International Council for
the Exploration of the Sea

Statistics Committee
C.M. 1991/D:38

A modular probability model for numerical simulation of fishery surveys

by
Joachim Gröger

Institut für Meereskunde Kiel
Abtlg. Fischereibiologie
Düsternbrooker Weg 20
W 2300 Kiel 1
Germany

1 Abstract

A modular probability model for simulation of fishery surveys is constructed and applied to the three arbitrary selected North Sea fish species *Gadus morhua*, *Melanogrammus aeglefinus* and *Merlangius merlangus*. Statistically it is based on the incomplete beta function

$$F_\beta(spec_{trans} \mid a,b) = \frac{1}{B(a,b)} \int_0^{spec_{trans}} u^{a-1} \, (1 - u)^{b-1} \, du$$

$$\underline{where} \quad a,\ b > 0 \ ,$$

$$spec_{trans} \in [0,1]$$

which is also known as the beta cdf or more precisely as the cumulative distribution function of a beta-distributed random variable. Because of its flexibility this function is appropriate for fitting empirical frequency distributions of various types. Its form is determined by the two shape parameters ***a*** and ***b***. Although ML estimators generally have good statistical properties unfortunately in this case the ML technique is computationally somewhat more difficult since

$$B(a,b) = \int_0^1 (spec_{trans})^{a-1} \, (1 - spec_{trans})^{b-1} \, d(spec_{trans})$$

$$for \quad a,b > 0$$

This means, that the beta pdf or density function $f_\beta\ (spec_{trans}\ ;\ a,b)$ has no simple closed form which could easily be solved with respect to ***a*** and ***b***. Hence, for estimating the parameters ***a*** and ***b*** the numerically simpler method of moments combined with an iterative version of the minimum chi-square method is used. Especially the later procedure is per definition more likely to give significant results with goodness-of-fit tests of the chi-square type than the ML method.

The probability model is called modular which means the incomplete beta function is fitted compartmentwise to frequency distributions of the above mentioned three fish species. These compartments are identified in earlier steps by using pattern recognition procedures for detection of variance-homogenous post-strata. The entire model is based on IYFS data from the ICES for the years 1983 to 1988.

Key words: ß-distribution, modular probability model, numerical simulation, survey sampling, re-stratification, pattern recognition, cluster, bootstrapping.

2 Introduction

To start with, a brief overview of the complete concept which hopefully sets the topic of the paper into the right context: the fitting process of the ß-distribution depicted in this paper is only one part of a more complex analysis strategy to optimize the sampling and analysis design of the International Young Fish Survey (IYFS). The catch data to be processed here come from the IYFS of the years 1983 to 1988 and are re-stratified by means of a variance optimizing pattern recognition procedure in an earlier step which results in variance homogenous post-strata. This re-stratification step is strongly based on formal definitions with respect to statistical terms of stratification as well as on catch information of 14 fish species which have been selected by means of an eco-statistical criterion, the so called constancy [Balogh, 1958]. This means that all following statistical operations will be carried out in the light of clustered catch positions and therefore lead to a modular probability model which is nothing more than a multi-subarea probability model. The entire analysis strategy is schematically summarized by Fig. 1.

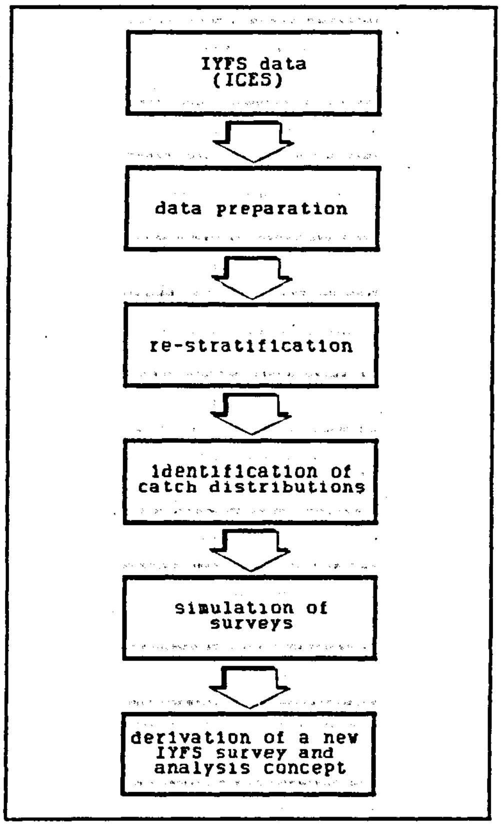

Fig. 1 General concept of the complete procedure of optimizing the IYFS.

In stock assessment, a knowledge of the expected stock size is of particular interest; statistically it is the expectation of the stock size which is in some sense similar to its mean. Such average stock numbers are however of no value without presenting an estimation error. This can be deducted on the one hand from the variability of the catches which is statistically their variance and on the other hand from the evaluation of the shape of the underlying empirical frequency distribution of the corresponding fish population. With this knowledge one can subsequently calculate the range of possible estimations of stock size with the help of specific measures and therefore the uncertainty of predictions.

