Potential Statistical Models for Describing Species Sensitivity Distributions
CCME Project # 382-2006

Prepared by:
Zajdlik & Associates Inc.
March 3rd, 2006

PN 1415

Table of Contents

List of Tables

List of Figures

List of Equations

Table 1: Acronym Definitions

Acronym	Definition
CDF	cumulative distribution function
ECDF	empirical cumulative distribution function
GOF	goodness of fit
HC_5	hazardous concentration to 5% of species
MTC	maximum tolerable concentration
NOEC	no observed effect concentration
PDF	probability density function
PP	probability- probability
QQ	quantile-quantile
SSD	species sensitivity distribution

Citation Format:

Zajdlik, B.A. 2006. Potential Statistical Models for Describing Species Sensitivity Distributions. Prepared for the Canadian Council of Ministers of the Environment, January 2006.

1 Introduction

The sensitivity of a species to a contaminant varies from species to species. Some species are very intolerant while others are more tolerant. A cumulative frequency histogram of these sensitivities, will exhibit a shape, often sigmoidal with a horizontal asymptote delimiting a concentration that affects all species. A low concentration will affect only a small proportion of the species. The concentration corresponding to the concentration that puts x% of the taxa at risk is known as the hazard concentration$_x$ or HC_x.

The HC_x is simply a percentile and may be estimated by ranking the species sensitivities and choosing the appropriate concentration. However for a variety of reasons, this non-parametric approach is less preferred than a parametric approach.

When using the parametric approach a model or distribution is fit to the cumulative species sensitivities. The word "fit" means to estimate the parameters of the distribution. The purpose of this document is to describe some models or statistical distributions that have been found useful in describing SSDs for freshwater aquatic organisms (Zajdlik, 2005). A dozen or more other potentially suitable models exist and might be profitably considered. All the primary[1] models in this document allow toxicity test endpoints to approach zero[2] (the origin) and infinity. This choice is synonymous with rejecting the concept of a threshold concentration for an ecosystem (i.e. Some species exist that are so sensitive to the contaminant that they will be affected by any contamination.)

At this point in time, jurisdictions other than Canada advocate the use of a single parametric model to describe all species sensitivity distributions (SSD). In the event that the prescribed model is unsuitable, a non-parametric fallback position is available. Canadian guidance differs in allowing the modeler to choose a suitable distribution from the myriad possible distributions. In the event that no parametric models adequately fit the SSD, guidelines may be developed using the conventional 1991 approach, assuming that there are enough data for this method.

Sections 7, 8, and 9 are extremely brief overviews of their respective subjects. These sections are in no way sufficient for those not already familiar with modelling statistical distributions to generate freshwater quality guidelines using the SSD approach. Readers not very familiar with these topics are encouraged to consult a statistician if developing freshwater quality guidelines.

The reader unfamiliar with probability density functions and cumulative distribution functions and the concepts of location and scale are encouraged to review the appendices and the document conventions, below, prior to reading model descriptions. Acronyms pertaining to probability density functions and cumulative distribution functions are first introduced in the appendices. A glossary expressing potentially unfamiliar terms in the context of this document is also provided.

---

[1] Some models such as the Burr Type III closely approximate the triangular distribution which is bounded. The use of bounded distributions implies belief in an ecosystem threshold of effect.
[2] Some models allow the domain (or concentrations in terms of SSDs) to include negative numbers if the ratio of the location to scale parameters is small.

Document Conventions:

- In some instances a logarithmic transformation is required or desirable. In this document the natural logarithm is explicitly stated in order to 1) be prescriptive and 2) ensure consistency among all users. Note that any reasonable logarithmic transformation could be used.

- Capital letters refer to a random variable while the lower case letter refers to a single realization of the random variable or observation.

- The Greek letters μ (mu) and σ (sigma) are generally reserved to describe the mean and standard deviation of a random variable. When speaking of the normal distribution the mean of the random variable also describes the location of the distribution and the standard deviation describes the scale of the distribution. For other distributions the location and scale of the distribution are NOT described by the mean and variance of the raw data, respectively. Therefore other symbols are used to describe the location and scale of these distributions. Hence, the seeming arbitrariness of symbols is a mechanism to ensure that the familiar concepts of means and variances are not mistakenly applied to estimate the parameters of the non-normal distributions discussed herein.

  To the extent possible, α is used as a location parameter, β is used as a scale parameter and λ is used as a shape parameter in this document.

2 Burr Type III Distribution

Burr (1942) presented a series of easy to use, flexible CDFs useful for approximation purposes. The Burr Type III distribution was used by Shao (2000) to describe SSDs. Shao (2000) reparameterized the original 2-parameter Burr Type III model to the following three parameter model:

Equation 1: Burr Type III CDF

$$F(X;k,\lambda,\beta)=\frac{1}{\left(1+\left(\frac{\beta}{x}\right)^{\lambda}\right)^{k}},\quad for\ \beta,k,\lambda,x>0.$$

where:

- x is a toxicity test endpoint or the logarithm of the toxicity test endpoint ;
- β is a scale parameter;
- λ is a shape parameter; and
- k is a location-shape parameter.

The CDF for the Burr Type III 3-parameter model henceforth referred to simply as the Burr Type III distribution or model, is shown below illustrating the effects of changing the scale and shape parameters in Figure 1 and Figure 2, respectively.

