

Computing Maximum-Likelihood Estimates for Parameters of the National Descriptive Model of Mercury in Fish

Open-File Report 2012–1181



Computing Maximum-Likelihood Estimates for Parameters of the National Descriptive Model of Mercury in Fish

By David I. Donato

Open-File Report 2012–1181

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2012

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <http://www.usgs.gov> or call 1-888-ASK-USGS

For an overview of USGS information products, including maps, imagery, and publications, visit <http://www.usgs.gov/pubprod>

Suggested citation:

Donato, D.I., 2012, Computing maximum-likelihood estimates for parameters of the National Descriptive Model of Mercury in Fish: U.S. Geological Survey Open-File Report 2012-1181, 16 p., at <http://pubs.usgs.gov/of/2012/1181>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Contents

1	Introduction	1
2	Maximum-Likelihood Estimates for a Linear Model	3
2.1	A Linear Model.....	3
2.2	The Likelihood Function for the Linear Model	3
2.3	The Log-Likelihood Function for the Linear Model	6
2.4	Maximizing the Likelihood and Log-Likelihood Functions for the Linear Model.....	6
3	Maximum-Likelihood Estimates for the National Descriptive Model of Mercury in Fish	7
3.1	Contributions to the Likelihood from Censored Observations	8
3.2	The Likelihood Function for the National Descriptive Model of Mercury in Fish	10
3.3	The Log-Likelihood Function for the National Descriptive Model of Mercury in Fish	11
3.4	Maximizing the Likelihood and Log-Likelihood Functions for the National Descriptive Model of Mercury in Fish	12
4	Numerical Computation of Parameter Estimates for the National Descriptive Model of Mercury in Fish.....	14
4.1	Newton-Raphson Iteration	14
5	Summary	15
	References Cited.....	16
	Appendix.....	16

Computing Maximum-Likelihood Estimates for Parameters of the National Descriptive Model of Mercury in Fish

By David I. Donato

1 Introduction

The National Descriptive Model of Mercury in Fish Tissue (NDMMF) is a statistical model (Wente, 2004) used to predict the concentration in fish tissue of methylmercury, a potent neurotoxin and a known health hazard for humans and wildlife. The utility of the NDMMF derives from its power to explain much of the observed variation in fish-tissue methylmercury concentrations as (1) variation by geographic location, (2) variation over time, and (3) variation by species and fish length due to bioaccumulation. Because it is based on a national database of methylmercury observations, this model has the potential to be used locally throughout the conterminous United States and in parts of Canada for planning ongoing sampling of methylmercury in fish, and for preparing public health warnings (fish consumption advisories) identifying specific species and sizes of fish from particular locations known or likely to be high in methylmercury content (Hearn and others, 2006).

The NDMMF may be expressed formally as follows:

$$y_{ijk} = \sum_{j=1}^n \alpha_j v_j + \sum_{k=1}^m \beta_k g_k + \epsilon_{ijk} \quad (1)$$

where

$$y_{ijk} = \ln(C_{ijk} + 1),$$

α_j = the parameter for the j th sampling event,

$$v_j = \begin{cases} 1 & \text{if the event is } j \\ 0 & \text{if the event is not } j, \end{cases}$$

β_k = the parameter for the k th species/cut combination, and

$$g_k = \begin{cases} \ln(\text{length}_{ijk} + 1) & \text{if the species/cut is } k \\ 0 & \text{if the species/cut is not } k \end{cases}$$

and where

- C_{ijk} is the observed methylmercury concentration for the i th sample from sampling event j with species/cut combination (or type) k ,
- n is the number of sampling events and m is the number of species/cut combinations,
- length_{ijk} is the length in inches of the fish for the i th sample from sampling event j with species/cut combination (or type) k ,
- some of the $\{C_{ijk}\}$ may be left censored, and

- ϵ_{ijk} is a random variable specifying the error for the i th sample from sampling event j with species/cut combination (or type) k .

The term **sampling event** as it applies to the NDMMF means the collection of samples from a particular location (such as a lake or part of a stream) at one particular time. The term **species/cut** in this context refers to the combination of a particular species of fish and a type of laboratory treatment (Wente, 2004). In this publication, the notations “ $\ln(a)$ ” or “ $\ln[a]$ ” are used for the natural logarithm of a and are equivalent to “ $\log_e(a)$ ”.

Before the NDMMF can be used to predict methylmercury concentrations in fish tissue, its parameters (the sets of values $\{\alpha_j\}$ and $\{\beta_k\}$) must be estimated on the basis of a dataset of observations, ordinarily for thousands of sampling events and hundreds of species/cut combinations. Among many statistical approaches that might be used to compute a set of parameters for this model in order to fit the parameters to the observed data, two widely used methods are those of **least-squares** and **maximum-likelihood** estimation (Cramér, 1946; Miller and Miller, 2004). Because some observations are left censored (that is, some are at or below some detection limit), the method of maximum-likelihood estimation is more suitable for estimating parameters for the NDMMF because it makes better use of the information provided by censored observations than the method of least-squares estimation (Helsel, 2004).

Computing the maximum-likelihood estimates of the parameters for the NDMMF requires substantial computer-processor time and memory because the NDMMF defines an unusually large number of parameters and variables (tens of thousands). Consequently, the task of fitting the parameters of the NDMMF to a growing database of fish-tissue mercury observations now strains or exceeds the computational capabilities of readily available statistical software packages and computer workstations. This report addresses this problem by deriving and presenting the mathematical and statistical results required to enable the development of custom computer software capable of fitting the parameters of the NDMMF more quickly and using less computer memory than general-purpose statistical software.

