PN-ABK 267
75232

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 64

MERGING AGGREGATE CATCH DATA with UNCERTAIN PRIOR KNOWLEDGE to APPROXIMATE AGE AND SIZE DISTRIBUTIONS and SELECTIVITY FUNCTIONS

by
Robert L. Burr
The University of Washington

December 1990

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 U.S. and developing countries' universities and institutions on fisheries stock assessment and management strategies.

This Working Paper has been produced by the collaborative efforts of The University of Washington and The University of Costa Rica - Centro de Investigacion en Ciencias del Mar y Limnologia (CIMAR). 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

T-SMC/18/6//23125

Merging Aggregate Catch Data with Uncertain Prior Knowledge to Approximate Age and Size Distributions and Selectivity Functions

Robert L. Burr

Reprinted from
IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS
Vol. 18, No. 6, 1988

Merging Aggregate Catch Data with Uncertain Prior Knowledge to Approximate Age and Size Distributions and Selectivity Functions

ROBERT L. BURR, MEMBER, IEEE

Abstract —The problem of determining fish age and size distributions and gear selectivity functions from aggregate data can be approached as an ill-posed inverse problem. Minimum cross-entropy inversion techniques allow the selection of a reasonable unique solution, directly incorporating background information of uncertain quality.

I. Introduction

In fisheries research and management it is often desirable to know the probability density function (PDF) of the age or length of a particular stock of fish. A detailed sampling program to determine the relative abundance of fish at different ages or lengths is usually prohibitively expensive. A method is proposed to merge limited and possibly indirect information from commercial and research fisheries catch reports with a reasonable hypothesized prior assessment to approximate the true PDF of fish length or age. This method is also employed to infer effective selectivity curves.

Fisheries managers are charged with the responsibility of optimizing the productivity of an aquatic natural resource while minimizing the risks for the participant fishermen and for the contextual ecosystem. Although fisheries management actions usually reflect political inputs more than the biological and bioeconomic status of the system, a class of rational models has been available for several decades to assist fisheries decisionmakers. These models quantify the resource at a very high level of generality, typically as a total biomass, and have been effective at describing the "tragedy of the commons" [1], the various kinds of overfishing that occur in common property fisheries. The inevitability of *economic overfishing* (increasing fishing effort until all of the participants are just breaking even) in open-access fisheries, and of *biological overfishing* (fishing the stock below biologically safe levels) when the ratio of price to cost of operation is high are prophetically indicated by these simple models [2].

A second generation of more elaborate models has emerged with the recognition of the limitations of the original simple assumptions. Several classes of overfishing that are dependent on the interaction between length or age distributions and gear

Manuscript received February 17, 1987; revised June 25, 1988.

R. L. Burr is with the Center for Quantitative Science in Fisheries, Forestry, and Wildlife, and the School of Nursing, University of Washington, Seattle, WA 98195.

IEEE Log Number 8823125.

0018-9472/88/1100-0996$01.00 ©1989 IEEE

selectivity phenomena have been noted. *Recruitment overfishing* reflects the fact that little fish come from big fish and that overharvesting the spawning stock depletes the next generation, regardless of the total stock biomass. *Growth overfishing* is an outcome of the fact that little fish become big fish, and that overharvesting young fish precludes their reaching full biomass potential (and spawning capacity). Thus the principal embellishment in recent fisheries models has been the introduction of age and size components.

A number of issues besides sexual maturity pertaining to age distributions are of interest to fisheries managers and biologists. Fish species have unique life histories, including age-dependent migrations, alterations in morphology, and changes in dietary components. From a population dynamics viewpoint, the age of a fish is the length of time it has been probabilistically available to be eaten or subject to other forces of mortality.

A comprehensive age distribution of a species may be expensive to acquire. In addition to the sampling program collection costs, the aging of each fish usually requires a microscopic examination by a highly trained specialist. Various parts of the fish, like the scales and bony structures, may reflect seasonal variations analogous to tree rings. In most species the best aging techniques use slices of the otoliths (ear bones) of the fish, prepared in a time-consuming procedure. Even these costly approaches to aging are not perfect, as the apparent rings are highly modified by variations in species life history, migrations, number of spawning cycles per year, and environmental disturbances like El Niño. Some fish species in tropical waters may experience few annual environmental variations affecting bone deposition, and most crustaceans shed all of their hard parts each time they molt, making direct determination of age essentially impossible. Small wonder that fisheries management entities put a considerable amount of effort into the construction and validation of age–length keys.

It certainly costs less to measure the length of an animal than to determine its age. Length distributions also have interest beyond their implications for aging. The size of a fish is probably the crucial determinant of its role in the ecosystem. Marine vertebrates can eat anything less than about one-tenth their own size, and are vulnerable to creatures roughly ten times larger than themselves. It is also the size and not directly the age of a fish that determines the vulnerability to geometrically selective entrapment devices like gillnets. Human fishing activity is the dominant source of mortality among mature fish in many exploited species and can cause dramatic alterations in the catch length profiles over time.

The preferred method for estimating age or length distributions of a stock of fish is to conduct a representative series of standardized research field sampling experiments, determine the age or length of each collected fish, and develop an empirical distribution using histogram, Parzen–Rosenblatt kernel, and more rarely, parametric probability density function estimating techniques. Only in the most important fisheries are sufficient resources available to support the boat time and personnel needed to carry out systematically such a research program. Particularly in developing countries, most fisheries managers must make the best inferences they can from current and historic aggregate commercial catch statistics. This highly summarized bioeconomic data about commercial boat landings are often collected at the primary transaction between the fisherman and the fish buyers, and rarely contain explicit length or age profile information. However, indirect information about age and length is coded into the commercial catch landing data because of the variations in the effective age/length selectivity of each boat's fishing effort.

It is well-understood that the configuration of the fishing gear and the behavior of the fishermen affect the size and age profile of the catch [3]. The size of the hook or the mesh of the net, the speed of the boat relative to the swimming speed of the fish, and the size and type of bait are all factors that almost mechanistically condition the catch length profile. Some representative gear-fishing behavior selectivity functions are shown in Fig. 1.