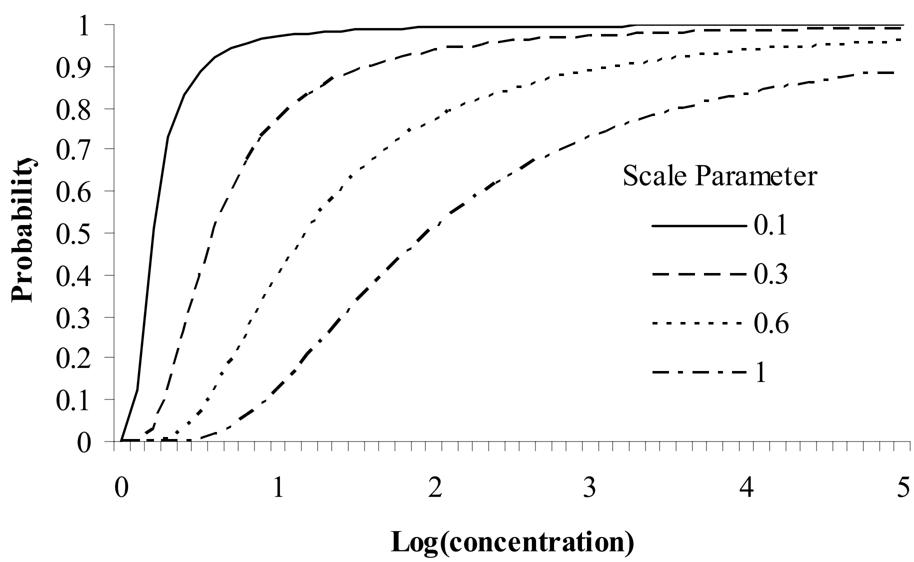

Figure 1: The Burr Type III Distribution Illustrating Changes in Scale Parameter

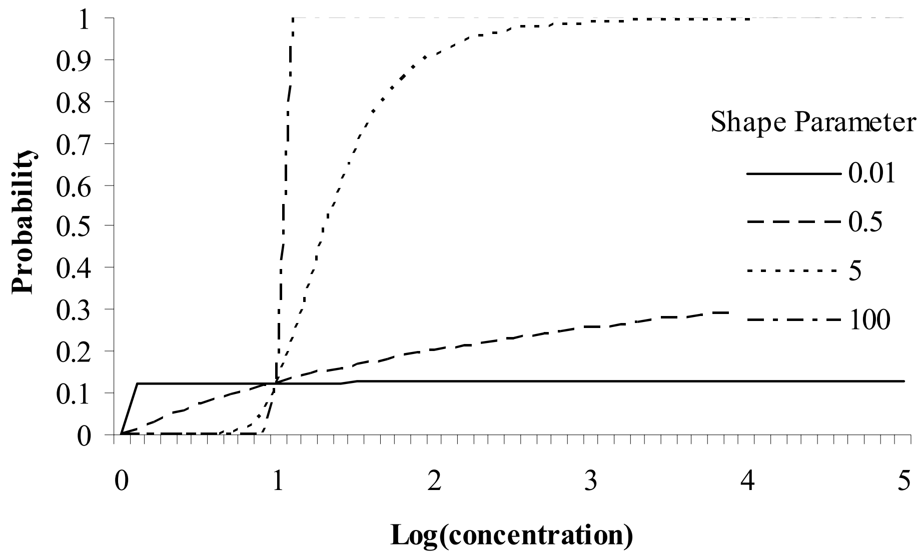

Figure 2: The Burr Type III Illustrating Changes in Shape Parameter

Note the extreme shape flexibility of the Burr Type III distribution as illustrated in Figure 2. The Burr Type III distribution was chosen as a candidate for describing SSDs for the following reasons:

- it is used in other jurisdictions such as Australia and New Zealand (ANZECC, 2000);
- includes the log-logistic distribution (which is used by OECD) as a limiting case;
- it approximates the log-normal and Weibull distributions which are also potential SSD descriptors;
- is mathematically tractable[3]; and;
- the sigmoidal shape is consistent with the differential sensitivity of species.

3 Extreme Value Distributions

Extreme value distributions are a family of distributions used to describe extrema (minimums or maximums) in a vast range of applied and theoretical sciences. Environmental examples include the frequency of forest fires, floods and earthquakes, prediction of maximum stream flows, wave heights, etc.

The Gumbel distribution (Gumbel, 1958) is one member of the extreme value distribution family. The Gumbel CDF is:

[3] The adjective "tractable" means that the model does not exhibit discontinuities, is easily differentiable, etc.

Equation 2: Gumbel CDF

$$F(x;\alpha,\beta) = \exp\left(-\exp\left(-\frac{x-\alpha}{\beta}\right)\right) \qquad for\ \beta > 0, -\infty < x, \alpha < \infty;$$

where:

- x is a toxicity test endpoint or the logarithm of the toxicity test endpoint ;
- α is the location parameter; and,
- β is the scale parameter.

Further information on the Gumbel distribution may be found in Gumbel (1958) and Johnston et al (1995). The Gumbel CDF illustrating the effect of changing the scale parameter is presented below.

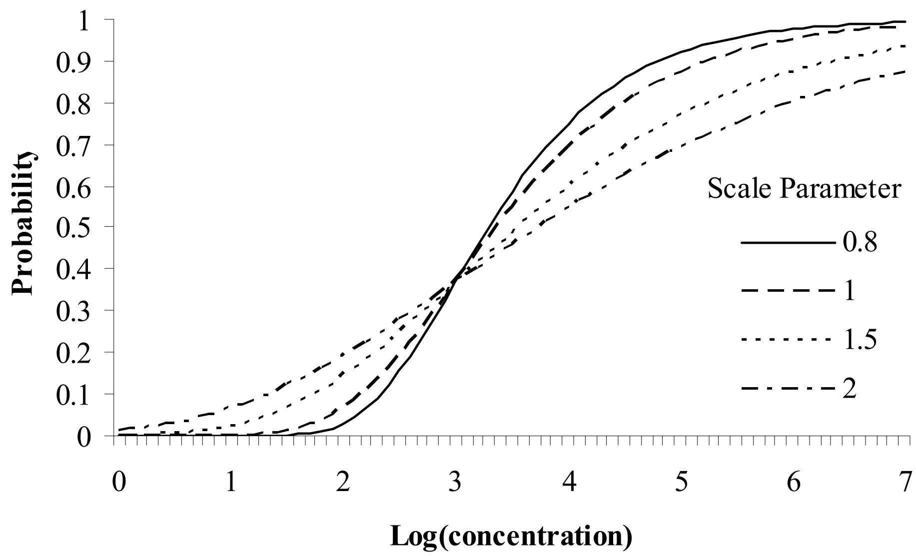

Figure 3: Gumbel Distribution Illustrating Changes in Scale Parameter

Changes in the location parameter merely shift the distribution to the right or left as illustrated for the logistic distribution in Figure 4. The Gumbel distribution was chosen as a candidate for describing SSDs for the following reasons:

- it was identified as a potentially useful empirical descriptor of SSDs by Zajdlik (2005);

- is mathematically tractable; and,

- the sigmoidal shape is consistent with the differential sensitivity of species.

WARNING: One potential problem with this distribution is that if the location parameter is too small, the model can predict negative concentrations.

4 Logistic Distribution

The logistic curve has been used both as a model and a statistical distribution. As a model, the logistic curve has been used to model growth from a lower to an upper asymptote. The shape of the logistic distribution is similar to the normal distribution. The logistic model closely approximates the Student T distribution with 9 degrees of freedom (Mudholkar and George, 1978.).

The mathematical formula describing the logistic distribution has long been used as an empirical model describing dose-response data (Berkson, 1944) and more recently discussed in van Ewijk and Hoekstra, (1993) and Environment Canada, (2005). The empirical description of dose response data using the logistic distribution makes "sense" in the same way that the normal cumulative distribution does when applied to such data; differing sensitivities of individuals give rise to a sigmoidal tolerance distribution. The logistic CDF has been empirically applied to the distribution of sensitivities of species for the same reason as it has been applied to individuals; species have differing sensitivities to a contaminant.

The method discussed by Aldenberg and Slob (1993) provides the basis for the derivation of environmental quality criteria in the Netherlands. In the Netherlands, the lower 50% confidence limit of the 5^{th} percentile of the distribution of NOECs (after natural logarithmic transformation) becomes the HC5. The HC5 so calculated is equivalent to the OECD MTC. (OECD, 1995).

The equation for the logistic CDF is presented below. Note that the variable x refers to ln(endpoint).

Equation 3: The Logistic CDF

$$F(x;\alpha,\beta) = \frac{1}{1+\exp\left(-\frac{x-\alpha}{\beta}\right)},$$

where:

- x is the ln(endpoint). Due to the logarithmic transformation, x is > 0.
- α is the location parameter; and,
- β is the scale parameter.