To define the shape of empirical histograms, scientists in many cases assume the poisson, the negativ binomial or related distributions as overall processes which may have generated the catch data [Taylor, 1953; Bliss, 1956; Bliss, 1971; Hairston *et al.*, 1971; Matérn, 1971; Taylor, 1971; Stiteler & Patil, 1971]. Or they cut off the zero catches and try to treat both categories of data (zero and non-zero catches) separately [Pennington & Grosslein, 1978]. Where means of catch numbers are calculated on the basis of statistical geographical strata the assumption of

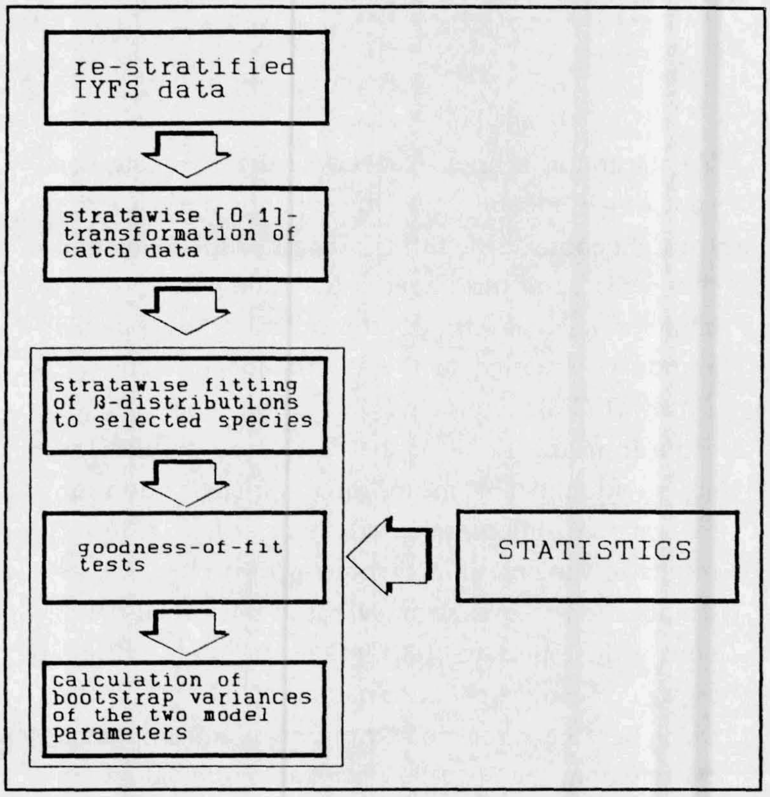

Fig. 2 The procedure of clusterwise fitting the ß-distribution to [0,1]-transformed catch data and verifying the fits.

a normal distribution of the strataspecific mean catches might be helpful since the central limit theorem (clt) can be applied. As the clt has asymptotic properties it is important that the means computed in this way must be statistically consistent [Hartung *et al.*, 1987]. This is however often not the case. For instance in the International Youngfish Survey (IYFS) on average in 33% of the cases the number of catch positions per ICES rectangle is below three. Stratawise arithmetic means therefore do not meet the asymptotic requirement of being calculated from large samples. Another condition of the clt is that the samples must be taken from non-degenerated and iid random variables. This means the samples must come from homogenous distributions and must be independent of each other as well as identically distributed. If one considers the artifical grid of regular ICES rectangles as an overlay of sampling units over the North Sea, the plausible impression intuitively arises that such an artifical structure would destroy truely underlying homogenous distributions. It is very likely that neighbouring samples taken on the basis of such strata are no longer independent of each other. Taking into account the different proportions of zero catches and other strataspecific characteristics one would also consider the stratawise empirical distributions as being non-identical. Plotted histograms of original non-transformed catch data illustrate this clearly. Taking all this into consideration, plotting the corresponding frequency distribution of the standardized strata means for one year and one species would not give the typical bell-shaped normal distribution. Goodness-of-fit-tests would confirm this result by rejecting the null hypothesis as being normally distributed.

To present a useful alternative to the theory of normally distributed catch data and/or means in this paper a very flexible distribution which is easy to compute will be proposed, the beta distribution (ß-distribution). This distribution meets the premise of generality since it is applicable to different species, years and areas. Its exceptional adaptability as well as the verification of its fits by optical impressions (plotting of histograms), goodness-of-fit-tests and boostrapping methods will be discussed here. The result is a modular probability or distributional model on which simulation studies can be performed. The concept of fitting and verifying the ß-distribution is schematically sketched in Fig. 2.The modular probability model provides more complete information about average catch numbers and their associated uncertainties. The latter can be achieved by constructing confidence limits about estimated mean catches, thus enabling one to rate the calculated mean; in other words, to get an idea of the range of possible mean values. These confidence limits according to eq. (1) are expected to be much broader than those of a superimposed normal or t-distribution.

This is due to an underestimation of the true variability of the catches and a misinterpretation of the shape of the real empirical frequency distribution in cases where the data are not normally or at least symmetrically distributed. Under these circumstances the normal distribution would not be suitable for representing the field situation, thereby resulting in misleading statements and interpretations.

$$[\overline{spec}_n - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}} ; \overline{spec}_n + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}] \qquad (1)$$

where

$u_{1-\frac{\alpha}{2}}$ = *quantile of* $f_\beta(spec|a,b)$ *at a given* α

Fig. 3 illustrates this exemplarily for *Gadus morhua*, the year 1986 and cluster 1. In other words, the true error around such a mean is much larger than usely thought. This can easily be shown by simply calculating the t- and the ß-confidence-limits corresponding to Fig. 3 and comparing its sizes. For the t-distribution an 95%-interval of size [12.86, 82.87] results and for the ß-distribution an 95%-interval of size [47.86, 5650.14] where the term in square brackets means

[lower limit , upper limit].

