WORKING PAPER SERIES

FISHERIES STOCK ASSESSMENT

TITLE XII
Collaborative Research Support Program

Fisheries Stock Assessment CRSP Management Office
International Programs, College of Agriculture
The University of Maryland, College Park, Maryland 20742

In cooperation with the United States Agency for International Development (Grant No. DAN-4146-G-SS-5071-00) the Fisheries Stock Assessment CRSP involves the following participating institutions:

The University of Maryland—Center for Environmental and Estuarine Studies
The University of Rhode Island—International Center for Marine Resource Development
The University of Washington—Center for Quantitative Sciences
The University of Costa Rica—Centro de Investigación en Ciencias del Mar y Limnología
The University of the Philippines—Marine Science Institute (Diliman)—College of Fisheries (Visayas)

In collaboration with The University of Delaware; The University of Maryland—College of Business and Management; The University of Miami; and The International Center for Living Aquatic Resources Management (ICLARM).

WORKING PAPER SERIES

Working Paper No. 46

"Inferring the Distribution of the Parameters of the von Bertalanffy Growth Model from Length Moments"

by
Robert L. Burr
University of Washington

December, 1988

Fisheries Stock Assessment
Title XII
Collaborative Research Support Program

The Fisheries Stock Assessment CRSP (sponsored in part by USAID Grant No. DAN-4146-G-SS-5071-00) is intended to support collaborative research between the U.S. and developing countries' universities and research institutions on fisheries stock assessment and management strategies.

This Working Paper was produced by the University of Washington and the University of Costa Rica-Centro de Investigacion en Ciencias del Mar y Limnologia (CIMAR) in association with the University of Delaware and the University of Miami. Additional copies are available from the CRSP Management Office and from:

Dr. Vincent Gallucci
Management Assistance for Artisanal Fisheries (MAAF)
Center for Quantitative Science, HR-20
School of Fisheries
3737 - 15th Avenue N.E.
Seattle, WA 98195

Reprinted from

Canadian Journal of Fisheries and Aquatic Sciences

Réimpression du

Journal canadien des sciences halieutiques et aquatiques

Inferring the distribution of the parameters of the von Bertalanffy growth model from length moments

R. L. Burr

Volume 45 • Number 10 • 1988

Pages 1779–1788

Canada

Fisheries and Oceans Pêches et Océans

Printed in Canada by The Runge Press Limited

Inferring the Distribution of the Parameters of the von Bertalanffy Growth Model from Length Moments

Robert L. Burr

Center for Quantitative Science in Fisheries, Forestry, and Wildlife and School of Fisheries, University of Washington, Seattle, WA 98195, USA

Burr, R. L. 1988. Inferring the distribution of the parameters of the von Bertalanffy growth model from length moments. Can. J. Fish. Aquat. Sci. 45: 1779–1788.

A theoretical approach is described for determining the joint distribution of the parameters of the von Bertalanffy growth model from statistical moments of length. The approach extends the work of K. J. Sainsbury, who had demonstrated that different mean parameter estimates are obtained by assuming that the von Bertalanffy equation applies to individual fish rather than to groups of fish. Sainsbury articulated the goal of studying the joint probability distributions of K and L_∞ in animal populations and developed a maximum likelihood procedure for estimating the parameters of particular distributional forms describing K and L_∞, which were assumed for mathematical convenience to be statistically independent. The primary goal of the present paper is to provide a framework for future research in generalizing Sainsbury's approach by considering (K, L_∞) to be a random vector described by a joint probability density function and by allowing broader classes of distributions to be considered. Minimum cross-entropy (MCE) inversion, an information–theoretic methodology for approximating probability distributions, is shown to be effective in selecting a reasonable and unique joint distribution corresponding to observable length moments. Appealing features of the MCE methodology include the ability to include prior knowledge of uncertain applicability and the capacity of the resulting approximate distribution to represent potential stochastic dependencies between the von Bertalanffy parameters. Several numerical examples, using simulated and historical data, are presented to illustrate how information about the variation and covariation of L_∞ and K can be inferred from a minimal set of length moments. The directions developed in this paper are far from a practical and useful methodology. The MCE inversion procedure is a "method of moments," with no statistical assessment of reliability. Further research is needed to make this promising pdf approximation scheme better suited for real fisheries problems.

Une approche théorique est décrite afin d'établir la distribution à plusieurs variables des paramètres du modèle de croissance de von Bertalanffy à partir de moments statistiques relatifs à la longueur. L'approche pousse plus loin les travaux de K. J. Sainsbury, qui a démontré que l'on peut obtenir différentes estimations moyennes de paramètres en supposant que l'équation de von Bertalanffy s'applique à chacun des poissons plutôt qu'à des groupes de poissons. Sainsbury a formulé le but de l'étude des distributions de probabilité à plusieurs variables des paramètres K et L_∞ chez des populations animales, et a élaboré une méthode du maximum de vraisemblance en vue d'évaluer les paramètres de formes particulières de distribution décrivant K et L_∞, qu'on a supposé statistiquement indépendants pour des raisons de commodité mathématique. Le but principal du présent article est de fournir un cadre de travail pour les futures recherches en généralisant l'approche de Sainsbury en considérant (K, L_∞) comme un vecteur aléatoire décrit par une fonction de densité de probabilité à plusieurs dimensions, et en permettant de tenir compte de classes de distributions plus larges. L'inversion d'entropie croisée minimale (ECM), méthode théorique pour évaluer approximativement des distributions de probabilités, s'est révélée efficace pour choisir une distribution à plusieurs variables raisonnable et particulière correspondant à des moments observables relatifs à la longueur. Les caractéristiques intéressantes de cette méthode comprennent la possibilité d'inclure des données antérieures d'applicabilité incertaine et la capacité de la distribution approximative résultante de représenter des dépendances stochastiques potentielles entre les paramètres de l'équation de von Bertalanffy. Plusieurs exemples numériques, faisant appel à des données simulées et antérieures, sont présentés afin de montrer comment il est possible d'inférer des données sur la variation et la covariation de L_∞ et de K à partir d'un ensemble minimal de moments relatifs à la longueur. Les orientations élaborées dans cet article sont loin de constituer une méthode utile et pratique. L'inversion ECM est une « méthode des moments » sans évaluation statistique de la fiabilité. Il faut effectuer d'autres recherches pour que cette méthode d'approximation de la fonction de distribution des probabilités convienne mieux aux vrais problèmes des pêches.

Received February 19, 1987
Accepted May 31, 1988
(J9148)

Reçu le 19 février 1987
Accepté le 31 mai 1988

A pervasive problem in quantitative natural resource management is how to infer some property of a complex natural system from indirect, fragmentary, and highly summarized evidence. Although background information is frequently abundant, most available statistical methods either rigidly build it into the analysis or pretend that it is not there. It is important to develop methods that can merge partial or aggregated information with contextual knowledge of indeterminable relevance into a reasonable picture about some aspect of the state of the fishery. The present paper begins the development of an approach to study the joint probability distribution of the parameters of the von Bertalanffy growth model. While the direction taken, based on asymptotic results from information theory, is far from a complete applied methodology, it explicitly models the available partial evidence while allowing the measured inclusion of prior or background information of unquantifiable validity. Although the focus here will be on inferring information about growth from a limited set of moments of length at several ages, it is hoped that this presentation will stimulate interest in the application of information–theoretic inversion procedures to other difficult inference problems in fisheries.

von Bertalanffy Growth

The von Bertalanffy model is a widely applied mathematical representation of growth in biological entities (von Bertalanffy 1938). While originally developed in the context of theoretical physiology, the model is now understood as a remarkably effective empirical approximation to the growth of individuals of many species and of their component organ systems.