The logistic CDF illustrating the effect of changing location and scale parameters is presented below in Figure 4 and Figure 5, respectively.

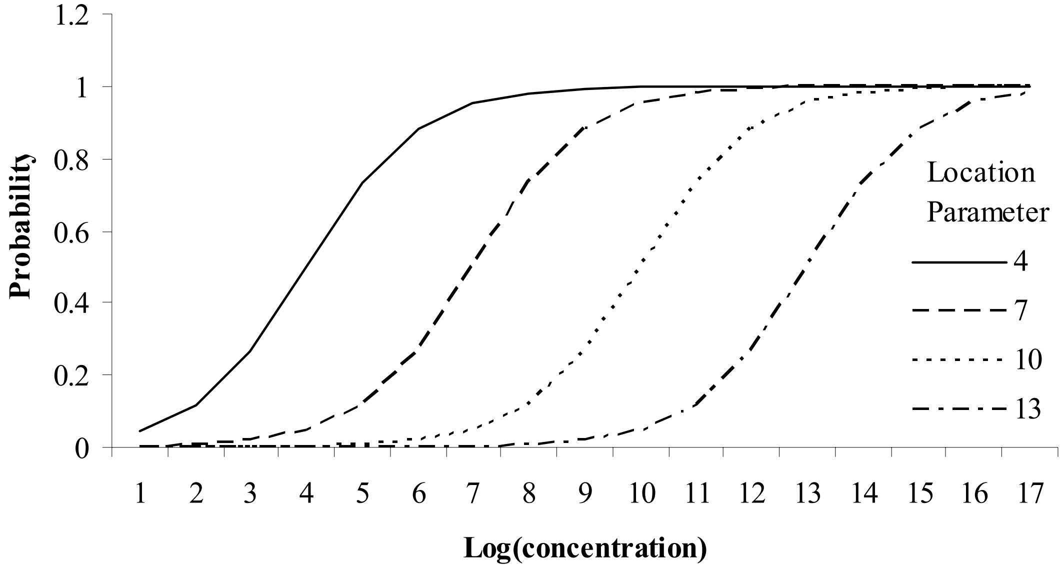

Figure 4: The Logistic Distribution, Change in Location Parameter

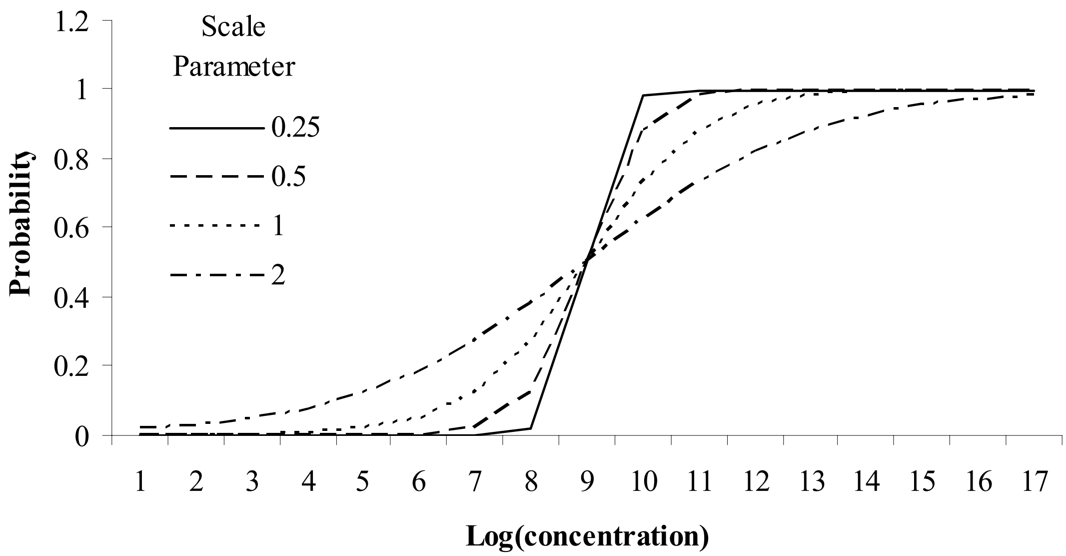

Figure 5: The Logistic Distribution, Change in Scale Parameter

The logistic model was chosen as a candidate for describing SSDs for the following reasons:

- is used in other jurisdictions including the Netherlands and by OECD (OECD. 1995);
- is mathematically tractable; and,
- the sigmoidal shape is consistent with the differential sensitivity of species.

Note however that the logistic distribution was never the best descriptor of the SSDs examined in Zajdlik (2005).

WARNING: One potential problem with this distribution is that if the location parameter is too small, the model can predict negative concentrations.

5 Lognormal and Normal Distributions

The normal distribution has a long history of usage and thus a consequent attendant richness of methodology and software. The normal distribution was applied to toxicological data as long ago as 1934 (Bliss, 1934) and has in Canada been until very recently (Environment Canada, 2005) the prescribed distribution when estimating endpoints from toxicity test results when the response is binary.

When data are positively skewed[4] a logarithmic[5] transformation may induce a normal distribution. In this case, the data on the original (untransformed) scale are lognormally distributed. Due to the abundance of tools for assessing normality and since estimation of parameters following logarithmic transformation follows those used to estimate the parameters of the normal distribution <u>the remainder of this section presents the normal distribution unless otherwise stated</u>.

The first application of the normal distribution to logarithm(toxicity test endpoints) in an SSD context we are aware of is by Wagner and Lokke (1991). A more recent example is Aldenberg <u>et</u> <u>al</u>, (2002). Generally, the similarly shaped logistic distribution has been used to describe SSDs due to its mathematical tractability, especially relative to the normal distribution.

[4] Skewness is a measure of asymmetry in a frequency histogram. A frequency histogram that is perfectly symmetrical has skewness = 0. A frequency histogram that has a larger tail to the right is positively skewed while a frequency histogram that has a larger tail to the left is negatively skewed.
[5] Refer to section 1 for a discussion of logarithmic transformations in the context of this document.

The equation for the normal CDF is presented below. Note that in the context of this section, the variable x refers to ln(endpoint).

Equation 4: The Normal CDF

$$F(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left\{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^{2}\right\}\partial t,$$

where:

- x is the ln(endpoint); Due to the logarithmic transformation, x is > 0.
- μ is the location parameter and mean of the ln(endpoints) ;
- σ is the scale parameter and standard deviation of the ln(endpoints).

The effect of changing scales on the normal CDF is illustrated below.

