
A Review of the Net Selectivity Problem and a Model for Apportioning Species Based on Size-Selective Sampling

Pham X. Quang and Harold J. Geiger

Reprinted from the Alaska Fishery Research Bulletin
Vol. 9 No. 1, Summer 2002

The Alaska Fisheries Research Bulletin can be found on the World Wide Web at URL:
<http://www.state.ak.us/adfg/geninfo/pubs/afrb/afrbhome.htm>

A Review of the Net Selectivity Problem and a Model for Apportioning Species Based on Size-Selective Sampling

Pham X. Quang and Harold J. Geiger

ABSTRACT: The theory of inference of a size distribution for fish captured with size-biased fishing gears is called the "net selectivity problem." This family of problems arises because of questions about the underlying distribution of the sampled fish, questions about the nature of the size bias, or perhaps questions about the mixtures of species of different sizes. We discuss these problems in the context of size-biased gillnets, although generalization to other kinds of fishing gear is straightforward. We estimate the length distribution of the captured fish, the size distribution of the fish population, and the parameters of the selectivity curves parametrically, using maximum likelihood. To avoid overparametrization we make strong assumptions about the relationship among the selectivity curves, so that fishing power cannot be estimated in this setup. This assumption is called "geometric similarity." The normal-normal and gamma-gamma combination models are considered, so called because of the shape of the fish-length distribution and the shape of the selectivity curves. The method is demonstrated with a worked simulation example and the American plaice *Hippoglossoides platessoides* data of Holst and Moth-Poulsen.

INTRODUCTION

The net-selectivity problem is a family of problems requiring inference from the number of fish captured in each of a series of selective or size-biased fishing gears. The probability of capture, modeled as a function of the attributes of the gear and the fish encountering the gear, is called the selectivity function. The framework of the problem always involves one or more populations of fish or other animals together with a family of selectivity functions; the nature of the inference changes depending on whether the question is about the gear or about the sampled fish populations. Usually the problem is discussed in terms of gillnet size-selectivity curves and body-size distribution of the fish population; Millar (1992, 1995) describes the similarities of the problems between gillnets, hooks, trawls, and other gear. Millar and Fryer (1999) distinguish among the population-selection functions, the available-selection functions, and the contact-selection functions. Our goal is to introduce a model that can be used to infer the proportional mix of different species, of differing sizes, captured in a series of gillnets. This member of the net-selectivity family of problems is called the species apportionment problem.

The normal, gamma, and skewed normal distributions have been used for a number of years as models for selectivity curves (Regier and Robson 1966). Henderson and Wong (1991) developed an encounter model and modeled walleye *Stizostedion vitreum* selectivity in gillnets with a gamma distribution. Millar and Holst (1997) used normal selectivity curves inside a larger log-linear modeling context. Catch numbers were assumed to follow Poisson distributions, with the Poisson mean modeled as a function of fishing power, size, abundance, and selectivity. Helsen et al. (1997) provided a nonlinear response surface regression.

Regier and Robson (1966) define 2 types of size-selectivity curves. Type-A curves give the probability of capture in a given gear as a function of fish size; type-B curves give the probability of capturing a fish of a given size as a function of the gear size or type. Regier and Robson also divide the problem into 2 classes of methods. Direct methods are used when the size-frequency of the study population is known or can be measured directly. For example, fisheries managers often use catch-per-unit of effort measures to keep fishing mortality in line with fish abundance (Quinn and Deriso 1999). A direct study might involve measuring the difference in size distribution from a catch in a se-

Authors: PHAM X. QUANG did this work while he was with the Department of Mathematical Sciences, University of Alaska Fairbanks. His current address is 1006 Cinnabar Way, Vacaville, CA 95687. E-mail: quang@stat.berkeley.edu. HAROLD J. GEIGER works for the Alaska Department of Fish and Game, Commercial Fisheries Division, P.O. Box 240020, Douglas, AK 99824. E-mail: hal_geiger@fishgame.state.ak.us

Sponsorship: Pham Quang was supported by a grant from the Alaska Department of Fish and Game, Division of Commercial Fisheries.

lective gear compared with a control gear or a non-selective fishing gear. A new gear's relative *fishing power*, or *catch-per-unit-of-effort* might be compared with a conventional gear. A direct study might also measure the effects of size selection on a population of fish. Indirect methods are used when the size frequency distribution of the study population must be inferred from the size-biased data collected with selective gear. In indirect studies, the object of interest is the size distribution of the target population of fish, which might vary from year to year due to the environment. If mean length and variance are to be estimated from a fish population migrating through a small area in a particular year, a series of nets may be used, each net with a different selection bias for a particular girth of fish. Because fish length is easy to measure, length measurements are usually available to the analyst. And because length and girth are strongly correlated, the nets can be thought of as length selective even though girth is more strongly related to the mechanical cause of capture.

A related problem is the species-apportionment problem. Here the analyst tries to determine the proportion of different species of fish that swim up a stretch of river. For example, 3 species of fish swim up a river in proportions p_1 , p_2 , and p_3 . The problem is to estimate p_i using length measurements of fish caught in gillnets of varying mesh sizes. Because the different species of fish have different size distributions, the species apportionment problem is solved by estimating the number of fish of each species that "would have been caught" if the fishing gear were equally effective at catching each species, and then using these numbers to estimate the apportionment proportions.

Selectivity models often require too many parameters because each gear variant has its own set of parameters. In these models, maximization of the likelihood function becomes difficult owing to numerical instability and overparametrization (Millar 1995; Millar and Fryer 1999). Several selectivity models can appear to provide similar fits to the data. Commonly, Baranov's assumption of geometric similarity (Baranov 1948) is introduced to reduce the number of parameters: the mode and the spread of a gillnet size-selection curve are assumed to increase proportionately to the size of the mesh.

Many authors have offered methods to estimate parameters of the selection. Only Hovgård (1996a) provides a method to also estimate the fishing power of the nets. To our knowledge, this methodology has not been applied to estimating the proportions of different species of fish in the population.

