Ideas on Superpopulation Models and Inference

Ed Stanek

Contents

I. What is Statistical Inference?
II. Is the Likelihood Principle Unacceptable for Survey Sampling?
III. There is an assumption about the structure of the basic population. Can it include counterfactuals? Does the population have to be potentially observable?
IV. Can Random Effects be recovered from an N dimensional Superpopulation (incomplete)

I. What is Statistical Inference?

Numerous authors have discussed the meaning of inference, and in particular, statistical inference. A dictionary definition is given by (Hinkelmann and Kempthorne 1994) ES 2288p26:

‘The making of an inference is the act of passing from one judgment to another, or from a belief or cognition to a judgment.’

This definition seems acceptable, but does not involve statistics. A definition that involves ‘statistics’ is given by (Daniel 1995) ES 2291p147:

‘Statistical inference is the procedure by which we reach a conclusion about a population on the basis of the information contained in a sample drawn from that population.’

This definition includes the ideas of a sample and population, but doesn’t specify the link between the two. This link is specified in the definition from (Sarndal, et al. 1992) ES 2289p514,

‘To make inference, in a statistician’s language, usually means to draw conclusions, with the aid of probability statements, from a sample to a larger universe. Confidence intervals and hypothesis tests are traditional tools of inference. Suppose we have data obtained by measuring the elements $k$ in a probability sample $s$ drawn from the finite population $U$. Two important types of inference are as follows:

a. Inference about the finite population $U$ itself.
b. Inference about a model or a superpopulation thought to have generated $U$.’
Probability is the connection between what is observed, and not observed. We consider this to be the basic feature of statistical inference.

Sarndal et al distinguish two types of inference. The second type of inference is directed at a superpopulation parameter as opposed to a finite population parameter. There is a distinction between these two parameters, as described by Sarndal. Consider the comments made by Sarndal et al (p514):

‘Case (a) has occupied a major part in this book. The objective is to estimate and in other ways make inferences about descriptive parameters that characterize \( U \). This is inference about the “now,” about the current state of the finite population. By constrast, case (b) poses questions about the process that underlies the finite population \( U \).’

There is a pecking order assumed for these types of inference. The parameters for case (a) are a special case of the correct model and parameters for case (b). This is supported by Sarndal’s (p514) further comments in reference to case (b):

‘The model builder is interested not in the finite population \( U \) at the present moment in time, but rather in the process or the causal system relating \( y \) to \( z \).’

If we understand the causal system, we can understand any special case. This is a prime motivator for the case (b) inference of Sarndal. Since the question for case (b) is more important, the answer should be more important!

However, it is valuable to focus on what is obtained by the ‘inference’ in case (b). The probability structure that underlies the model for the superpopulation is taken as an assumption. Inference is limited to parameters in the assumed model. Thus, the second inference question does not validate the understanding of the probability structure that defines the causal system, but simply characterize that system. If one believes the causal system is the correct causal system, then it is correct. What the statistical inference provides is an estimate of parameters in that system.

It is noteworthy that ‘cause’ is brought into the problem here. While cause is of interest, it does not follow that the statistical inference for case (b) is more compelling. It is only more compelling if the statistical inference provides insight on cause. A more extended discussion of cause is given in Chapter 1 by (Hinkelmann and Kempthorne 1994) ES 2288. These authors also comment on these two approaches to inference, while not identifying the approaches explicitly. In defending Sarndal’s inference approach case (a) (classical approach), Hinkelmann says (p29):

‘We cannot assume that our data are a realization of some convenient stochastic process, e.g., a Gaussian linear model. We shall use randomization and rely on randomization tests of significance and inversions thereof to obtain intervals of uncertainty about effects of treatments.’
Hinkelmann’s argument seems to be that if one uses a mechanism that allegedly follows a probability structure (such as a SRS) to determine what will be realized (ie. the sample), then the probability structure must be accepted, and is not an ‘assumption’. I would argue (as did Basu 1971) ES 1880 that use of the probability structure still requires an assumption, but one which may be more readily believed.

From Rao in the context of survey sampling, (Rao 1975) ES 1881(ES1881, p489),

‘the basic problem of statistical inference is indeed the forming of opinions about a finite population by observing only a part of it.’

This is the perspective that I have taken, with the exception that I would like to allow the ‘finite population’ to be very large- essentially infinite.

**Statistical Inference: Induction or Deduction?**

Hinkelmann and Kempthorne think of the process of statistical inference as one of induction (Hinkelmann and Kempthorne 1994) ES 2288(es2288, p6-7). The idea is that we observe something, and we use what we observe to draw some conclusions about what we don’t observe. However, to link what we observe with what we don’t observe, we need some connection. This connection is provided by probability- which ultimately must be ‘assumed’, thus leading to deduction. I think that Basu would agree with this, as indicated by the following comment:(Basu 1971) ES 1880p242, ES1880.

