SAMPLING FISH POPULATIONS

BU-154-M D. S. Robson

Paper presented in the session "Sampling for Zoologists" at the 1962 AAAS meetings

ABSTRACT

Fish sampling methods, facilitated by the great variety of nets, traps, electrical shocking devices and selective poisons to which fish are vulnerable, are nevertheless plagued by the difficulties inherent in all ecological sampling procedures aimed at selecting a "representative" sample from a natural population. Each sampling device operates selectively with respect to such factors as age, size, and behavior patterns, with the result that the population characteristics of major interest are non-identifiable without supplemental information concerning the degree of selectivity. The necessary supplemental information is also obtained by selective sampling methods, raising new problems of identifiability.

*No. BU-154-M in the Biometrics Unit, Cornell University, Mimeo Series

SAMPLING FISH POPULATIONS*

BU-154-M

D. S. Robson

Paper presented in the session "Sampling for Zoologists" at the 1962 AAAS meetings

Since the problem of sampling fisheries is contained within the broader problem of ecological sampling, and since special cases usually lead to stronger results we might expect that fish sampling methods should include some highly developed techniques peculiar to the circumstances of fish populations. While this is true to some extent, particularly with respect to methods of capture, most of the difficulties described by Dr. Eberhardt also beset the fishery biologist, with the evident complication that a fish population is distributed in a three dimensional habitat as well as through the fourth dimension of time.

Probably the most conspicuous feature distinguishing fish sampling from general ecological sampling is the extensive use of nets in the collection of samples. The great variety of nets, traps, and also electrical devices and selective poisons to which fish are vulnerable places the fishery ecologist in an enviable position despite the additional dimension of depth. These blessings are mixed, however, because of the selectivity of these devices with respect to age, size, and behavioral characteristics of individual fish; also, fatalities due to these methods of capture and to any prolonged exposure to air, which is often preceded and accompanied by violent activity on the part of the fish being handled, exert a further selective force on a sample to be captured and released, often eliminating a major fraction of those captured and hence potentially resulting in a highly select sample for release.

Another distinct characteristic of fish populations which accrues to the fishery biologist is the relatively well defined spacial boundary imposed on the habitat. The distinctness of these boundaries ranges, however, from the unmistakeable confines of a small impoundment such as a spring-fed farm pond where the fish population is too small to be of any considerable interest in itself, to the domain ofmarine fishes which for some species may cover a major portion of the globe. The well defined small pond populations are of interest only in aggregates, so the boundary problem reappears in determining aggregates, and the more expansive fresh water and marine environments, despite their acclaimed uniformity, do contain subpopulations overlapping in indistinct boundary zones. Thus, shoreline, bottom, and water surface which physically bound the acquatic habitat, serve

---

* Biometrics Unit, Plant Breeding Department, Cornell University

merely to define the space of reference within which, because of inaccessibility, the boundary problem is even more difficult to approach than that of terrestrial organisms.

Because of the economic importance of fisheries as a source of food and recreation, the vast majority of fisheries investigations are economically oriented and motivated, the major objective being to achieve some near optimum type and rate of productivity to balance with the corresponding near optimum type and rate of exploitation, the latter optimum determining the former. Efforts at regulating the population consist mainly of controls restricting the methods, timing, type and amount of exploitation, supplemented when physically and economically feasible with other measures such as watershed and stream improvement, pollution control, predator control, and planting of hatchery-reared fish.

The fish hatchery itself represents the extreme form of fish management, with rigid control over nutrition, mating and exploitation, and some degree of control over a multitude of environmental factors, while management of our major food species falls at the other end of the scale where, particularly with marine forms, control of exploitation is presently the only feasible practice. Although hatchery management decisions are also based on information obtained through sampling, many problems of the field biologist are eliminated under hatchery conditions, where fish movement is highly restricted and mortality is directly observable. Even sampling fish from a small tank is not a trivial operation, however, because of vertical stratification according to size and individual variation in ability or inclination to avoid capture. Repeated samples from the lot of fish, allowing a between-sample rest period, reveal pronounced individual differences in vulnerability to capture. Once recognized, this difficulty is readily overcome in the hatchery by using somewhat more elaborate sampling methods which automatically assign equal or known probabilities of capture. In the field, no fully satisfactory solution has been found for this sampling problem; the simultaneous use of several different methods of capture tends to compensate for the unequal vulnerabilities associated with any one capture-method, but at the same time creates the new problem of weighting the samples obtained by these several methods which operate at different and unknwon levels of efficiency. Commercial fisheries pose this problem on a grand scale since each fishing vessel on each trip represents a distinct method of capture operating on the fish population at some unknown level of efficiency. In this case, differential vulnerability with respect to fish size is an apparent consequence of mesh size of the fishing nets; in fact, regulation of mesh size is one of the principal methods of managing a commercially exploited fish population. When regarded as a sample, the total commercial catch for the season, or any part thereof, is of highly doubtful value to the fishery ecologist because of the unknown properties of the sampling procedure. Inferences from this sample back to the population carry with them the same hazzard that ruined the Literary Digest following its now classical illustration of how not to sample public opinion for predicting the outcome of a presidential election.