We solve the indirect problem by deriving the distribution of the lengths of fish captured by a given net

as the distribution of the population weighted by the selectivity of the net. We have not found such a distribution in the fisheries literature, although a similar distribution is well-known in sampling theory. Our approach is parametric, using the maximum likelihood method to estimate population and selectivity parameters. However, our formulation does not allow estimation of fishing power. We will show how to estimate the number of fish encountering all of the nets, and propose a method to estimate the relative abundance of different species of fish. Programs written in S-Plus 2000 (S-Plus 2000) in support of this paper are available by e-mail from either author.

Investigating gillnet selectivity

An experimental study consists of the researcher fishing with several gillnets of various mesh sizes, and recording the lengths of the captured fish. The data is reduced to the frequencies of catches in each length class. More precisely, assume J nets of respective mesh sizes m_1, \dots, m_J are fished. The fish-length range is partitioned into I intervals of equal width Δ ; l_i is the midpoint of the i th interval, and the number of catches from the j th net in the i th interval is n_{ij} . The data are the matrix of frequencies n_{ij} .

The (length-based) selectivity of the j th net is the function $r_j(x)$, defined as the probability that a fish is captured by the j th net, given that it is of length x and that it contacted the net. Our definition corresponds to the contact-selection curve or retention curve of Millar and Fryer (1999). Millar and Fryer also introduce the population-selection curve (the probability any fish in the population of length x is captured) and the available-selection curve (the probability that any fish of length x that is available to the gear is captured). We do not consider these curves in our formulation. Thus, the population under consideration for our purposes consists of all of the fish of a given species that encounter the nets.

METHODS

Consider the distribution of the lengths of captured fish. Catch frequencies depend on both the selectivity of the nets and the length distribution of the fish population (e.g., Hovgård 1996a; Hovgård et al. 1999; Hamley 1975, Figure 1, that Hamley attributes to Baranov). The classical estimation method of McCombie and Fry (1960) and Kitahara (1970), with selectivity estimated by graphically adjusting catches from individual length groups to fit a common master curve, is an implicit

acknowledgement that size distribution of the captured fish depends on the selectivity of the net (Hovgård 1996a; Hovgård et al. 1999).

The probability density function (PDF) $f_j(x)$ of lengths of fish captured by the j th net depends on the selectivity $r_j(x)$ and the PDF $\mathbf{j}(x)$ of population lengths as (proof in Appendix A)

$$f_j(x) = \frac{r_j(x)\mathbf{j}(x)}{p_j}, \quad (1)$$

where

$$p_j = \int_0^\infty r_j(t)\mathbf{j}(t) dt. \quad (2)$$

The quantity p_j can be interpreted as the average selectivity of the j th net for the given fish population. An equivalent form of Eq. (1) has been used in line transect sampling (e.g., Quang 1991). Because catch lengths are observed, the expression of $f_j(x)$ allows us to simultaneously estimate population and selectivity parameters by the maximum likelihood method. All notations are listed and redefined in Appendix B.

We assume parameterized functional forms $\mathbf{j}(x|\mathbf{q}_0)$ and $r_j(x|\mathbf{q}_j)$ for fish length and net selectivity, respectively. Here \mathbf{q}_0 and \mathbf{q}_j denote vectors of parameters. Then $f_j(x)$ has the functional form $f_j(x|\mathbf{q})$, where \mathbf{q} denotes the vector of all parameters involved in the problem. Let

$$q_{ij}(\mathbf{q}) = \int_{l_i - \frac{1}{2}\Delta}^{l_i + \frac{1}{2}\Delta} f_j(x|\mathbf{q}) dx,$$

then (n_{1j}, \dots, n_{lj}) are multinomially distributed with total count $n_j = \sum_{i=1}^l n_{ij}$, cell probabilities are $q_{1j}(\mathbf{q}), \dots, q_{lj}(\mathbf{q})$, and the log-likelihood is

$$L_1(\mathbf{q}) = \sum_{j=1}^J \sum_{i=1}^l n_{ij} \log[q_{ij}(\mathbf{q})]. \quad (3)$$

If Δ is small, then $q_{ij}(\mathbf{q}) \approx \Delta f_j(l_i|\mathbf{q})$, and the log-likelihood is approximately

$$L_2(\mathbf{q}) = \sum_{j=1}^J \sum_{i=1}^l n_{ij} \log[f_j(l_i|\mathbf{q})], \quad (4)$$

maximizing the log-likelihood yields estimates of all parameters.

The species apportionment problem

Knowing the relative abundance of different species of migrating fish is important for fisheries management. For example, sonar is used to estimate the total number of migrating salmon moving up a river over a unit of time. Because the fishery managers are trying to manage each individual breeding stock separately, they must have estimates of the species proportions to apply to the total passage estimate for the sonar to be useful. These relative abundances can be estimated as ratios of the number of individuals belonging to the different species and contacting the nets. The number of individuals of a given species contacting the nets is $N = \sum_{j=1}^J N_j$ and N_j , the number of individuals contacting the j th net, can be estimated as follows.

Let Y_{jk} be the length of the k th individual contacting the j th net, and let \mathbf{d}_{jk} be its catch indicator:

$$\mathbf{d}_{jk} = \begin{cases} 1 & \text{if the } k\text{th fish is caught by the } j\text{th net} \\ 0 & \text{if not} \end{cases}$$

Then \mathbf{d}_{jk} has the Bernoulli distribution with mean $p_j(\mathbf{q})$. Indeed, letting $E(\cdot)$ denote mathematical expectation,

$$\begin{aligned} E(\mathbf{d}_{jk}) &= P(\mathbf{d}_{jk} = 1) = P(\mathbf{d}_{jk} = 1, 0 \leq Y_{jk} < \infty) \\ &= \int_0^\infty P(\mathbf{d}_{jk} = 1 | Y_{jk} = y) \mathbf{j}(y|\mathbf{q}) dy \\ &= \int_0^\infty r_j(y|\mathbf{q}) \mathbf{j}(y|\mathbf{q}) dy \\ &= p_j(\mathbf{q}). \end{aligned}$$

If net j captures n_j individuals then $n_j = \sum_{k=1}^{N_j} \mathbf{d}_{jk}$, and

$$E(n_j) = \sum_{k=1}^{N_j} E(\mathbf{d}_{jk}) = N_j p_j(\mathbf{q}).$$

This result suggests estimating N_j by

$$\hat{N}_j = \frac{n_j}{p_j(\hat{\mathbf{q}})} \quad (5)$$

once an estimated \mathbf{q} becomes available.