Immediately it can be seen that the ß-confidence-interval is approximately 80 times larger than that of the t-distribution. It should be emphasized that in case of the ß-distribution both quantiles *u* must be computed since the histogram in Fig. 3 is

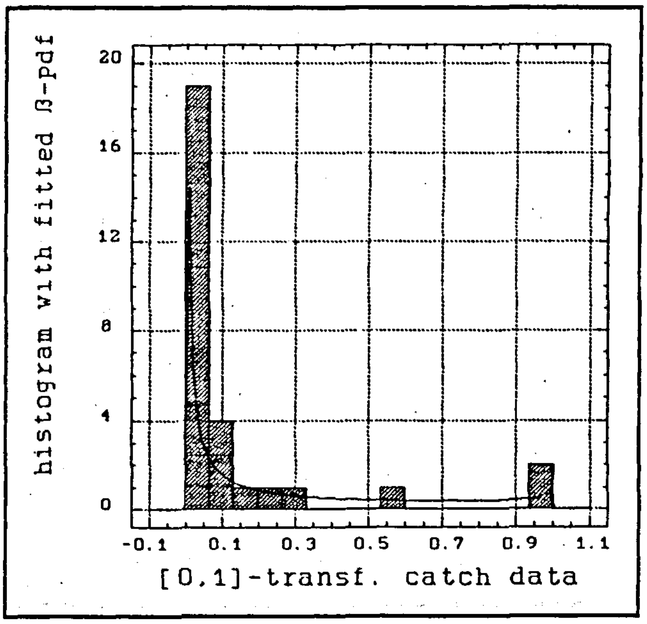

Fig. 3 Histogram with fitted ß-distribution of *Gadus morhua* (Cluster of catch positions no. 3, 1986, IYFS).

obviously not symmetrical.

Beside the application of the ß-distribution there are other possibilities of describing the underlying distribution, for instance by means of re-sampling methods or kernel and density functions, respectively. The principle advantage of the probability approach is its ability to get integrated into many statistics over maximum likelihood techniques.

Finally a comment to the mathematical notation in this paper: in order to be more transparent within equations "***spec***" stands as a shortform for catch numbers of any species.

3 Characteristic properties of the ß-distribution

The ß-distribution (sometimes called "incomplete beta function") is a continuous distribution. Its density function (pdf) has the form of eq. (2) [Rohatgi, 1976]

$$f_{\beta}(spec_{trans} \mid a,b) = \begin{cases} \frac{1}{B(a,b)} (spec_{trans})^{a-1} (1 - spec_{trans})^{b-1} & 0 < spec_{trans} < 1 \; ; \quad a,\, b > 0 \\ 0 & otherwise \end{cases} \tag{2}$$

and is obviously determined by the two shape parameters ***a*** and ***b***, whereby both have to be greater than 0. The two coefficients are named form parameters since they are responsible for its flexibility and its various possible shapes. In this respect it differs from most of the other distributions (eg. the normal distribution). The normal distribution is fixed by its "mean" (expectation) and standard deviation which describe its location and the width of its tails, respectively. It never differs greatly from the principle form of a symmetric bell-shaped curve. The ß-distribution however changes its principle type depending on the catch data. In other words, it actually adapts itself so close to real world data that it gives a good indication of the underlying process which has generated these data, i.e. an idea about the "true" population of the data. Within simulation studies it is essential to use definite functions. Obviously, in many cases the pdf can not be definite. Therefore for Monte Carlo experiments it is necessary to use the cumulative version of the pdf, the so called cumulative distribution function (cdf) which is given in eq. (3).

$$F_{\beta}(spec_{trans} \mid a,b) = \begin{cases} 0\,, & spec_{trans} \leq 0 \\ \frac{1}{B(a,b)} \int_{0}^{spec_{trans}} u^{a-1} (1 - u)^{b-1} \, du\,, & 0 < spec_{trans} < 1 \\ 1\,, & spec_{trans} \geq 1 \end{cases} \tag{3}$$

In both functional types (pdf as well as cdf) the term *B(a,b)* occurs. This expression gives the ß-distribution its name: it is the beta function which unfortunately does not allow a simple analytical solution of the estimation problem in terms of maximum likelihood methods (ML estimation). It can be seen from eq. (4) that the pdf has no closed form [Rohatgi, 1976].

$$
\begin{aligned}
B(a,b) &= \int_0^1 (spec_{trans})^{a-1}\,(1 - spec_{trans})^{b-1}\, d(spec_{trans}) \\
&= \frac{\Gamma(a)\,\Gamma(b)}{\Gamma(a+b)}
\end{aligned}
\qquad (4)
$$

Eq. (4) furthermore shows that the ß-distribution is connected to the gamma distribution over the expression *Γ(.)*. The gamma pdf is a general version of the exponential distribution. It can be seen that there are many analytical relationships to other theoretical probability distributions, like for instance the F, the binomial, the negative binomial and the uniform distribution [Boswell *et al.*, 1979]. Boswell *et al.* [1979] even claim that every univariate distribution can be fitted by the ß-distribution. The only prerequisite is that it has to be unimodal. Under certain circumstances it can also be fitted to bimodal histograms in the case where both modi are close to the one and to the other end. Fig. 4 demonstrates the exceptional variability of different shapes which are depictable by the ß-distribution.