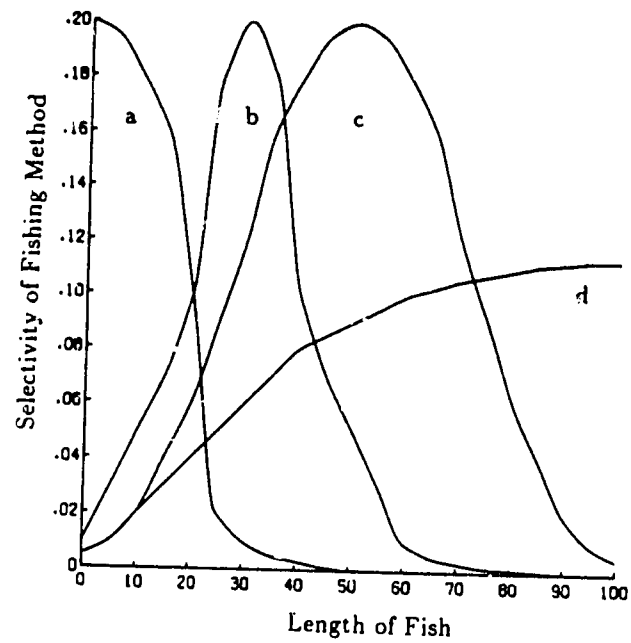

Fig. 1. Examples of selectivity functions for fishing gear. (a) Small trap. (b) Small mesh gillnet. (c) Large mesh gillnet. (d) Trawl.

Suppose aggregate catch reports are received from M fishermen, each using different gear or fishing methods with known selectivity functions $\mathrm{sel}_i(l)$. Typical information recorded about fish caught might include the total number of fish, the total weight of fish, and mean and mean square of length, or possibly a coarse histogram or count of fish in aggregate length categories. It will be assumed that each of these quantities can be at least asymptotically expressed as an expected value or moment of the true but unknown length PDF $q^{\mathrm{true}}(l)$ for the population under study.

For example, if the catch is large enough, we may be comfortable in describing the relationship between the number of fish caught and the unknown length PDF $q^{\mathrm{true}}(l)$ by means of the expected value integral constraint

$$N_i = C_i \int \mathrm{sel}_i(l)\, q^{\mathrm{true}}(l)\, dl \tag{1}$$

where C_i is a sampling scale factor incorporating fishing effort and absolute abundance. Similarly, the total landed weight W_i for fisherman i can be expressed as

$$W_i = C_i \int A l^3 \mathrm{sel}_i(l)\, q^{\mathrm{true}}(l)\, dl \tag{2}$$

where A is a constant for a particular species of fish.

The mean and mean square of length are asymptotically

$$M_i = C_i \int l \cdot \mathrm{sel}_i(l)\, q^{\mathrm{true}}(l)\, dl \tag{3}$$

and

$$MS_i = C_i \int l^2 \cdot \mathrm{sel}_i(l)\, q^{\mathrm{true}}(1)\, dl. \tag{4}$$

The number of fish $B_{i,k}$ expected in length bin or category k, composed of fish longer than l_k and shorter than or equal to l_{k+1}, is

$$B_{i,k} = C_i \int_{l_k}^{l_{k+1}} \mathrm{sel}_i(l)\, q^{\mathrm{true}}(l)\, dl. \tag{5}$$

To emphasize further the integral constraint nature of these

equations, let all the constant factors be included into the selectivity function $\mathrm{sel}_i(l)$, resulting in the integral expressions

$$N_i = \int \mathrm{sel}_i(l)\, q^{\mathrm{true}}(l)\, dl \tag{6}$$

$$W_i = \int A l^3 \cdot \mathrm{sel}_i(l)\, q^{\mathrm{true}}(l)\, dl \tag{7}$$

$$M_i = \int l \cdot \mathrm{sel}_i(l)\, q^{\mathrm{true}}(l)\, dl \tag{8}$$

$$MS_i = \int l^2 \cdot \mathrm{sel}_i(l)\, q^{\mathrm{true}}(l)\, dl \tag{9}$$

$$B_{i,k} = \int_{l_k}^{l_{k+1}} \mathrm{sel}_i(l)\, q^{\mathrm{true}}(l)\, dl. \tag{10}$$

For each report i, these values can be considered a well-defined summarization or mapping of the unknown continuous function $q^{\mathrm{true}}(l)$ through an integral kernel into a single scalar number or moment. The integral kernel, which is strongly influenced by the selectivity function $\mathrm{sel}_i(l)$, acts as a window through which the unknown PDF $q^{\mathrm{true}}(l)$ is indirectly perceived.

We would like to solve the inverse problem, that is, to deduce a reasonable approximation of the continuous function $q^{\mathrm{true}}(l)$ given a finite set of any of N_i, W_i, M_i, MS_i, or $B_{i,k}$. This is a well-known form of ill-posed inverse problem with an infinite convex set of solution PDF's $r(l)$, each consistent with the given finite set of integral constraints [4]. The aggregate data can also be regarded as constraints on the age distribution, a concept at least as interesting to population dynamicists as the length distribution. For simplicity, assume that age and length are continuous variables. The number of fish in length category k, composed of fish longer than l_k and shorter than or equal to l_{k+1} is

$$C_k = A \int_{l_k}^{l_{k+1}} \mathrm{sel}(l)\, q^{\mathrm{true}}(l)\, dl. \tag{11}$$

However, the length PDF can be related to the underlying age PDF $q^{\mathrm{true}}(a)$,

$$q^{\mathrm{true}}(l) = \int p(l|a)\, q^{\mathrm{true}}(a)\, da \tag{12}$$

where $p(l|a)$ is the conditional probability of length given age. Thus C_k can be written as

$$C_k = A \int_{l_k}^{l_{k+1}} \mathrm{sel}(l) \left(\int p(l|a)\, q^{\mathrm{true}}(a)\, da \right) dl. \tag{13}$$

Interchanging integrals, this becomes

$$C_k = \int \left(A \int_{l_k}^{l_{k+1}} \mathrm{sel}(l)\, p(l|a)\, dl \right) q^{\mathrm{true}}(a)\, da \tag{14}$$

or

$$C_k = \int A_k(a)\, q^{\mathrm{true}}(a)\, da \tag{15}$$

where

$$A_k(a) = A \int_{l_k}^{l_{k+1}} \mathrm{sel}(l)\, p(l|a)\, dl. \tag{16}$$

Equation (15) expresses the catch in length bin k as an integral constraint on the true but unknown age distribution $q^{\mathrm{true}}(a)$. It is assumed that the selectivity function $\mathrm{sel}(l)$ is known and also that the conditional probability $p(l|a)$ can be determined from theoretical considerations or empirical age–length keys, then (15) is an integral equality constraint bearing on the inverse problem of determining the age distribution. Of course, other length-related catch summary statistics can be similarly expressed as constraints on the age density function as well.