Parametric models of net selectivity and fish length distribution

The length distribution of the fish population is sometimes ignored in the literature on net selectivity, possibly because it is not the focus in net selectivity studies. However, a variety of parametric models of net selec-

tivity have been considered: the normal model (Baranov 1948; Holt 1957), the skewed normal model (McCombie and Fry 1960; Regier and Robson 1966), the binormal (Sechin 1969; Kawamura 1972; Hovgård 1996a), the Pearson Type I curve (Hamley and Regier 1973), and the lognormal, gamma, and inverse gaussian models (Millar and Fryer 1999).

We now analyze the combination normal population length model and normal selectivity model (the normal-normal model), and also the combination gamma-gamma model. The advantage of the normal-normal model is that most calculations are explicit, but the gamma-gamma model seems to be more realistic. Other model combinations can be worked out similarly. Of course, all combination models risk overparametrization, bad fit, or nonconvergence of the optimizer.

Normal-normal model

The normal model of length distribution is

$$j(x | \mathbf{m}, \mathbf{s}) = \frac{1}{\sqrt{2\pi\mathbf{s}}} \exp\left(-\frac{(x - \mathbf{m})^2}{2\mathbf{s}^2}\right), \quad (6)$$

and the normal model of net selectivity is

$$r_j(x | h_j, \mathbf{m}_j, \mathbf{s}_j) = h_j \exp\left(-\frac{(x - \mathbf{m}_j)^2}{2\mathbf{s}_j^2}\right). \quad (7)$$

The parameter h_j in the selectivity model equals the maximum height of the curve $r_j(x)$. Hovgård (1996b) calls h_j the fishing power of the j th net and points out that failure to incorporate fishing power in selectivity models may lead to various misinterpretations of fishing results.

Unless a non-selective net is also fished, h_j is confounded with fishing effort and fish abundance (Millar and Fryer 1999). In this application we assume $h_j=1$. Manipulating Eq. (1), (6), and (7) gives (see proof in Appendix A)

$$f_j(x | \mathbf{q}) = \sqrt{\frac{a_j}{\mathbf{p}}} \exp\left[-a_j \left(x - \frac{b_j}{a_j}\right)^2\right], \quad (8)$$

with

$$a_j = \frac{1}{2\mathbf{s}^2} + \frac{1}{2\mathbf{s}_j^2}; \quad b_j = \frac{\mathbf{m}}{2\mathbf{s}^2} + \frac{\mathbf{m}_j}{2\mathbf{s}_j^2} \quad (9)$$

and

$$p_j = \frac{h_j}{\sqrt{2a_j\mathbf{s}}} \exp\left(\frac{b_j^2 - a_j c_j}{a_j}\right) \quad (10)$$

with $c_j = \mathbf{m}^2/(2\mathbf{s}^2) + \mathbf{m}_j^2/(2\mathbf{s}_j^2)$. Note that h_j canceled out of Eq. (8), hence that is why h_j cannot be estimated by maximum likelihood. The remaining $2J + 2$ parameters in Eq. (8) are $\mu, \mathbf{s}, \mu_1, \dots, \mu_J, \mathbf{s}_1, \dots, \mathbf{s}_J$. In Eq. (8) the lengths X_{j1}, \dots, X_{jm} of fish caught by the j th net are normally distributed with mean $\mathbf{x}_j = b_j/a_j$ and variance $\mathbf{t}_j^2 = 1/(2a_j)$. Maximum likelihood estimates of \mathbf{x}_j and \mathbf{t}_j^2 are:

$$\hat{\mathbf{x}}_j = \frac{\hat{b}_j}{\hat{a}_j} = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ji},$$

$$\hat{\mathbf{t}}_j^2 = \frac{1}{2\hat{a}_j} = \frac{1}{n_j} \sum_{i=1}^{n_j} (X_{ji} - \hat{\mathbf{x}}_j)^2. \quad (11)$$

Next, we would solve for the $2J$ quantities \hat{a}_j and \hat{b}_j from Eq. (11), and then solve for $\hat{\mathbf{m}}, \hat{\mathbf{s}}, \hat{\mathbf{m}}_1, \hat{\mathbf{s}}_1, \dots, \hat{\mathbf{m}}_J, \hat{\mathbf{s}}_J$ from Eq. (9). Because there are $2J+2$ parameters but only $2J$ equations, we cannot obtain unique estimates of the parameters. Also, \hat{N}_j in Eq. (5) cannot be unequivocally determined because of the presence of h_j in Eq. (10). We deal with overparameterization by assuming geometric similarity (Baranov 1948). That is, $h_1 = \dots = h_J = h$, $\mathbf{m}_j = k_1 \mathbf{m}_j$, and $\mathbf{s}_j = k_2 \mathbf{m}_j$, and the number of parameters is reduced to 5: $\mathbf{m}, \mathbf{s}, h, k_1$, and k_2 . Discussions regarding the impact of that assumption can be found in Hamley (1975), and more recently, in Millar and Fryer (1999). The log-likelihood becomes

$$L_2(\mathbf{m}, \mathbf{s}, k_1, k_2) = \sum_{i=1}^I \sum_{j=1}^J n_{ij} \left[\frac{1}{2} \log(a_j) - a_j \left(l_i - \frac{b_j}{a_j} \right)^2 \right], \quad (12)$$

where

$$a_j = \frac{1}{2\mathbf{s}^2} + \frac{1}{2\mathbf{m}_j^2 k_2^2}; \quad b_j = \frac{\mathbf{m}}{2\mathbf{s}^2} + \frac{k_1}{2\mathbf{m}_j k_2^2}.$$

Maximizing Eq. (12) with a numerical optimizer gives estimates of μ, \mathbf{s}, k_1 , and k_2 . Again, h cannot be determined as it is confounded with fishing effort and fish abundance. We have simulated data that provide a good fit to the normal-normal model, although we have not found real-world data that this model fits.

Gamma-gamma model

The gamma probability density function is not as well known or as straightforward to work with as the normal distribution. For fish length, X , this distribution takes the form

$$j(x|\mathbf{a},\mathbf{b}) = \frac{1}{\Gamma(\mathbf{a})\mathbf{b}^{\mathbf{a}}} \exp\left(-\frac{x}{\mathbf{b}}\right) x^{\mathbf{a}-1}, \quad (13)$$

where \mathbf{a} is the shape parameter and \mathbf{b} is the scale parameter. The mean of this distribution is given by $E(X)=\mathbf{a}\mathbf{b}$, and the variance is $V(X)=\mathbf{a}\mathbf{b}^2$.