The von Bertalanffy equation

$$(1)\quad l(t) = L_\infty(1 - e^{-K(t-t_0)})$$

describes the tendency of the rate of an animal's growth in length to decrease with age. It is a solution of the simple linear first-order differential equation

$$(2)\quad \frac{dl(t)}{dt} = K(L_\infty - l(t))$$

under the initial condition that

$$(3)\quad l(t_0) = 0.$$

The von Bertalanffy model is determined by three parameters: L_∞, the asymptotic length, K, the Brody growth constant (Brody 1945) (sometimes incorrectly referred to as the growth rate), and t_0, the initial time. The asymptotic length L_∞ is the maximum size the animal can theoretically attain if it is allowed to grow indefinitely. The Brody growth constant K is one of the factors in determining the change of length with time, the other factor being the difference between the current length and the asymptotic length. The initial time t_0 is a parameter controlling the horizontal placement of the von Bertalanffy curve. It can be viewed as an empirical initial condition that needs to be fit, the time at which a fish would have had zero length had it grown along the von Bertalanffy curve for its whole life. Because the von Bertalanffy model often does not fit particularly well near the time origin, most researchers hesitate to invest too much biological interpretation in this quantity. While there is probably no single species that is perfectly described by the von Bertalanffy model, it has been and will probably continue to be an important means by which the growth of biological organisms is summarized and compared (Pitt 1970; Green 1973; Daan 1974; Bowering 1976; Ralph and Maxwell 1977). Many of the alternative growth models in use today can be viewed as embellishments, extensions, or reactions to this venerable representation (Richards 1959; Silliman 1967; Pauly and Gaschutz 1979; Gaschutz et al. 1980; Schnute 1981).

Almost universally, the parameters of the von Bertalanffy equation are estimated in practice from data representing a group made up of many individual animals. Questions have been raised about the efficacy of estimating the von Bertalanffy parameters (Knight 1968; Bayley 1977; Roff 1980) and using them to contrast subpopulations (Gallucci and Quinn 1979). Sainsbury (1980) has shown that there is a crucial difference between assuming that the von Bertalanffy equation applies to individual animals and assuming that it applies to a group of animals.

Sainsbury's Model

Sainsbury (1980) has presented a model in which individual animals are assumed to grow in accordance with the von Bertalanffy growth equation with parameters that are fixed with respect to each fish but which are allowed to vary randomly between fish. This approach was motivated by concerns that a mathematical description of the growth of an individual animal might be inappropriate for describing the typical growth for a population of animals and vice versa. Sainsbury's model for the population essentially becomes the characterization of the probability distribution of K and the probability distribution of L_∞ for the group.

With the assumptions that K and L_∞ are probabilistically independent, that L_∞ is normally distributed, and that K follows a gamma distribution, Sainsbury derived expressions for the expected value and variance of the length at age T, as well as the first two moments of the growth increment I_T. He also presented a maximum likelihood method of estimating the parameters of the assumed probability density functions from collected data.

Focusing on the distinction between $E[e^{-KT}]$ and $e^{-E[K]T}$, Sainsbury demonstrated that different parameter estimates are obtained by assuming that the von Bertalanffy equation applies to individual fish rather than to groups of fish, an observation that should apply to other nonlinear growth models as well.

A sympathetic criticism of Sainbury's model concerns the strong assumptions made about the form of the probability distributions characterizing K and L_∞. As that author pointed out, the selected distributions were chosen as a compromise between reality and analytic tractability. If L_∞ were truly Gaussian, then there is a finite probability of observing a fish with a negative length. While the gamma distribution employed by Sainsbury to describe the variability of K admits a broad class of densities, it also is a function with just a few degrees of freedom, and there are distributional shapes that it fits poorly.

But it is the assumption of probabilistic independence of K and L_∞ that is the most serious limitation of the model. There is considerable evidence that estimates of these parameters significantly covary in natural populations (Knight 1968; Gallucci and Quinn 1979), with empirical assessments of the correlations between K and L_∞ ranging as high as -0.999. Instead of being independent, these parameter estimates are so nearly dependent on each other that it is even possible to consider reparameterizing the growth equation using just one parameter rather than two, an approach initiated by Gallucci and Quinn (1979). It is not clear what inferences about the correlation of the parameters themselves can be made from the strong empirical correlation of the parameter *estimates*, but the possibility of nonindependence merits further study. The primary goal of the present paper is to provide a framework for generalizing Sainsbury's approach by considering (K, L_∞) to be a random vector described by a joint probability density function and by allowing broader classes of distributions to be considered. Of particular interest will be distributional forms that permit the estimation of the degree of dependence of the two parameters.

An Ill-Posed Inverse Problem

In this paper we explore the possibility of inferring a general joint probability density function $q(L_\infty,K)$ from measured length moments of several age classes. For clarity, we will first assume that the initial time t_0 is known by other means, and we will lose no generality in setting it to zero. Later in the paper, we will relax this assumption and explicitly consider t_0 as a random quantity.

Suppose we have evidence about the distribution of lengths at a succession of ages, in particular, the first two (noncentral) moments. The moments for age class T can be related to the underlying joint density function by means of the integral equations

$$(4)\quad E(l_T) = \iint L_\infty(1-e^{-KT})q^{\text{true}}(L_\infty,K)dL_\infty dK$$

and

$$(5)\quad E(l_T^2) = \iint L_\infty^2(1-e^{-KT})^2 q^{\text{true}}(L_\infty,K)dL_\infty dK.$$

These equations are well-defined from right to left, since if we know the underlying pdf $q^{\text{true}}(L_\infty,K)$, then we can compute expected values unambiguously. Each expected value can be considered an integral constraint on the true pdf, a well-defined mapping of the continuous function of two variables $q^{\text{true}}(L_\infty,K)$ into a single scalar number or moment.

We would like to solve the inverse problem, that is, to deduce a reasonable approximation of $q^{\text{true}}(L_\infty,K)$ given a finite set of first- and second-order moments at several ages. The moments contain diffuse information about the underlying pdf, with the integral kernel acting as a window through which the unknown pdf $q^{\text{true}}(L_\infty,K)$ is indirectly perceived.

This is an example of a class of well-known ill-posed inverse problems with an infinite convex set C of solution pdf's $r(L_\infty,K)$, each consistent with the given set of integral constraints in the form of the measured moments.

Sainsbury turned this ill-posed problem into a well-posed one by assuming a particular form for the solution pdf that has a relatively small number of parameters. Anyone who does this runs the risk of being criticized for choosing a form that does not capture all the qualities of reality that someone else might think important. For example, Sainsbury's model is inadequate for the study of the covariation of L_∞ and K because he assumed that they are independent. In a sense, he has incorporated possibly spurious prior information into the problem in a manner that his estimation technique cannot overcome.

Prior Knowledge

Applied statisticians often express dissatisfaction with formal estimation procedures in statistics because background information has to be either ignored or rigidly adhered to (Hodges and Lehman 1952; Blum and Rosenblatt 1967; Jaynes 1968; Kashyap 1971). Neither position is desirable, nor does it model the processes of human understanding. Contextual knowledge is unquestionably relevant to the applied problem, but difficult to merge gracefully with new information in the form of actual measurements of the system. If we had no new measurements at all, we would base our predictions on our experience with similar systems, or on our experience with the behavior of this particular system in the past. If we had a limited amount of information about the actual system under study, we would want a solution consistent with both the current data and our prior understanding. It would seem reasonable to give precedence to the new accurate knowledge and then resolve any remaining inferential ambiguities by appealing to the prior knowledge base. Kullback's Principle of minimum cross-entropy (Kullback 1959) provides a rule for picking a unique solution using both the new system measurements and the background knowledge. It states that from a set of possible solutions, we should choose the one most similar to our prior information.

Cross-Entropy Minimization

In fisheries management, there is often a wealth of prior information of uncertain applicability about a particular fish stock, coming perhaps from historical records, experience with similar species, or theoretical principles. It would seem reasonable to use Kullback's Principle to resolve the ambiguity of the solution set C by picking the element in C most similar to an assumed prior $p(L_\infty,K)$ that somehow represents our prior understandings. If we have absolutely no insight into the problem area, we would specify a uniform pdf, and our minimum cross-entropy (MCE) procedure would reduce to the well-known maximum entropy (ME) formalism. No matter how the prior density is specified, the selected element must still satisfy all of the given moments because it is chosen from a set composed only of elements that meet all of the moment constraints. Our prior knowledge is never allowed to contradict or restrict the evidence of the current data but guides us to a well-defined solution when the current data is insufficient.