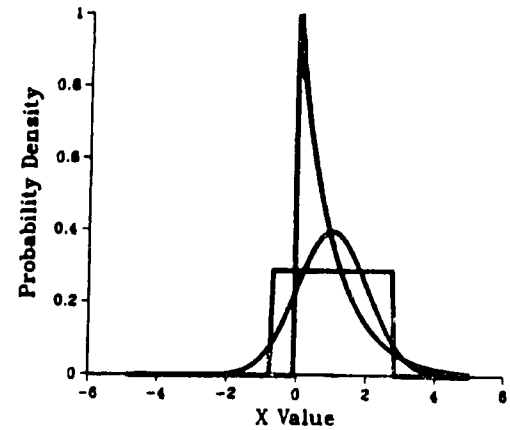

Fig. 2. Members of convex set composed of densities with same mean and variance.

II. Cross-Entropy Minimization

A given set of moments may not uniquely define a probability density function, although the density function completely determines the set of all possible moments. In general, as infinite class $\mathscr{C}$ of density functions is consistent with a finite set of moment constraints. For example, three densities with the same first two moments are shown in Fig. 2. In Appendix I, the class $\mathscr{C}$ defined by general expected value constraints is shown to be a convex set.

Some fisheries biometricians may not admit the need to pick a unique PDF from the class of density functions consistent with the known moments, feeling that it is more honest just to describe the class $\mathscr{C}$ of possible PDF's as best one can so as to point out the nonuniqueness of the solution to the stated inverse problem. However, a succinct description of the solution class is rarely possible or useful in practice. While all elements of the solution class are possible, some are extremely unlikely looking, and unrepresentative of the class. Picking a unique PDF is equivalent to specifying all possible moments. The various methods of PDF estimation differ in the criterion used to make an optimal choice of this specification.

Applied statisticians often express dissatisfaction with formal estimation procedures in statistics because background information either has to be ignored or rigidly adhered to. 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 [5] 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.

In fisheries management, a wealth of prior information of uncertain applicability about the length distribution of a particular fish stock is often available, coming perhaps from historical records, experience with similar species, or theoretical principles. For example, in many fish species it is possible to postulate a reasonable prior length PDF $p(l)$ based on the joint assumption of an exponential mortality with age and the von Bertalanffy

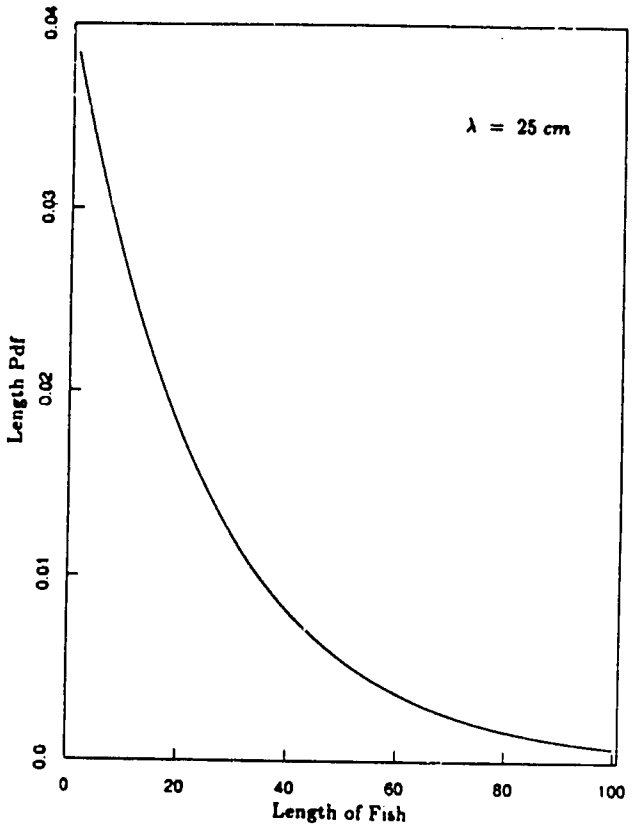

Fig. 3. Hypothetical true length PDF for Example 1.

age–length relation. Regardless of how the description of the background information is obtained, it would seem reasonable to use Kullback's principle to resolve the ambiguity of the solution set $\mathscr{C}$ by picking the element in $\mathscr{C}$ most similar to the assumed prior $p(l)$. No matter what prior density is specified, the selected element will still satisfy all of the given moments.

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

$$H[q(l), p(l)] = \int q(l) \log \frac{q(l)}{p(l)}\, dl. \tag{17}$$

Also known as the directed divergence 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.

Kullback's principle would have us find the posterior PDF $q(l)$ that minimizes $H[q(l), p(l)]$ while exactly satisfying the set of constraint equations

$$m_j = \int f_j(l) q(l)\, dl, \qquad j = 1, \cdots, M. \tag{18}$$

Note that the expected value constraints corresponding to our data are in this form.

The general problem of PDF approximation using the minimum cross-entropy (MCE) criterion has been studied, and a solution for the posterior PDF $q(l)$ for nonpathological integral kernels $f_j(l)$ is well-known.

$$q(l) = p(l) \exp\left\{ - \sum_{j=0}^{M} \beta_j f_j(l) \right\} \tag{19}$$

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

$$m_j^{\text{measured}} = \int f_j(l) p(l) \exp\left\{ - \sum_{k=0}^{M} \beta_k f_k(l) \right\} dl, \qquad j = 1, \cdots, M, \tag{20}$$

along with the normalizing constraint

$$1 = \int p(l) \exp\left\{ - \sum_{k=0}^{M} \beta_k f_k(l) \right\} dl. \tag{21}$$

The latter constraint comes about because the posterior $q(l)$ 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 II. For a more detailed consideration of the conditions under which this result exists and is unique, the careful reader is referred to [6] and [7]. A review of successful applications of MCE inversion techniques can be found in [8].

III. Examples

Several examples will now be developed to demonstrate how diverse aggregate data and detailed prior assumptions can be used to approximate the probability density functions of age and length.

A. Example 1

Suppose that the unknown true length PDF of a species of fish in a particular fishery is an exponential distribution,

$$q^{\text{true}}(l) = (1/\lambda) \exp\{-l/\lambda\} \tag{22}$$

with $\lambda = 25$ cm, as in Fig. 3.[1] Further, suppose that we have information about the number of fish caught by five sets of boats each using gillnets with different size mesh. The normalized selectivity functions of these hypothetical gear types are shown in Fig. 4 and model the rule of thumb that the selectivity of a gillnet is approximately Gaussian, with a standard deviation or spread parameter that is about 20 percent of the expected mean length.