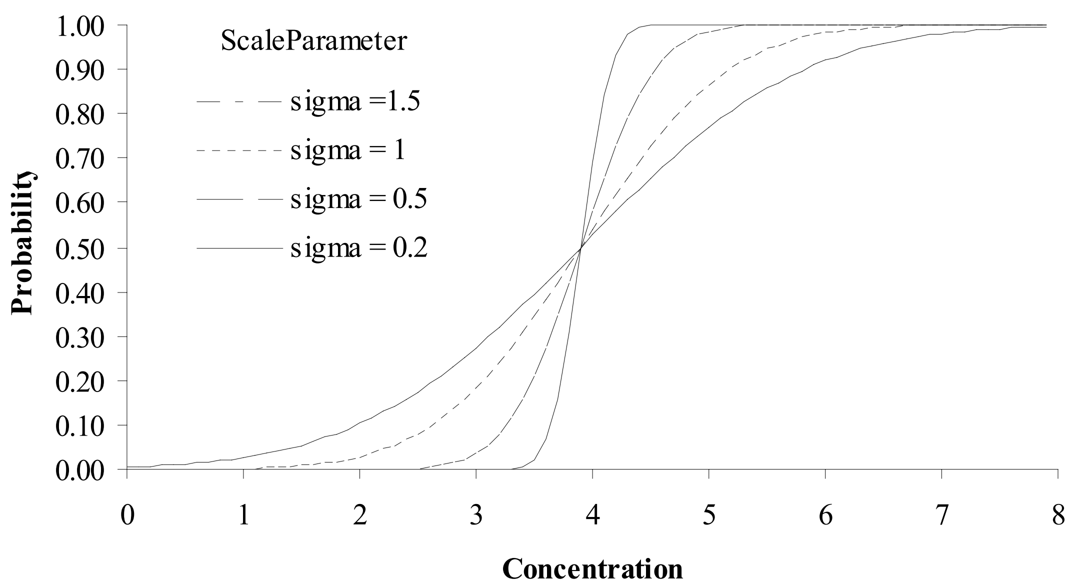

Figure 6: Effect of Changing Scale on Normal CDF

A comparison of the logistic and normal CDFs illustrates the similarities and differences between the two CDFs.

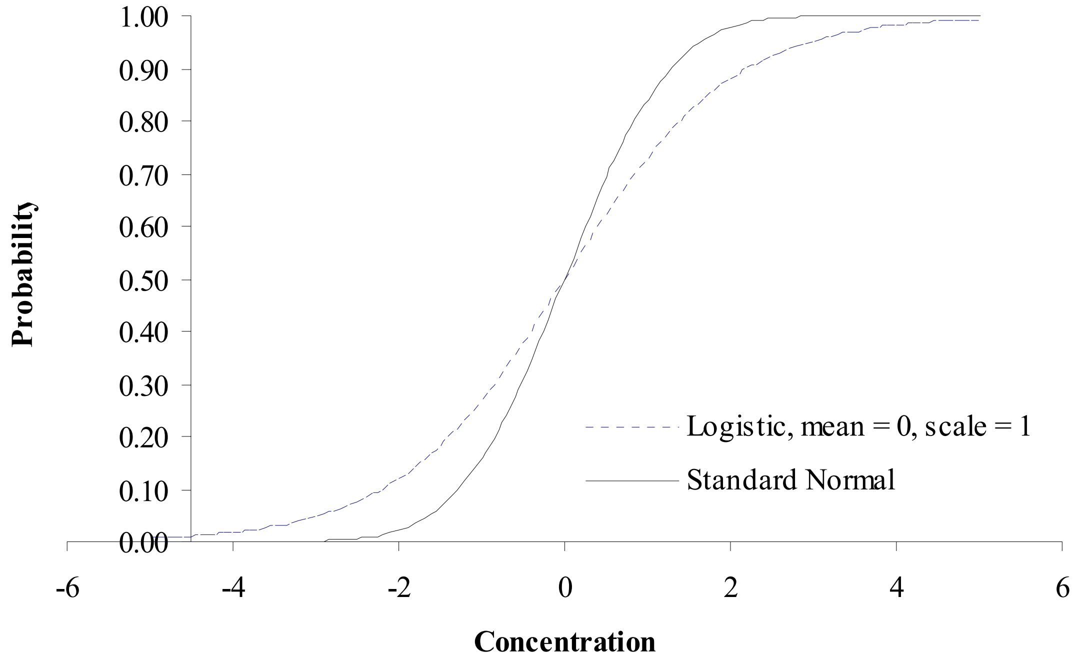

Figure 7: Logistic versus Normal Cumulative Distribution

Note the similarity in shape between the normal and logistic CDFs but also the larger tails of the logistic distribution. Assuming that the logistic distribution is the best descriptor of a species sensitivity dataset, erroneously choosing the normal distribution (applied to ln(endpoints)) will bias the estimated HC_5 upward, leaving the environment unprotected at the anticipated level of protection.

The normal distribution was chosen as a candidate for describing SSDs for the following reasons:

- the sigmoidal shape is consistent with the differential sensitivity of species; and,
- the lognormal distribution was selected as the best descriptor of one of the data sets examined by Zajdlik (2005).

6 Weibull Distribution

The Weibull distribution is a member of the family of extreme value functions for minima. It may be thought of as a “weakest link” function. The Weibull distribution with shape parameter = 1 is equivalent to the exponential distribution. Both the exponential and Weibull distributions are asymmetric about their means and may be used to model asymmetric tolerance distributions.

Both these distributions are used extensively in modelling time-to-failure data (Lawless, 1982, Klein and Moeschberger, 2003 and Lee and Wang, 2003). Christensen (1984) advocated using the Weibull distribution in lieu of the probit (cumulative normal) method to describe dose responses generated by aquatic toxicity tests.

To the best of our knowledge the first application of the Weibull distribution to species sensitivity distributions is by Zajdlik (2005) although Shao (2000) notes that the reciprocal Weibull is a limiting case of the Burr type III distribution when applying that distribution to SSDs. Further information on the Weibull distribution may be found in Johnston et al (1995).

The cumulative distribution function for the Weibull distribution is:

Equation 5: The Weibull CDF

$$F(x;\beta,\lambda) = -1 - \exp\left(-\left(\frac{x}{\beta}\right)^{\lambda}\right)$$

where:

- x is a toxicity test endpoint or the logarithm of the toxicity test endpoint ;
- β is a scale parameter; and
- λ is a shape parameter.

The Weibull CDF illustrating the effect of changing shape parameters is presented below. Note that the Weibull distribution with shape parameter = 1 is the exponential distribution.