The detailed mathematical explanation of the application of the method of maximum likelihood to the NDMMF in this publication includes the derivation of the mathematical expressions used in computing maximum-likelihood estimates of the parameters of the NDMMF. These details are included in order to do the following:

1. Provide mathematical and computational expressions suitable for use in implementing custom software for fitting parameters to the NDMMF in the maximum-likelihood sense; and
2. Present the derivation of maximum-likelihood-estimation (MLE) expressions for the current NDMMF in sufficient detail to facilitate future derivation of revised MLE expressions in case the NDMMF is revised.

To aid understanding of the derivation of the MLE expressions for the NDMMF, this publication first presents in **section 2** the derivation of MLE expressions for a much simpler but structurally similar model.¹ Then **section 3** derives and presents the mathematical MLE expressions for the

¹This report uses the symbols α and β differently than the seminal report on the NDMMF (Wente, 2004) in order to clarify and emphasize the structural similarity between a simple linear statistical model and the NDMMF. The linear model described in **section 2** illustrates concepts in a simple setting; these concepts are then applied to the more complex NDMMF in **section 3**.

NDMMF. Finally, **section 4** discusses the actual computation of maximum-likelihood estimates for NDMMF parameters using the results of **section 3**.

2 Maximum-Likelihood Estimates for a Linear Model

This section applies the method of maximum-likelihood estimation to a particular linear statistical model involving left-censored data. The derivation of maximum-likelihood estimates for this simple model illustrates the basic techniques and principles to be applied in the following section for the more complicated model—the NDMMF.

2.1 A Linear Model

The particular linear statistical model used for illustration in this section is:

$$y = a + bx + \epsilon, \quad (2)$$

where

- $\epsilon \sim N(0, \sigma)$,
- some observed values of y may be left censored and, thus, known only to be at or below a detection limit that may differ from observation to observation,²
- y is the response variable,
- x is the independent variable, and
- a , b , and σ are the unknown parameters to be estimated.

2.2 The Likelihood Function for the Linear Model

In general, a **likelihood function** is the joint probability or probability density of a set of observations (that is, of a sample of observed values of the independent and response variables) construed as a function of the unknown parameters of a statistical model; in general, finding the **maximum-likelihood estimates** of the unknown parameters means finding a set of values for the unknown parameters that maximizes this joint probability (or probability density) of the observations under the model. Like least-squares estimation, maximum-likelihood estimation fits the values of the parameters of a statistical model to a set of observed values of independent variables and response variables (Cramér, 1946; Miller and Miller, 2004).

For this particular linear model, the likelihood function $\mathcal{L}(a, b, \sigma)$ is the joint probability density of the uncensored observations multiplied by the joint probability of the censored observations, with this entire product regarded as a function of the unknown parameters a , b , and σ . Thus, the likelihood function for this linear model is:

$$\mathcal{L}(a, b, \sigma) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2} \frac{(y_i - (a + bx_i))^2}{\sigma^2} \right] \right) \cdot \prod_{j=1}^m P\{y_j \leq D_j \mid a, b, \sigma, x_j\}, \quad (3)$$

where

²That is, if y_j is a left-censored observation and D_j is its associated detection limit, then $y_j \leq D_j$.

- $P\{y_j \leq D_j | a, b, \sigma, x_j\}$ is the probability that y_j is less than or equal to D_j given particular values for a, b, σ , and x_j ,
- D_j is the detection limit for censored observation j ,
- n is the number of uncensored observations, and
- m is the number of left-censored observations.

The likelihood function is an application of the general multiplicative formula of probability theory. The general multiplicative formula states that when each member of a set of observations is **independently** drawn from a population with a known probability distribution, then the joint probability (or joint probability density) of the set of observations is the product of the probabilities (or probability densities) of the individual observations (Chung, 1974; Wilks, 1962). The likelihood function “ $\mathcal{L}(a, b, \sigma)$ ” defined here follows this principle in that it is the product of probability densities for all n uncensored observations and of probabilities for all m censored observations.

To understand the first repetitive product on the right-hand side of equation 3, first recall that the probability density function (PDF) of a normal distribution³ is

$$f_{\mu, \sigma}(w) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{(w - \mu)^2}{\sigma^2}\right], \quad (4)$$

where w is the independent variable and μ and σ are the mean and standard deviation of the normal distribution, respectively (Miller and Miller, 2004).

Then observe that equation 2 can be rewritten as

$$y - (a + bx) = \epsilon \quad (5)$$

to emphasize that the difference between the observed response y and the predictor $(a + bx)$ is equivalent to the error term ϵ . Since ϵ is a normally distributed random variable with $\mu = 0$, equation 5 indicates that $y - (a + bx)$ is also a stochastic quantity following the normal probability distribution with a mean of zero. Therefore,

$$(y - (a + bx)) - 0 = y - (a + bx) \quad (6)$$

can be substituted for $(w - \mu)$ in equation 4 to express the probability density for each uncensored observation (x_i, y_i) as a function of the unknown parameters (a, b, σ) . The probability density thus derived,

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{(y_i - (a + bx_i))^2}{\sigma^2}\right], \quad (7)$$

has the properties required for the likelihood function. This probability density tends to be relatively large when the error term $(y_i - (a + bx_i))^2$ is near zero, and it tends to be relatively small when the error term is further from zero. This means that choices of (a, b, σ) that produce small error terms (errors of estimation near zero) also produce higher values of the likelihood function. Intuitively, higher likelihood corresponds to a "better" fit of the model to the data.