To implement this optimization procedure, it is necessary to be precise about how to measure the dissimilarity between two pdf's $p(L_\infty,K)$ and $q(L_\infty,K)$. In information theory this distance is commonly quantified by Kullback's cross-entropy functional

$$(6)\quad H[q(L_\infty,K),p(L_\infty,K)] = \iint q(L_\infty,K) \times \log \frac{q(L_\infty,K)}{p(L_\infty,K)}\, dL_\infty dK.$$

Also known as the directed divergence, the minimum discrimination information, or the Kullback–Leibler number, this distortion measure on the space of probability density functions can be interpreted as the expected value of a log-likelihood ratio. Cross-entropy has been used previously as a distance measure in fisheries application by several authors (MacDonald and Pitcher 1979; Schnute and Fournier 1980).

Kullback's Principle would have us find the posterior pdf $q(L_\infty,K)$ that minimizes $H[q(L_\infty,K),p(L_\infty,K)]$ while exactly satisfying the set of constraint equations

$$(7)\quad m_j = \iint f_j(L_\infty,K)q(L_\infty,K)dL_\infty dK,\quad j=1,\ldots,M.$$

That is, we are treating m_j as the average value of the scalar kernel function $f_j(L_\infty,K)$ when the true probability density function is $q(L_\infty,K)$. For example, the kernel corresponding to the mean length at age 5 is $f_5(L_\infty,K) = L_\infty(1-e^{-K5})$. It should be noted that the expected value constraints (equations 4 and 5) corresponding to our data are in the form of Eq. 7.

The general problem of pdf approximation using the MCE criterion has been studied, and a solution for the posterior pdf

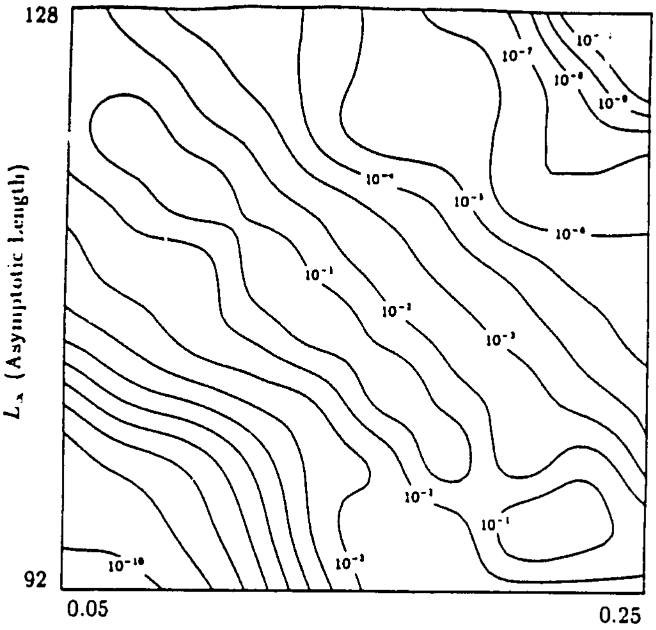

FIG. 1. MCE posterior pdf for Example 1.

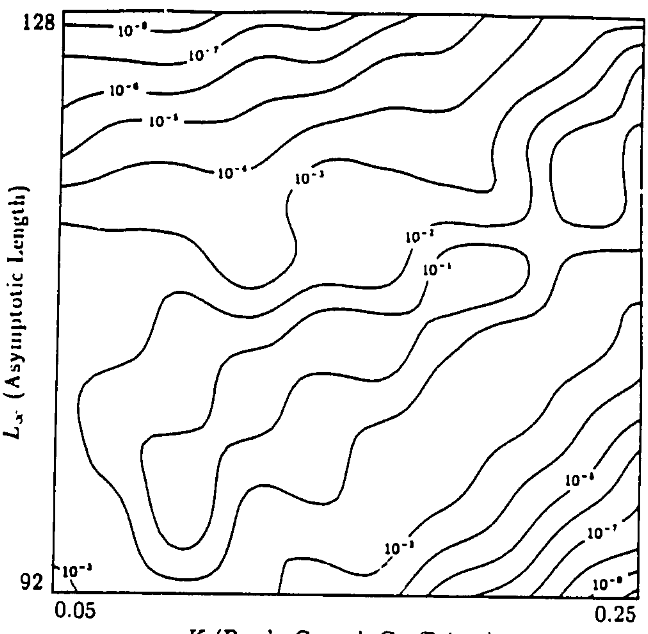

FIG. 2. MCE posterior pdf for Example 2.

$q(L_\infty,K)$ for nonpathological integral kernels $f_i(L_\infty,K)$ is well-known (Shore and Johnson 1981):

(8) $q^{mce}(L_\infty,K) = p(L_\infty,K)\exp\{-\sum_{i=0}^{M} \beta_i f_i(L_\infty,K)\}$

where the $\{\beta_i\}$ are Lagrange multipliers whose values are made consistent with the measured moments m_j by solving the set of nonlinear equations

(9) $m_j^{measured} = \iint f_j(L_\infty,K)p(L_\infty,K)\exp\{-\sum_{i=0}^{M} \beta_i \times f_i(L_\infty,K)\}dL_\infty dK \quad j=1,\ldots,M$

along with the normalizing constraint

(10) $1 = \iint p(L_\infty,K)\exp\{-\sum_{i=0}^{M} \beta_i f_i(L_\infty,K)\}dL_\infty dK.$

The latter constraint comes about because the posterior $q(L_\infty,K)$ is a probability density function and hence must integrate to unity. In practice we generally have to solve this system of nonlinear equations using numerical methods such as the Newton–Raphson procedures.

A nonrigorous derivation of the form of the MCE posterior density is presented in Appendix B. A detailed consideration of the conditions under which this result exists and is unique, which is beyond the scope of the present paper, can be found in Csiszár (1975) and Johnson (1979). While there are kernel functions f_i for which no unique MCE solution exists, this possibility is rarely relevant in practice. Existence and uniqueness are guaranteed if the kernel functions can be written as a multivariate power series, or if the kernel functions, no matter how discontinuous, are bounded. A review of successful applications of MCE inversion techniques can be found in Shore (1984).

We will now apply this procedure for selecting a unique pdf to the problem of inferring the joint pdf of L_∞ and K from length moments of several age classes, information that is commonly collected, for example, in the determination of age–length keys. Suppose we knew the mean and the mean square of length at three specific ages. We could then write seven equations constraining the true pdf $q^{true}(L_\infty,K)$:

(11) $1 = \iint q^{true}(L_\infty,K)dL_\infty dK$

(12) $m_1 = \iint L_\infty(1-e^{-Kt_1})\, q^{true}(L_\infty,K)dL_\infty dK$

(13) $ms_1 = \iint L_\infty^2(1-e^{-Kt_1})^2 q^{true}(L_\infty,K)dL_\infty dK$

(14) $m_2 = \iint L_\infty(1-e^{-Kt_2})q^{true}(L_\infty,K)dL_\infty dK$

(15) $ms_2 = \iint L_\infty^2(1-e^{-Kt_2})^2 q^{true}(L_\infty,K)dL_\infty dK$

(16) $m_3 = \iint L_\infty(1-e^{-Kt_3})q^{true}(L_\infty,K)dL_\infty dK$

(17) $ms_3 = \iint L_\infty^2(1-e^{-Kt_3})^2 q^{true}(L_\infty,K)dL_\infty dK.$

From equation 8 we can directly write the form of the MCE posterior solution:

(18) $q^{mce}(L_\infty,K) = p(L_\infty,K)\cdot\exp\{\beta_0 + \beta_1\cdot L_\infty(1-e^{-Kt_1}) + \beta_2\cdot L_\infty^2(1-e^{-Kt_1})^2 + \beta_3\cdot L_\infty(1-e^{-Kt_2}) + \beta_4\cdot L_\infty^2(1-e^{-Kt_2})^2 + \beta_5\cdot L_\infty(1-e^{-Kt_3}) + \beta_6\cdot L_\infty^2(1-e^{-Kt_3})^2\}.$

The MCE posterior density is completely determined when we fit the Lagrange multipliers β_i to reproduce the measured moments, usually with an iterative numerical procedure. A number of nonlinear minimization techniques seem to work equally well, including Newton–Raphson methods (Gokhale and Kullback 1978), the MINPACK minimization routines (More et al 1980), and the Nelder–Mead simplex procedure (Nelder and Mead 1965). The following examples with illustrate how effective the MCE procedure is at recovering the latent pdf of the von Bertalanffy parameters.

