LARGE MARINE ECOSYSTEMS :

EXPLORATION AND EXPLOITATION FOR SUSTAINABLE DEVELOPMENT AND CONSERVATION ON FISH STOCKS

Edited by

Dr.V.S.Somvanshi

February 2004

Fishery Survey of India
Government of India
Ministry of Agriculture
(Deptt. of Animal Husbandry & Dairying)
Mumbai-400 001

BOOTSTRAP EVALUATION OF THE SAMPLING SCHEME TO ESTIMATE THE MARINE FISH LANDINGS – A CASE STUDY

M. Srinath,
Central Marine Fisheries Research Institute, Cochin-14

ABSTRACT

Usefulness of the Bootstrap methodology in complex surveys has been demonstrated with the help of an application to the sampling scheme to estimate the marine fish landings. The methodological approach and evaluation of the sampling design with reference to the coefficient of variation of the estimate has been indicated. The optimum sample size (number of days) for a desired level of precision has also been estimated.

Introduction

Exploited fish stocks are assessed with the help of micro analytic and macro models (Alagaraja, 1990). Catch in numbers (age specific or length specific) or catch in weight and the corresponding fishing effort expended are the main inputs to the fish population models. For a proper evaluation of the stock, statistics of catch and effort along with the relevant biological characteristics are essential over time and space. The quality of this input data governs the performance (predictive or interpolative) of the models and determines the relevance of management options derived from the stock assessment .

Catches usually are estimated from sampling of commercial landings. These sampling schemes are often complex and multistage in nature. In India marine fish catch statistics are collected by the Central Marine Fisheries Research Institute, Cochin through a sampling system based on the theory of sampling (Banerji, 1971). Most of the catches are from the inshore regions and landed at about 1400 landing centres spread all along the coast line in the various maritime states of India. Keeping pace with the changing pattern of the fishery the mode of collection has also undergone change periodically without any significant alterations in the basic structure of the sampling design.

The sampling design followed by the Central Marine Fisheries Research Institute during 1970s and before was explained by Kutty *et.al.*(1973). With the spurt in the implementation of mechanization in the fishing industry the quantity and quality of data to be collected increased tremendously. Taking this into account the concept of Single Centre Zone was introduced meaning a particular centrrt which there was intense mechanized activity. The mode of collection during the late 1970s and early part of 1980s was described by Jacob *et.al.*(1983). Later, the mode of collection underwent slight change with respect to selection of crafts and the modified sampling scheme was given by Alagaraja (1990).

The estimation of the catch is straightforward and does not need any elaboration. It is known that if the first stage units are selected with replacement and the second stage units are selected systematically then the estimate of the variance reduces only due to that among the first stage units (Sukhatme and Sukhatme, 1970). However, in this case the first stage units are the landing centre days which are not selected with replacement but only the landing centres. Thus, in this case estimation of variance poses a problem. Another important aspect is the sample size. It is important to know the optimum sample size for a desired level of precision. Are the currently observed number of days and the boats selected on the selected day adequate enough for estimating the total catch for a specified level of precision? This question can be answered if an estimate of the variance is available with us (Here the total cost of the survey is not considered). This paper presents an approach in an attempt to answer the above question. This is done through a case study of the sampling scheme followed at a Single Centre Zone. Cochin Fisheries Harbour one of the most important landing centres in Kerala where large number of mechanized boats operate is chosen. The data were collected during January 1992 to December 1993 formed the material for anaysis. At this center catches from trawlers, drift-gillnetters, hooks & lines, ring-seine and purse-seine are landed. Of these, the catches from the trawlers formed the major component of the total landings and thus only trawl catches were considered for estimation of variance and also for determining the optimum sample size.

The Monte Carlo Bootstrap methodology was applied to evaluate the sampling scheme in terms of estimates of the coefficient of variation and determining the sample number of days for observation. *Kimura and Balsiger (1985) pointed out that one could spend considerable time and effort fitting these data into classical sampling theory. Alternatively, the bootstrap method uses the well-defined structure of the survey to define an empirical* *process. This sample is processed repeatedly using Monte Carlo methods and the resulting variability analyzed.* According to Efron (1982) the important theme of resampling methods such as Bootstrap is the substitution of computational power for theoretical analysis. The bootstrap can routinely answer questions which are too complicated for traditional statistical analysis.