³Although it is conventional to use the variable x in expressing the PDF and cumulative distribution function (CDF) of the normal distribution, w is used here to avoid confusion with the specific censored observations $\{x_i\}$.

To understand the second repetitive product on the right-hand side of equation 3, consider what it means for an observation to be **left censored**. An observation is **left censored** if we know that its value lies somewhere on the real number line (or perhaps on the nonnegative side of the real number line) at or below a particular value D_j , generally referred to as the **detection limit**⁴ (Helsel, 2004). Thus, for the m left-censored observations, the multiplicative contribution of observation j to the likelihood is the probability that observation j is at or below the detection limit D_j . Since $\epsilon \sim N(0, \sigma)$, we can express the contribution of left-censored observations to the likelihood in terms of the cumulative distribution function (CDF) of the normal distribution. Let $F_{0,\sigma}$ denote the CDF of a normal distribution with mean $\mu = 0$ and standard deviation σ .

$$F_{0,\sigma}(A) = \int_{-\infty}^A \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{(w-0)^2}{\sigma^2}\right] dw \quad (8)$$

denotes the probability of observing a value less than or equal to some specific value A .

Since

$$y_j \leq D_j \quad \rightarrow \quad a + bx_j + \epsilon_j \leq D_j \quad \rightarrow \quad \epsilon_j \leq D_j - (a + bx_j), \quad (9)$$

then

$$P\{y_j \leq D_j \mid a, b, \sigma, x_j\} = \int_{-\infty}^{D_j - (a + bx_j)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{w^2}{\sigma^2}\right] dw. \quad (10)$$

Thus, equation 3 can be rewritten as follows:

$$\mathcal{L}(a, b, \sigma) = \prod_{i=1}^n \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{(y_i - (a + bx_i))^2}{\sigma^2}\right] \right) \cdot \prod_{j=1}^m \left(\int_{-\infty}^{D_j - (a + bx_j)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{w^2}{\sigma^2}\right] dw \right). \quad (11)$$

For many physical measurements, values are constrained to be nonnegative. Laboratory measurements of concentrations of chemicals in samples of organic material are an example of nonnegative values that may be subject to detection limits. A measurement that is both below a detection limit and also constrained to be nonnegative can be described as **interval censored**. An observation is interval censored if we know that its value lies in an interval on the real number line between an upper and a lower limit. The term **left censored** is, however, widely applied to censored observations of concentrations; therefore, when the term **left censored** is applied to concentrations, the restriction to nonnegative values is implicit. When the response variable y_j for the linear model represents a concentration value, its multiplicative contribution to the likelihood differs from that of an unconstrained left-censored observation. To determine the contribution to likelihood when y_j is

⁴A **detection limit** is both a lower bound and an upper bound. It is a lower bound for meaningful point measures. When a measurement procedure (such as a laboratory analysis of a chemical concentration) cannot provide a measurement below a certain value, that value is the detection limit. In such a case, a measurement below the detection limit has no meaning as a point value, but it does still provide quantitative information that can be used in statistical analysis: the measurement conveys the information that a value has been observed that is less than or equal to the detection limit. The detection limit is also the upper bound for the unknown value that could not be measured as a point value. A measurement at or below the detection limit is referred to as a **left-censored** observation.

constrained to be nonnegative, this constraint on y_j must be used to find the constraint on the error term ϵ . As

$$y_j \geq 0 \quad \rightarrow \quad a + bx_j + \epsilon_j \geq 0 \quad \rightarrow \quad \epsilon_j \geq -(a + bx_j), \quad (12)$$

then in the case of censored concentrations,

$$P\{0 \leq y_j \leq D_j \mid a, b, \sigma, x_j\} = \int_{-(a+bx_j)}^{D_j-(a+bx_j)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2} \frac{w^2}{\sigma^2}\right] dw \quad (13)$$

is the contribution to the likelihood for each observation.

For simplicity, the following two subsections on the linear model treat censored observations as general left-censored observations that may potentially have negative values.

2.3 The Log-Likelihood Function for the Linear Model

Since the logarithm function is defined for all positive real numbers and is monotone increasing on the positive side of the real line, the logarithm of the likelihood function, or the **log-likelihood function**, will be maximized by the same parameter set as the likelihood function. For some models, a third option of minimizing the **negative log-likelihood function** will find the same parameter set and may prove to be either analytically or computationally easier than either of the other two equivalent methods of maximum-likelihood estimation for the model. Because of this equivalence of optimization problems, we can choose to solve whichever is easiest. In general, numerical solutions for maximum-likelihood problems tend to use either the positive or negative log-likelihood function—especially for models (like the NDMMF) that involve large numbers of parameters—because machine computations involving logarithms are less prone to the problems of overflow, underflow, and loss of precision.

For the linear model with some left-censored observations, the log-likelihood function is the following:

$$\ln(\mathcal{L}(a, b, \sigma)) = \ln\left(\frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}}\right) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (a + bx_i))^2 + \sum_{j=1}^m \ln\left[F_{0,\sigma}(D_j - (a + bx_j))\right]. \quad (14)$$

Equation 14 is derived from equations 8 and 11. Notice that the natural logarithm of an indexed product is an indexed sum. The log-likelihood function is easier to use than the likelihood function itself in finding the optimal parameter set for the linear model.

2.4 Maximizing the Likelihood and Log-Likelihood Functions for the Linear Model

In order to maximize the log-likelihood function (and the likelihood function as well), we must find the ordered triple $(\hat{a}, \hat{b}, \hat{\sigma})$, if it exists, defined so that for any ordered triple (a', b', σ') it will be true that $\ln(\mathcal{L}(\hat{a}, \hat{b}, \hat{\sigma})) \geq \ln(\mathcal{L}(a', b', \sigma'))$, or equivalently, $\mathcal{L}(\hat{a}, \hat{b}, \hat{\sigma}) \geq \mathcal{L}(a', b', \sigma')$.