By choosing several different prior distributions we can illustrate the sensitivity of the MCE posterior to the assumed background information. If we assume the prior is also exponential with parameters $\lambda_1 = 10$ cm, $\lambda_2 = 25$ cm, and $\lambda_3 = 90$ cm, respectively, as portrayed in Fig. 5(a), (c) and (e), then the resulting MCE posteriors are depicted in Fig. 5(b), (d), and (f). Clearly, with just five aggregate statistics we can produce a good approximation of the true density even when the prior is dramatically different on a biological scale. When the prior density corresponds to the true density (Fig. 5(c)), the MCE posterior perfectly matches the true density (Fig. 5(d)). In general, the discrepancies between the MCE posteriors and the true density are mainly located at short lengths, where only limited information is available from the extremely selective small mesh gillnets.

When the background information forces the imposition of a very inappropriate prior density, artificial features can be induced in the mixture of background and current information that is the MCE posterior density. For example if we assume the hump-shaped prior density in Fig. 5(g), the MCE posterior (Fig. 5(h)) is a good approximation for the larger animals, but evi-

[1] While this example is artificially simple, it is a first-order approximation of the length distributions of Corvina reina (*Cynoscion* albus) and Coliamarilla (*Cynoscion* stoltzmanni), two commercially important species the author is studying in a small-scale artisanal fishery in the Gulf of Nicoya on the Pacific coast of Costa Rica. These fish grow to approximately a meter in length in four years.

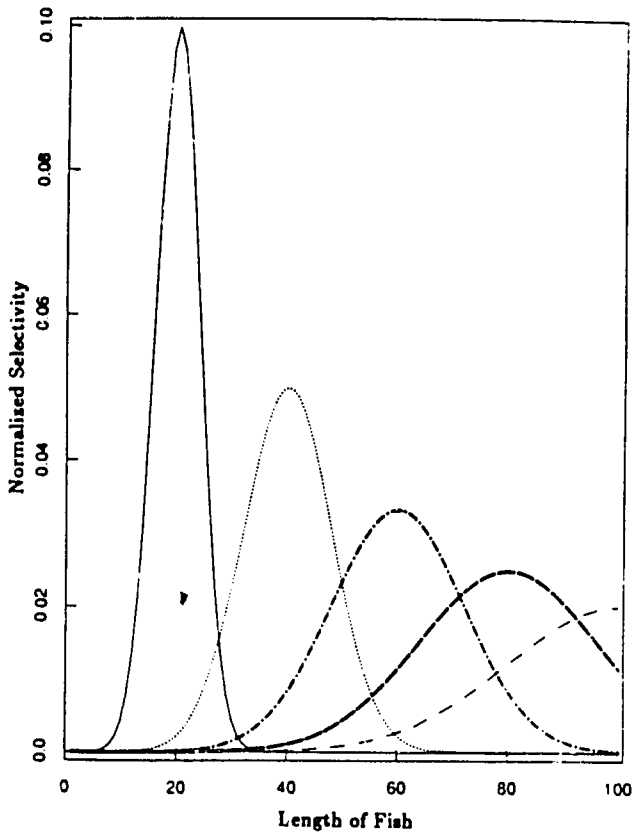

Fig. 4. Normalized selectivity functions for five hypothetical gear types assumed in Example 1.

dences features in the shorter lengths that are artifacts of the combination of the inappropriate prior and the specialization of the information from the current data about smaller fish. Superficial evaluation could easily result in a misleading biological interpretation and possibly undesirable management actions. Of course, acquiring more comprehensive data for the smaller fish would allow the MCE method to overcome the mis-specified prior density.

This simulated example underscores the critical role of prior knowledge in the specification of MCE limited information problems. The resulting posterior will be dominated by the prior density when the current information is scanty. The MCE posterior is the optimal unique solution of a given prior density, but will in general be different for distinct prior descriptions. It is somewhat disquieting to employ a method that can be manipulated by subjective factors. However, it should be noted that all competitive methods require making assumptions (often quite stringent and generally hidden) to force a unique solution, and that the MCE methodology merely requires the practitioner to be explicit. As scientific and editorial standards for reporting MCE applications develop, it is clear that precise statement of the prior density will be necessary, since the adequacy of the MCE posterior can not be interpreted in its absence.

B. Example 2

Let $p(l|a)$ be defined as Gaussian, with

$$\mu_l(a) = L_\infty(1 - e^{-ka}) \tag{23}$$

and

$$\sigma_l^2(a) = \mu_l(a) = L_\infty(1 - e^{-ka}) \tag{24}$$

where the von Bertalanffy parameters are arbitrarily chosen to be $L_\infty = 100$ cm and $k = 0.15$. Suppose further that the selectivity function for a particular gillnet has a Gaussian shape centered at $\mu = 50$ cm with a scale parameter $\sigma = 20$ cm as in Fig. 6.

If the hypothetical true age density $q^{\text{true}}(a)$ is as shown in Fig. 7, then the resulting catch length histogram, summarized into 5-cm bins by the equation

$$C_i = \int A \int_{i \cdot 5\,\text{cm}}^{(i+1)\cdot 5\,\text{cm}} \text{sel}(l)\, p(l|a)\, dl\, q^{\text{true}}(a)\, da \tag{25}$$

is shown in Fig. 8. It is apparent that the probabilistic diffusion of length with age and the summarization of the data into bins has significantly blurred the latent age density function.

With the MCE methodology we can attempt to recover the age density function from the hypothetical catch-at-length data summarized in Fig. 8. From our knowledge of the natural history and population dynamics of the particular fish species we might be led to postulate the prior density $p(a)$ given in Fig. 9. The form of the MCE posterior density is

$$q^{\text{MCE}}(a) = p(a) \cdot \exp\left\{ \lambda_0 + \sum_{i=1}^{K} \beta_i A_i(a) \right\} \tag{26}$$

where

$$A_i(a) = \int_{i \cdot 5\,\text{cm}}^{(i+1)\cdot 5\,\text{cm}} A \cdot \text{sel}(l)\, p(l|a)\, da. \tag{27}$$

The numerically computed posterior density $q^{\text{MCE}}(a)$ having the form given in (26) is shown in Fig. 10 and is a close approximation to the hypothetical true density $q^{\text{true}}(a)$.

IV. Inferring Selectivity: A Dual Problem

In the development so far it has been assumed that the selectivity functions $\text{sel}_i(l)$ are explicitly known and that the underlying length or age PDF needs to be determined. The same methodology can be used to consider the dual problem of approximating the normalized nonnegative length selectivity function $\text{sel}(l)$ for a particular fishing method from moments gathered with respect to various known length or age PDF's.