Review of evaluation of CMFRI sampling scheme

Except for a study by Kutty *et.al.,* (1973) there had been no attempt to evaluate the sampling design of CMFRI in terms of the precision of the estimates and deriving optimum sample size. They tried to answer (1) to what extent the catch statistics at the all India level and at the state level were accurate (2) whether any improvement in the sampling procedure was possible and (3) whether the sampling fraction which depended on the number of survey staff should be increased. They arrived at certain conclusions based on the then existing sampling scheme and some simplistic assumptions on the primary stage units. They concluded that the sampling design followed by CMFRI was scientifically sound; the procedure gave fairly reliable estimates of the total all India fish landings; the state-wise estimates though less accurate were nevertheless realistic; and suggested redistribution of field staff on the basis of optimum allocation. Their study concentrated mainly on the sampling coverage on all India basis and allocation of field staff to the east and west coast of India based on the survey results from 1966 to 1970. Alagaraja and Srinath (1980) attempted to estimate the reliability of the estimates of marine fish landings in India by regressing the estimated landings with the quantity exported during the corresponding year. They concluded that the precision of the estimates obtained through sample survey of the CMFRI were within the acceptable range.

Bootstrap application to fishery surveys

Studies on bootstrap evaluation of complex surveys in fisheries have not been many and there has not been any such attempt in India. Kimura and Balsiger (1985) applied the bootstrap methods to evaluate sable fish pot index surveys in the north east Pacific Ocean. The goal of the pot index surveys was to provide estimates of average annual catch per set which could be followed through the time to provide indices of inter-annual relative abundance. The purpose of the study was to evaluate the sable fish pot index survey, analyse statistically and make recommendations concerning future design and sample sizes. Analysis of variance (ANOVA) and the Monte Carlo bootstrap (Efron 1982) were used to evaluate the survey data base. ANOVA was used to examine the statistical significance of surveys design variables. The Monte Carlo bootstrap was used to evaluate the effect of varying the number of locations sampled and the number of sets made within each depth stratum at a location. This was done in two ways. First bootstrap estimates of the coefficient of variation of the annual average catch per set were calculated. These estimates gave some indicators of how well mean annual abundance would be measured at various sampling levels. Second, the bootstrap was used to estimate a Z-statistic (between years) which indicated the statistical significance of observed mean differences at various sampling levels. The bootstrap process proceeded by first arranging a given year's data so that they were indexed by location, depth and set within location and depth, then the following steps were taken. 1. A location was randomly selected 2. Depths then were systematically sampled. At each depth within the selected location, the required numbers of sets were randomly selected and the observed catch per set recorded. 3. Steps (1) and (2) were repeated until the required number of locations were sampled. Using the bootstrap, sampling was always with replacement.

The results indicated that the increasing the number of locations could effectively reduce the estimated C.V whereas increasing the number of sets had remarkably little effect. From these they concluded that more locations should be sampled with fewer sets made at each location. They had also made comparison of bootstrap estimate of CV with the estimate for two stage sampling theory. This revealed excellent comparison of bootstrap estimates with the estimates derived from the two stage sampling. They further observed that when the between location variability was large there was little benefit from increasing the number of sets sampled within a location. Pelletier and Gros (1991) studied the propagation of sampling errors in catches to a yield model using virtual population analysis for which catch at age data was essential. The errors were assessed from three techniques, the delta method, Gaussian approximation and the bootstrap. The age specific catches were estimated from sampling of commercial landings and a detailed description of the sampling procedure was given. The three approaches were then compared in terms of required initial assumptions , types of results and probable accuracy of variance estimator. Their analysis indicated that bootstrap provided lower coefficient of variation values than the delta method. The bootstrap was more informative than the analytical approach and it provided the distribution of yield per recruit replicates in addition to mean values and variance estimates. They also found that the variance estimates from bootstrap and Gaussian approximation were quite close. In respect of reliability of variance estimators they contended that the bootstrap results were likely to be closer to reality. Stanely (1992) used the bootstrap analysis to examine the variance in trawl catch per unit effort(CPUE) for four fisheries along the Canada's Pacific coast. The resulting confidence limits based on the bias corrected percentile method indicated that the variance in CPUE varied widely among the fisheries. Depending on the fishery 20 to 100 randomly selected observations per year were required to provide minimally sufficient precision for stock assessment.