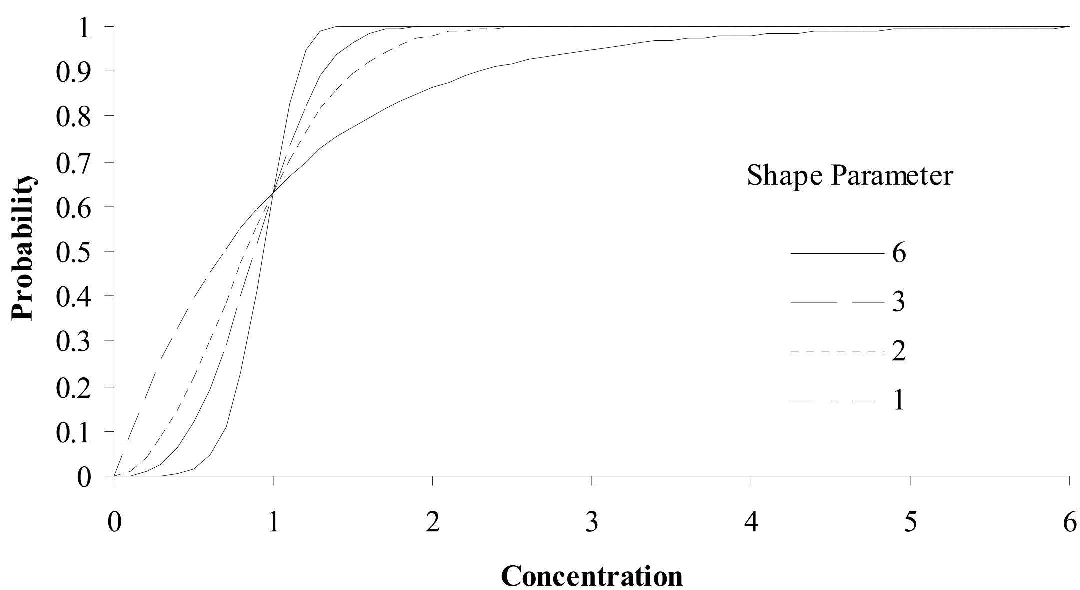

Figure 8: Weibull Distribution CDFs with Changing Shape Parameter

It is surprising that the Weibull distribution has not previously been considered as a descriptor of SSDs. The Weibull distribution was chosen as a potential descriptor of SSDs for the following reasons:

- the sigmoidal shape is consistent with the differential sensitivity of species;
- it is widely used in interpreting time-to-failure data;
- it has the desirable property of an asymmetric CDF which may be important when the distribution of sensitivities to a contaminant are reasonably expected to be asymmetric due to the mode of toxic action. An example is an herbicide where the mode of toxic action is generally markedly different between plants and animals;
- it is mathematically tractable; and
- from a potential choice of normal, lognormal, logistic, log-logistic, inverse Gaussian, gamma, Weibull, exponential, extreme value and Laplace distributions to describe 6 different SSDs (2,4D, atrazine, copper, formaldehyde, nitrate, uranium and zinc), the Weibull distribution was the best descriptor of 3 of the 6 data sets. (Zajdlik, 2005).

7 Assessing Goodness of Fit

This section on assessing goodness of fit was written to address the implication of the Canadian position to allow choice when modelling SSDs to generate environmental quality guidelines. Modelers must be able to decide among potentially suitable models as the choice of model can have a very large effect on an environmental quality guideline. This section is not a comprehensive evaluation of the literature and provides only a very brief introduction to the topic. Readers not confident in their ability to evaluate models to describe SSDs are strongly encouraged to consult a statistician.

There are two general classes of tools for assessing the goodness of fit of a tentative model; graphical methods and formal tests. Graphical methods for <u>assessing</u> goodness of fit are widely used and often available in statistical packages. <u>Tests</u> for goodness of fit are often formalizations of graphical procedures; in that departures from a specific shape are measured and summed in a numeric way to generate a test statistic. Some formal tests may be generically applied to a variety of distributions; other formal tests are designed specifically for a given distribution. Tests designed for a specific distribution other than the normal distribution are not usually available in statistical software.

7.1 Graphical Techniques

Graphical assessments of GOF:

- may be especially informative about lack of fit in one particular region of an ECDF that might be missed by a formal test; and,
- are available in many statistical software packages or straightforward to generate given availability of functions for PDFs and CDFs of the distributions being entertained.

Sole reliance on graphical techniques is not advocated since random variation, especially in small samples may lead to incorrectly dismissing a distribution.

7.1.1 Frequency Histograms

Empirical frequency histograms (empirical PDFs) may be compared to the PDF of a candidate model. Empirical PDFs provide robust estimates of location, dispersion and skewness. Empirical PDFs are not recommended for assessing distributions unless the modeler is experienced as 1) the choice of the width of the classes (boxes) in the histogram, and 2) random variation in small samples, can markedly affect the appearance of the histogram.

7.1.2 ECDF and CDF Based Plots

CDF- based plots are preferred over the frequency histogram (or empirical PDF) for the following reasons:

- the complexity of the ECDF or CDF is not dependent upon the number of observations; and
- there is no effect of bin width on the appearance of the graphic.

Two CDF-based plotting methods are presented below.

7.1.2.1 Probability-Probability Plots

The ECDF provides the probability of observing an observation less than some value as described in

Appendix 2: Cumulative Probability Distribution Functions. After fitting a hypothesized model the cumulative probabilities corresponding to the ordered raw data are plotted against the empirical cumulative probabilities. If the hypothesized model adequately describes the data set, a straight line should result.

The following PP plot uses the fish mass data set presented in Appendix 1: Probability Density Functions. The hypothesized distribution is the normal.

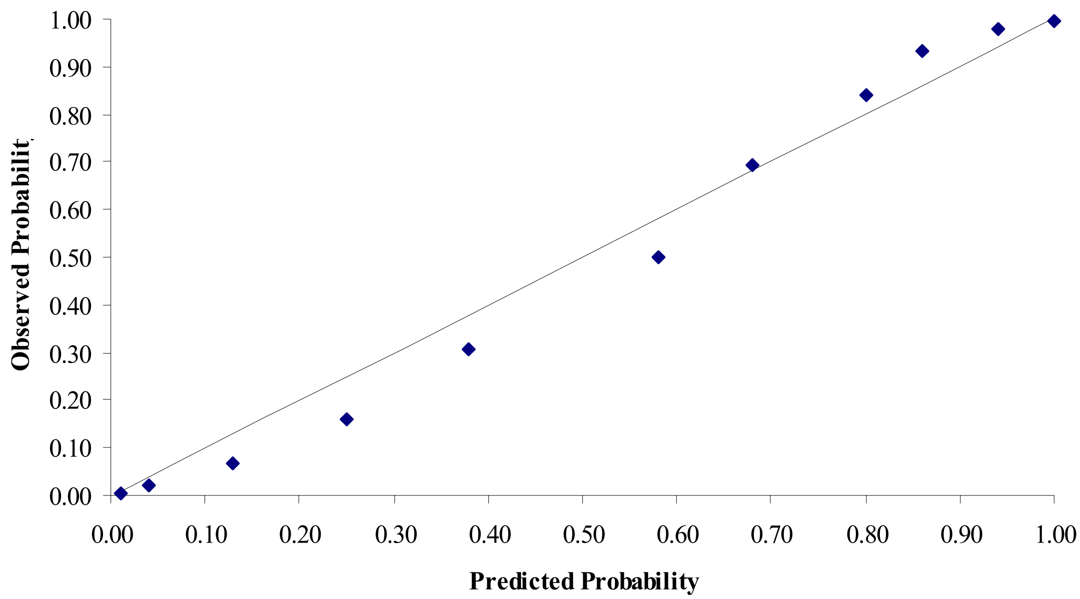

Figure 9: PP Plot for Fish Mass Data

The straight line in Figure 9: PP Plot for Fish Mass Data, represents perfect agreement between predicted and observed probabilities. The square symbols represent the observed versus predicted probability pairs and should fall along the straight line if the data are generated by the hypothesized (normal in this case) distribution. Figure 9 reasonably supports the normal model for this data set although there seems to be a lack of fit in the middle of the distribution. PP plots typically exaggerate the lack of fit in the middle of a distribution.