If a maximizing ordered triple $(\hat{a}, \hat{b}, \hat{\sigma})$ exists, it will satisfy the three simultaneous conditions determined by equating to zero (0) each of the three partial derivatives of the log-likelihood function with respect to a , b , and σ^2 (Miller and Miller, 2004). The three simultaneous equations that must be satisfied by $(\hat{a}, \hat{b}, \hat{\sigma})$ are:

$$\frac{\partial}{\partial a}(\ln[\mathcal{L}(a, b, \sigma)]) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - (a + bx_i)) - \sum_{j=1}^m \frac{f_{\sigma}(D_j - (a + bx_j))}{F_{0,\sigma}(D_j - (a + bx_j))} = 0, \quad (15)$$

$$\frac{\partial}{\partial b}(\ln[\mathcal{L}(a, b, \sigma)]) = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - (a + bx_i))x_i - \sum_{j=1}^m \frac{f_{\sigma}(D_j - (a + bx_j))}{F_{0,\sigma}(D_j - (a + bx_j))}(x_j) = 0, \text{ and} \quad (16)$$

$$\frac{\partial}{\partial \sigma^2}(\ln[\mathcal{L}(a, b, \sigma)]) = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - (a + bx_i))^2 + \sum_{j=1}^m \frac{\frac{d}{d\sigma^2}(F_{0,\sigma}(D_j - (a + bx_j)))}{F_{0,\sigma}(D_j - (a + bx_j))} = 0. \quad (17)$$

In each of equations 15 and 16, the contribution to the partial derivative of the log-likelihood function from uncensored observations equals the contribution from the left-censored observations. Equation 17 showing the partial derivative of the log-likelihood function with respect to σ^2 could be replaced by an equation involving the partial derivative with respect to σ ; either choice is acceptable analytically, but the equation involving the partial derivative with respect to σ^2 is preferable for computation.

In equation 17, the expression

$$\frac{d}{d\sigma^2}(F_{0,\sigma}(D_j - (a + bx_j)))$$

denotes the derivative of the normal CDF with respect to σ^2 evaluated at the point $D_j - (a + bx_j)$; in this expression, the normal CDF is construed as a function of σ^2 . Thus,

$$F_{0,\sigma}(D_j - (a + bx_j)) = \int_{-\infty}^{D_j - (a + bx_j)} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\frac{w^2}{\sigma^2}\right] dw \quad (18)$$

where the variable of integration is denoted as w in order to avoid notational confusion with the observed values of the independent variable x_j . (Please see equation 4.) In solving the system of equations 15, 16, and 17 for a, b , and σ , this expression for the derivative of the normal CDF with respect to σ^2 must be evaluated numerically. The Newton-Raphson Method (also known as Newton's Method) may be used to solve this system of three simultaneous equations numerically. The Newton-Raphson Method for finding maximum-likelihood parameters is described in **section 4.1** (Weisstein, 1999).

3 Maximum-Likelihood Estimates for the National Descriptive Model of Mercury in Fish

The methods used in **section 2** to estimate the three unknown parameters of a simple linear model in the maximum-likelihood sense may now be specialized and applied to estimate the $n + m + 1$ parameters of the NDMMF in the maximum-likelihood sense. Comparison of the linear model (as follows)

$$y = a + bx + \epsilon$$

with the NDMMF (as follows)

$$y_{ijk} = \sum_{j=1}^n \alpha_j v_j + \sum_{k=1}^m \beta_k g_k + \epsilon_{ijk}$$

shows some structural similarity between the the linear model and the NDMMF but significant dissimilarity as well. In particular, although the NDMMF is structurally linear, it is effectively nonlinear because the **covariate** term $g_k = \ln(\text{length}_{ijk} + 1)$ is a nonlinear function of the physically measured covariate (fish length) and because the response variable $y_{ijk} = \ln(C_{ijk} + 1)$ is a nonlinear function of the measured response C_{ijk} .

Before discussing the fitting of the parameters of the NDMMF, it is instructive to consider how the NDMMF—once its parameters have been determined—is used to estimate or predict a fish-tissue methylmercury concentration. The simple linear model can be used to compute a response y from a value for the single independent variable x along with values for the two parameters (a, b) . In comparison, the NDMMF also requires three items of data in order to compute an estimate of a fish-tissue methylmercury concentration: the α parameter associated with a particular sampling event, the β parameter associated with a particular species/cut combination, and a fish length (in inches). The linear model and the NDMMF are similar in that any single estimate makes use of exactly two parameters. An estimate made with the linear model makes use of the parameters a and b , while an estimate made with the NDMMF makes use of exactly one α_j and exactly one β_k because, except for the selected event j and the selected species/cut k , the **indicator variables** “ v_j ” and “ g_k ” are zero for all other sampling events and for all other species/cut combinations, respectively. Although NDMMF model parameters are used only two at a time for prediction, all parameters are involved in fitting the NDMMF to observed data in the maximum-likelihood sense. This is shown by the equations and systems of equations for maximizing the likelihood and log-likelihood functions of the NDMMF presented below.