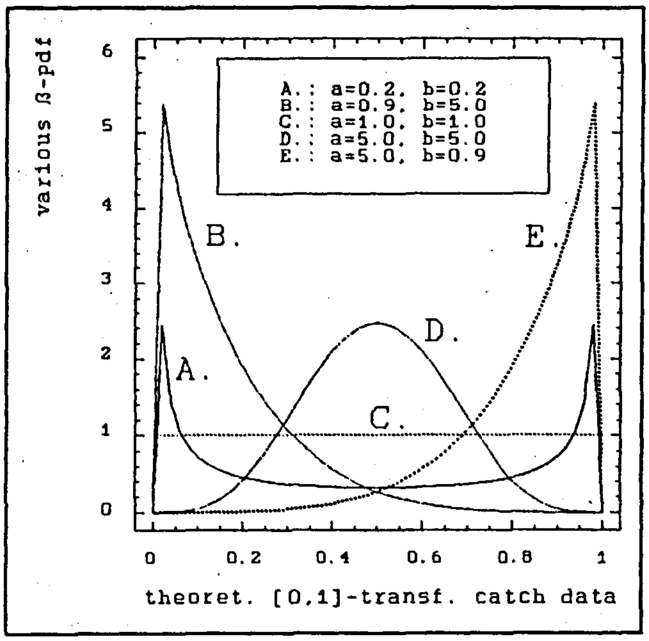

Fig. 4 Diverse possible shapes of a fitted ß-distribution [Boswell *et al.*, 1979].

It can be inferred from this figure that if both *a* and *b* are larger than 1 and *a* = *b* (Fig. 4(D.)) the curve results in a symmetrical shape - very similar to the normal (continuous) or the binomial pdf (discrete). On the other hand if *a* and *b* are smaller than 1 and *a* = *b* (Fig. 4(A.)) it yields a bimodal distribution; it then has the form of a halved bathtub. If *a* = *b* = 1 (Fig. 4(C.)) then it follows the uniform pdf *U(0,1)* (discrete or continuous). If *a* > *b* (Fig. 4(E.)) it produces a more right-skewed distribution; *b* > *a* (Fig. 4(B.)) results in a more left-skewed curve - similar to that of an exponential (continuous), poisson (discrete) or negative binomial type pdf (discrete). In other words, only the position of *a* and *b* relative to each other tells one something about the general shape of the curve, thereby providing more information than one can expect from calculating only the parameters. This is in some sense similar but less of an effort than comparing mean, mode and median of any empirical frequency distribution and reading its skewness.

The normal pdf is also depictable by the ß-distribution which is shown in Fig. 5. This histogram based on 100 simulated data from a ***N(0,1)*** is fitted by a ß-distribution without much loss of information. The fit is confirmed by the corresponding chi-square-test on a 5% significance level at 9 degrees of freedom ($\chi^2 = 15.7771$, $p = 0.0716846$); fitting the normal distribution to these data does not lead to a much better result at a 5% significance level ($\chi^2 = 11.0541$, $p = 0.198654$, $df = 8$).

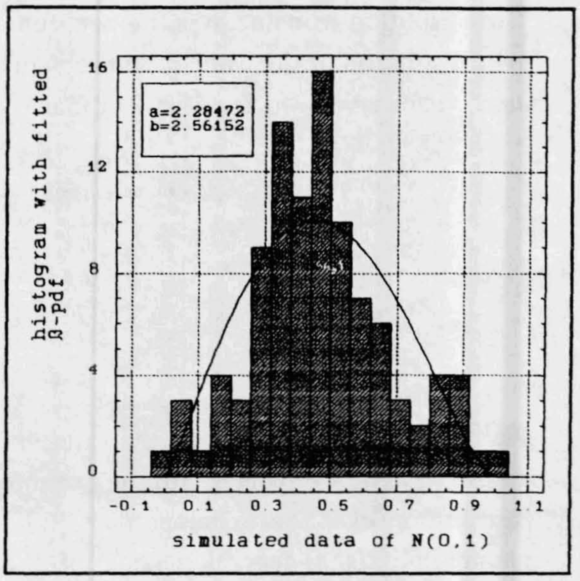

Fig. 5 A ß-distribution fitted to 100 simulated data of N(0,1).

When talking about the properties of any distribution in statistics then the depiction of characteristic quantities is meant. Two such important quantities are the expectation (eq. (5)) and the theoretical variance (eq. (6)).

$$E[spec_{trans}] = \frac{a}{a + b} \tag{5}$$

$$VAR[spec_{trans}] = \frac{a\,b}{(a + b + 1)\,(a + b)^2} \tag{6}$$

The expectation (often called first moment) of the distribution indicates the centre or the theoretical mean of the distribution whereby the variance describes the range of possible values around the expectation. In fitting a model on the basis of the moment estimation technique it is necessary to also have the second moment of the distribution (eq. (7)) [Burkhart & Strub, 1974].

$$E[(spec_{trans})^2] = \frac{(a + 1)\,a}{(a + b + 1)\,(a + b)} \tag{7}$$

The ß-distribution is not that well known and therefore not frequently applied in the biologically oriented sciences. A possible reason for this is that the catch data have to lie between 0 and 1 making a linear transformation necessary to shift them into the [0,1]-domain. This simple operation will be shown later. The estimation of the model coefficients *a* and *b* is also time consuming and requires more effort when one is using the ML technique or the minimum chi-square method. The modular fitting of the ß-distribution to real stratified catch data, the verification of the goodness-of-fit and the estimation of the variance over bootstrapping techniques will be discussed in the following section.