Example 1

The first two numerical examples, using simulated data, are designed to show how information about the covariation of L_∞ and K can be inferred from a minimal set of length moments. The data, while fictitious, were created to emulate some of the growth characteristics of a slow-growing marine organism like

4-

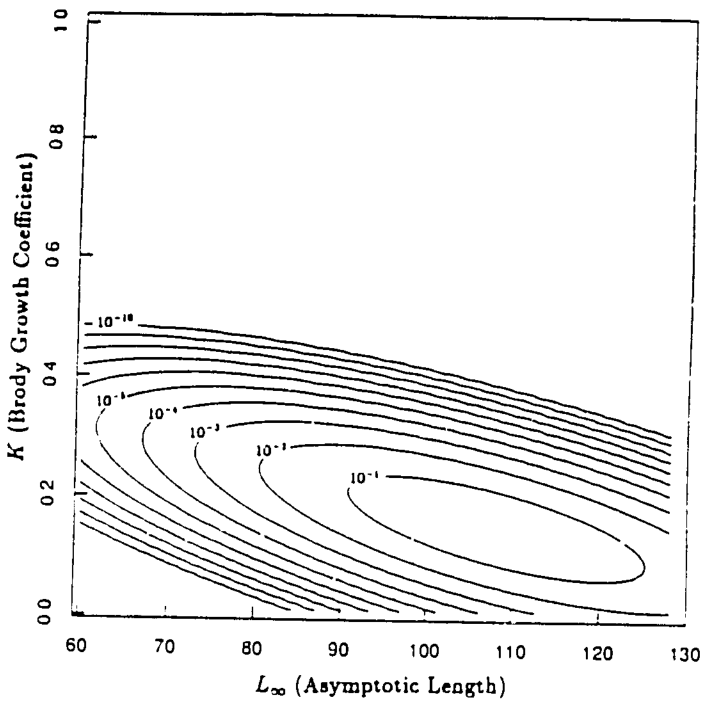

FIG. 3. Simulated true density for Example 3.

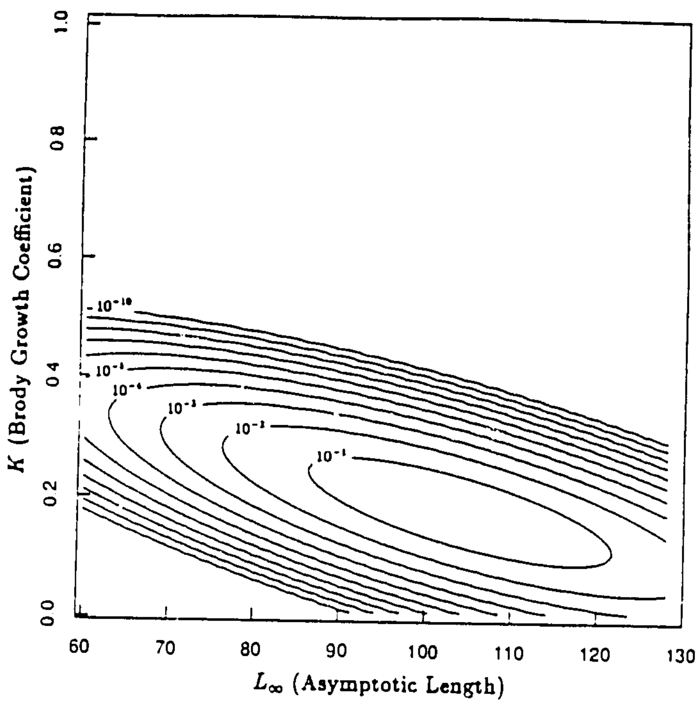

FIG. 5. MCE posterior density for Example 3.

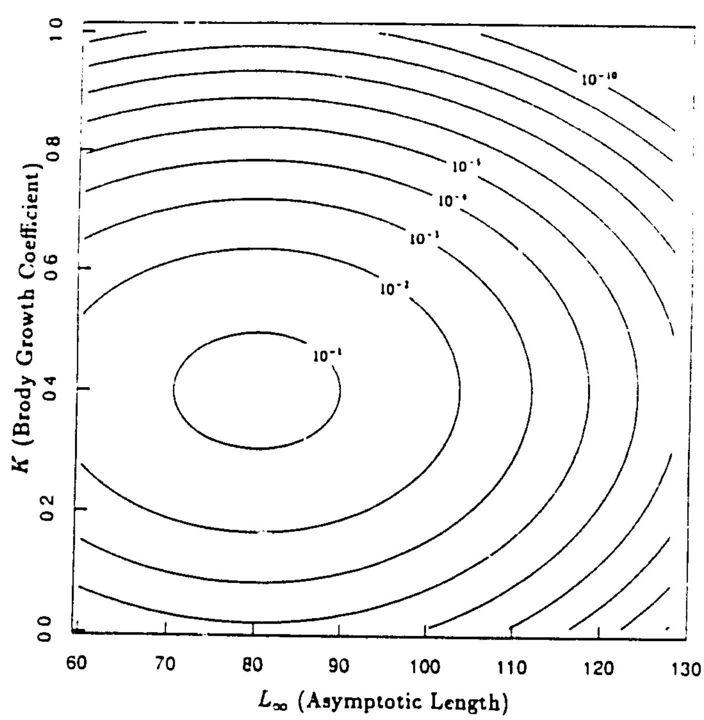

FIG. 4. Assumed prior density for Example 3.

the turbot, which may continue to grow into the second decade of its life. One hundred (L_∞, K) pairs were generated from a bivariate Gaussian distribution such that μ_{L_∞} = 108 cm, σ_{L_∞} = 10 cm, μ_k = 0.15, σ_k = 0.05, and $\rho_{(L_\infty,K)}$ = −0.99. One sample pair thus generated had a K-element less than zero and was discarded. von Bertalanffy curves were computed for each of the remaining pairs and the ensemble mean and mean square "collected" for lengths at ages 1, 10, and 20 yr. While moments at these three ages nicely bracket the initial, intermediate, and asymptotic portions of the growth curves, this example is somewhat artificial in that if we actually collected information at these ages, we would probably also have data about the lengths at all the intermediate ages as well. Nevertheless, this hypothetical example is designed to show that we can deduce a great deal from a very limited set of information. To emphasize the recovery of covariance phenomena, the prior pdf $p(L_\infty, K)$ was assumed to have the same mean and standard deviation in each component as the generating distribution, but to be independent in dimensions, that is $\rho_{(L_\infty,K)}$ = 0.0.

The Lagrange multipliers $\beta_i, i=0,\ldots,6$, were fit with the Nelder–Mead simplex algorithm to the six estimated moments and the normalizing constraint

$$1 = \iint q(L_\infty, K; \{\beta_i\})\, dL_\infty dK. \tag{19}$$

Figure 1 illustrates the logarithmically spaced equiprobability contours of the MCE posterior pdf for this problem. The information about the interaction of the von Bertalanffy parameters, diffusely encoded in the measured moments, is recovered by the MCE posterior density. It is obvious from inspection of equation 18 that the posterior pdf is not Gaussian, even though the assumed prior density and the true density being approximated are both Gaussian.

TABLE 1. Porgy data (values inferred by Tanaka (1962) from a length-frequency distribution).

Age	Mean	SD
I	10.99	0.8
II	15.26	1.2
III	19.84	1.4
IV	23.50	1.2
V	26.82	1.4

Example 2

Another data set was generated, under the same conditions as given in Example 1, except that the correlation between L_∞ and K was set to ρ = +0.99. The same independent prior density was assumed as well.