‘If we define mathematics as the art and science of deductive reasoning-an effort at deducing theorems from a set of basic postulates, using only the three laws of logic- then statistics (the art and science of induction) is essentially anti-mathematics. A mathematical theory of statistics is, therefore, a logical impossibility!’

If we assume that a probability framework holds that represents simple random sampling, results are deductive rather than inductive, and statistics is possible. Basu had this statement in mind, as is clear by his statement (p242).

‘Mr. Chairman, you have always been telling us that the ultimate decision is an ‘act of will’ on the part of the decision (inference) maker. Isn’t it equally true that the choice of the probability model for the observable $X$ is also an act of will on the part of the statistician?’

**More Formal Definitions of Statistical Inference**

Cassel et al describe the fixed population inference problem as follows (Cassel, et al. 1977) ES 2290p33 ES2290:
‘In traditional statistical theory, an inference problem is often described in the following terms: Let \( X \) be a random quantity to be observed. Let \( \mathcal{X} \) be the sample space of \( X \), and let \( \mathcal{A} \) be a \( \sigma \)-algebra of subsets of \( \mathcal{X} \). Let

\[ \mathcal{Q} = \{ P_\theta : \theta \in \Omega \} \]

be a family of probability measures \( P_\theta \) on \( \mathcal{A} \), such that each probability measure is indexed by a parameter \( \theta \) belonging to a specified parameter space \( \Omega \). We observe \( X \) without knowing the true value of \( \theta \in \Omega \), and the observation is to serve as a basis for inference about \( \theta \).’

They also describe inference under superpopulation models as follows (p108):

‘The data \( d = \{(k, y_k) : k \in s\} \) form the starting point for the inference, as in previous chapters, but the outlook is different: The sample is given, and what remains for the statistician to do is to predict the values of \( y_k \) for the \( N - n(s) \) unobserved coordinates of \( y = (y_1, y_2, \ldots, y_N) \). This will lead to

\[ \bar{y} = \frac{\sum_{k=1}^{N} y_k}{N} \]

the prediction of \( \bar{y} \). The superpopulation may be indexed by an unknown parameter \( \theta \), so that the prediction of \( y_k \), for \( k \notin s \), may involve estimation, in the classical or Bayesian sense, of the unknown \( \theta \), through the observed \( y_k \).’

**Finiteness of the population**

I would like to think of statistical inference in a survey context as including all of statistical inference. Although superpopulation models do not necessarily have an infinite number of elements (as for example the exchangeable superpopulation models on p85 of (Cassel, et al. 1977) ES 2290, they are often described as infinite (Hartley and Sielken 1975) ES 2272. Basu limits his comments on inference in survey sampling to a survey set-up, which he defines as a finite population. I would like not to have this limitation.

We should ask whether a ‘survey set-up’ is sufficiently general to encompass what is meant by statistical inference. Basu is careful to say “no” to this premise: (Basu 1971) ES 1880 He considers the population to be finite, and that it is possible to establish a frame. Thus (see p204), Basu says

‘We are thus excluding from the survey theory such populations as, for example, the insects of a particular species in a particular area or the set of all color-blind adult males in a particular country. Such populations as above can, of course, be the subject matter of a valid statistical inquiry but the absence of a sampling from makes it impossible for such populations to be surveyed in the sense we understand the term survey here.’

I argue that the frame can be conceptual, and is not needed in practice unless one wants to use a random number generator to identify labeled units. The theory and methods don’t
require the actual labeling of the units, but rather that units could be conceptually labeled. I believe the real question to be whether one believes the probability model. If the units are labeled, and a random number generator is used to select labels (and hence units in a sample), then the simple random sampling probability model is often believed. If instead, a group of subjects “happen to be handy” (Sardal, et al. 1992) ES 2289, one can still ‘assume’ a simple random sampling probability model, but not all may agree with the assumption. The fact that more people agree with an assumption does not make it a fact.

II. Is the Likelihood Principle Unacceptable for Survey Sampling?

(comments by Godambe to Basu (Basu 1971) ES 1880 p235:
There is a problem mentioned by Godambe involving estimation of a probability in a finite population, or an infinite population. Estimates in two similar settings differ based on the likelihood principal, and lead Godambe to reject the likelihood principle. Note first that the Likelihood Principle is described by (Cassel, et al. 1977) ES 2290p46 as:

‘The likelihood principle says, roughly, that if two different outcomes (possibly generated under different experimental circumstances) give likelihood functions that differ at most by a multiplicative factor not dependent on the unknown parameters, then the two outcomes should give rise to the same inference.’