4 The estimation of model parameters and their variance as well as the evaluation of the qoodness-of-fit

Various techniques of estimation can be used in model fitting. Due to their good properties the ML methods are thus often preferred. Unfortunately in the current case the ML estimation is numerically too time consuming since no analytical solution of the problem exists. Iterative procedures were therefore applied. Although the variance of the ML estimators is slightly smaller [Strub, 1972; Burkhart & Strub, 1974] than when applying for instance moment estimators, further advantages of the ML estimation are not significant enough to justify its exclusive use.

For the current work the numerically much simpler moment estimation method seem to be more suitable since there are approximately 270 model fittings for the period 1983 to 1988, for the on average 15 corresponding yearly strata (- clusters of catch positions) and the three species *Gadus morhua*, *Melanogrammus aeglefinus* and *Merlangius merlangus*. 2700 model fittings will also be added due to variance estimations based on bootstrapping.

The reason why parameter estimation and goodness-of-fit tests are usually two totally different approaches is that the first is normally based on the entire catch sample whereas the latter is performed on grouped catch data which may lead to divergent results [Gurland & Hinz, 1971]. The minimum chi-square method may also be appropriate since its principle is very similar to that of a chi-square goodness-of-fit test: they both minimize the same test criterion or statistic. In other words, it may lead to more matching results of fitting and testing than other estimation techniques since it is also based on grouped data. Like the ML technique the minimum chi-square method is computionally somewhat difficult and would only be applied should the moment technique fail.

Before estimating the parameters *a* and *b* the catch data must be standardized. As mentioned above this is the linear process of shifting them into the [0,1]-domain and will be achieved by use of eq. (8) [Burkhart & Strub, 1974].

$$spec_{trans} = \left[\frac{spec - spec_{min}}{spec_{max} - spec_{min}}\right] \qquad (8)$$

Fig. 6 shows that this transformation generally does not change the principle shape of the frequency distribution. Otherwise more complicated transformation techniques would have to be used because of an actual change in the variable [Mood *et al.*, 1988].

To come back to the moment estimation technique and the derivation of the two moment estimators, for the ß-distribution the first empirical moment will be needed. This is relatively easy to be done by equally setting the expectation with the arithmetic mean of the transformed catch data as in eq. (9) [Burkhart & Strub, 1974].

$$E[spec_{trans}] = \overline{spec_{trans}} = \frac{\hat{a}}{\hat{a} + \hat{b}} \qquad (9)$$

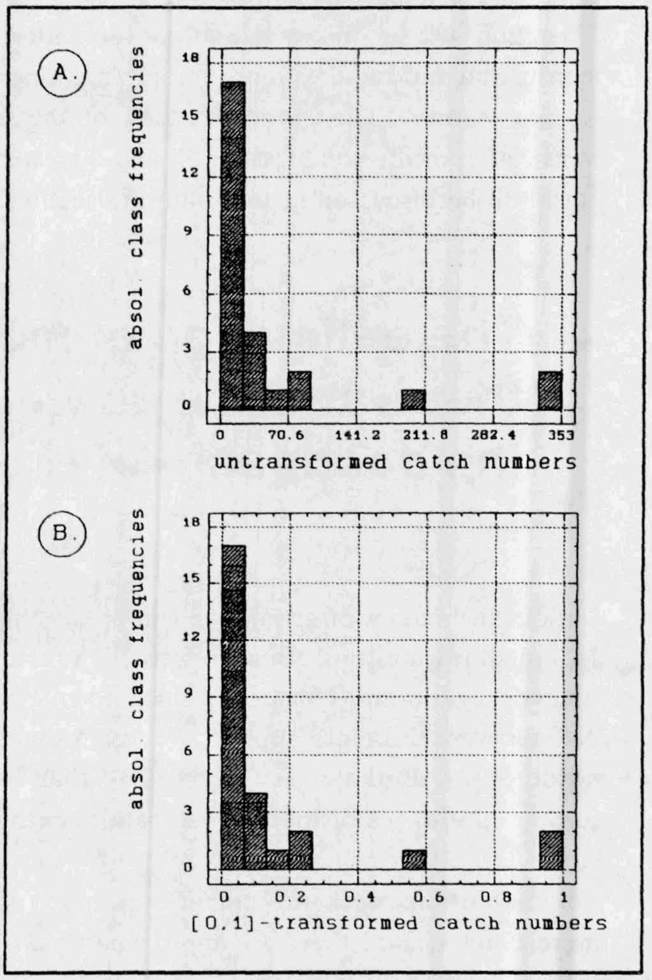

Fig. 6 A comparison concerning histograms of (A.) real and (B.) [0,1]-transformed catch data (1986, cluster of catch positions no. 3, *Gadus morhua*).