Once again the MCE posterior density for this example, as depicted in Fig. 2, shows the influence of measured length

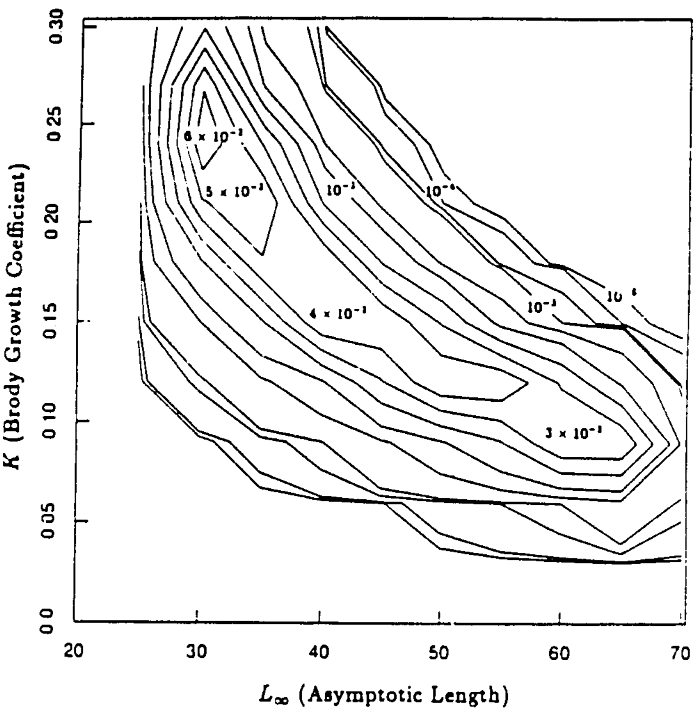

FIG. 6. MCE posterior density for Example 4 (Tanaka's (1962) porgy data).

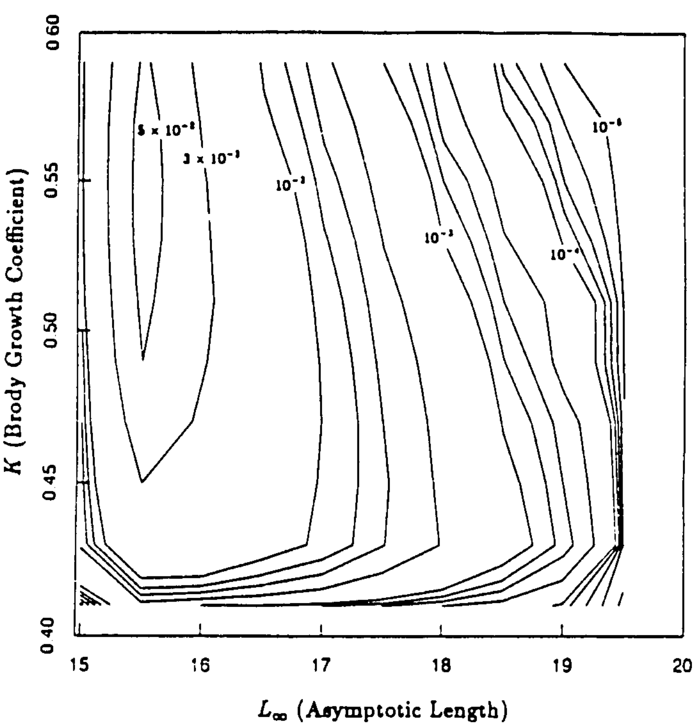

FIG. 7. MCE posterior density for Example 5 (Sund's (1911) sprat data, as reported in Ricker (1969)).

TABLE 2. Sprat data (Sund's (1911) sprat (*Clupea sprattus*) data, as reported by Ricker (1969), composed of back-calculated lengths of the 1903 year-class)

Age	Mean	SD
I	5.86	1.124
II	10.29	1.387
III	12.79	1.056
IV	13.92	0.950
V	14.65	0.868

moments. The interaction of L_∞ and K is strongly apparent in this posterior density.

Example 3

In the preceding examples we have specified a prior pdf with the correct mean and variance in each dimension in order to emphasize the recovery of the covariance information. Example 3 is presented to demonstrate that the MCE procedure can overcome the effects of an unrepresentative prior distribution. One thousand (L_∞,K) pairs were generated from a bivariate Gaussian density, trimmed at zero in both dimensions to preclude negative asymptotic length and negative growth, with parameters $\mu_{L_\infty} = 108$ cm, $\sigma_{L_\infty} = 10$ cm, $\mu_k = 0.15$, and $\rho_{(L_\infty,K)} = -0.70$. A contour plot of this density is depicted in Fig. 3. The prior pdf $p(L_\infty,K)$ was assumed to have parameters $\mu_{L_\infty} = 80$ cm, $\sigma_{L_\infty} = 20$ cm, $\mu_k = 0.4$, $\sigma_k = 0.1$, and $\rho_{(L_\infty,K)} = 0.0$ and is displayed in Fig. 4. Forty moments were "collected" from the simulated data, representing the mean and mean square of length at ages 1–20. A contour plot of the MCE posterior pdf corresponding to this 41-variable nonlinear minimization problem is shown in Fig. 5. Obviously the MCE procedure has recovered the important characteristics of the true density from the length moments, in spite of the choice of prior.

Example 4

Few attempts have been made to apply this MCE method to real fisheries data, in part because generalizable software has not yet been written. Hence, every application currently requires the development of a custom program. Two preliminary analyses will now be presented.

Tanaka (1962) inferred the moments of length-at-age for porgy (*Taius tumifrons*) from a length–frequency distribution collected in 1950. His estimates of the mean and standard deviation for each age are presented in Table 1.

The MCE posterior density, computed using a uniform (noninformative) prior, is presented in Fig. 6, and evidences negative diagonal structure.

Example 5

Ricker (1969) summarized back-calculated length data for a sample of sprats (*Clupea sprattus*) collected in 1908, and analyzed by Sund in 1911. The mean and standard deviation of back-calculated length at various ages for the 1903 year-class are summarized in Table 2.

The MCE posterior density corresponding to these constraints, computed using the uniform prior pdf, is presented in Fig. 7. Negative covariation is less apparent in this example, although there is an interesting slight bulge of probability toward higher L_∞ values for lower K values in this contour plot.

In the preceding discussion we have explicitly assumed that the initial conditions parameter t_0 in the von Bertalanffy model is identically zero, or constant, for all the animals in the group. In many applications, this is not a credible assumption, and we must fit t_0 as well. The general method we have outlined above for two latent variables can be trivially extended to infer a trivariate posterior pdf $q(L_\infty,K,t_0)$ from length moments. In the next section we will consider a special trivariate application that can be reduced to the bivariate formalism just considered.

TABLE 3. Eel data (mark–recapture data for eels (*Anguilla australis*) in the Doyleston Drain, New Zealand, summarized from Fig. 2 in Burnet (1969)).

Mean length at first capture	46.41
SD of length at first capture	12.29
Mean length 1 yr later	50.40
SD of length 1 yr later	11.10
Correlation of length over 1-yr interval	0.9845

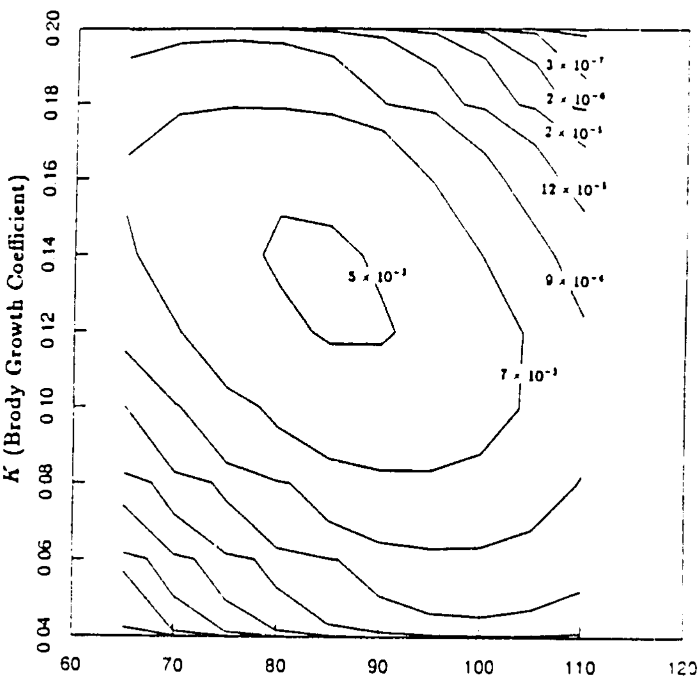

FIG. 8. MCE posterior density for Example 6 (Burnet's (1969) eel data).