Note that unlike the normal distribution, the mean and variance of the gamma distribution are functionally related. Because the gamma distribution is non-negative and is right-skewed, it may represent fish length distribution better than the normal distribution. Because large fish may be caught in a narrow-mesh net by tangling of their maxillaries and opercula, rather than being wedged by their gills, the thick right tail of the gamma selectivity curve may reflect this phenomenon better than the normal selectivity curve. Even though the gamma distribution is not as well known as the normal distribution, it has been used in fisheries and wildlife applications. For example, it is used to model detection probability in line-transect sampling (Quang and

Becker 1997). In Figure 1 we show 3 gamma densities with the same mean of 4, and variances of 1, 4 and 8.

The gamma model of net selectivity is

$$r_j(x|h_j,\mathbf{a},\mathbf{b}_j) = h_j \left(\frac{x}{(\mathbf{a}-1)\mathbf{b}_j} \right)^{\mathbf{a}-1} \exp\left(\mathbf{a}-1-\frac{x}{\mathbf{b}_j}\right). \quad (14)$$

Again h_j is the maximum height of the curve $r_j(x)$. As in the normal-normal, the gamma-gamma model is overparameterized and h_j cannot be determined by maximum likelihood. The number of parameters is reduced by assigning the same shape parameter \mathbf{a} to length distribution and selectivity, and by assuming geometric similarity.

This leaves $h_1 = \dots = h_j = h$, and $\mathbf{b}_j = km_j$, so that the parameters are h , \mathbf{a} , and k . The length distribution of fish caught by the j th net is now

$$f_j(y|\mathbf{q}) = \frac{\mathbf{l}_j^{2\mathbf{a}-1} y^{2\mathbf{a}-2} \exp(-\mathbf{l}_j y)}{\Gamma(2\mathbf{a}-1)}$$

with $\mathbf{l}_j = 1/\mathbf{b} + 1/\mathbf{b}_j$. The log-likelihood is

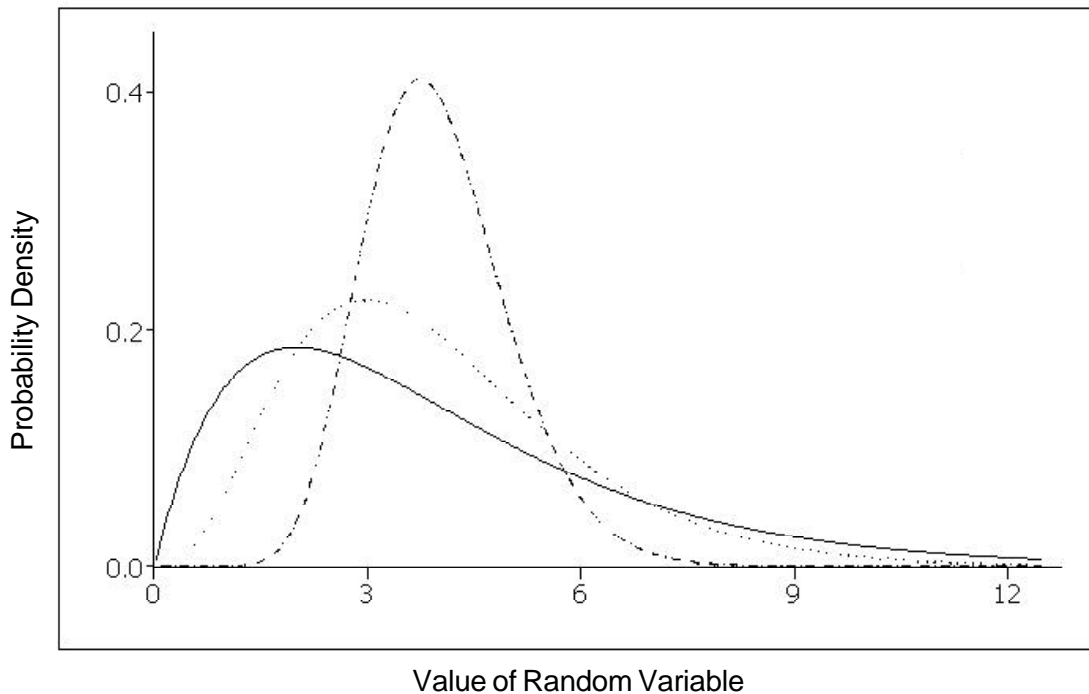


Figure 1. Three gamma distributions with the same mean and differing variables.

$$L_2(\mathbf{a}, k) = -n \log[\Gamma(2\mathbf{a} - 1)] + (2\mathbf{a} - 1) \sum_{i=1}^I \sum_{j=1}^J n_{ij} \log(I_j) + (2\mathbf{a} - 2) \sum_{i=1}^I \sum_{j=1}^J n_{ij} \log(l_i) - \sum_{i=1}^I \sum_{j=1}^J I_j n_{ij} l_i. \quad (15)$$

Again we need a numerical optimizer to maximize Eq. (15).

Test of goodness of fit

With net selectivity models, simple tests for goodness of fit may be misleading because of low power and poor approximation of test statistics to large-sample distributions. The usual test of goodness of fit, which is based on Pearson's X^2 statistic, provides a poor approximation for sparse tables. All numerical examples presented in the next section are in this category because too many of the n_{ij} s are less than 5. But the X^2 statistic itself is also approximately normally distributed, and this approximation may be used instead of the usual chi-square test, even when individual counts drop below 5.