Suppose we had one or several sites where we independently knew the length or age PDF's. Then the jth information summary collected on ith site can be expressed as either

$$M_{ij} = \int M \cdot f_{ij}(l)\, q_i(l)\, \text{sel}^{\text{true}}(l)\, dl \tag{28}$$

if $q_i(l)$ is known, or

$$K_{ij} = \int K \cdot G_{ij}(l)\, \text{sel}^{\text{true}}(l)\, dl \tag{29}$$

if $q_i(a)$ is known, where

$$G_{ij}(l) = \int f_{ij}(l)\, p_i(l|a)\, q_i(a)\, da. \tag{30}$$

In both cases it is assumed $\text{sel}^{\text{true}}(l)$ is a normalized strictly positive continuous function.

Then the true selectivity function could be approximated by an MCE posterior function,

$$\text{sel}(l) = \text{psel}(l) \exp\left\{ \beta_0 + \sum_{i=1}^{I} \sum_{j=1}^{J} M\beta_{ij} f_{ij}(l)\, q_i(l) \right\} \tag{31}$$

for the case where $q_i(l)$ is known, or

$$\text{sel}(l) = \text{psel}(l) \exp\left\{ \beta_0 + \sum_{i=1}^{I} \sum_{j=1}^{J} K\beta_{ij} G_{ij}(l) \right\} \tag{32}$$

for the case where $q_i(a)$ is known, $G_{ij}(l)$ as above. If the prior

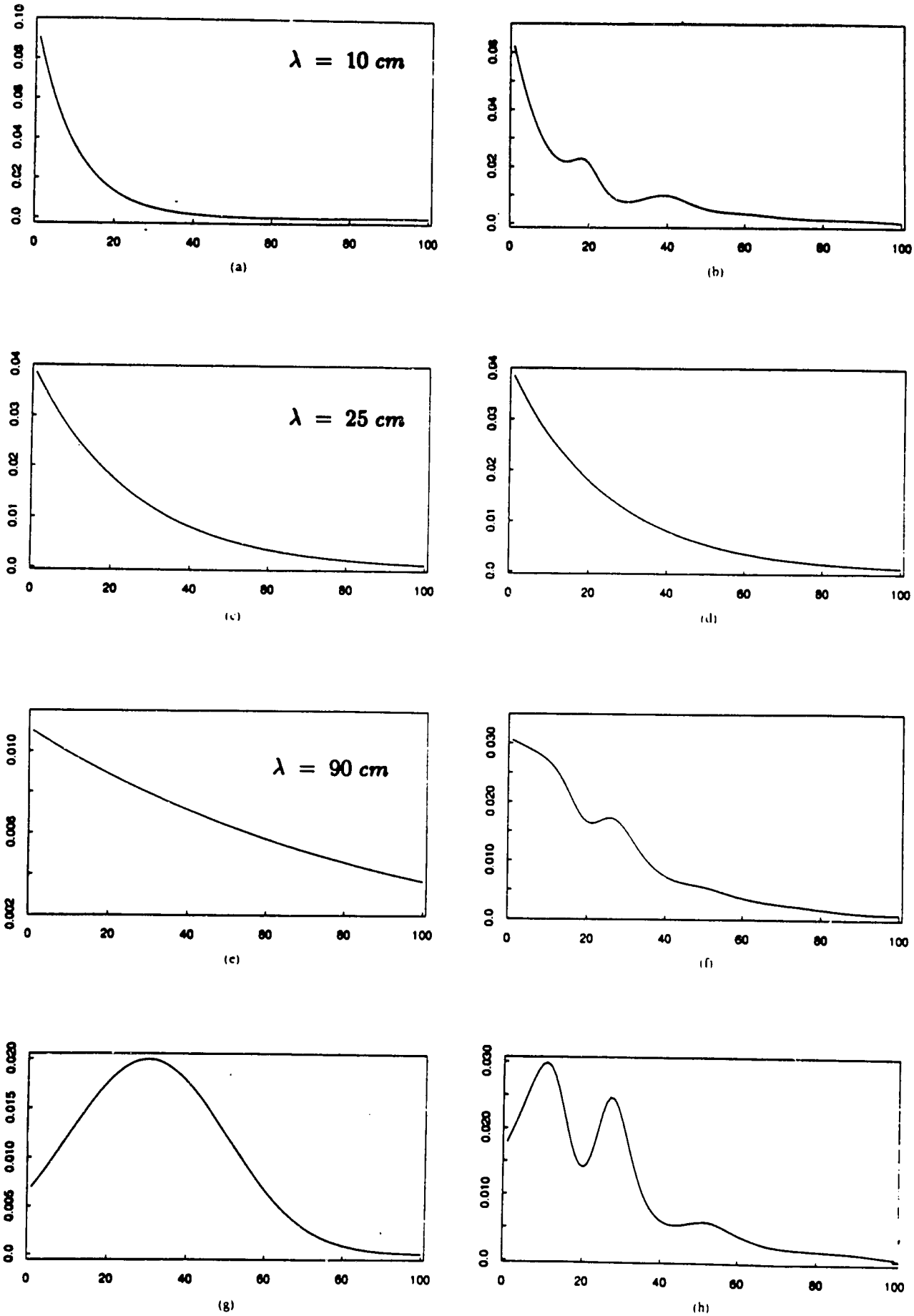

Fig. 5. (a), (c), (e), (g) Prior densities for Example 1. (b), (d), (f), (h) Corresponding posterior densities

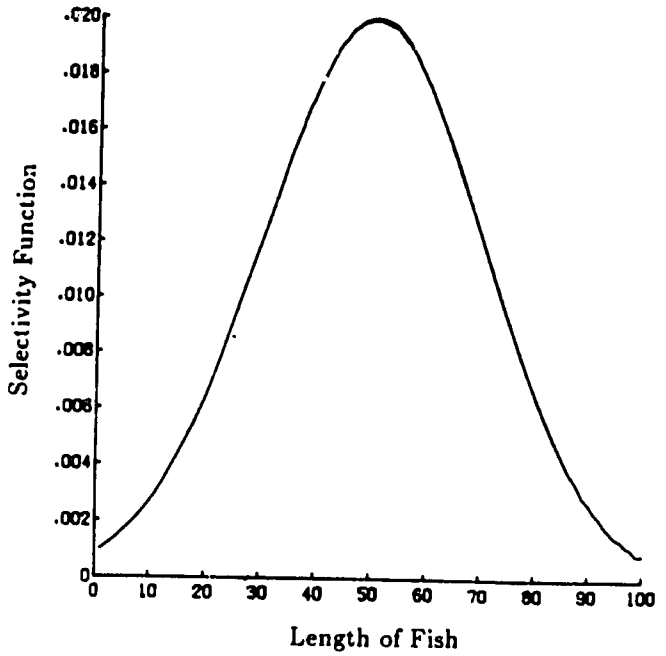

Fig. 6. Assumed selectivity function for Example 2.