3.1 Contributions to the Likelihood from Censored Observations

As previously explained in **section 2.2**, the term **left censored** is applied to observations known to be at or below some detection limit D_k . In some cases, an observation described as left censored may be subject to additional constraints, such as the restriction of chemical concentrations to be nonnegative. The related term **interval censored** refers to an observation known to fall in an interval on the real number line between a lower limit and an upper limit. Left censoring is a special case of interval censoring with an implicit (rather than explicit) lower limit for the censoring interval. The implicit lower limit may depend on either, or both, of the form of the statistical model and the physical characteristics of the observed data. The response variable y for the linear model considered in **section 2** is an example of a response constrained neither by the form of the model nor by the physical characteristics of the observations; this response variable y can assume any value on the real number line, so this model’s implied lower limit⁵ for a left-censored observation is $-\infty$. In contrast to the simple linear model, the response variable for the NDMMF is implicitly constrained both by the form of the model and by the physical characteristics of chemical concentrations. The functional form of the NDMMF response variable $y_{ijk} = \ln(C_{ijk} + 1)$ constrains the measured value C_{ijk} to be strictly greater than -1 in order that the argument of the \ln function also be strictly greater than zero (as it must be in order for the logarithm function to be defined for the argument). Ignoring for the moment that the concentration C_{ijk} must be

⁵As this example illustrates, the interval applied in interval censoring may be an improper interval.

nonnegative, left censoring for the NDMMF is equivalent to interval censoring with $-1 < C_{ijk} \leq D_k$. (The detection limit is designated here by D_k because it can vary from observation to observation.)

For the sake of generality, and also because a number of studies have estimated NDMMF parameters assuming interval censoring with $-1 < C_{ijk} \leq D_k$, the likelihood and log-likelihood functions for the NDMMF are expressed in the following subsections in terms of this least restrictive form of interval censoring. These generalized expressions can be readily specialized by constraining concentrations to be nonnegative, as explained in the final paragraph of **section 3.2** below.

The likelihood function for the NDMMF presented in the following **section 3.2** depends in part on the probabilities of the censored observations. Determining the probability of a censored observation requires that the upper and lower limits for measured concentrations be used in finding the corresponding upper and lower limits on the stochastic error term ϵ_{ijk} . The upper and lower limits for ϵ_{ijk} determine the probability of a censored observation as the integral of the PDF of the error term ϵ_{ijk} from the minimum error value to the maximum error value; or, equivalently, as the CDF of the error term evaluated at its upper limit minus the CDF of the error term evaluated at its lower limit, as shown below in equation 21 in **section 3.2**.

Throughout the remainder of **section 3**, let D_c^U denote the upper limit for a censored concentration and let D_L denote the lower limit.⁶ The upper and lower limits for the error term ϵ_{ijk} can be determined from the upper and lower bounds D_c^U and D_L for censored concentrations as follows:

Finding the upper bound for ϵ_{ijk} :

$$y_{ijk} = \ln(C_{ijk} + 1) \quad \rightarrow \quad e^{y_{ijk}} = C_{ijk} + 1 \quad \rightarrow \quad C_{ijk} = e^{y_{ijk}} - 1.$$

Therefore

$$C_{ijk} \leq D_c^U \quad \rightarrow \quad e^{y_{ijk}} - 1 \leq D_c^U \quad \rightarrow \quad e^{y_{ijk}} \leq D_c^U + 1 \quad \rightarrow \quad y_{ijk} \leq \ln(D_c^U + 1).$$

Substituting $\alpha_c v_c + \beta_c g_c + \epsilon_{ijk}$ for y_{ijk} in the inequality above results in

$$\alpha_c v_c + \beta_c g_c + \epsilon_{ijk} \leq \ln(D_c^U + 1) \quad \rightarrow \quad \epsilon_{ijk} \leq \ln(D_c^U + 1) - (\alpha_c v_c + \beta_c g_c).$$

Taking the natural logarithm of both sides of the inequality is valid (preserves the inequality) because the natural logarithm function is monotone increasing over its entire domain.

Finding the lower bound for ϵ_{ijk} :

$$C_{ijk} \geq D_L \quad \rightarrow \quad e^{y_{ijk}} - 1 \geq D_L \quad \rightarrow \quad e^{y_{ijk}} \geq D_L + 1 \quad \rightarrow \quad y_{ijk} \geq \ln(D_L + 1).$$

So, therefore, substituting $\alpha_c v_c + \beta_c g_c + \epsilon_{ijk}$ for y_{ijk} as before results in

$$\alpha_c v_c + \beta_c g_c + \epsilon_{ijk} \geq \ln(D_L + 1) \quad \rightarrow \quad \epsilon_{ijk} \geq \ln(D_L + 1) - (\alpha_c v_c + \beta_c g_c).$$

⁶ This notation indicates that the upper censoring limit may vary from observation to observation; however, the lower limit is the same for all censored observations.

3.2 The Likelihood Function for the National Descriptive Model of Mercury in Fish

In its general form, the NDMMF does not necessarily restrict the form of the various random error terms $\{\epsilon_{ijk}\}$. For purposes of determining the likelihood function for the NDMMF, however, the model will be restricted by the simplifying assumption that all members of the set of random error terms $\{\epsilon_{ijk}\}$ are **independently and identically distributed** as normal variates with mean $\mu = 0$ and standard deviation σ (where σ is an unknown parameter of the NDMMF), as follows:

$$\{\epsilon_{ijk}\} \text{ are i.i.d. with } \epsilon_{ijk} \sim N(0, \sigma) \text{ for all } \{ijk\}.$$