Godambe writes in response to Basu’s paper:
‘I find it difficult however to agree with him in one respect. The likelihood principle, which does not permit the use of the sampling distribution generated by randomization for inference purposes, is unacceptable to me in relation to survey sampling. It seems as though the likelihood principle has different implications for two intrinsically similar situations: for the coin tossing experiment the likelihood principle allows the use of binomial distributions while inferring about the binomial parameter but if the experiment is replaced by one of the drawing balls from a bag containing black and white balls, the likelihood principle does not allow the use of corresponding binomial (or hypergeometric if sampling is without replacement) distribution to infer the unknown proportion of white balls in the bag.’

This last conclusion I believe is because the likelihood function under SRS is flat. Since the likelihood function is flat, it doesn’t lead to a single estimate, and hence contradicts the likelihood principle. Basu comments on Godambe’s comment on page 241:
‘Now, let me turn to Professor Godambe’s objection to the likelihood principle in the context of survey sampling. It will be easier for us to understand Godambe’s point if we examine the following example. In a class there are 100 students. An unknown number \( \tau \) of these students have visited the musical show Hair. Suppose we draw a simple random sample of 20
students and record for each student, not his (or her) name, but only whether he has seen *Hair*. The likelihood is then a neat (hypergeometric) function involving only the parameter of interest $\tau$. Godambe likes this likelihood function. However, if we had also recorded the name of each of the selected students, then the likelihood function would have been a lot messier. It would no longer have been a direct function of $\tau$, but would have been a function of the state of nature $\theta = (Y_1, Y_2, \ldots, Y_N)$, where $Y_j$ is 1 or 0 according as the student $j$ has or has not seen *Hair*. Godambe does not know how to make any sense of this likelihood function. My advice to Professor Godambe will be this: “If the names (labels) are ‘not informative’, if there is no way that you can relate the labels to the state of nature $\theta$, then do not make trouble for yourself by incorporating the labels in your data”. After all, isn’t this what we are doing all the time? When we toss a coin several times to determine the extent of its bias, do we record for each toss the exact time of the day or the face that was up when the coin was stationary on the thumb? We throw out such details from our data in the belief that they are not relevant (informative). Statistics is both a science and an art.”

Basu’s comment doesn’t address how to relate the likelihood principle to superpopulation sampling. Is it possible that the apparent conflict can be resolved by different projections of the superpopulation model of Stanek?

Likelihood is discussed by others. Section 5.6 in (Cassel, et al. 1977) ES 2290p128-129 discusses a manner attributed to Royall for using likelihood in superpopulation sampling. The idea is that Royall assumes that the superpopulation has an $N$-dimensional normal distribution. Also, (Bolfarine and Zacks 1992) ES 2292 in Chapter 4 discusses likelihood in superpopulation sampling.
III. There is an assumption about the structure of the basic population. Can it include counterfactuals? Does the population have to be potentially observable?

In a completely randomized one way experimental study, the potentially observable population consists of responses to all subjects under all treatments. We distinguish between counterfactuals (conditional on not being selected in the study) and potentially observables (including all potentially observable responses). The superpopulation and arguments for inference can be constructed in this context. However, the existence of the population is a starting assumption. In a completely randomized study, this assumption may be readily accepted. However, in other contexts, it may not be accepted.

Consider a study on smoking, where the population is defined to be potentially observable response to all subjects given they are smoking, and given they are not smoking. Some may object to the assumption that there is a potential non-smoking response for a current smoker. Thus, there may be disagreement as to how to define the population.

The discussion by Royall of Basu’s paper is relevant here (Basu 1971) ES 1880p 238:

‘It seems frequently to be true that at some time before the values $y_1, y_2, \ldots, y_N$ are fixed it is natural and generally acceptable to consider these numbers as values, to be realized, of the random variables $Y_1, Y_2, \ldots, Y_N$. For instance, these might be the numbers of babies born in each of the $N$ hospitals in the state during the next month. What particular values will appear is uncertain, and this uncertainty can be described probabilistically. Although subjectivists would presumably accept these statements, in many finite populations such models are precisely as objective as those used everyday by frequentists. If such a model is appropriate before the $y$’s are realized, it seems to be equally appropriate after they are fixed but unobserved. If a fair coin is flipped, the probability that it will fall heads is one half; if the coin was flipped five minutes ago, but the outcome has not yet been observed, my statement that the probability of heads is one half is no less objective now than it was six minutes ago. The state of uncertainty is not transformed from objective to subjective by the single fact that the event which determines the outcome has already occurred.

It can be argued that since the event has already occurred, the outcome should be treated as a fixed but unknown constant (so that now the probability of heads is one if the fixed but unknown outcome is heads and otherwise is zero). Such an argument leads back to the conventional model but rests on an unduly restrictive notion of the scope of objective probability theory.’

**Logic for Inference:**

1. Assume a population.
   i. Define parameters in terms of potentially observables
   ii. Assume population size is $N$
2. Assume a probability structure that links the population to a super-population.
   i. Define the expected value and variance of the superpopulation.
ii. Define a projection of the superpopulation to the population.

IV. Can Random Effects be recovered from an N dimensional Superpopulation
References