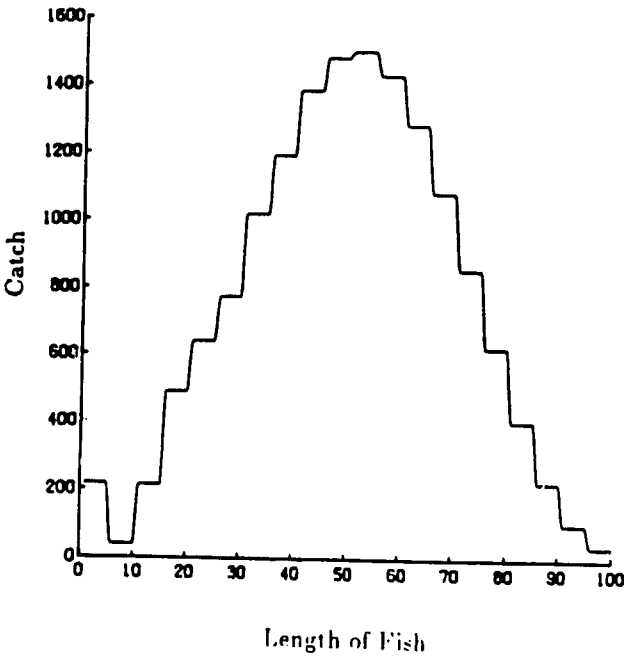

Fig. 8. Catch length histogram for Example 2.

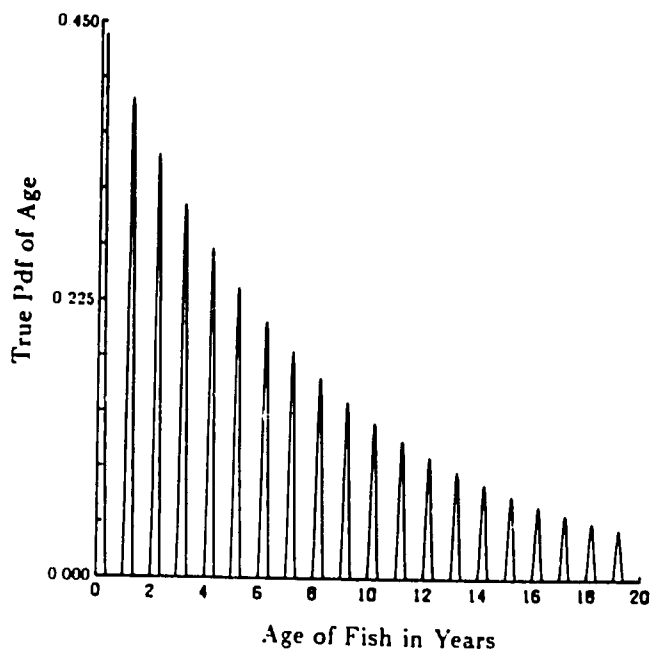

Fig. 7. Hypothetical true PDF for Example 2.

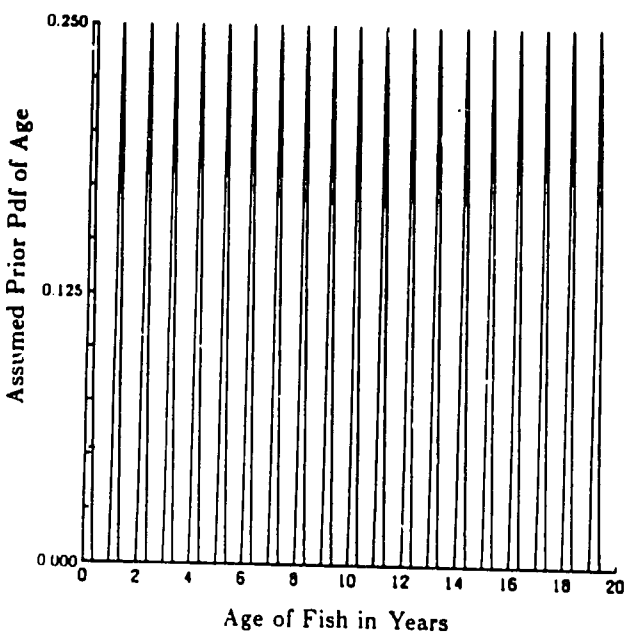

Fig. 9. Prior PDF for Example 2.

assessment psel(l) is reasonable, the approximating posterior function sel(l) will converge rapidly to $\text{sel}^{\text{true}}(l)$ as the number of distinct moments represented in the problem increases.

As an example of inferring the selectivity function given knowledge of the age density function, the conditional probability $p(l|a)$, and catch information, consider a slight variation of Example 2. If we assume that the age distribution is known and that we have a prior assessment of the normalized selectivity function, psel(l), then the posterior approximation sel(l) can be deconvolved from the catch-at-length histogram with the aid of (11). In particular, let psel(l) have a Gaussian shape with $\mu = 40$ cm and $\sigma = 10$ cm as represented in Fig. 11. The numerically determined MCE posterior function corresponding to the catch-at-length profile given in Fig. 8 is displayed in Fig. 12. The dashed line in the same figure depicts the hypothetical true selectivity function.

V. Remarks

The MCE method, as applied for example to length PDF approximation, capitalizes on the variations in the information about $q^{\text{true}}(l)$ implied by projection through the different integral kernel selectivity functions $\text{sel}_i(l)$. 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 (19) 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

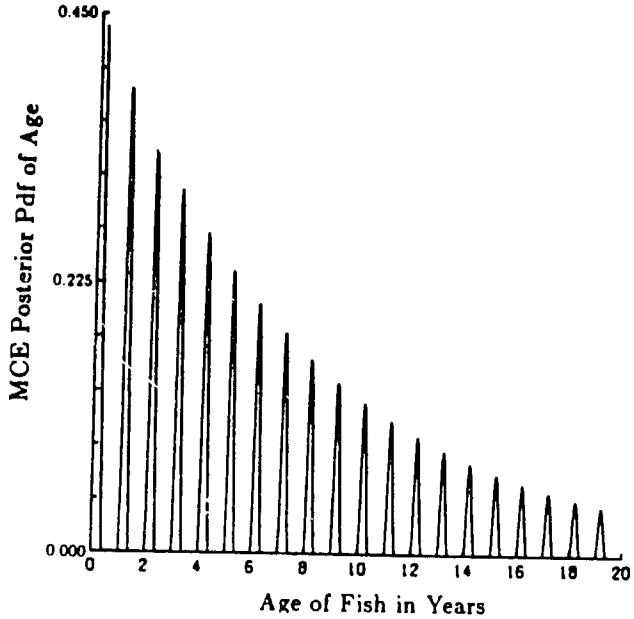

Fig. 10. Minimum cross-entropy posterior PDF for Example 2.