7.1.2.2 Quantile-Quantile Plots

Quantile-quantile plots are similar to probability-probability plots but quantiles of the observed data are plotted against quantiles of the hypothesized distribution. Again, the parameters of the hypothesized distribution are estimated from the data. For QQ plots the fitted model is used to predict the quantiles at points corresponding to the ranked position[6] of the ordered raw data.

The following QQ plot uses the fish mass data set presented in Appendix 1: Probability Density Functions. The hypothesized distribution is the normal.

---

[6] There are various ways to determine the ranked position or "plotting position". A discussion of the relative merits of these plotting positions is beyond the scope of this document.

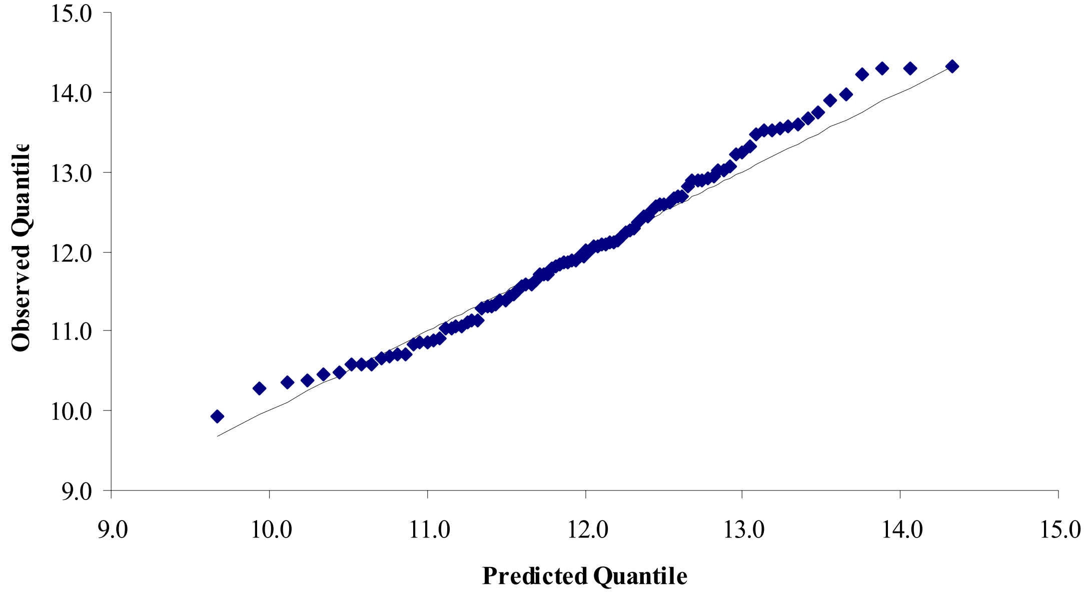

Figure 10: QQ Plot for Fish Mass Data

The straight line in Figure 10 represents perfect agreement between predicted and observed quantiles. The square symbols represent the observed versus predicted quantile pairs and should fall along the straight line if the data are generated by the hypothesized (normal in this case) distribution. Figure 10 reasonably supports the normal model for this data set although there seems to be a lack of fit in the tails of the distribution. QQ plots typically exaggerate the lack of fit in the tails of a distribution. Hence, QQ plots are preferred over PP plots when graphically assessing potential SSD models due to interest in the tails of the distribution.

7.2 Formal Tests

Formal GOF tests:

- should be relied upon to reject a distribution for data sets with sample sizes < 15[7];
- should not be relied upon to reject a distribution for large data sets
- the failure to reject a distribution should not lead to automatic acceptance of that distribution. Final acceptance must be augmented by graphical assessments and/or power analyses.

Sole reliance on formal GOF test is not advocated due to the lack of power in small to moderate samples, excessive discrimination for large samples and potential de-emphasis on a good fit in the lower tail.

[7] The recommendation to use the number 15 is based on professional judgment only and is provided to give non-statisticians a clear path forward. The number chosen is likely very conservative; sophisticated users can use power calculations to sidestep this issue.

Goodness of fit (GOF) tests are used to determine whether a postulated model adequately describes the observed or available data. The adverb “adequately” may be defined using statistical concepts. In this document, adequately refers to the level of significance. **The CCME requires using the generally accepted 5% level of significance.**

WARNING: The most common software implementations of GOF tests require that the parameters of the hypothesized distribution be known and NOT estimated from the data set being tested. Users should be aware of this restriction.

D’Agostino and Stephens (1986) provide modifications to the critical values and sometimes test statistics, to deal with the problem of simultaneously estimating parameters and conducting goodness of fit tests for the normal, exponential, extreme value and logistic distributions for both the Kolmogorov-Smirnov (section 7.2.1.3) and Anderson-Darling (section 7.2.1.1) tests. Monte Carlo analyses may be used to address this problem when considering other distributions.

7.2.1 Generic Tests

7.2.1.1 Anderson-Darling

The Anderson-Darling test is less widely used or available than the Kolmogorov-Smirnov test. The Anderson-Darling test, like the Kolmogorov-Smirnov test, compares the ECDF and theoretical CDFs. However the test statistic is a weighted average of the squared differences between the two distributions. Due to the weighting used, the Anderson-Darling test is more sensitive to discrepancies in the tails of distributions. Finally, the Anderson-Darling test generally exhibits more power than other ECDF-based tests when parameters are estimated (D’Agostino and Stephens, 1986).

The Anderson-Darling test is recommended for general use when assessing the GOF for SSDs in the absence of more powerful tests.

7.2.1.2 Chi-Square

The chi-square test squares the differences between the frequencies in an empirical histogram and the expected frequencies. One weakness of this approach is also that of frequency histograms; the shape of the histogram and by extension the chi-square test statistic, is a function of the number of bins. Also, the chi-square test will not likely be applicable for the majority of small to medium SSD data sets as the rule of thumb for its defensible application is that the number of observations within each arbitrary class is 5. Finally, the power of this test is generally low.

Note that the suggestion that the chi-square test may be adjusted to compensate for estimated parameters is commonly encountered. This suggestion is a matter of debate among researchers.

The chi-square test is not recommended for general use when assessing the GOF for SSDs.

7.2.1.3 Kolmogorov-Smirnov

The Kolmogorov-Smirnov test is a widely used and available nonparametric comparison of the ECDF and theoretical CDF. The test statistic is the absolute value of the maximum difference between the two distributions. The Kolmogorov-Smirnov tests is generally more powerful than the chi-square test and is more sensitive to location shifts (lateral shifts in distributions) than differences in the tails of distributions. However the Kolmogorov-Smirnov test exhibits poor power relative to other ECDF-based tests when parameters must be estimated.

The Kolmogorov-Smirnov test is grudgingly recommended when assessing the GOF for SSDs in the absence of more suitable generic or distribution-specific tests such as the Anderson-Darling test.