Sampling scheme and Data base

A month was divided into 3 ten day groups. From the first ten day group from among the first five days, a day was selected at random. Starting from this day, 3 clusters of two days each were formed. From the remaining two ten day groups , the clusters were selected with interval of 10 days. For example, from the first five days if the day selected was 3, then the three clusters in the first ten day group were (3,4),(5,6) and (7,8). Then from the next two ten day-groups, clusters would be (13,14), (15,16), (17,18), (23,24), (25,26) and (27,28). Thus we have 9 clusters of 2 days each accounting for 18 days. As the trawlers usually land their catches only in the after noons, the time of observation for all selected days is fixed as 1200 to 1800 hours and each day was considered as a single observation day as against the landing centre day already mentioned earlier. Thus we will have 18 days of observation in a month and these 18 days could be considered as a simple random sample from the days in month drawn without replacement. On each selected day a certain number of boats were selected to observe the catch depending upon the number of boats landing (Alagaraja *op.cit)*. Here also it was assumed that the boats were selected without replacement though in practice they were usually selected systematically.

The monthwise number of fishing days (Number of days) and the observed number of days at the Cochin Fisheries Harbour during 1992 and 1993 are given Table 1. From the table it is clear that the number of fishing days in month is varying so also the number of days observed. This is due to many reasons such as Sundays and some festival days being fishing holidays and some self imposed closed holidays by the fishermen. Although 18 days per month were selected, observations could not be made on some days due to various reasons and only the effective number of days observed were considered for the study. Ideally one would expect the sampling scheme to be uniform in all the months but due to the peculiar nature of the population being covered the uniformity could not be ensured. However, the evaluation of the sampling scheme would still be valid because the basic structure was not disturbed. Thus the scheme of collection of catch statistics

Table 1 : Number of fishing days and number of days observed at Cochin Fisheries Harbour during 1992 and 1993

Month	1992		1993	
	NF	NDS	NF	NDS
January	27	18	26	13
February	25	11	24	16
March	26	16	27	16
April	26	15	24	15
May	26	16	26	12
June	18	9	11	8
July	17	7	23	12
August	26	15	25	15
September	26	14	26	15
October	27	8	26	15
November	25	17	26	13
December	27	15	26	13

NF: Number of fishing days in a month
NDS: Number of days observed

for the purpose of this study can be assumed to be that of a classical two-stage design with the days forming the first stage units and the boats landing their catches being the second stage units.

Using the classical estimate of variance found in theory of two-stage sampling (Sukhatme and Sukhatme, 1970), the optimum sample size can analytically obtained using the appropriate formulation found in theory of two-stage sampling, based on the estimated variations at each stage of sampling. However, strictly in the implementation stage and at selection stage there might be some deviations from the theoretical approach and this may preclude estimation of optimum sample size using the classical formulation. It may be noted in the case study undertaken there is not much of complexity and the computations would be very straightforward. However on a larger scale which is the case with the All India sample survey for fish catch statistics, the calculations could be very complex because the nature of sampling scheme might vary from region to region. So the data from this case study is utilized for demonstrating the usefulness of bootstrap evaluation of the complex sample survey.

Results and conclusions

Ideally, the bootstrap evaluation in this case should be carried out in two stages one for the days and the other on the number of boats on the selected day. However, the bootstrap sampling was done only among the first stage units because on analysis it was found the percentage contribution of the variance due to the second stage units to the total variance was not large enough to be considered and the major contribution to the total variance was from among the first stage units only. Using the estimate of variance formula as given in Sukhatme and Sukhatme(1971) the coefficient of variation in the estimated average catch per day by considering only the first stage units and that by including the second stage units also is summarized in the Table 2. From the table it is clear that major contribution to the total variance is mainly from variation among the first stage units only and hence for the remainder of the study the variance of the first stage units were only considered and the variance due to the second stage units was ignored. Similar approach was followed by Kutty *et.al.*, (1973). Therefore, the bootstrap was done only for the first stage and the C.V. was estimated for different

Table 2 : Percentage coefficient of variation during 1992 and 1993.

Month	1992		1993	
	I	II	I	II
January	5.72	7.77	9.66	12.26
February	12.81	14.52	6.38	7.92
March	6.21	7.59	5.05	5.54
April	4.97	6.95	7.30	8.07
May	10.30	11.74	11.09	11.84
June	14.21	14.69	19.40	19.97
July	38.76	38.86	21.24	21.99
August	11.39	11.95	8.09	9.22
September	10.95	11.45	9.12	10.03
October	9.82	10.66	9.26	10.64
November	13.80	15.26	5.46	7.05
December	11.34	11.91	9.33	11.31

I : C.V. by considering primary stage only
II : C.V. by considering both the stages

bootstrap sample sizes. Only those months where the number of observations is more than 8 days were considered for analysis.