The basic impediment to the effective utilization of samples from a natural population is that with admittedly unknwon and unequal probabilities of selection, the total number of unknowns is of a higher order of magnitude than the population size. In addition to those unknown quantities which would numerically characterize the composition of the population, a vast set of new unknowns is generated by the sampling procedure itself, which assigns to every possible sample of every possible size some unknown probability of selection. Faced with this situation, the statistical analyst can but assign arbitrary numerical values or restrictions on a sufficient number of unknwons to enable him to solve for those remaining, and his solution, of course, is then no less arbitrary than his restrictions. The message contained in the sample is thus ambiguous, its interpretation depending upon subjectively imposed restrictions. In a strict sense, this situation obtains in every statistical problem in every field of research, the arbitrary assignment of probabilities or restrictions being couched in the assumptions concerning the statistical model, and the successes and failures of the statistical method can only be ascribed to the insight of the subject matter specialist who is responsible for formulating assumptions which approximate reality. In the fish sampling problem where every possible sample conceivably has a different probability of selection, the simplification called for to reduce the number of unknowns generated by the sampling process is a partition of all possible samples into a relatively small number of subsets within which each sample has approximately equal or approximately known relative probability of selection. For example, the biologist might believe that relative probabilities of selection depend only on fish size, so that within size classes all fish have approximately equal probabilities of selection; this, however, would still leave open the question of differential vulnerability between size-classes, which must either be guessed or in some way objectively estimated before statistical analysis of the sample is possible.

Once a sufficient number of assumptions have been made to eliminate all ambiguity, any further simplifying assumptions actually reduce the amount of pertinent information contained in the sample. The more tightly the problem is specified by assumptions, the greater the amount of sample information that becomes surplus, until finally all of the pertinent information in the sample can be summarized in a single statistic, and in the limiting case of complete specification the entire sample provides no new information. A simple example to illustrate this phenomenon is a coin tossing experiment specified to the extent that successive tosses are independent and the probability of heads, though unknown, is the same at every toss. Under these assumptions, the only unknown quantity in the model is the probability of heads, and in any fixed number of tosses the only information relevant to this unknown is the observed number of heads. The remaining information in the sample - namely, the order in which these heads occurred - is entirely irrelevant. If the further assumption is made that the probability of heads is one-half then the entire experiment is pointless and contributes nothing to our understanding of the coin-tossing process. On the other hand, if we regard the assertion that probability of heads is one-half as a hypothesis rather than an assumption, the number or proportion of heads in the experiment is again relevant and provides a criterion for testing this hypothesis. By the same token, if we go back to the assertion that successive tosses of the coin are independent and regard this as a hypothesis rather than assumption then the ordering of the heads in the sample becomes relevant and provides a criterion for testing independence. Only one major assumption then remains; the original problem included the specification that the probability of heads was the same for every toss. Since the same coin was being tossed each time, the subject matter specialist - in this case, a coin tossing expert - might well argue that this assumption must be very nearly true, and he would regard the test of ordering as truly a test of independence. A real skeptic, however, acknowledges that either or both of these assertions might be incorrect, and if the test of ordering proved significant he would conclude that either the tosses were not independent, or the probability of heads was not constant, or both.

A very similar example can be devised in the context of fish sampling if the problem is specified to the extent that all fish in the population have an equal and independent chance of being selected and if the population is assumed to be in a stationary state with a constant but unknown annual mortality rate. Under these conditions, average age is the only information in a fixed size sample which is relevant to the unknown annual mortality rate. If any one of these assumptions is doubtful and therefore regarded as a hypothesis then the entire age distribution in the sample becomes pertinent and provides a test criterion. Again, statistical significance in this test would not specifically identify the incorrect assumption but would indicate only that at least one of the assumptions is false.

This principle applies in general to all statistical methods and is useful in eliminating grossly oversimplifying assumptions. In the fish sampling problem, however, these oversimplifications may be already eliminated on a priori grounds, merely from consideration of the sampling techniques employed. On a priori grounds, the sample, instead of carrying surplus information, is actually inadequate to unambiguously estimate the unknown characteristics of the population, even under reasonable simplifying assumptions. With this impediment another source of statistical information is necessary to eliminate the ambiguities that cannot be assumed away. Unfortunately, any new source of statistical information is, in principle, subject also to a set of weaknesses which must somehow be assumed away. A source selected by the fishery biologists is the fish marking program, with all of its attendant problems. In theory, the existence of an identifiable subpopulation of known or estimable size and composition permits the estimation of sampling efficiency, at least with respect to this marked population, thus moving the researcher a step closer to non-ambiguity. If the marked population is annually sampled and replenished with new marked releases then, provided certain assumptions are fulfilled, the vital statistics of the marked population can be estimated directly, along with the estimation of sampling efficiency, and if certain other assumptions are fulfilled then the dynamics of the entire population are estimated by those of the marked population.