7.2.2 Distribution Specific Tests

7.2.2.1 Shapiro-Wilks

The Shapiro-Wilks test (Shapiro and Wilk, 1965) tests the strength of the correlation between the predicted and observed probabilities in a PP plot (section 7.1.2.1). This distribution-specific test is preferred over the generic test.

7.3 Recommendations

The following recommendations should be considered in light of the caveat prefacing this section; a comprehensive evaluation of the goodness of fit literature was not conducted.

- Until such time as specialized methods and/or further guidance are developed to address the particular needs of those using SSDs to generate environmental quality guidelines, **the CCME requires using formal tests for goodness of fit in conjunction with graphical assessments of goodness of fit.** This policy will obviate one potential source of controversy but does not discourage the use of graphical techniques. Readers not confident in their ability to evaluate models to describe SSDs are strongly encouraged to consult a statistician.

- Use formal goodness of fit tests specially designed for a distribution rather than generic tests unless the literature or more comprehensive guidance provides evidence to the contrary. Consultation with a statistician may be required to determine if a formal GOF test exists for a specific distribution.

- Use the Shapiro-Wilk test to test for normality.

- Use quantile-quantile plots to graphically assess the goodness of fit over other graphical methods, whenever possible.

- Do not use the chi-square test.

- The Kolmogorov-Smirnov test should be avoided.

- The Anderson-Darling test is generally preferred over other ECDF-based tests when parameters are estimated.

8 Criteria for Choosing Between Models

This section like the previous, is a thumbnail sketch of the literature and provides only general guidance. Readers interested in further guidance on this topic should consult the literature. Currently, the most definitive textbook on the subject is D'Agostino and Stephens (1986). Readers not comfortable with choosing between potential distributions would be prudent to consult a statistician when estimating environmental quality guidelines, particularly in the small sample case.

One numeric criterion that could be used to choose between contending models is a comparison of a test statistic for the various candidate models. When using the Kolmogorov-Smirnov or Anderson-Darling tests the model corresponding to the smallest value indicates the "best" fitting model. However the term "best" is a function of the test. The Kolmogorov-Smirnov test is best at detecting location shifts; therefore the "best" model using this criterion may not be the best model for choosing a model for estimating environmental quality guidelines where a good fit of the tails is paramount.

Readers should be aware that even the best fitting model may be a poor descriptor of the SSD and failure to reject a model should not be interpreted as "This model is correct" but rather "There was insufficient evidence to reject the current model". Strong reliance on the failure to reject a hypothesis might be countenanced if suitable power studies are available.

Modelers should use graphical techniques in conjunction with formal methods and understanding of the formal methods, in order to choose the best fitting model and also to decide when even the best-fitting model inadequately represents an SSD. When in doubt regarding contending models for the purposes of generating an environmental quality guideline, modelers should consult a statistician.

9 Softwar e

The parameters of a distribution (or model) must be estimated before the model may be used to estimate an HC_5 and indeed even before settling upon a final model. In some cases closed form solutions are available and parameters may be estimated manually. An example is the normal distribution where the sample location and scale parameters are respectively, the mean and standard deviation of the SSD dataset. Estimating parameters for most distributions requires software.

Software may take two forms: 1) software that directly estimates the parameters of a distribution or 2) software that allows the user to input a function to be optimized. Of the two forms only the first is generally approachable. The second form requires advanced knowledge of statistics and optimization. Every major statistical package has the 2nd capability while direct estimation of parameters of a distribution is generally available in most major statistical packages for a generous variety of distributions.

Once parameters are estimated, graphical assessments of GOF are required. Empirical frequency histograms and ECDFs may be plotted manually or using widely available graphical routines in spreadsheets such as Microsoft Excel. Some statistical software automatically generates PP or QQ plots for commonly encountered distributions such as the normal distribution. Other distributions will require the modeler to construct their own PP or QQ plots. Generating PP or QQ plots requires probabilities or quantiles for each posited distribution. These could be tediously extracted from tables but programmed functions are highly desirable. Programmed PDF and CDF functions are generally available in most major statistical packages.

Some GOF tests are generally available. These include the less desirable chi-square and Kolmogorov-Smirnov test. The Anderson-Darling test is less widely available. All software we are aware of [8]requires that the parameters be known and NOT estimated from the data. Therefore SSD modelers will be required to suitably adjust test results.

Using the Burr Type III distribution will require greater efforts by modelers. Programmed PDF and CDF functions for the Burr Type III model are not widely available nor is software specifically designed to estimate its parameters. One exception is the software BurrliOZ offered by CSIRO (2000). It fits the Burr Type III distribution and provides an easy-to-use graphical interface for non-statisticians. One concern with this software is that there does not seem to be an option to reject the Burr Type III distribution in the event that it is unsuitable.

Programmed PDF and CDF functions for the generalized F distribution are not widely available nor is software specifically designed to estimate its parameters.

[8] Information in this section was that available to the author without reviewing software or searching the literature.

10 Citations

Aldenberg, T. and W. Slob. 1993. Confidence limits for hazardous concentrations based on logistically distributed NOEC data. Ecot. Env. Safety. 25:48-63.

Aldenberg, T., J.S. Jaworska and T.P. Traas. 2002. Normal species sensitivity distributions and probabilistic ecological risk assessment. In: Sensitivity Distributions in Ecotoxicology, Eds: L. Posthuma, G. W. Suter II and T. P. Traas. Lewis Publishers, Boca Raton.

ANZECC, 2000. Australian and New Zealand Guidelines for Fresh and Marine Water Quality. Paper No. 4.

Berkson, J. 1944. Application of the logistic function to bioassay. J. Amer. Stat. Assoc. 39:357-365.

Bliss, C.I. 1934. The method of probits. Science. 79:38-39.

Burr I.W. 1942. Cumulative frequency functions. Annals of Mathematical Statistics 13: 215-232.

Christensen, E.R. 1984. Dose-response functions in aquatic toxicity testing and the Weibull model. Wat. Res. 18:213-221.

CSIRO, 2000. Burrlioz, ver. 1.0.14.

D'Agostino, R.B. and M.B. Stevens. 1986. Goodness of Fit Techniques, Marcel Deker

Environment Canada. 2005. Guidance Document on Statistical Methods for Environmental Toxicity Tests. EPS 1/RM/46.

Gumbel, E.J. 1958. Statistics of Extremes. Columbia University Press, New York.

Johnson, N.L., S. Kotz and N. Balakrishnan. 1995. Continuous Univariate Distributions, Vol. 2. John Wiley and Sons, New York.

Klein, J.P and M.L. Moeschberger. 2003. Survival Analysis. Springer-Verlag, New York.

Lawless, J. F. 1982. Statistical models methods for lifetime data. Wiley and Sons, Toronto. 580pp.

Lee, E.T. and J.W. Wang. 2003. Statistical Methods for Survival Data Analysis, Third Edition. John Wiley and Sons, New York.

Mudholkar, G.S. and E.O. George. 1978. A remark on the shape of the logistic distribution. Biometrika, 65:667-668.

OECD. 1995. Guidance Document for Aquatic Effects Assessment. OECD Environment Monographs No 92, Organisation for Economic Co-operation and Development, Paris.