With this simplifying assumption about the error terms, the likelihood function for the NDMMF may be expressed as:

$$\mathcal{L}(\{\alpha_j\}, \{\beta_k\}, \sigma) = P_u \cdot P_c, \quad (19)$$

where

$$P_u = \prod_{u=1}^{n_u} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{(y_u - \alpha_u v_u - \beta_u g_u)^2}{\sigma^2}\right],$$

$$y_u = y_{ijk},$$

$$\alpha_u = \alpha_j,$$

$$\beta_u = \beta_k, \quad (20)$$

$$v_u = v_j,$$

$$g_u = \ln(\text{length}_{ijk} + 1), \text{ and}$$

$$n_u = \text{the number of uncensored observations.}$$

and

$$P_c = \prod_{c=1}^{n_c} P\{D_L \leq C_{ijk} \leq D_c^U \mid \{\alpha_j\}, \{\beta_k\}, \sigma\},$$

$$= \prod_{c=1}^{n_c} \left[F_{0,\sigma}(\ln[D_c^U + 1] - \alpha_c v_c - \beta_c g_c) - F_{0,\sigma}(\ln[D_L + 1] - \alpha_c v_c - \beta_c g_c) \right],$$

$$\alpha_c = \alpha_j,$$

$$\beta_c = \beta_k,$$

$$v_c = v_j, \quad (21)$$

$$g_c = \ln(\text{length}_{ijk} + 1),$$

$$n_c = \text{the number of censored observations,}$$

$$D_c^U = \text{the upper detection limit for the censored observation } c, \text{ and}$$

$$D_L = \text{the lower limit for the censoring interval for the censored observation } c.$$

In equation groups 20 and 21, each index u or c identifies a particular uncensored or censored observation, respectively, which is associated with exactly one event j and exactly one species/cut k .

In the derivation of the log-likelihood function, it is convenient to express P_u in the following alternative but equivalent form:

$$P_u = \frac{1}{\sigma^{n_u} (2\pi)^{\frac{n_u}{2}}} \exp \left[-\frac{1}{2\sigma^2} \left(\sum_{u=1}^{n_u} [y_u - \alpha_u v_u - \beta_u g_u]^2 \right) \right]. \quad (22)$$

The likelihood function for the NDMMF is the product of two composite factors contributing to the likelihood: a contribution from uncensored observations and another contribution from censored observations. The contribution P_u from uncensored observations is a product of probability densities, and the contribution P_c for censored observations is a product of probabilities. Each factor in the indexed product P_u is the probability density for a particular observation of an uncensored mercury concentration and fish length in association with a specific event j and a specific species/cut k for any given values of the parameters $\{\alpha_j\}$, $\{\beta_k\}$, and σ . Each factor in the indexed product P_c is the probability that a censored mercury concentration was in the range of possible values below the detection limit D_c^U and simultaneously above the lower censoring limit D_L for any given values of the parameters.

Equation 21 can be readily specialized for two different interpretations of censored observations of concentrations. **Case 1:** As stated previously in the first paragraph of **section 3.1**, the functional form of the NDMMF requires that $C_{ijk} > -1$; this is equivalent to requiring that the lower limit of the censoring interval, D_L , approach (but not reach) -1 from above. As $D_L \rightarrow -1$, the term $F_{0,\sigma}(\ln[D_L + 1] - \alpha_c v_c - \beta_c g_c)$ in equation 21 approaches $F_{0,\sigma}(-\infty) = 0$ because as $D_L + 1 \rightarrow 0$, then $\ln[D_L + 1] \rightarrow -\infty$. Therefore, when $D_L \rightarrow -1$, the term $F_{0,\sigma}(\ln[D_L + 1] - \alpha_c v_c - \beta_c g_c)$ can be dropped from equation 21. **Case 2:** If C_{ijk} is constrained to be nonnegative, then $D_L = 0$ and $\ln[D_L + 1] = \ln[1] = 0$. In this latter case, the second term in equation 21 becomes $F_{0,\sigma}(-\alpha_c v_c - \beta_c g_c)$.

3.3 The Log-Likelihood Function for the National Descriptive Model of Mercury in Fish

As explained previously for the linear model presented in **section 2.1**, the likelihood function will be maximized by the same parameter set that maximizes the log-likelihood function.⁷ In this sense, the problem of maximizing the likelihood is equivalent to the problem of maximizing the log-likelihood. For the NDMMF, as for many maximum-likelihood estimation problems, it is advantageous both analytically and computationally to work with the log-likelihood function rather than the likelihood function.

For the NDMMF, the log-likelihood function is:

$$\begin{aligned} \ln(\mathcal{L}(\{\alpha_j\}, \{\beta_k\}, \sigma)) &= \ln(P_u) + \ln(P_c) \\ &= \ln \left(\frac{1}{\sigma^{n_u} (2\pi)^{\frac{n_u}{2}}} \right) - \frac{1}{2\sigma^2} \left(\sum_{u=1}^{n_u} [y_u - \alpha_u v_u - \beta_u g_u]^2 \right) \\ &\quad + \sum_{c=1}^{n_c} \ln \left[F_{0,\sigma}(\ln[D_c^U + 1] - \alpha_c v_c - \beta_c g_c) - F_{0,\sigma}(\ln[D_L + 1] - \alpha_c v_c - \beta_c g_c) \right]. \end{aligned} \quad (23)$$

⁷Although a maximizing parameter set can be found for reasonably consistent datasets, there are datasets for which a maximizing parameter set may not exist or may be difficult to compute using conventional fixed-precision machine numbers. This potential problem is beyond the scope of this publication.