Some of the complications arising in a marking program have already been mentioned. The stress associated with handling and marking may, and in some instances has, been observed to result in an immediate high mortality following release, either due to the fact that the fish were already moribund when released or due to a concentration of predators attracted by the handout of near helpless fish. The further possibility has been suggested but not proven that any conspicuous mark distinguishes that one fish out of a crowd as a convenient target for a pursuing predator, thus subjecting the marked fish to a higher natural mortality rate long after release. Any fish which loses its mark is thereby eliminated from the marked population and is effectively a mortality or emigrant from that population while adding one to the unmarked population. Tag loss does occur in practice, as evidenced by residual traces of tagging on recaptured fish, and fish marked by mutilation have often been observed to regenerate the mutilated parts, becoming indistinguishable from naturally injured or even normal fish, presumably as a result of an improperly executed operation. Multiple marks and tags have been used to some extent to estimate the rate of tag losses, but a fish bristling with a variety of tags may be suspected of following a unique way of life. A tagging program which relies on tag returns by fishermen is further plagued by unreliable reporting; if a reward is offered as incentive for tag returning then the tags have been known to serve as a medium of exchange among commercial fishermen, accumulating in the hands of tavern keepers in the fishing ports. A variety of incentive schemes have been employed, including fairly substantial rewards and annual tag number lotteries, but the information gleaned from this type of reporting is of very dubious value, and tag recoveries by research vessels or research personnel stationed on fishing vessels remains by far the most effective method of reporting.

Once all such technical difficulties have been overcome, and the mark-recapture problem has been reduced in principle to an equivalent ball and urn problem, the statistical properties of the estimation procedure are still rather discouraging to the fishery biologist, since the sampling variance of his estimators is of the same order of magnitude as the population size, and noamount of statistical sophistry can alter this fact. Furthermore, an often overlooked fact is that sampling variance tells only half the story of the variability which concerns the fishery manager. Because of the many factors influencing population dynamics, the true composition of the population at any point in time is a chance outcome, never to be repeated, so the concern of the fishery manager lies rather with the basic constants governing this stochastic process, and the effect which his management practices might have on these constants. From this strategic point of view, the error in a mark-recapture estimator includes not only the difference between the estimate and the true population size, but also the difference between the true population size at that particular time and the average population size which would be generated by the stochastic process in operation. Both of these errors are subject to a variance of the order of magnitude of the average population size.

A similar hierarchy obtains with respect to the catch from the fish population, which is a chance outcome affecting the population size. The constants governing the stochastic fishing process are of strategic concern to the fishery manager while the exact catch obtained at any particular time by a commercial or sports fishery is of only tactical interest. Total catch from a commercial fishery can usually be measured directly and, by subsampling techniques, the detailed age and size composition of the catch can be estimated. For strategic purposes, however, the actual catch must be adjusted as mentioned earlier to a standardized fishing effort.

The age or size distribution of the catch has a typical unimodal form, increasing to a maximum frequency for the youngest age group attaining vulnerable size and then steadily decreasing with age. The selective pressure against survival of the faster growing young fish, revealed in the unimodal catch curve, has implications which have not yet been fully investigated. The effect of decreasing the apparent average growth rate in the population has long been recognized and serves as one explanation for the so-called Lee's Phenomenon, where the older fish in a population exhibit early scale annuli which are smaller than average, suggesting that these survivors were initially slow growers. The more permanent and hence more important effect of this selective pressure on the genetic constitution of the population remains to be studied. In fish hatchery stock, selection pressure in the opposite direction - selection for fast growing young fish - has, over the years, developed strains with substantially higher growth rates than wild stock reared under the same conditions, the conclusion suggested being that domestic and wild stocks are proceeding genetically in opposite directions with respect to growth rate. Pronounced behavioral differences are also manifested by domestic and wild stock, even at the fry stage of development, the wild stock being much warier and showing greater endurance and, of course, a higher survival rate when placed under natural conditions.

A growing interest in the experimental study of these differences has led to some perplexing sampling problems of a type common to all fields of experimental ecology. In order to compare two strains with respect to endurance, for example, a pair of fish, one of each strain, might be placed in a controlled fast current and observed to determine which fish first surrenders the battle with the current. Endurance in this sense, however, depends in part on the size of the fish. When matched as to size, the wild fish almost invariably displays the greater endurance.

but in this case the wild member of the pair is considerably older than the domestic fish, so the contest would still seem unfair. The same difficulty, incidentally, would arise in a comparison between sexes within the same s train, since males and females normally show different growth rates.

The question of how to sample for the purposes of this type of comparison falls within the domain of experimental design. The known facts point to a functional dependence of response on certain other variables, here age and size, which themselves are functionally related and known to have a different functional relation to the two populations. In light of these known facts, the most informative sample will be one deliberately selected to represent a wide range of the two variables age and size, within each of the two populations, in order to estimate the response function over the range covered by both populations. Response measured quantitatively would, of course, convey more information than the dichotomous response described earlier in the paired experiment - which, regardless of the criteria of pairing, compares the imcomparable.

This rather disconnected array of problems touched upon here serves mainly to illustrate the general principle that any sampling problem in existence anywhere has a homologue in fishery biology.