Since two estimators must be derived the second empirical moment of the ß-distribution will also be needed [Burkhart & Strub, 1974]. Eq. (10) below explains how.

$$E[(spec_{trans})^2] = \overline{spec^2_{trans}} = \frac{(\hat{a} + 1)\,\hat{a}}{(\hat{a} + \hat{b} + 1)\,(\hat{a} + \hat{b})} \qquad (10)$$

Simple reshaping and inserting of the formulas (9) and (10) into each other gives eq. (11) and eq. (12) i.e. the moment estimators for ***a*** and ***b*** [Burkhart & Strub, 1974]. In other words, for computing the parameters ***a*** and ***b*** i.e. determining or fixing the form of the ß-distribution it is only necessary to compute the formulas (11) and (12).

$$\hat{a} = \left[\frac{(\overline{spec_{trans}})^2 - \overline{spec_{trans}}\;\overline{spec^2_{trans}}}{\overline{spec^2_{trans}} - (\overline{spec_{trans}})^2}\right] \tag{11}$$

$$\hat{b} = \left[\frac{\overline{spec_{trans}} - \overline{spec^2_{trans}} - (\overline{spec_{trans}})^2 + \overline{spec_{trans}}\;\overline{spec^2_{trans}}}{\overline{spec^2_{trans}} - (\overline{spec_{trans}})^2}\right] \tag{12}$$

Also note, the hats on the ß-distribution coefficients denote statistically that these are not the theoretical parameters but their estimators calculated from the transformed catch data. This notation might be comparable with the difference between theoretical expectation and the arithmetic mean computed from real world data.

After having fit the model to the transformed catch data the quality of the fit must be evaluated. This can generally be done by optical inspection or on the basis of goodness-of-fit tests. The first is a graphical possibility to inspect the histogram and to see whether the drawn in curve of the estimated ß-distribution describes the frequency distribution best [D'Agostino & Stephens, 1986]. The latter is a set of statistical tests eg. the tests of Kolmogorow & Smirnow, of Shapiro & Wilks and finally the tests of the chi-square type.

$$\chi^2_{corr.} = \sum_{i=class\ 1}^{k} \left(\frac{(\,|c.l.^{a}_{i} - c.l.^{e}_{i}| - 0.5\,)^2}{c.l.^{e}_{i}}\right) \tag{13}$$

where

$c.l.^{a}_{i}$ = *upper class limit calculated from the catch data of the i-th frequency class*

$c.l.^{e}_{i}$ = *to* $c.l.^{a}_{i}$ *corresponding value of the* β*-distribution*

The disadvantage of the first two tests is that their test statistic is not independent of the distribution for which they check the goodness-of-fit. Unfortunately, to the authors knowledge no tables for these two tests concerning the ß-distribution exist. Also D'Agostino & Stephens [1986] only present the usual test statistics and tables of testing the fit for the normal and the exponential distributions. Whereby especially the Shapiro & Wilks test is designed for small samples, the dilemma of the chi-square goodness-of-fit test is its asymptotic test statistic. This means that this test is particularly constructed for large samples. To overcome this deficit i.e. to make the test more robust against errors associated with smaller samples, the Yates correction will in accordance with eq. (13) additionally be carried out [Bleymüller *et al.*, 1979]. If either the uncorrected or the corrected chi-square test statistic indicates a rejection of the null hypothesis (i.e. the transformed catch data would be considered as not ß-distributed) this is taken as the final and definite result. As mentioned earlier in this section, the principle problem with this type of test is that it works with grouped catch data. This can generally lead to divergent fitting and verification results since the process of grouping the catch data is in some sense relatively arbitrary (also in case of a visual inspection of histograms) [D'Agostino & Stephens, 1986; Gurland & Hinz, 1971]. For pure pragmatical reasons and since there is no completely appropriate test, the goodness-of-fit test of the chi-square type will be applied in its corrected and uncorrected versions. The afore mentioned together with visual checks of the fit between histogram and estimated ß-distribution curve forms a valid set of verification instruments.

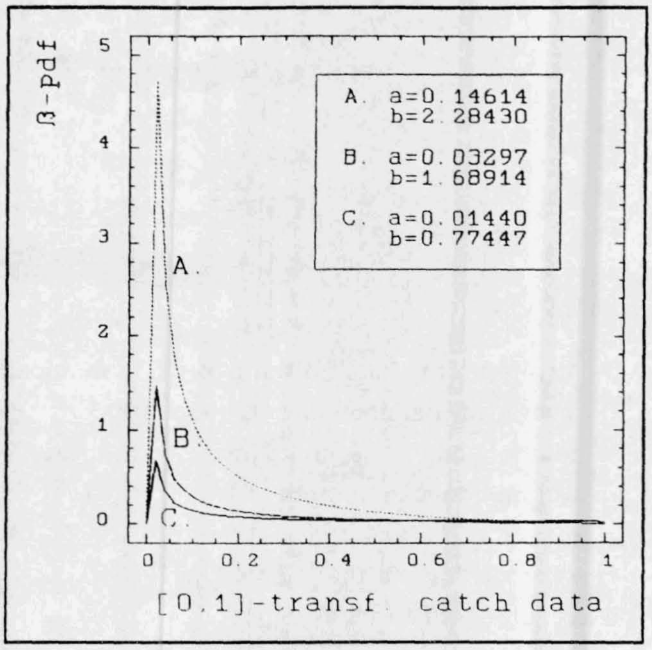

Fig. 7 Bootstrapping: graph with (A.) maximum parameters, (B.) original parameters and (C.) minimum parameters (Cluster of catch positions no. 1, 1983, *Gadus morhua*).