3.4 Maximizing the Likelihood and Log-Likelihood Functions for the National Descriptive Model of Mercury in Fish

To maximize the likelihood and log-likelihood functions for the NDMMF, it is necessary to determine a collection of $n + m + 1$ parameters $\{\{\hat{\alpha}_j\}, \{\hat{\beta}_k\}, \hat{\sigma}\}$ such that

$$\mathcal{L}(\{\hat{\alpha}_j\}, \{\hat{\beta}_k\}, \hat{\sigma}) \geq \mathcal{L}(\{\alpha'_j\}, \{\beta'_k\}, \sigma') \quad (24)$$

for every ordered $(n + m + 1)$ -tuple of parameters $(\{\alpha'_j\}, \{\beta'_k\}, \sigma')$ or equivalently such that

$$\ln[\mathcal{L}(\{\hat{\alpha}_j\}, \{\hat{\beta}_k\}, \hat{\sigma})] \geq \ln[\mathcal{L}(\{\alpha'_j\}, \{\beta'_k\}, \sigma')] \quad (25)$$

for every ordered $(n + m + 1)$ -tuple of parameters $(\{\alpha'_j\}, \{\beta'_k\}, \sigma')$.

If a maximum for the likelihood and log-likelihood functions exists, it will satisfy the $n + m + 1$ simultaneous conditions determined by equating to zero each of the $n + m + 1$ partial derivatives of the log-likelihood function with respect to each of the n α_j 's, the m β_k 's, and σ^2 (Miller and Miller, 2004).

The partial derivatives of the log-likelihood function have three forms, as shown below:

1. **For the α_j 's:**

$$\begin{aligned} \frac{\partial}{\partial \alpha_j} \left(\ln \left[\mathcal{L}(\{\alpha_j\}, \{\beta_k\}, \sigma) \right] \right) = \\ \frac{1}{\sigma^2} \sum_{u=1}^{n_u} (y_u - \alpha_u v_u - \beta_u g_u) \\ + \sum_{c=1}^{n_c} \frac{[-f_{0,\sigma}(\ln[D_c^U + 1] - \alpha_c v_c - \beta_c g_c) + f_{0,\sigma}(\ln[D_L + 1] - \alpha_c v_c - \beta_c g_c)]}{[F_{0,\sigma}(\ln[D_c^U + 1] - \alpha_c v_c - \beta_c g_c) - F_{0,\sigma}(\ln[D_L + 1] - \alpha_c v_c - \beta_c g_c)]} \\ = 0. \end{aligned} \quad (26)$$

The summations in the j th equation implied by equation 26 only pick up values for the observations for event j .

2. **For the β_k 's:**

$$\begin{aligned} \frac{\partial}{\partial \beta_k} \left(\ln \left[\mathcal{L}(\{\alpha_j\}, \{\beta_k\}, \sigma) \right] \right) = \\ \frac{1}{\sigma^2} \sum_{u=1}^{n_u} (y_u - \alpha_u v_u - \beta_u g_u) g_u \\ + \sum_{c=1}^{n_c} \frac{[-f_{0,\sigma}(\ln[D_c^U + 1] - \alpha_c v_c - \beta_c g_c) + f_{0,\sigma}(\ln[D_L + 1] - \alpha_c v_c - \beta_c g_c)]}{[F_{0,\sigma}(\ln[D_c^U + 1] - \alpha_c v_c - \beta_c g_c) - F_{0,\sigma}(\ln[D_L + 1] - \alpha_c v_c - \beta_c g_c)]} g_c \\ = 0. \end{aligned} \quad (27)$$

The summations in the k th equation implied by equation 27 only pick up values for the observations involving species/cut k .

3. **For σ^2 :**

$$\begin{aligned} \frac{\partial}{\partial \sigma^2} \left(\ln \left[\mathcal{L}(\{\alpha_j\}, \{\beta_k\}, \sigma) \right] \right) = \\ -\frac{n_u}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{u=1}^{n_u} (y_u - \alpha_u v_u - \beta_u g)^2 \\ + \sum_{c=1}^{n_c} \frac{d}{d\sigma^2} \frac{[F_{0,\sigma}(\ln[D_c^U + 1] - \alpha_c v_c - \beta_c g_c) - F_{0,\sigma}(\ln[D_L + 1] - \alpha_c v_c - \beta_c g_c)]}{[F_{0,\sigma}(\ln[D_c^U + 1] - \alpha_c v_c - \beta_c g_c) - F_{0,\sigma}(\ln[D_L + 1] - \alpha_c v_c - \beta_c g_c)]} = 0. \end{aligned} \quad (28)$$

The summations in equation 28 pick up values from all observations for all sampling events and for all species/cut combinations.

4 Numerical Computation of Parameter Estimates for the National Descriptive Model of Mercury in Fish

The simple linear model of **section 2** illustrates in principle what is involved in computing maximum-likelihood estimates of the parameters for a statistical model. Equations 15, 16, and 17 constitute a system of three simultaneous equations in the unknown parameters (a, b, σ) . The three parameters can be determined by solving the system of three equations. In principle, model parameters for the NDMMF are computed as they are for the linear model, by solving a system of simultaneous equations. The computational problem and task for the NDMMF is, however, on a very different scale from that of the linear model. Equations 26, 27, and 28 imply a system of $n + m + 1$ equations in as many unknown parameters with n equaling the number of sampling events and m equaling the number of species/cut combinations. In practical cases, this means that finding the maximum-likelihood estimates of the parameters of the NDMMF requires solving a system of tens of thousands of equations in as many unknowns.

Since there is no analytical solution for the system of $n + m + 1$ equations, the system of equations must in practice be solved numerically. A numerical solution will be computationally feasible only if the system of equations converges from an initial set of parameter estimates to a solution within a reasonably small number of iterations. The Newton-Raphson Method is an obvious choice for this problem because it generally offers relatively rapid convergence in finding the zeroes of functions.