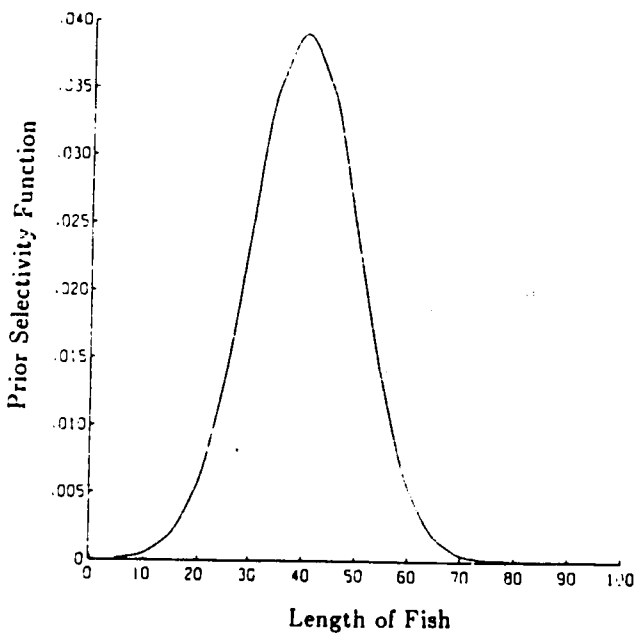

Fig. 11. Assumed prior selectivity function.

be generally noted that if the hypothesized prior $p(l)$ happens to be identical to the true PDf $q^{\text{true}}(l)$, then $\beta_i = 0$, $\forall\ i$ and $q(l) = p(l) = q^{\text{true}}(l)$.

As the number of reported moments m_i increase, the MCE procedure can overcome any misspecified prior PDF so long as $p(l) > 0, \forall\ l > 0$. In the information theory literature this appealing behavior of the MCE inverse is termed "washing out" old uncertain information with new facts. Whenever an inconsistency arises between the prior $p(l)$ and the actual data, the new data takes 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 deleted from the argument of the exponential function in (19).

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 $q(l)$ and manifests itself as unlikely detail.

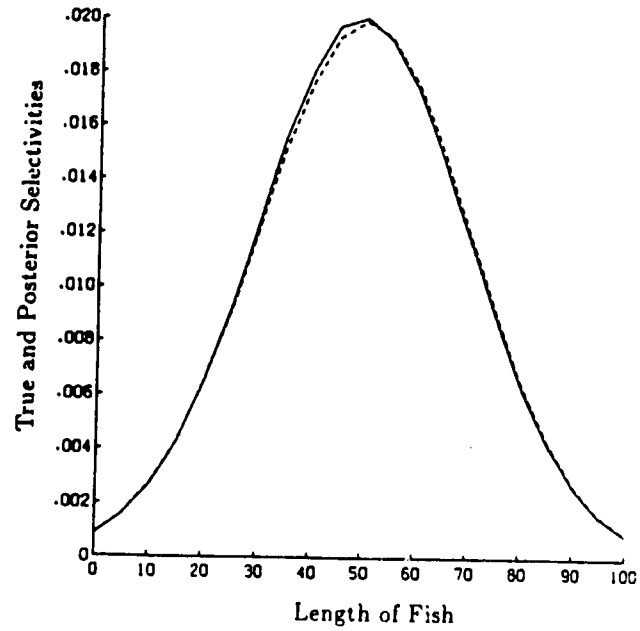

Fig. 12. True selectivity function (dashed line) MCE posterior selectivity function (solid line)

In the extreme, if two selectivity functions are equal, $\text{sel}_i(l) = \text{sel}_j(l)$ but the corresponding measured moments are not equal because of sampling variations, the set $\mathcal{C}$ of consistent solutions is empty and no MCE solution is defined. This situation can be avoided by allowing only one moment for each unique kernel function, averaging inconsistent evidence before the MCE inversion. This may not be possible with overlapping indirect evidence. More will be said about sampling variability in the next section.

VI. Estimation Issues

It has been stressed that the MCE inversion procedure is a method of moments, where summaries of sampled data are assumed rather 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 elegant PDF approximation scheme better suited for practical problems is an active research topic.

One approach to apprehend the uncertainty is to assume the Gaussian approximate sampling distribution for the estimated moments, then determine the multivariate PDF of the Lagrange parameters that is implied by the deterministic multivariate nonlinear function mapping the moments to the model parameters.

Another procedure would be to treat the form of the posterior PDF as a parametric model, then form a maximum likelihood estimate of the parameters $\{\beta_i\}$ based on the raw data. However, this paper has focused on situations where only aggregate data are available from which to work. In any event the optimality of the MCE posterior is only with respect to a particular limited set of available moments.

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

$$m_i = \int f_i(l)\, q^{\text{true}}(l)\, dl, \tag{33}$$

8

two inequality constraints might be written as

$$m_i + 2\frac{\sigma_i}{\sqrt{N_i}} > \int f_i(l)\, q^{\text{true}}(l)\, dl \tag{34}$$

$$m_i - 2\frac{\sigma_i}{\sqrt{N_i}} < \int f_i(l)\, q^{\text{true}}(l)\, dl. \tag{35}$$

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 PDF's satisfying the given inequality constraints. It is also possible in some MCE problems to incorporate equality constraints on the sampling distribution itself into the fundamental calculus of variations problem.

VII. Conclusion

Fisheries managers must increasingly make inferences about the age and length distributions of the aquatic natural resource stocks they oversee. Comprehensive field sampling surveys for each commercial species are very expensive and are generally out of the question for developing countries and small-scale artisanal fisheries. The present correspondence has attempted to develop an approach for inferring length and age information from highly summarized data about the operating commercial fishery.

The problem of determining fish age and size distributions and gear selectivity functions from aggregate data can be approached as an ill-posed inverse problem. Minimum cross-entropy inversion techniques allow the selection of a reasonable unique solution, directly incorporating background information.

This ability to include contextual, background, or subjective information allows the MCE approach to define good approximations from very indirect and limited evidence. The subjectivity of the prior can be a source of concern, since very different approximations can be obtained with different assumed prior densities. However, the MCE approach prioritizes current evidence with respect to the background knowledge, and a sufficiency of new information will overcome an unrepresentative assumed prior density.