Mark–Recapture Experiments

Until this point, the symbol t has really referred to the age of individual animals and not to chronological time. In a population of animals of mixed ages, it is possible to refer to t as actual time by allowing $-t_0$ for each organism to refer to that animal's age at time slice 0. However, to avoid confusion with the entrenched interpretation of t_0, we will introduce a new symbol τ to denote age at time slice 0, $\tau = -t_0$.

In some mark–recapture studies, the age of an animal cannot be determined without sacrificing it, although its length can be determined with an acceptable handling risk. It is of interest to infer the distribution of the von Bertalanffy parameters from the moments of length at two or more specific time points. For example, an investigator might dig up a number of clams of various but unknown ages, measure their lengths, mark each one with a unique identification code, release the animals in their original habitat, and then recapture and remeasure Δ years later.

The mean and mean square of length at each measurement time slice can be expressed as a constraint on the trivariate pdf $q^{true}(L_\infty,K,\tau)$:

$$(20)\quad E(l_0) = \iiint L_\infty(1-e^{-K\tau})q^{true}(L_\infty,K,\tau)dL_\infty dKd\tau$$

$$(21)\quad E(l_0^2) = \iiint L_\infty^2(1-e^{-K\tau})^2 q^{true}(L_\infty,K,\tau)dL_\infty dKd\tau$$

$$(22)\quad E(l_\Delta) = \iiint L_\infty(1-e^{-K(\Delta+\tau)})q^{true}(L_\infty,K,\tau)dL_\infty dKd\tau$$

$$(23)\quad E(l_\Delta^2) = \iiint L_\infty^2(1-e^{-K(\Delta+\tau)})^2 q^{true}(L_\infty,K,\tau)dL_\infty dKd\tau.$$

Because we have assigned a unique code to each animal, we can also estimate from our pair of length measurements the powerful joint moment

$$(24)\quad E(l_0\cdot l_\Delta) = \iiint L_\infty^2(1-e^{-K\tau}) \times (1-e^{-K(\Delta+\tau)})q^{true}(L_\infty,K,\tau)dL_\infty dKd\tau.$$

Using the MCE methodology to infer a unique posterior pdf $q^{mce}(L_\infty,K,\tau)$, we can immediately write the form of the MCE posterior density:

$$(25)\quad q^{mce}(L_\infty,K,\tau) = p(L_\infty,K,\tau)\cdot\exp\{\beta_0 + \beta_1\cdot L_\infty(1-e^{-K\tau}) + \beta_2\cdot L_\infty^2(1-e^{-K\tau})^2 + \beta_3\cdot L_\infty(1-e^{-K(\Delta+\tau)}) + \beta_4\cdot L_\infty^2(1-e^{-K(\Delta+\tau)})^2 + \beta_5\cdot L_\infty^2(1-e^{-K\tau})(1-e^{-K(\Delta+\tau)})\}.$$

The MCE posterior probability density function of the von Bertalanffy parameters is explicitly determined when the Lagrange multipliers are adjusted so as to exactly reproduce the measured moments corresponding to the constraints.

General MCE inversion methods can become computationally expensive in dimensions higher than two. It is of practical interest to consider when this trivariate problem can be reduced to two dimensions. In many mark–recapture experiments, the distribution of ages at time zero can be considered to be independent of the other two von Bertalanffy parameters and perhaps even known or manipulated.

If it can be justified that the distribution of ages at time zero is independent of the other parameters and can be known a priori, the five constraints corresponding to the measured moments can be rewritten

$$(26)\quad E(l_0) = \iint L_\infty(1-\nu_1(K))q^{true}(L_\infty,K)dL_\infty dK$$

$$(27)\quad E(l_\Delta) = \iint L_\infty(1-e^{-K\Delta}\nu_1(K))q^{true}(L_\infty,K)dL_\infty dK$$

$$(28)\quad E(l_0^2) = \iint L_\infty^2(1-2\nu_1(K)+\nu_2(K))q^{true}(L_\infty,K)dL_\infty dK$$

$$(29)\quad E(l_\Delta^2) = \iint L_\infty^2(1-2e^{-K\Delta}\nu_1(K)+e^{-2K\Delta}\nu_2(K)) q^{true}(L_\infty,K)dL_\infty dK$$

$$(30)\quad E(l_0\cdot l_\Delta) = \iint L_\infty^2(1-\nu_1(K)-e^{-K\Delta}\nu_1(K)+e^{-K\Delta}\nu_2(K)) \times q^{true}(L_\infty,K)dL_\infty dK$$

where

$$(31)\quad \nu_1(K) = E_\tau(e^{-K\tau})$$

and

$$(32)\quad \nu_2(K) = E_\tau(e^{-2K\tau})$$

which by our assumptions are known quantities.

For example, if the age distribution at time slice 0 is uniformly distributed between ages a^- and a^+, then

$$(33)\quad \nu_1(K) = \frac{(e^{-Ka^+}-e^{-Ka^-})}{K(a^+-a^-)}$$

and

$$(34)\quad \nu_2(K) = \frac{(e^{-2Ka^+}-e^{-2Ka^-})}{2K(a^+-a^-)}.$$

In any case the MCE posterior pdf takes the form

$$(35)\quad q^{mce}(L_\infty,K,\tau_0) = p(L_\infty,K)\cdot q_\tau^{known}(\tau)\cdot\exp\{\beta_0 + \beta_1\cdot L_\infty(1-\nu_1(K)) + \beta_2\cdot L_\infty^2(1-2\nu_1(K)+\nu_2(K)) + \beta_3\cdot L_\infty K(1-e^{-K\Delta}\nu_1(K)) + \beta_4\cdot L_\infty^2(1-2e^{-K\Delta}\nu_1(K)+e^{-2K\Delta}\nu_2(K)) + \beta_5\cdot L_\infty(1-\nu_1(K)-e^{-K\Delta}\nu_1(K)+e^{-K\Delta}\nu_2(K))\}$$

and as before, we must adjust the Lagrange multipliers until the MCE posterior density reproduces the given moments.

Example 6

Burnet (1969) has studied the growth of freshwater eels in New Zealand with a mark–recapture methodology. Data for *Anguilla australis* was hand-digitized from Fig. 2 in Burnet's paper to estimate the mean, standard deviation, and correlation of the length at first capture, and the length a year later, and are summarized in Table 3. Under the assumptions of a uniform prior density and a hypothesized negative exponential age distribution at first capture (with mean age = 8.5 yr), the estimated pdf of the von Bertalanffy parameters is presented in Fig. 8, in which some negative diagonal structure can be observed.

Discussion

The MCE formalism, which we have here applied to the problem of determining the distribution of the von Bertalanffy parameters, exploits the variations in the information about $q^{true}(L_\infty,K,t_0)$ implied by projection through different integral kernel functions. The concept developed in this paper can be extended to any deterministic growth model that can represent the growth of an individual animal.

The strengths and weaknesses of the MCE approach lie in the ability to insert background knowledge of unknown applicability into the problem by way of the prior pdf. Examination of equation 11 shows that the posterior pdf is in the form of the prior pdf multiplied by an exponential distortion factor. If the prior pdf is a good guess, then the magnitude of the Lagrange parameters will be small and the analytic degrees of freedom of the model will be spent "fine-tuning" the posterior pdf, explaining what is not already known about the system under study. If the prior pdf is not a good guess, the magnitude of the Lagrange parameters will be large as the degrees of freedom of the distortion function are spent overcoming the unrepresentativeness of the prior pdf. It should be generally noted that if the hypothesized prior $p(L_\infty,K,t_0)$ happens to be identical to the true pdf $q^{true}(L_\infty,K,t_0)$, then $\beta_i = 0$, $\forall_i$ and $q^{mce}(L_\infty,K,t_0) = p(L_\infty,K,t_0) = q^{true}(L_\infty,K,t_0)$.

As the number of measured moments m_i increases, the MCE procedure can overcome any specification of the prior pdf so long as $p(L_\infty,K,t_0) > 0$ everywhere. In the information theory literature, this appealing behavior of the MCE inverse is termed the "washing out" of old uncertain information with new facts. Whenever there is an inconsistency between the prior $p(l)$ and the actual data, the new data take precedence.