Let the X^2 statistic be calculated as

$$X^2 = \sum_{j=1}^J \sum_{i=1}^I \frac{[n_{ij} - n_j q_{ij}(\hat{\mathbf{q}})]^2}{n_j q_{ij}(\hat{\mathbf{q}})} \quad (16)$$

Following Koehler and Larntz (1980) and Agresti and Yang (1987), let

$$t^2(\hat{\mathbf{q}}) = 2IJ + \sum_{j=1}^J \sum_{i=1}^I \frac{1 - I^{-1} q_{ij}(\hat{\mathbf{q}})}{n_j q_{ij}(\hat{\mathbf{q}})}, \quad (17)$$

then the statistic $Z = (X^2 - IJ) / t(\hat{\mathbf{q}})$ is approximately standard normal, and large values of Z (e.g., $Z > 1.96$), are taken as evidence of bad fit. Such an approximation is valid when n , IJ , and $n^2/(IJ)$ are all large. Because power may be low in these goodness-of-fit tests, direct examination of Pearson's residuals

$$e_{ij} = \frac{n_{ij} - n_j q_{ij}(\hat{\mathbf{q}})}{\sqrt{n_j q_{ij}(\hat{\mathbf{q}})}} \quad (18)$$

may be a more useful guide to judge goodness of fit than a simple hypothesis test.

Examples

Simulated data for the gamma-gamma model

We simulated the catch frequencies, n_{ij} (Table 1), from the gamma-gamma model as follows. We generated $N=10,000$ independent "fish" lengths from a gamma distribution (Eq. (13) with $\mathbf{a} = 3$, and $\mathbf{b} = 4.5$ cm). We chose 5 nets (mesh sizes $m_j=10, 12, 14, 16$, and 18 cm) with gamma-shaped selectivity $r_j(y)$ under the geometric selectivity hypothesis and full fishing power (Eq. (14) with $h = 1$, $\mathbf{a} = 3$, and $k = 0.6$). $N_j = 2,000$ fish encountered each net. We assumed no fish encountered a net more than once.

We simulated numbers Y_k representing the lengths of the 2,000 fish encountering the j th net, generated by

Table 1. Simulated catch frequencies based on the gamma-gamma model. Recall y denotes an index of the midpoint of length classes. Numbers within the table denote the numbers of fish caught by the mesh size.

y	Mesh Size (cm)					y	Mesh Size (cm)					y	Mesh Size (cm)				
	10	12	14	16	18		10	12	14	16	18		10	12	14	16	18
1	4	5	2	0	0	11	122	106	94	97	80	21	17	18	24	34	24
2	26	14	16	10	9	12	92	84	106	82	60	22	11	18	24	18	16
3	42	54	25	15	18	13	73	77	91	75	88	23	8	12	14	14	17
4	72	58	60	47	35	14	71	71	56	67	66	24	13	9	9	10	16
5	91	92	82	59	45	15	68	67	64	66	58	25	6	5	9	16	10
6	117	88	93	69	60	16	45	51	59	64	47	26	2	5	6	7	10
7	126	108	108	96	70	17	37	43	37	44	48	27	2	8	5	5	8
8	104	117	116	79	77	18	27	34	33	44	52	28	6	6	5	2	5
9	120	133	94	95	84	19	31	24	39	41	20	29	3	4	3	4	3
10	113	136	111	84	88	20	28	24	30	27	23	30	1	3	3	4	4

the gamma distribution (Eq. (13)), with probability of capture $r_j(Y)$. To simulate random captures, we generated 2,000 independent uniform-(0,1) random numbers U_k , and we declared the k th fish captured if $U_k < r_j(Y_k)$ and we tallied it in the appropriate length class ($I=30$ classes, width: $\Delta = 1$ cm, class midpoints: $l_i = 1, 2, \dots, 30$ cm).

Calculations were done in S-Plus (S-Plus 2000), using the function `nlminb` to maximize the log-likelihood (Eq. (4)) under the constraint that all parameters must be positive. Large-sample estimates of standard errors (SE) of the parameters \mathbf{a} and k were obtained as square roots of the hessian matrix (Ripley and Venable 1994). Large-sample estimates of functions of parameters, such as N_j and N , were obtained by the delta method (Agresti 1990).

Table 2. Estimated net encounters from data displayed in Table 1, together with estimated SE. The column labeled n_j provides total catches by mesh size.

Mesh Size (cm)	N_j	SE	n_j
10	1932.50	35.18	1478
12	2019.34	84.44	1474
14	2051.74	24.35	1418
16	1956.39	148.55	1275
18	1858.15	164.75	1141

In this simulation, $n = 6,787$ fish were caught. The estimated total encounters was $\hat{N} = 9,818$ fish (SE = 557). This estimate is systematically low because the estimation process assumes $h = 1$. The estimated parameters are $\hat{\mathbf{a}} = 2.58$ (SE = 0.03), $\hat{\mathbf{b}} = 3.83$ (SE = 0.17), and $\hat{k} = 0.71$ (SE = 0.07). Table 2 gives the estimates of per-net encounters. Pearson's chi-square value is $X^2 = 153.1$ at 146 degrees of freedom, corresponding to a P -value of 0.33. The Z -approximation for X^2 equals 0.175, corresponding to a P -value of 0.569. Pearson's residuals are shown in Figure 2. The largest residual was 3.05. The fit is acceptable by the usual criteria.

American Plaice data

The data are catches of American plaice *Hippoglossoides platessoides* using multimono trammel nets with inner mesh sizes $m_j = 9.83, 10.84, 11.94, 12.87, 13.97, \text{ and } 15.07$ cm. These data were published by Holst and Moth-Poulsen (1995) and have been reworked by Millar and Fryer (1999). Length classes are 1-cm with midpoints at 20, 21, ..., 55 cm. The catch frequency matrix $\{n_{ij}\}$ has 216 entries, of which 68 are zeros and 136 are ≤ 5 . A total of 4,306 fish were captured. We again consider the gamma-gamma model and

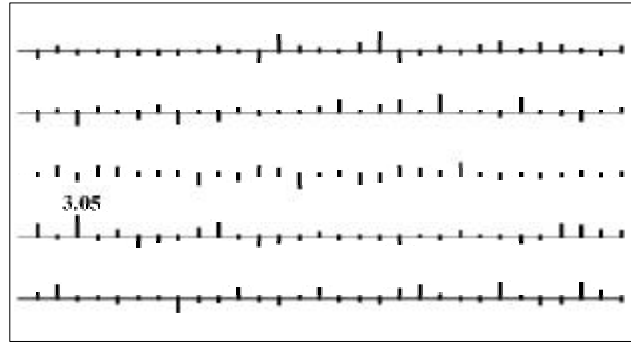


Figure 2. Pearson's residuals of simulated data.

assume geometric similarity. The estimated number of total encounters was $\hat{N} = 10,811$ (SE = 2,557). The estimated parameters were $\hat{\mathbf{a}} = 49.48$ (SE = 0.41), $\hat{\mathbf{b}} = 0.64$ (SE = 0.02), and $\hat{k} = 0.05$ (SE < 0.01).