It is difficult to assess the practicality of the MCE approach described in this paper and to compare it with potential competitive methods. It is not a statistical estimation procedure but an exact fit of a just-identified model crafted for asymptotic information-theoretic optimality, hence the usual questions of bias and efficiency are not applicable. It is a computationally expensive approach, with each problem statement essentially requiring a unique program to solve a specific large set of nonlinear equations.

The appropriateness of its use will have to be reconsidered for each application. It would seem to be indicated for problem environments where current evidence is limited but of good quality, where reasonable prior knowledge can be assumed and where the nontrivial cost of computation is less than the cost of collection of additional data. Many fisheries inference problems have these information-limited characteristics, and the MCE approach developed here may provide a useful solution framework for these difficult situations.

Appendix I
Convexity

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

$$p_1(\vec{x}) \neq p_2(\vec{x}). \tag{36}$$

By the definition of membership in $\mathscr{C}$,

$$m_k = \int f_k(\vec{x})\, p_1(\vec{x})\, d\vec{x} \tag{37}$$

and

$$m_k = \int f_k(\vec{x})\, p_2(\vec{x})\, d\vec{x}. \tag{38}$$

To show convexity, it is sufficient to demonstrate that

$$p(\vec{x}) = a \cdot p_1(\vec{x}) + (1-a) \cdot p_2(\vec{x}) \in \mathscr{C} \qquad \forall\, a \ni 0 \leqslant a \leqslant 1 \tag{39}$$

but

$$\begin{aligned}\int f_k(\vec{x})\, p(\vec{x})\, d\vec{x} &= a \int f_k(\vec{x})\, p_1(\vec{x})\, d\vec{x} + (1-a) \int f_k(\vec{x})\, p_2(\vec{x})\, d\vec{x} \\ &= a \cdot m_k + (1-a) \cdot m_k \\ &= m_k, \qquad 0 \leqslant k \leqslant m.\end{aligned} \tag{40}$$

Therefore,

$$p(\vec{x}) = a \cdot p_1(\vec{x}) + (1-a) \cdot p_2(\vec{x}) \in \mathscr{C} \qquad \forall\, a \ni 0 \leqslant a \leqslant 1. \tag{41}$$

Appendix II
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 $q^t(x)$ 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

$$H[q(\vec{x}), p(\vec{x})] = \int q(\vec{x}) \log \frac{q(\vec{x})}{p(\vec{x})}\, d\vec{x} \tag{42}$$

while exactly satisfying the set of constraint equations

$$m_j = \int f_j(\vec{x}) \cdot q(\vec{x})\, d\vec{x}, \qquad j = 1, \cdots, M. \tag{43}$$

A. An Isoperimetric Calculus of Variations Problem

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

$$m_j = \int g_j[q(\vec{x})]\, d\vec{x} \tag{44}$$

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

$$F_{\text{meas}}[q(\vec{x})] = \int I_{\text{meas}}[q(\vec{x})]\, d\vec{x} \tag{45}$$

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

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 [4], forming the equation

$$Q[q(\vec{x}); \vec{\beta}] = I_{\text{meas}}[q(\vec{x})] + \sum_{j=1}^{M} \beta_j g_j[q(\vec{x})]. \tag{46}$$

We now minimize this equation with respect to $q(\vec{x})$. Taking the

derivative and setting it equal to zero,

$$\frac{dQ}{dq} = 0, \tag{47}$$

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

B. Application to the MCE Problem

For the minimum cross-entropy problem we can identify

$$I_{\text{meas}} = q(\vec{x}) \log \frac{q(\vec{x})}{p(\vec{x})} \tag{48}$$

and

$$g_j = f_j(\vec{x}) \cdot q(\vec{x}). \tag{49}$$

Therefore,

$$Q\left[q(\vec{x}); \vec{\beta}\right] = q(\vec{x}) \log \frac{q(\vec{x})}{p(\vec{x})} + \sum_{j=1}^{M} \beta_j f_j(\vec{x}) q(\vec{x}) + \lambda_0 q(\vec{x}). \tag{50}$$

The last term reflects that we usually have the constraint

$$\int q(\vec{x})\, d\vec{x} = 1 \tag{51}$$

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

$$\frac{dQ}{dq} = \log \frac{q(\vec{x})}{p(\vec{x})} + 1 + \lambda_0 + \sum_{j=1}^{M} \beta_j f_j(\vec{x}) \tag{52}$$

and for the second derivative we obtain

$$\frac{d^2Q}{dq^2} = \frac{1}{q(\vec{x})}. \tag{53}$$

Setting the first derivative equal to zero

$$\log q(\vec{x}) = \log p(\vec{x}) - 1 - \lambda_0 - \sum_{j=1}^{M} \beta_j f_j(\vec{x}). \tag{54}$$

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

$$\log q(\vec{x}) = \log p(\vec{x}) - \sum_{j=0}^{M} \beta_j f_j(\vec{x}). \tag{55}$$

Therefore,

$$q(\vec{x}) = p(\vec{x}) \cdot \exp\left\{ - \sum_{j=0}^{M} \beta_j f_j(\vec{x}) \right\} \tag{56}$$

which is the classical minimum cross-entropy posterior density [9], [10].

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.

References

[1] G. Hardin, "The tragedy of the commons," *Science*, vol. 162, pp. 1243–1247, 1968.

[2] H. S. Gordon, "The economic theory of a common property resource: The fishery," *J. Polit. Econ.*, vol. 62, pp. 124–142, 1954.

[3] W. E. Ricker, "Computation and interpretation of biological statistics of fish populations," *Bull. Fish. Res. Board Can.*, vol. 191, pp. 70–73, 1975.

[4] R. Weinstock, *Calculus of Variations with Applications to Physics and Engineering*. New York: McGraw-Hill, 1952.

[5] S. Kullback, *Information Theory and Statistics*. New York: Dover, 1969.

[6] I. Csiszár, "I-divergence geometry of probability distributions and minimization problems," *Ann. Prob.*, vol. 3, no. 1, pp. 146–158, 1975.

[7] J. E. Shore, and R. W. Johnson, "Properties of cross-entropy minimization," *IEEE Trans. Inform. Theory*, vol. IT-27, no. 4, pp. 472–482, July 1981.

[8] J. E. Shore, "Inversion as logical inference: Theory and applications of maximum entropy and minimum cross-entropy," *SIAM-AMS Proc.*, vol. 14, 1984.

[9] R. W. Johnson, "Axiomatic characterization of the directed divergences and their linear combinations," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 6, pp. 709–716, Nov. 1979.

[10] J. E. Shore and R. W. Johnson, "Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 1, pp. 26–37, Jan. 1980.

0018-9472/88/1100-1005$01.00 ©1989 IEEE

10