The MCE inverse methodology can be viewed as a formal way to deal with missing information problems by adapting the form of the model to the available moments. The kernels corresponding to missing moments are simply not present in the argument of the exponential function in equations 8 and 18.

It is a "method of moments" inverse technique. That is, the Lagrange parameters $\{\beta_i\}$ are defined, not statistically estimated. Sampling variability in the measured moments will be propagated through to the posterior density $q(l)$.

Practical implementation of the MCE methodology generally relies on some kind of iterative nonlinear minimization procedure. Convergence can be accelerated if either $p(L_\infty,K,t_0)$ is a very good guess or if a good starting value for $\vec{\beta}$ is chosen. One method of quickly estimating a reasonable starting value is to assume that $\vec{\beta}$ is small enough that the multiplicative exponential distortion factor in equation 11 can be replaced with a Taylor series linearization about the origin. We can then write an approximate expression for the MCE posterior as

$$(36)\quad q^{mce}(\vec{x};\vec{\beta}) = p(\vec{x})[1+\sum_i^M \beta_i f_i(\vec{x})]$$

which can be fit to the moments using linear mathematics.

This form is suggestive because it closely resembles a class of pdf estimators based on orthogonal function expansions that may be computed very efficiently. If $p(\vec{x})$ is multivariate Gaussian, and $f_i(\vec{x})$ are appropriately chosen multivariate Hermite polynomials, then equation 36 is the well-known Hermite orthogonal expansion of the pdf. Similarly, if $p(\vec{x})$ is multivariate exponential, multivariate Laguerre polynomials form a convenient orthogonal expansion. A future research direction is to explore the adequacy of these pdf estimates of the von Bertalanffy parameters, which are suboptimal approximations in the cross-entropy sense, but which can be estimated with significantly less computation than the general MCE form.

Estimation Issues

It has been stressed that the MCE inversion procedure is a "method of moments" where summaries of sampled data are assumed somewhat arbitrarily to be equivalent to asymptotic expected values, from which the Lagrange parameters are defined rather than statistically estimated. Anyone who has participated in fisheries research data collection or has had the responsibility of summarizing such data would be justifiably concerned about this suppression of uncertainty. The attempt to make this pdf approximation scheme better suited for practical problems is an active research topic.

The MCE posterior density is the optimal solution to a calculus of variations problem where the moments are represented as integral equality constraints on the unknown true density. It is also possible to formulate this problem using integral inequality constraints. For example, instead of using the equality constraint

$$(37)\quad m_i = \int f_i(L_\infty,K,t_0)q^{true}(L_\infty,K,t_0)dL_\infty dK dt_0,$$

two inequality constraints might be written:

$$(38)\quad m_i + 2\frac{\sigma_i}{\sqrt{N_i}} > \int f_i(L_\infty,K,t_0)q^{true}(L_\infty,K,t_0)dL_\infty dK dt_0$$

$$(39)\quad m_i - 2\frac{\sigma_i}{\sqrt{N_i}} < \int f_i(L_\infty,K,t_0)q^{true}(L_\infty,K,t_0)dL_\infty dK dt_0.$$

That is, confidence intervals based on some reasonable assessment of the variability of the moments due to sampling are employed to constrain the class of consistent densities. Kullback's Principle can still be applied to the now larger convex set of pdfs satisfying the given inequality constraints.

Until a satisfactory method is found to analytically determine the effect of moment estimation error on the posterior pdf, computationally expensive resampling methods of assessing the variability, such as jackknifing and bootstrapping (Efron 1982), can be employed.

Conclusion

The problem of determining of the general joint distribution of the parameters of the von Bertalanffy growth model can be approached as an ill-posed inverse problem. MCE inversion techniques allow the selection of a reasonable unique solution, directly incorporating background information when available.

Acknowledgements

This manuscript was prepared while the author was a postdoctoral fellow affiliated with the USAID-funded Management Assistance for Artisanal Fisheries (M.A.A.F.) Project at the University of Washington. The support and encouragement provided by the director of M.A.A.F., Dr. Vincent F. Gallucci, is gratefully acknowledged.

References

BAYLEY, P. 1977. A method for finding the limits of application of the von Bertalanffy growth model and statistical estimates of the parameters. J. Fish. Res. Board Can. 34: 1079–1084.

BLUM, J. R., AND J. ROSENBLATT. 1967. On partial a priori information in statistical inference. Ann. Math. Stat. 38: 1671–1678.

BOWERING, W. 1976. Distribution, age and growth, and sexual maturity of witch flounder (*Glyptocephalus cynoglossus*) in Newfoundland waters. J. Fish. Res. Board Can. 33: 1574–1584.

BRODY, S. 1945. Bioenergetics and growth. Reinhold Publishing Corporation, New York, NY.

BURNET, A. M. R. 1969. The growth of New Zealand freshwater eels in three Canterbury streams. N.Z. J. Mar. Freshwater Res. 3: 376–384.

CSISZAR, I. 1975. I-Divergence geometry of probability distributions and minimization problems. Ann. Prob. 3(1): 146–158.

DAAN, N. 1974. Growth of North Sea cod (*Gadus morhua*). Neth. J. Sea Res. 8: 27–48.

EFRON, B. 1982. The jackknife, the bootstrap, and other resampling plans. CMBS-NSF Reg. Comb. Ser. Appl. Math. SIAM No. 38: 92 p.

GALLUCCI, V. F., AND T. J. QUINN II. 1979. Reparameterization, fitting, and testing a simple growth model. Trans. Am. Fish. Soc. 108: 14–25.

GASCHUTZ, G., D. PAULY, AND N. DAVID. 1980. A versatile BASIC program for fitting weight and seasonally oscillating growth data. I.C.E.S. C.M. 1980/D:6, Statistics Committee.

GOKHALE, D. V., AND S. KULLBACK. 1978. The information in contingency tables. Marcel Dekker, New York, NY.

GREEN, R. 1973. Growth and Mortality in an Arctic intertidal population of *Macoma balthica* (Pelecypoda, Tellinidae). J. Fish. Res. Board Can. 30: 1345–1348.

HODGES, J. L., AND E. L. LEHMAN. 1952. The use of previous experience in reaching statistical decisions. Ann. Math. Stat. 23: 396–402.

JAYNES, E. T. 1968. Prior probabilities. IEEE Trans. Syst. Sci. Cybern. Vol. SSC-4, No. 3: 227–241.

JOHNSON, R. W. 1979. Axiomatic characterization of the directed divergences and their linear combinations. IEEE Trans. Inf. Theor. Vol. IT-25, No. 6: 709–716.

KASHYAP, R. L. 1971. Prior probabilities and uncertainty. IEEE Trans. Info. Theor. Vol. IT-17, No. 6.

KNIGHT, W. 1968. Asymptotic growth: an example of nonsense disguised as mathematics. J. Fish. Res. Board Can. 25: 1303–1307.

KULLBACK, S. 1959. Information theory and statistics. Wiley (1959), Dover (1969), New York, NY.

MACDONALD, P. D. M., AND T. J. PITCHER. 1979. Age groups from size-frequency data: a versatile and efficient method of analyzing distribution mixtures. J. Fish. Res. Board Can. 36: 987–1001.

MORE, J. J., B. GARBOW, AND K. HILLSTROM. 1980. User's guide for MINPACK-1. Argonne National Laboratory Technical Report ANL-80-74, DOE W-31-109-Eng-38.

NELDER, J. A., AND R. MEAD. 1965. A simplex method for function minimization. Comput. J. 7: 308–313.

PAULY, D., AND G. GASCHUTZ. 1979. A simple method for fitting oscillating length growth data, with a program for pocket calculators. I.C.E.S. C.M. 1979/G:24, Demersal Fisheries Committee.

PITT, T. 1970. Distribution, abundance, and spawning of yellowtail flounder, *Limanda ferruginea*, in the Newfoundland area of the Northwest Atlantic. J. Fish. Res. Board Can. 27: 2261–2271.

RALPH, R., AND J. MAXWELL. 1977. Growth of two Antarctic lamellibranchs: *Adamussium colbecki* and *Laternula elliptica*. Mar. Biol. 42: 171–175.