Table 3. Results from the American Plaice *Hippoglossoides platessoides* data of Holst and Moth-Poulsen. N is the estimated number contacting the net. The number caught in the net is n_j .

Mesh Size (cm)	N_j	SE	n_j
9.83	3791.29	1479.41	759
10.84	3011.62	814.59	1194
11.94	1871.87	282.78	1029
12.87	1086.30	62.98	675
13.97	661.60	29.05	424
15.07	388.35	53.55	225

Table 3 gives estimates of the number of fish encountering each of the 6 nets. Pearson's goodness-of-fit statistic is $X^2 = 155,312.6$ at 206 degrees of freedom. The Z -value for X^2 is 0.222, which does not indicate a poor fit. Fourteen Pearson residuals were larger than 10 in absolute value. In Figure 3, all Pearson's residuals are truncated to the range -10 to 10. The largest residual (305.3) is also indicated. It corresponds to the catch of two 44-cm plaice by a net with 9.83-cm meshes.

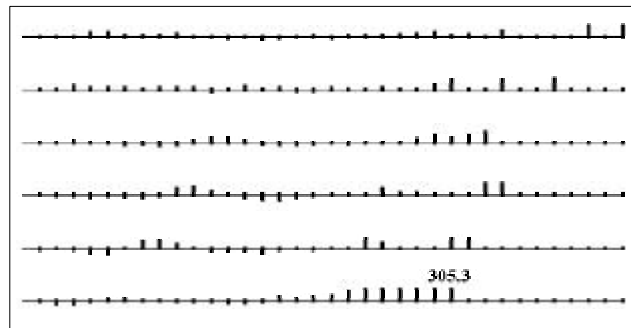


Figure 3. Pearson's residuals of American Plaice data.

DISCUSSION

As other authors have noted—and we wish to stress this—selectivity models are generally overparameterized. This has led to assumptions that prevent the estimation of h_j , and to models that often produce unstable parameter estimates in the maximum likelihood setting. The usual prescription for this problem would be to simplify the model and reduce the number of parameters. However, with our selectivity model it may be possible to partially fix the problem of overparameterization by incorporating additional information through Bayesian methods.

Our model provides an obvious starting point for a Bayesian approach in either the direct or indirect situation. For example, suppose fishing studies are repeated over several years to estimate the size distribution of a population of fish, or maybe to apportion species that differ in size to a distribution of sonar targets. Then selectivity parameter estimates from previous years might form the basis for informative prior distributions for parameters of the selectivity curves. Because the normal-normal model is simpler to consider, suppose μ_j s and \mathbf{s}_j s were estimated in several previous years, then an average and standard error of the previous year's estimates will help form choices for the parameters of a prior distribution for the selectivity distributions. However, because the length distribution of the population of fish changes from year to year, the prior parameters should be chosen to provide a relatively noninformative prior for that distribution. Similarly, in the direct case, if the fish length from the population was sampled with nonselective gear, the mean and standard deviation from the sample could provide a basis for a prior for μ and \mathbf{s} . In this direct case, the prior for the μ_j s and \mathbf{s}_j s should be relatively noninformative. The Bayesian machinery is straightforward and well described in the statistical literature for the normal-normal model (e.g., Box and Tiao 1973), although this normal-normal model may not be very useful for real-world data. The Bayesian machinery is not quite as straightforward for the gamma-gamma model. The inverse gamma distribution is sometimes used as a prior distribution for a rate parameter, but analysts would

probably want to consider a computer-intensive analysis that would not require the traditional conjugate prior-posterior distribution relationship.

Our formulation may be helpful for riverine sonar operators. Although net selectivity models have been used in connection with riverine sonar for years, we know of no instance where specific methods for riverine sonar have been published. When a portion of the river is hard to ensonify, the sonar operators may wonder if the sonar is detecting fish in that hard-to-reach area. Sonar operators often fish in those areas, and consider a "large" catch to be an indication that the sonar may be underestimating salmon passage. Usually when nets are used this way, operators only report whether they caught a large number or a small number of fish, without doing any kind of statistical expansion. For the gamma-gamma model, once good estimates of the parameters have been generated, the number of fish encountering nets can be quantified, as we have shown. Quantifying the number of fish encountering the nets would be an improvement over simply reporting the number of fish caught.

Finally, we note a lack of published gillnet selectivity studies given the possible importance of the long-term influence of net selectivity on fish populations. For example, the largest gillnet fishery for sockeye salmon is in Bristol Bay, Alaska. The gillnet catch exceeded 44 million fish in 1995, and catches have exceeded 10 million fish every year since 1978. Mathisen (1971) measured the effects of mesh size on the biology of Bristol Bay sockeye from 1908 to 1966. With the exception of Mathisen's study, Bue (1986) did not find any other studies of this issue in Bristol Bay at the time of his work. As far as we can tell, no studies of this issue have been conducted in Bristol Bay since Bue (1986). Ricker (1981) proposed that gillnet-size selectivity might have altered size and age distributions nearly 20 years ago. Large changes in Pacific salmon body sizes and age distributions are increasingly generating scientific interest (e.g., Ishida et al. 1993; Bigler et al. 1996; Helle and Hoffman 1998). Recent efforts have been aimed at hypothesized mechanisms based on density dependence and ocean-climate change. The role of net selectivity in controlling the body size of spawning salmon needs additional study.