4.1 Newton-Raphson Iteration

The Newton-Raphson Method (also called Newton's Method) is a well-known numerical algorithm for finding the zero or root of a function or equation (Weisstein, 1999). The zero or root of a function of a real variable is the value of the independent variable at which the value of the function is equal to zero. The Newton-Raphson Method requires that the function be continuous and differentiable. In essence, the Newton-Raphson Method makes use of a few low-order terms of the Taylor Series representation of the function within a neighborhood of a root in order to find that root. If the function has few or no points of inflection in the neighborhood and is otherwise well behaved within the neighborhood of the root, the Newton-Raphson Method tends to converge rapidly to an acceptably precise solution within a few iterations.

Computing maximum-likelihood estimates of the numerous parameters of the NDMMF requires that the Newton-Raphson Method be applied repeatedly to each equation in the system of $n + m + 1$ equations implied by equations 26, 27, and 28. A single pass through all $n + m + 1$ equations, in order to find a parameter for each equation through a separate Newton-Raphson iteration for each equation, does not suffice to determine the values of all parameters because of the dependencies among the equations in the system of $n + m + 1$ equations. In the n equations implied by equation 26, the α_u and α_c parameters depend on the current values of the β_u and β_c parameters. Similarly, in the m equations implied by equation 27, the β_u and β_c parameters depend on the current values of the α_u and α_c parameters. In equation 28, the value of σ^2 depends on the values of all of the α and β parameters. Thus, as each successive Newton-Raphson iteration determines a value for one parameter, the values of many other parameters are affected and must be corrected through another pass. Multiple passes through the $n + m + 1$ equations (applying the Newton-Raphson Method to each equation in sequence during each pass) are required in order to achieve convergence to the overall collection of parameter values that solves the system of $n + m + 1$ equations and, thereby, computes the set of parameters that maximizes the likelihood and log-likelihood functions.

Despite the abundance of opportunity during this computation for the dependencies among the parameters to prevent convergence towards an overall solution, experience shows that convergence does in fact take place in practice when the starting set of initial parameter estimates is sufficiently close to the maximum-likelihood estimates and the data meet minimal requirements for consistency and completeness. Experience also shows that when the starting estimates of parameters used in a Newton-Raphson iteration are those determined by ordinary least-squares estimation, these estimates are usually sufficiently close to the maximum-likelihood estimates to enable convergence of the multiple-pass, sequential Newton-Raphson iteration described in this section.

5 Summary

This report explains how the method of maximum-likelihood estimation (MLE) can be applied to find the parameters of the National Descriptive Model of Mercury in Fish (NDMMF), a statistical model used to predict the concentration of methylmercury in fish tissue. Computing the parameters for the NDMMF in the maximum-likelihood sense is a complicated procedure because the NDMMF often involves a large number of parameters (tens of thousands) and because the NDMMF makes use of the information provided by left-censored observations of methylmercury concentrations. Because of the large number of parameters and because of the additional computation required to make use of left-censored observations, the task of computing parameters for the NDMMF strains or exceeds the computational capabilities of readily available statistical software packages and computer workstations. Consequently, researchers and practitioners need custom software to fit NDMMF parameters in the MLE sense more quickly using less computer memory.

The two main purposes of this publication are the following:

1. To provide mathematical and computational expressions suitable for use in implementing custom software for fitting parameters to the NDMMF in the maximum-likelihood sense; and
2. To present the derivation of maximum-likelihood estimation (MLE) expressions for the current NDMMF in sufficient detail to facilitate derivation of revised MLE expressions for any future revisions of the NDMMF.

As an aid to understanding, this publication illustrates the mathematics of maximum-likelihood parameter estimation for a simple linear model before deriving the mathematical results for the more complex NDMMF. Finally, this publication explains in general terms how the Newton-Raphson Method can be used in computations to find the set of best-fit parameters for the NDMMF in the maximum-likelihood sense.

References Cited

- Chung, Kai Lai, 1974, Elementary probability theory with stochastic processes: New York, Springer-Verlag, 325 p.
- Cramér, Harald, 1946, Mathematical methods of statistics: Princeton University Press, 575 p.
- Hearn, Paul, and others, 2006, EMMMA; a Web-based system for environmental mercury mapping, modeling, and analysis: U.S. Geological Survey Open-File Report 2006–1086, 13 p.
- Helsel, D.R., 2004, Nondetects and data analysis; statistics for censored environmental data: Wiley-Interscience, 268 p.
- Miller, Irwin, and Miller, Marylees, 2004, John E. Freund's mathematical statistics with applications (7th ed.): Upper Saddle River, N.J., Pearson Prentice Hall, 614 p.
- Weisstein, E.W., 1999, "Newton-Raphson Method" and "Newton's Method" in CRC concise encyclopedia of mathematics: Chapman & Hall/CRC, 1969 p.
- Wente, S.P., 2004, A statistical model and national data set for partitioning fish-tissue mercury concentration variation between spatiotemporal and sample characteristic effects: U.S. Geological Survey Scientific Investigations Report 2004–5199, 15 p.
- Wilks, S.S., 1962, Mathematical statistics: John Wiley & Sons, Inc., 644 p.

Appendix

This publication includes the source code for custom software to compute the best-fit parameters for the National Descriptive Model of Mercury in Fish (NDMMF) in the maximum-likelihood sense. The source code is provided as a plain-text (ASCII) file of C code. This C code is based directly on the mathematical expressions presented in this publication, with obvious correspondences between key variable names in the C code and the terms and symbols used in this report. The software is provided subject to the disclaimer contained within the source-code file.