Having calculated the empirical estimators of the ß-distribution coefficients ***a*** and ***b*** it is necessary to get an idea about their possible ranges. In other words, it would be very helpful to estimate their variances and to be able to compute mean parameters. If there is only one sample for each year, cluster and species re-sampling techniques could be one aid to solve this problem. In particular the bootstrapping method has some quite interesting properties which makes it a more powerful instrument than the jackknife procedure. The trick is to copy the sample several times and to take subsamples with replacement from this multiplied sample. The "sampling with replacement" ensures that all subsamples are independent of each other. Against this background the fact that all subsamples can be considered as equally likely enables one to use only a simple random number generator based on the uniform distribution $U(l,u)$ for taking subsamples [Beran, 1982; Efron, 1983; Diaconis & Efron, 1983; Boos & Monahan, 1986; Hinkley & Schechtman, 1987]. The quantity l in the symbol means "lower limit" and u stands for "upper limit". The main advantage of bootstrapping is that as many subsamples of unlimited size can be taken. In comparison the jackknife subsamples would be strongly delimited to $(n - 1)$ subsamples of maximum size $(n - 1)$ which is especially a crucial point in small sized samples. The expression n in the term $(n - 1)$ denotes the number of yearly catches (catch positions) per cluster. Fig. 7 depicts the graphical idea behind the bootstrapping for *Gadus morhua*, cluster no. 1 and the year 1983. The above curve (A.) represents the upper limit of the range (= maximum parameters), the middle curve (B.) shows the original model of the ß-distribution (= original parameters) and finally the curve (C.) gives the lower limit of the possible range (= minimum parameters).

Eq. (14) shows the principle construction of the corresponding bootstrap variance which is based on the bootstrapping samples examplarily for the parameter *a*. In case of *b*, *a* must be replaced by *b* within the formula. From this equation it can easily be seen that for each bootstrap sample the ß-distribution will be fitted to the bootstrapped catch data so that a sequence of estimated parameters *a* and *b* will be created. This sequence is used to calculate a bootstrap mean and variance for each of the two parameters.

$$VAR_{boot}(\hat{a}) = \frac{1}{N_{boot}} \sum_{i=1}^{N_{boot}} (\hat{a}_{boot_i} - \overline{\hat{a}_{boot}})^2$$

where

$$\hat{a}_{boot_i} = \text{calculated parameter } \hat{a}_i \text{ of the } i\text{-th bootstrap sample} \tag{14}$$

$$\overline{\hat{a}_{boot}} = \frac{1}{N_{boot}} \sum_{i=1}^{N_{boot}} \hat{a}_{boot_i} = \text{mean of all } \hat{a}_{boot_i}$$

$$N_{boot} = \text{number of bootstrap samples}$$

5 General summary of main results and conclusions

It can be shown that the ß-distribution could be satisfactorily fitted to the stratified and [0,1]-transformed catch data of *Gadus morhua* in 78% of the clusters, of *Melanogrammus aeglefinus* in 80% of the cases and of *Merlangius merlangus* in 94% of the clusters for the years 1983 to 1988 (all percentages rounded). This will be confirmed by the corresponding goodness-of-fit tests at a 5% significance and the associated degrees of freedom level, and by optical inspection of the different histograms. At least the premise of generality seems to be fulfilled. This result may lead to the intuitive imagination that especially for *Merlangius merlangus* the different underlying populations per stratum might be ß-distributed. Also in case the aggregation structure of fish populations may change in time (as Cochran [1977] suspects and it probably will be) the adaquacy of the ß-distribution seems very high.

Secondly it should be emphasized that without plotting the curve a lot information with regard to the shape of the empirical distribution can be inferred only from the size constellation of the two parameters *a* and *b*. A convenient conclusion is to interpret the parameter constellation as a statistical index or coefficient for the spatial aggregation pattern of any fish species. Especially *Merlangius merlangus* shows the strata-internal tendency to spread more regularly over the clusters because in approximately 27% of the cases *a* is nearly as large as *b* and both parameters are smaller than 1. The percent rates for the other two species are: 14% (*Gadus morhua*) and 16% (*Melanogrammus aelgefinus*). It can also be shown that there might be a relationship between the tendency to right-skewed graphs and the increasing uncertainty about the parameter *b* since in those cases the coefficient of variation of *b* will decrease relatively to that of *a*.

A further interesting result is that the computation of ß-distribution confidence limits gives approximately 95% larger intervalls than by means of the t-distribution. That is, the obvious variability of catches is much larger (and also not symmetric !) than usually assumed.

Simulation studies on the basis of the modular ß-distribution model show that for the two species *Gadus morhua* and *Melanogrammus aeglefinus* a sample size of 50 per stratum (cluster) is adaquate enough. For *Merlangius merlangus* the sample size could be reduced to 20 per stratum. Since within the IYFS exclusive fishing of a particular fish species is usually not possible (due to the composition of species within the catch) a practical consequence of the results concerning the sample size is to have at least 50 catch positions per stratum. The simulation studies also indicate that one should not consider the North Sea to be one huge homogenous system but to split it up into smaller subunits which could then be considered as being representative for natural life areas of fish species. This latter result will be confirmed by re-stratification experiments performed one step earlier (see Fig. 1).