LITERATURE CITED

- Agresti, A. 1990. Categorical data analysis. John Wiley and Sons.
- Agresti, A., and M. C. Yang. 1987. An empirical investigation of some effects of sparseness in contingency tables. *Computational Statistics and Data Analysis* 5:9–21.
- Baranov, F. I. 1948. Theory and assessment of fishing gear. Chapter 7 in *Theory of fishing with gillnets*, 2nd ed. (Translation from Russian by Ontario Department of Lands and Forests, Maple, Ontario).
- Bigler, B. S., D. W. Welch, and J. H. Helle. 1996. A review of size trends among North Pacific salmon (*Oncorhynchus* spp.). *Canadian Journal of Fisheries and Aquatic Sciences* 53:455–465.
- Box, G. P., and G. C. Tiao. 1973. Bayesian inference in statistical analysis. Addison-Wesley, Reading, Massachusetts.
- Bue, B. G. 1986. Effects of gill net selectivity on sockeye salmon in the Egegik and Naknek-Kvichak Districts, Bristol Bay, Alaska. Master's thesis. University of Alaska, Fairbanks.
- Hamley, J.M. 1975. Review of gill net selectivity. *Journal of the Fishery Research Board of Canada* 32:1943–1969.
- Hamley, J. M., and H. A. Regier. 1973. Direct estimates of gillnet selectivity to walleye (*Stizostedion vitreum vitreum*). *Journal of the Fishery Research Board of Canada* 30:817–830.
- Helle, J. H., and M. S. Hoffman. 1998. Changes in size and age at maturity of two North American stocks of chum salmon (*Oncorhynchus keta*) before and after a major regime shift in the North Pacific Ocean. *North Pacific Anadromous Fisheries Commission Bulletin* 1:81–89.
- Henderson, B. A., and J. L. Wong. 1991. A method for estimating gillnet selectivity of walleye (*Stizostedion vitreum vitreum*) in multimesh multifilament gill nets in Lake Erie, and its applications. *Canadian Journal of Fisheries and Aquatic Sciences* 48:2420–2428.
- Helser, T. E., J. P. Geaghan, and R. E. Condrey. 1997. Estimating gillnet selectivity using nonlinear response surface regression. *Canadian Journal of Fisheries and Aquatic Sciences* 55:1328–1337.
- Holst, R., and T. Moth-Poulsen. 1995. Numerical recipes and statistical methods for gillnet selectivity. *International Council for the Exploration of the Sea* 1995, B:18.
- Holt, S. J. 1957. A method for determining gear selectivity and its applications. *International Council for the Exploration of the Sea. ICNAF-ICES-FAO Joint Scientific Meeting, Lisbon, Paper no. S-15.*
- Hovgård, H. 1996a. A two-step approach to estimating selectivity and fishing power of research gill nets used in Greenland waters. *Canadian Journal of Fisheries and Aquatic Sciences* 53:1007–1013.
- Hovgård, H. 1996b. Effect of twine diameter on fishing power of experimental gill nets used in Greenland waters. *Canadian Journal of Fisheries and Aquatic Sciences* 53:1014–1017.
- Hovgård, H., H. Lassen, N. Madsen, T. M. Poulsen, and D. Wileman. 1999. Gillnet selectivity for North Sea Atlantic cod (*Gadus morhua*): Model ambiguity and data quality are related. *Canadian Journal of Fisheries and Aquatic Sciences* 56:1307–1316.
- Ishida, Y., S. Ito, M. Kaeriyama, S. McKinnell, and K. Nagasawa. 1993. Recent changes in age and size of chum salmon (*Oncorhynchus keta*) in North Pacific Ocean and possible causes. *Canadian Journal of Fisheries and Aquatic Sciences* 50:290–295.
- Kawamura, G. 1972. Gill-net mesh selectivity curve developed from length-girth relationship. *Bulletin of the Japanese Society of Scientific Fisheries* 38:1119–1127.
- Kitahara, T. 1970. On selective curve of gillnet. *Nippon Suisan Gakkaishi* 37:289–296.
- Koehler, K. J., and K. Larntz. 1980. An empirical investigation of goodness of fit statistics for sparse multinomials. *Journal of the American Statistical Association* 75:336–344.
- Mathisen, O. A. 1971. Escapement levels and productivity of the Nushagak sockeye salmon run from 1908 to 1966. *Fisheries Bulletin* 69:747–763.
- McCombie, A. M., and F. E. Fry. 1960. Selectivity of gill nets for lake whitefish (*Coregonus clupeaformis*). *Transactions of the American Fisheries Society* 8:176–184.
- Millar, R. B. 1992. Estimating the size-selectivity of fishing gear by conditioning on the total catch. *Journal of the American Statistical Association* 87:962–968.
- Millar, R. B. 1995. The functional form of hook and gillnet selection curves cannot be determined from comparative catch data alone. *Canadian Journal of Fisheries and Aquatic Sciences* 52:883–891.
- Millar, R.B., and R. Holst. 1997. Estimation of gillnet and hook selectivity using log-linear models. *International Council for the Exploration of the Sea Journal of Marine Science* 54:471–477.
- Millar, R. B., and R. J. Fryer. 1999. Estimating the size-selection curves of towed gears, traps, nets and hooks. *Reviews in Fish Biology and Fisheries* 9:89–116.
- Quang, P. X. 1991. A non-parametric approach to size-biased line transect sampling. *Biometrics* 47:269–279.
- Quang, P. X., and E. F. Becker. (1997). Combining line transect and double count sampling technique for aerial surveys. *Journal of Agriculture, Biological, and Environmental Statistics* 2:1–14.
- Quinn, T. J., and R. B. Deriso. 1999. *Quantitative Fish Dynamics*. Oxford University Press, New York.
- Regier, H. A., and D. S. Robson. 1966. Selectivity of gill nets, especially to Lake whitefish. *Journal of Fishery Research Board of Canada* 23:423–454.
- Ricker, W. E. 1981. Changes in the average size and average age of Pacific salmon. *Canadian Journal of Fisheries and Aquatic Sciences* 38:1636–1656.
- Ripley, W. N., and B. D. Venable. 1994. *Modern Applied Statistics with S-Plus*. Springer-Verlag, New York.
- Sechin, Y. T. 1969. A mathematical model for the selectivity curve of gillnet. *Rybn. Khoz.* 45(9):56–58. (Translation from the Russian by International Pacific Halibut Commission, Seattle, Washington)
- S-PLUS 2000. 2000. MathSoft, Inc. Seattle, Washington.

Appendix A. Proofs.

