021	1.4874
0.7842	0.0667	0.5451
3.9898	0.0868	0.274
0.1967	0.8909	0.6352
0.8103	0.1124	0.4455
	2.8492	1.5651
	1.0417	0.9681
	0.2068	0.2442
	4.6191	0.3844
	1.9741	0.7742
	1.5957	0.5659
	1.6158	0.2304
	0.5045	2.8271
	1.3013	0.5061
		2.9904
		0.8786
		1.1874
		0.1343
		4.1991
		0.264
		0.0296
		0.01
		0.2372
		0.824
		0.6965
		1.542
		0.4408
		1.1393
		0.0407
		0.3193
		0.8868
		0.295
		1.317
		0.8212
		0.0689

Continued on next page.

Table B-9. Continued.

		0.3808
		1.5485
		0.1753
		0.464
		2.0116
		1.5744
		0.4989
		0.4622
		0.993
		0.5531

Table B-10. Data Sets of Gamma Distribution for Confidence Interval Test for Gamma and Empirical Distributions

μ=1, CV=0.5, n=50	μ=1, CV=1, n=50[a]	μ=1, CV=2, n=50
0.3495	0.0601	0.1240
0.8061	0.3119	0.4576
0.7780	0.4481	1.9215
2.9950	2.3902	0.0019
0.5772	4.5431	0.0990
1.0095	1.1284	0.0002
1.2915	0.4152	0.0042
1.3076	0.3355	0.9951
1.3639	2.9557	0.0170
0.9038	0.1026	5.2902
0.8323	0.2100	0.6898
0.6116	2.4059	0.7764
1.0571	0.7057	0.2163
0.8644	0.2634	3.7640
0.6304	1.0877	7.0598
0.5235	1.4813	0.0002
0.9455	1.1096	0.1302
1.0915	0.1061	0.2326
0.6012	1.1857	0.5471
1.6019	2.0174	1.6067
1.1438	0.5683	7.8384
1.0575	0.2362	0.1253
0.4541	0.4795	0.0730
0.3622	1.1851	0.0087
0.8123	0.2408	0.0393
2.1055	0.7300	0.3317
0.6358	0.7693	0.0081
1.5140	1.1722	0.1039
1.4096	0.5189	0.1621
1.7865	0.4107	0.0583
0.6721	0.2586	1.1142
0.2636	0.4835	0.1813
1.1551	0.1659	0.0799
1.3607	3.0559	0.7581
1.2498	0.6218	0.0305
1.4574	2.9759	0.0000[b]
0.7495	0.8157	0.0379
1.0114	0.7441	8.4097
1.4722	3.9121	2.5678
0.6928	0.5513	0.0000[b]

Continued on next page.

Table B-10. Continued.

1.6130	2.2717	0.0000[b]
0.9870	1.1144	0.4993
0.7406	0.2313	0.0535
0.8659	3.0870	0.0099
1.4405	1.4994	0.7027
0.9251	0.1262	0.0002
0.8307	0.8866	0.0008
0.6517	0.9366	0.0078
0.2690	2.3969	2.0831
0.3600	1.9045	0.0138

[a] This column of random numbers was also used for empirical distribution test

[b] 10^{-8} instead of 0 was used to input into AuvTool.

Table B-11. Data Sets of Weibull Distribution for Confidence Interval Test

μ=1, CV=0.5, n=50	μ=1, CV=1, n=50[a]	μ=1, CV=2, n=50
1.9039	0.3153	5.2990
0.5978	0.0127	2.6816
1.0927	0.5815	2.1355
0.9301	1.2358	0.2639
1.6499	1.1584	0.0378
1.3412	1.4805	0.9145
0.8921	1.0239	2.2486
0.1698	0.3815	2.5698
1.4625	1.5667	0.1570
0.8771	0.2888	4.4419
1.1049	2.2432	3.2990
1.3993	0.5836	0.2601
1.7622	0.2578	1.4947
1.2979	0.1610	2.9431
0.5172	0.3411	0.9560
0.8274	0.1958	0.6328
1.8242	0.5276	0.9334
1.7421	0.4313	4.3881
0.8332	0.6515	0.8599
1.6576	1.0983	0.3747
0.2951	0.0667	1.7927
0.7599	0.1024	3.1135
1.4442	0.2091	2.0364
0.1256	3.4456	0.8604
0.4571	1.5845	3.0833
0.557	0.1596	1.4497
0.5511	0.2757	1.3807
1.0884	0.1158	0.8723
0.6542	0.4753	1.9223
0.5512	1.1204	2.2647
0.1549	0.0292	2.9721
1.3131	2.5217	2.0243
0.8776	1.0073	3.6727
1.8067	0.8420	0.1433
0.9044	0.3588	1.6671
0.8439	0.1500	0.1541
1.5217	0.0749	1.3050
0.9813	1.1035	1.4243
0.5569	0.9749	0.0663
1.1891	1.4446	1.8357
1.5017	1.7104	0.2947

Continued on next page.

Table B-11. Continued.

0.1748	2.4092	0.9286
1.2034	1.2038	3.1467
0.794	0.2840	0.1393
1.4865	0.1130	0.6214
0.952	0.7812	4.1090
1.2488	1.7273	1.1997
0.8569	0.1702	1.1323
0.6973	3.3622	0.2623
0.5376	3.0228	0.4174

Table B-12. Random Samples Used for Reliability Test

0.4854	0.2303	0.0814	0.3035	1.7358
0.9021	0.0667	0.0868	0.8909	0.1124
2.8492	1.0417	0.2068	4.6191	1.9741
1.5957	1.6158	0.5045	1.3013	1.6154