Shao, Q. 2000. Estimation for hazardous concentrations based on NOEC data: an alternative approach. Environmetrics. 11:583-595.

Shapiro, S.S. and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples) Biometrika. 52:591-611.

Van Ewijk, P.H. and J.A. Hoekstra. 1993. Calculation of the EC50 and its confidence interval when subtoxic stimulus is present. Ecotox. Env. Saf. 25:25-32.

Wagner, C. and H. Lokke. 1991. Estimation of ecotoxicological protection levels from NOEC toxicity data. Wat. Res. 25(10):1237-1242.

Zajdlik, B.A. 2005. Statistical Analysis of the SSD Approach for Development of Canadian Water Quality Guidelines. Prepared for the Ontario Ministry of Environment and the Canadian Council of Ministers of the Environment.

Appendix 1: Probability Density Functions

A frequency histogram may be used to illustrate some useful concepts regarding statistical distributions.

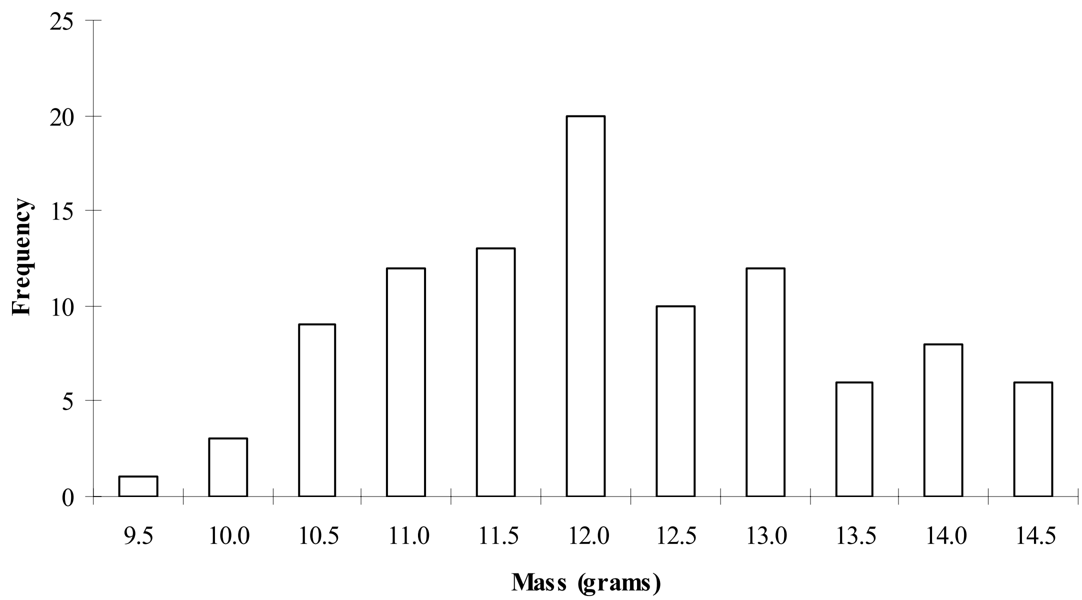

Figure 11: Frequency Histogram of Fish Weights

Each box in the frequency histogram represents a range of fish weights. The first box represents the number of fish with weights between 9.25 and 9.75 grams with a mid-point of 9.5 grams. If this number is divided by the total number of fish, the results are expressed as a proportion or probability. Figure 11, above represents 100 fish, therefore 1/100 fish fall into the first size class and the probability of falling into the first size class is 0.01.

Given these size classes, and this probability histogram a potential question is: "How likely would it be to randomly select a fish (from the same group of fish from which the original 100 fish were randomly selected) with a weight greater than 13.75 grams?" The answer is 14/100 or 14% (8 fish in the 13.75-14.25g size class + 6 fish in the 14.25 -14.75g size class). The answer 14%, is a function of the distribution of the fish weights across the size classes. Another question is "How likely is it that a randomly selected fish weighs between 11.75 and 12.25 grams? The answer is 12%. Note that if the shape of the distribution changes so do our answers.

Fish weight can take on any value within a certain minimum and maximum. Thus it is an example of a continuous random variable. The probability of observing any single value is zero for continuous data because in theory we can always increase the precision of the measurement to the point where the probability of observing that specific value becomes vanishingly small. However the probability that a random variable falls between two values can be estimated from a frequency histogram as shown above. A probability distribution generated from data as described above is an empirical probability distribution function. A mathematical function describing the density of probability is known as a probability density function or pdf. The pdf for the normal distribution is given below.

Equation 6: The Normal Probability Density Function

$$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \mathit{for}\ \sigma>0, -\infty<x<\infty, \mu<\infty,$$

where:

- x is the observed value;
- μ is the population mean. This is known as a location parameter as it describes where on the "x"-axis, the center of the distribution is.; and,
- σ is the population standard deviation; This is known as a scale parameter as it describes how spread out, on the "x"-axis, the distribution is.

Both μ and σ are parameters of this distribution. They are estimated from the data.

This pdf produces the familiar bell shaped curve encountered in many introductory statistics courses. Other pdfs corresponding to other statistical distributions have other characteristic shapes. How can we use this information to determine what distribution might represent a given data set?

As shown above, the frequency histogram (scaled to represent probability) is an empirical representation of a pdf. If we plot a pdf and it matches the pdf from a known distribution we have empirical evidence that the observed data corresponds to the putative distribution. Formal tests are also available to determine whether a dataset follows a specific distribution. Most of these tests involve the cumulative distribution function as described in

Appendix 2: Cumulative Probability Distribution Functions.

Appendix 2: Cumulative Probability Distribution Functions

A pdf describes the probability that a random variable falls between two specified values. What if we are interested in the probability that a random variable is less than some value? A practical method is to develop a formula that integrates the area under the probability density function up to the value of interest.

This function is known as the cumulative distribution function. Cumulative distribution functions are used to describe species sensitivity distributions. Members of a species exposed to a given concentration of a substance will elicit some average effect[9]. This average effect will range anywhere from 0% (no individuals were affected to 100% (all individuals were affected). The average percent affected for a species can be translated to the probability of being affected simply by dividing by 100% as discussed in Appendix 1: Probability Density Functions.

Now consider the sensitivities of different species to the same substance. Some species will be very sensitive while others will be less sensitive. Following the description in Appendix 1: Probability Density Functions, an empirical pdf can be generated. The pdf can be used to generate an empirical CDF or ECDF. The ECDF is used in goodness of fit tests (section 7).

Appendix 3: Glossary

location parameter: a parameter of a distribution that describes where the data are located along the measurement scale. The arithmetic mean is the location parameter for the normal distribution and locates the center of the data within the measurement scale. (See also *scale* and *shape parameter*.)

scale parameter: a parameter of a distribution that describes where the spread of data along the measurement scale. The standard deviation around the mean is the scale parameter for the normal distribution and describes the spread of data along the measurement scale. (See also *location* and *shape parameter*.)

shape parameter: a parameter of a distribution that describes the shape of the distribution. The common normal distribution does not have a shape parameter. (See also *location* and *scale parameter*.)

[9] The term effect may represent any effect such as mortality, a percent inhibition etc.