1. Proof of Eq. (1).

Let Y_{jk} be the lengths of fish contacting the j th net, let X_{jk} be the lengths of fish caught by the j th net, and let $F_j(x)$ be the common CDF of the X_{jk} s, then

$$F_j(x) = P(X_{jk} \leq x) = P(Y_{jk} \leq x \mid \mathbf{d}_{jk} = 1) = \frac{P(Y_{jk} \leq x, \mathbf{d}_{jk} = 1)}{P(\mathbf{d}_{jk} = 1)} = \frac{1}{p_j} \int_0^x g_j(y) \mathbf{j}(y) dy.$$

Then take the derivative of $F_j(x)$.

2. Proof of relations among Equations 8, 9, and 10.

$$f_j(x \mid \mathbf{q}) = \frac{h_j}{\sqrt{2\mathbf{p}\mathbf{s}} p^2} \exp\left[-\frac{(x-\mathbf{m})^2}{2\mathbf{s}^2} - \frac{(x-\mathbf{m}_j)^2}{2\mathbf{s}_j^2}\right] = \sqrt{\frac{a_j}{\mathbf{p}}} \exp\left[-a_j \left(x - \frac{b_j}{a_j}\right)^2\right],$$

where

$$a_j = \frac{1}{2\mathbf{s}^2} + \frac{1}{2\mathbf{s}_j^2}; \quad b_j = \frac{\mathbf{m}}{2\mathbf{s}^2} + \frac{\mathbf{m}_j}{2\mathbf{s}_j^2} \quad \text{and} \quad c_j = \frac{\mathbf{m}^2}{2\mathbf{s}^2} + \frac{\mathbf{m}_j^2}{2\mathbf{s}_j^2}; \quad p_j = \frac{h_j}{\sqrt{2a_j\mathbf{s}}} \exp\left(\frac{b_j^2 - a_j c_j}{a_j}\right).$$

We have

$$\begin{aligned} A &= \frac{(x-\mathbf{m})^2}{2\mathbf{s}^2} + \frac{(x-\mathbf{m}_j)^2}{2\mathbf{s}_j^2} = \left(\frac{1}{2\mathbf{s}^2} + \frac{1}{2\mathbf{s}_j^2}\right)x^2 - 2\left(\frac{\mathbf{m}}{2\mathbf{s}^2} + \frac{\mathbf{m}_j}{2\mathbf{s}_j^2}\right)x + \left(\frac{\mathbf{m}^2}{2\mathbf{s}^2} + \frac{\mathbf{m}_j^2}{2\mathbf{s}_j^2}\right) \\ &= a_j x^2 - 2b_j x + c_j = a_j \left[\left(x - \frac{b_j}{a_j}\right)^2 - \frac{b_j^2 - a_j c_j}{a_j^2} \right] = a_j \left(x - \frac{b_j}{a_j}\right)^2 - \frac{b_j^2 - a_j c_j}{a_j}, \end{aligned}$$

$$f_j(x \mid \mathbf{q}) = \frac{h_j}{\sqrt{2\mathbf{p}} \mathbf{s} p_j} \exp\left(\frac{b_j^2 - a_j c_j}{a_j^2}\right) \exp\left[-a_j \left(x - \frac{b_j}{a_j}\right)^2\right].$$

Let $a_j = 1/(2t^2)$. We then have

$$f_j(x \mid \mathbf{q}) = \frac{h_j t}{\mathbf{s} p_j} \exp\left(\frac{b_j^2 - a_j c_j}{a_j^2}\right) \frac{1}{\sqrt{2\mathbf{p}t}} \exp\left[-\frac{\left(x - \frac{b_j}{a_j}\right)^2}{2t^2}\right] = \frac{h_j}{p_j \sqrt{2a_j \mathbf{s}}} \exp\left(\frac{b_j^2 - a_j c_j}{a_j^2}\right) \frac{1}{\sqrt{2\mathbf{p}t}} \exp\left[-\frac{\left(x - \frac{b_j}{a_j}\right)^2}{2t^2}\right].$$

Because $f_j(x \mid \mathbf{q})$ is a probability density function, $p_j = \frac{h_j}{\sqrt{2a_j \mathbf{s}}} \exp\left(\frac{b_j^2 - a_j c_j}{a_j^2}\right)$.

Appendix B. List of Notations.

a_j, b_j, c_j	functions of the parameters
e_{ij}	Pearson's residual
$f_j(x)$	PDF of lengths of fish caught by the j th net
I	number of length classes
J	number of gillnets
k, k_1, k_2	parameters used for geometric similarity
l_i	midpoint of length class i
m_j	mesh size of j th net
n_{ij}	catch frequency, i.e., number of class- i fish caught by j th net
n_j	number of fish caught by j th net
n	number of fish caught by all nets
N_j	number of fish encountering the j th net
N	number of fish encountering all nets
p_j	average selectivity of the j th net for a given fish population
q_{ij}	probability that a fish captured by the j th net falls in the i th size class
$r_j(x)$	selectivity curve of the j th net
X^2	Pearson's chi-square statistic
X_j	random variable representing the length of a fish captured by the j th net
Y_j	random variable representing the length of a fish encountering the j th net
$\mathbf{q}, \mathbf{q}_0, \mathbf{q}_j, \mathbf{a}, \beta, \mu, \mathbf{s}$	parameters
$\mathbf{t}(\mathbf{q})$	asymptotic standard deviation of X^2
$L_1(\mathbf{q}), L_2(\mathbf{q})$	log-likelihood functions
d_{jk}	catch indicator of the k th fish for the j th net
Δ	width of a length class
$\mathbf{j}(x)$	PDF of population lengths

The Alaska Department of Fish and Game administers all programs and activities free from discrimination based on race, color, national origin, age, sex, religion, marital status, pregnancy, parenthood, or disability. The department administers all programs and activities in compliance with Title VI of the Civil Rights Act of 1964, Section 504 of the Rehabilitation Act of 1973, Title II of the Americans with Disabilities Act of 1990, the Age Discrimination Act of 1975, and Title IX of the Education Amendments of 1972.

If you believe you have been discriminated against in any program, activity, or facility, or if you desire further information please write to ADF&G, P.O. Box 25526, Juneau, AK 99802-5526; U.S. Fish and Wildlife Service, 4040 N. Fairfield Drive, Suite 300, Arlington, VA 22203 or O.E.O., U.S. Department of the Interior, Washington DC 20240.

For information on alternative formats for this and other department publications, please contact the department ADA Coordinator at (voice) 907-465-4120, (TDD) 907-465-3646, or (FAX) 907-465-2440.