Finally, it appears that the ß-distribution is a readily applicable instrument which gives a realistic image of underlying distribution and probability processes, respectively. Furthermore the ß-distribution is not only a valid approach for forming confidence limits but can also be more or less easily integrated into many other statistical methods over ML techniques as for example ANOVA, regression analysis, or where the normal distribution is usually applied.

6 References

Balogh, J. (1958): Lebensgemeinschaften der Landtiere. Akademie Verlag, Berlin. 3

*Beran, R. (1982): Estimated sampling distributions: the bootstrap and competitors. Annals of Statistics, **10**: 212–225. . . . . 12*

Bleymüller, J., Gehlert, G., Gülicher, H. (1979): Statistik für Wirtschaftswissenschaftler. Verlag Franz Vahlen, München. 11

Bliss, C. I. (1956): The analysis of insect counts as negative binomial distributions. Proceedings Tenth International Congress of Entomology, 2: 1015–1032. 3

*Bliss, C. I. (1971): The aggregation of species within spatial units. Statistical Ecology, **Vol. 1**. Spatial Patterns and Statistical Distribution (Eds. G. P. Patil, E. C. Pielou, W. E. Waters). The Pennsylvania State University Press. 311–335. . . . . 3*

*Boos, D. D., Monahan, J. F. (1986): Bootstrap methods using prior information. Biometrika, **73**: 77–83. . . . . 12*

*Boswell, M. T., Ord, J. K., Patil, G. P. (1979): Chance mechanisms underlying univariate distributions. Statistical Ecology, **Vol. 4**: Statistical Distributions in Ecological Work (Eds. J. K.*

Ord, G. P. Patil, C. Taillie). International Co-operative Publishing House, Maryland, U.S.A. 3-158. 7

Burkhart, H. E., Strub, M. R. (1974): A model for simulation of planted loblolly pine stands. ***Vol. 30:*** *Growth Models for Tree and Stand Simulation (Ed. J. Fries). Royal College of Forestry, Sweden. 128-135. .... 8-10*

Cochran, W. G. (1977): Sampling techniques. Wiley Series in Probability and Statistics, New York. 13

D'Agostino, R. B., Stephens, M. A. (1986): Goodness of fit techniques. Marcel Dekker Inc., New York. 11, 12

Diaconis, P., Efron, B. (1983): Statistik per Computer: der Münchhausen-Trick. Spektrum der Wissenschaft, Juli 1983: 56-71. 12

Efron, B. (1983): Estimating the error rate of a prediction rule: improvement on cross-validation. Journal of the American Statistical Association, 78: 316-331. 12

Gurland, J., Hinz, P. (1971): Estimating parameters, testing fit, and analyzing untransformed data pertaining to the negative binomial and other distributions. Statistical Ecology, ***Vol. 1:*** *Spatial Patterns and Statistical Distributions (Eds. G. P. Patil, E. C. Pielou, W. E. Waters). The Pennsylvania State University Press. 143-194. .... 9, 12*

Hairston, N. G., Hill, W. H., Ritte, U. (1971): The interpretation of aggregation patterns. Statistical Ecology, ***Vol. 1:*** *Spatial Patterns and Statistical Distributions (Eds. G. P. Patil, E. C. Pielou, W. E. Waters). The Pennsylvania State University Press. 337-356. .... 3*

Hartung, J., Elpelt, B., Klösener, K.-H. (1987): Statistik. R. Oldenbourg Verlag, München. 4

Hinkley, D., Schechtman, E. (1987): Conditional bootstrap methods in the mean-shift model. Biometrika, 74: 85-93. 12

Matérn, B. (1971): Doubly stochastic poisson processes in the plane. Statistical Ecology, ***Vol. 1.*** *Spatial Patterns and Statistical Distributions (Eds. G. P. Patil, E. C. Pielou, W. E. Waters). The Pennsylvania State University Press. 195-213. .... 3*

Mood, A. M., Graybill, F. A., Boes, D. C. (1988): Introduction to the theory of statistics. McGraw-Hill, New York. 10

Pennington, M. R., Grosslein, M. D. (1978): Accuracy of abundance indeces based on stratified random trawl surveys. ICES C.M. 1978/D:13. 3

Rohatgi, V. K. (1976): An introduction to probability theory and mathematical statistics. Wiley, New York. 6, 7

Stiteler, W. M., Patil, G. P. (1971): Variance-to-mean ratio and Morisita's index as measures of spatial patterns in ecological populations. Statistical Ecology, ***Vol. 1:*** *Spatial Patterns and Statistical Distributions (Eds. G. P. Patil, E. C. Pielou, W. E. Waters). The Pennsylvania State University Press. 421-459. .... 3*

Strub, M. R. (1972): A beta distribution approach to diameter class frequencies in planted loblolly pine stands. M.S. thesis, Virginia Poly. Inst. and State Univ. Blacksburg, Virginia, U.S.A. 68 p. 9

Taylor, C. C. (1953): Nature of variability in trawl catches. U.S. Fish Wildlife Service, Fishery Bulletin, ***54:*** *145–166.* 3

Taylor, L. R. (1971): Aggregation as a species characteristic. Statistical Ecology, ***Vol. 1:*** *Spatial Patterns and Statistical Distributions (Eds. G. P. Patil, E. C. Pielou, W. E. Waters). The Pennsylvania State University Press. 357–377.* 3