RICHARDS, F. J. 1959. A flexible growth model for empirical use. J. Exp. Bot. 10(29): 290–300.

RICKER, W. E. 1969. Effects of size selective mortality and sampling bias on estimates of growth, mortality, production, and yield. J. Fish. Res. Board Can. 26: 479–541.

1975. Computation and interpretation of biological statistics of fish populations. Bull. Fish. Res. Board Can. 191: 70–73.

ROFF, D. A. 1980. A motion for the retirement of the von Bertalanffy function. Can. J. Fish. Aquat. Sci. 37: 127–129.

SAINSBURY, K. J. 1980. Effect of individual variability on the von Bertalanffy growth equation. Can. J. Fish. Aquat. Sci. 37: 241–247.

SCHNUTE, J. 1981. A versatile growth model with statistically stable parameters. Can. J. Fish. Aquat. Sci. 38: 1128–1140.

SCHNUTE, J., AND D. FOURNIER. 1980. A new approach to length-frequency analysis: growth structure. Can. J. Fish. Aquat. Sci. 37: 1337–1351.

SHORE, J. E. 1984. Inversion as logical inference: theory and applications of maximum entropy and minimum cross-entropy. SIAM-AMS Proceedings. Vol. 14.

SHORE, J. E., AND R. W. JOHNSON. 1980. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. IEEE Trans. Inf. Theor. Vol. IT-26, No. 1: 26–37.

1981. Properties of cross-entropy minimization. IEEE Trans. Inf. Theor. Vol. IT-27, No. 4: 472–482.

SILLIMAN, R. P. 1967. Analog computer models of fish populations. Fish. Bull. (Wash., DC) 66: 31–46.

TANAKA, S. 1962. A method of analyzing a polymodal frequency distribution and its application to the length distribution of the Porgy, *Taius tumifrons*. J. Fish. Res. Board Can. 19: 1143–1159.

VON BERTALANFFY, L. 1938. A quantitative theory of organic growth. Human Biol. 10(2): 181–213.

WEINSTOCK, R. 1952. Calculus of variations with applications to physics and engineering. McGraw-Hill, New York, NY.

Appendix A: Convexity

It is easily shown that the class C of probability density functions consistent with a given set of moment constraints is a convex set. Suppose that there exist two densities, $p_1(\vec{x}) \in C$ and $p_2(\vec{x}) \in C$, such that

$$(40) \quad p_1(\vec{x}) \neq p_2(\vec{x}).$$

By definition of membership in C:

$$(41) \quad m_k = \int f_k(\vec{x})p_1(\vec{x})\,d\vec{x}$$

and

$$(42) \quad m_k = \int f_k(\vec{x})p_2(\vec{x})d\vec{x}.$$

To show convexity, it is sufficient to demonstrate that

$$(43) \quad p(\vec{x}) = a\cdot p_1(\vec{x}) + (1-a)\cdot p_2(\vec{x}) \in C \quad \forall a \ni 0 \leq a \leq 1.$$

But

$$\begin{aligned} \int f_k(\vec{x})p(\vec{x})d\vec{x} &= a\int f_k(\vec{x})p_1(\vec{x})d\vec{x} + \\ &\quad (1-a)\int f_k(\vec{x})p_2(\vec{x})d\vec{x} \\ &= a\cdot m_k + (1-a)\cdot m_k \\ &= m_k \quad 0 \leq k \leq m. \end{aligned}$$

Therefore

$$(45) \quad p(\vec{x}) = a\cdot p_1(\vec{x}) + (1-a)\cdot p_2(\vec{x}) \in C \quad \forall a \ni 0 \leq a \leq 1.$$

Appendix B: Derivation of the Form of the MCE Posterior Density

Cross-entropy minimization is a general procedure for approximating a true but unknown probability density function $q'(x)$ given a set of moments and a prior assessment $p(x)$. The approximating posterior $q(x)$ is chosen such that of all distributions consistent with the known moments, we select the one most similar to the prior model. If the assumptions are specific and the set of measured moments is not internally contradictory, the posterior $q(x)$ thus obtained is unique.

The logic of Kullback's Principle would have us find the function $q(\vec{x})$ that minimizes

$$(46) \quad H[q(\vec{x}),p(\vec{x})] = \int q(\vec{x})\log \frac{q(\vec{x})}{p(\vec{x})} \, d\vec{x}$$

while exactly satisfying the set of constraint equations

$$(47) \quad m_j = \int f_j(\vec{x})\cdot q(\vec{x})d\vec{x} \quad j=1,\ldots,M.$$

An Isoperimetric Calculus of Variations Problem

All of the MCE problems addressed in this paper have a common structure, which may be addressed with calculus of variations techniques. We are given a set of constraints $\{m_j\}$ which are known to satisfy the following definite integral equation:

$$(48) \quad m_j = \int g_j[q(\vec{x})]d\vec{x}$$

where $q(\vec{x})$ is a function to be determined. We are also given a measurement functional, again a definite integral:

$$(49) \quad f_{meas}[q(\vec{x})] = \int I_{meas}[q(\vec{x})d\vec{x}$$

and we want to obtain the $q(\vec{x})$ that extremizes $F_{meas}[q(\vec{x})]$.

In optimization theory, applications with this form are called "isoperimetric" calculus of variations problems. The name derives from the classical problem of finding the function with the maximum enclosed area given a fixed length boundary or perimeter.

To solve this problem, we use the Lagrange multiplier method (Weinstock 1952), forming the equation

$$(50) \quad Q[q(\vec{x});\vec{\beta}] = I_{meas}[q(\vec{x})] + \sum_{j=1}^{M} \beta_j g_j[q(\vec{x})].$$

We now minimize this equation with respect to $q(\vec{x})$. Taking the derivative and setting it equal to zero,

$$(51) \quad \frac{dQ}{dq} = 0,$$

we derive an expression that $q(\vec{x})$ can only satisfy at an extremum of $Q[q(\vec{x});\beta]$. We must then verify that the extremum is in fact a minimum by checking the higher order derivatives.

Application to the MCE Problem

For the MCE problem, we can identify

$$(52) \quad I_{meas} = q(\vec{x})\log \frac{q(\vec{x})}{p(\vec{x})}$$

and

$$(53) \quad g_j = f_j(\vec{x})\cdot q(\vec{x}).$$

Therefore

$$(54) \quad Q[q(\vec{x});\vec{\beta}] = q(\vec{x})\log \frac{q(\vec{x})}{p(\vec{x})} + \sum_{j=1}^{M} \beta_j f_j(\vec{x})q(\vec{x}) + \lambda_0 q(\vec{x}).$$

The last term reflects the fact that we usually have the constraint

$$(55) \quad \int q(\vec{x})d\vec{x} = 1$$

as well because $q(\vec{x})$ must be a valid normalized probability density function.

Taking the first derivative with respect to $q(\vec{x})$, we derive

$$(56) \quad \frac{dQ}{dq} = \log \frac{q(\vec{x})}{p(\vec{x})} + 1 + \lambda_0 + \sum_{j=1}^{M} \beta_j f_j(\vec{x})$$

and for the second derivative we obtain

$$(57) \quad \frac{d^2Q}{dq^2} = \frac{1}{q(\vec{x})}.$$

Setting the first derivative equal to zero:

$$(58) \quad \log q(\vec{x}) = \log p(\vec{x}) - 1 - \lambda_0 - \sum_{j=1}^{M} \beta_j f_j(\vec{x}).$$

Calling $\beta_0 = \lambda_0 + 1$ and $f_0(\vec{x}) = 1$ for all $\vec{x}$, we can write

$$(59) \quad \log q(\vec{x}) = \log p(\vec{x}) - \sum_{j=0}^{M} \beta_j f_j(\vec{x}).$$

Therefore

$$(60) \quad q(\vec{x}) = p(\vec{x})\cdot\exp\left\{-\sum_{j=0}^{M} \beta_j f_j(\vec{x})\right\}$$

which is the classical MCE posterior density (Johnson 1979; Shore and Johnson 1980).

Inserting this solution into the expression for the second derivative, we see that the positivity of the prior density $p(\vec{x})$ implies the positivity of $q(\vec{x})$, which guarantees that the second derivative is positive at the solution point. Hence, the solution is a minimum as desired.