A bootstrap experiments was carried out with 1000 bootstraps. The bootstrap software for this study was developed in C language. The monthly coefficient of variation for different sample sizes starting from sample size of 2 days during 1992 and 1993 ranged from 40% to 8% for 2 days to 18 days observation respectively. In most of the months the coefficient of variation ranged between 10 to 15% for 10 or more days of observation per month. If it is assumed that a precision level of 10 to 15 % for estimating the total landings from a centre is assumed to be satisfactory, it could be concluded in general, 10 days observation would be sufficient to estimate the catch statistics. The conclusions about the optimum number of observations for a desired level of precision cannot obviously, be generalized to all the single centre zones in the country. Besides, these results are applicable only to the trawl fishery of the selected centre and hence same conclusions may not be valid for covering all other types of fisheries such as gillnet, ring-seine , purse-seine fishery. Because the resource constraint in time and money the investigator could concentrate only one type fishery which is of course the most dominant and an important fishery. Another important point to be considered is the fact that the trawl fishery is multi species in nature, though primarily targeted to exploit the shrimps. It was mentioned that the CMFRI provides species / group-wise estimates of landings. The review revealed usefulness of the bootstrap methodology in complex surveys as the sample surveys for collection of marine fish catch statistics. Since in the Indian context no study was carried to assess the fishery sampling scheme, the present work could be considered a precursor to the ensuring studies in evaluating the fishery surveys. Bootstrap technique was applied in the present case making certain assumptions on the sampling scheme which from the practical point of view seem quite tenable. Present study aims to study only the trawl fishery in its entirety by considering the total catch and not the individual items of the catch. It may be mentioned here that because of apparent complexity in estimating coefficient of variation item-wise it would be impracticable to advise sampling schemes for individual species or groups. The main purpose of the study was to demonstrate the usefulness of the bootstrap methodology in the complex fishery survey along the Indian coast, whose merits have already been indicated earlier. In conclusion, the following observation of Efron and Tibshirani (1986) sums up the bootstrap analysis. Even for relatively simple problems computer intensive methods like bootstrap are an increasingly good data analytic bargain in era of exponentially declining computational costs.

References

Alagaraja, K. and M. Srinath. 1980. Marine fish landings in India – Estimates and precision. *Indian J. Fish.* 27 (1&2): 155-160.

Alagaraja, K. 1990. Fisheries Resource Assessment – A Glance. *Industrial Fisheries Association Annual*, 7 : 5-11.

Banerji, S.K. 1971. Fishery Statistics. *Food Agricultural Organisation of the United Nations*. Publ. No.FC/DEV/71/5: 15p.

Efron, B. 1979. Bootstrap methods: another look at the jackknife. *The Annuals of Statistics*. &(1): 1-26.

Efron, B. 1982. The jackknife, the bootstrap and other resampling plans. *SIAM, CBMS-NSF Regional Conference Series in Applied Mathematics*, No.38: Philadelphia, Pennsylvania, USA.

Efron, B. and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistical Science*, 1: 54-77.

Efron, B. 1991. More efficient Bootstrap computations. *Journal of American Statistical Association, 85*: 78-89.

Jacob, T., K. Alagaraja and K.N. Kurup. 1983. Marine fisheries statistics in India – Present status. In Proceedings of the Workshop on Acquisition and Dissemination of data on marine living resources of Indian seas. *Mar.Fish.Infor.Ser.T&E.Ser.*No.46.

Kimura, D.K. and J.W. Balsiger. 1985. Bootstrap methods for evaluating Sablefish pot index surveys. *North American Journal of Fisheries Management*, 5 : 47-56.

Kutty, D.K., A.K.K. Nair and S.Z. Quasim. 1973. An evaluation of the sampling design adopted by the Central Marine Fisheries Research Institute for estimating marine fish production in India. *Indian J. Fish.*, 22 : 16-34.

Pelletier, D. and P. Gros. 1991. Assessing the impact of sampling error on model based management advice: comparison of equilibrium yield per recruit variance estimators. *Can.J.Fish.Aquat.Sci.*, *48*: 2129-2139.

Stanely, R.D.1992. Bootstrap calculation of CPUE variance from trawl log books. Do fisheries generate enough observation for stock assessment. *North American Journal of Fisheries Management, 12*: 19-27.

Sukhatme, P.V. and B.V. Sukhatme, 1970. *Sampling Theory of Surveys with Applications*, Ames. I A: Iowa State University Press.