

Using modern bone assemblages to estimate ancient populations

Nick Winder

Archaeomedes Project, Janus House, 46-48 St Andrew's Street, Cambridge, CB2 3BH, U.K.

Summary

Archaeologists trying to characterise palaeoeconomies sometimes wish to work backwards, using populations of bones persisting in the modern lithosphere to infer the sizes, or at least, relative sizes of bone populations associated with human subsistence activities. This raises a serious inferential problem; unless we can make allowances for consistent variation in dismemberment strategy, dumping behaviour and survival probabilities both within and between species, we will get the wrong answer.

It is over 25 years since Krantz (1966) published a statistic which, he argued, solved this problem. The theoretical underpinnings of Krantz's method appeared shaky and many more years elapsed before Fieller and Turner (1982) put it on a firmer base. Both Krantz and Fieller and Turner required that we be able to match left-right paired bones reliably. This paper will show there is no reason why the sorts of inferential methods conceived by Krantz, Fieller and Turner among others could not be modified and applied to assemblages with less than perfect pairing data and to some assemblages with no left-right pairing data at all. It is even possible to obtain estimators from populations of potsherds and chipped stones as well as bones.

The methods described here were developed in the course of a fixed-term research contract funded by SERC. Although case studies are in preparation, many more will be needed before their practical usefulness can be evaluated and no further work is envisaged in the immediate future. This interim description has been prepared to bring them to the attention of interested colleagues.

Sediments as traps

A body of sediments is a sort of 'deadfall trap' removing objects from the biosphere as they fail and are discarded. Some of these discarded objects will 'escape' by being re-deposited, rendered unrecognisable, destroyed or missed in excavation. Others will remain in the trap and form the archaeological *assemblage*. Each sedimentary unit will 'access' a limited range of space-time within which populations of objects, including animals and plants, reside; this is the *catchment* of the sediments. The populations of objects within the catchment of a designated sedimentary unit are called the *target populations*.

A deer that fell dead on a site will obviously be within its catchment but a deer that died half a mile away may not. A deer killed half a mile away and brought back for consumption will have been within its catchment. The catchment is not something we could draw a ring round on a map. Target populations are sets of objects which have an

appreciable probability of being recovered from a given body of sediments. Archaeologists often talk about target populations without realising that these can only be defined with reference to a specific body of sediments and class of artefacts or other remains. Sedimentary units that are close to each other may have had identical, or at least very similar catchments, while the catchments of units further apart may have been profoundly different.

When we start thinking about sediments as traps it is natural to think about the trapping of animals and the analytical tricks used to estimate population size from the numbers trapped. One of the best known of these is the Lincoln Index. In its simplest form, the Lincoln Index requires that a sample of size S_1 be caught, marked and returned to the (closed) population of size N . A new sample of size S_2 is then caught of which M will be marked. Assuming the distribution of marks in the population to be random and that no animal will learn to avoid capture during the

course of the experiment, we might suppose the proportion of marks in S_2 to approximate the proportion of marks in the whole population:

$$S_1/N \approx M/S_2$$

which, on re-arranging gives:

$$N \approx S_1 * S_2 / M \dots \dots \dots (1).$$

This method can be applied to archaeological data whenever analogues of S_1 , S_2 and M can be recognised.

Archaeozoologists take the number of right and left hand bones as analogues of S_1 and S_2 and the number of matched pairs as M . This gives us a *Probable Number of Individuals* (PNI) statistic based on the Lincoln Index:

$$PNI = \frac{Rt * Lt}{Pr} \dots \dots \dots (2)$$

where Rt is the number of rights, Lt the number of lefts and Pr the number of left-right pairs. The Lincoln Index is a *Maximum Likelihood Estimator* (MLE) of N subject to the assumptions that the capture probability of a given individual is fixed and that all capture events are statistically independent. A Maximum Likelihood Estimator is that value most likely to have resulted in the observed data given the validity of the probabilistic model. Although sometimes biased, MLEs are generally very efficient estimators when sample size is large.

Bailey (1951) has shown that, when the number 'recaptured' individuals is small (less than ten or so) a less biased release-recapture statistic can be constructed. Converting Bailey's modified formula into an archaeozoological form gives:

$$PNI = \frac{(Rt+1) * Lt}{Pr+1} \dots \dots \dots (3).$$

There are clearly two alternatives for Bailey's formula. We could also use:

$$PNI = \frac{Rt * (Lt+1)}{Pr+1} \dots \dots \dots (4).$$

Averaging the two gives:

$$PNI = \frac{(Rt+1) * Lt + (Lt+1) * Rt}{2(Pr+1)} \dots \dots (5).$$

The difference between the modified and unmodified versions are small when numbers are large, so Bailey's formula can be used in any circumstance.

Using either method on animal bones raises two obvious problems. Firstly, the probabilistic assumptions are unrealistic unless the likelihood of recovering a right hand bone is unaltered by the capture or non capture of its left hand pair. This may not be the case in practice. If a dog finds a chicken carcass it is likely to eat all of it. If a dog fails to find a chicken carcass, it eats none of it. We can not expect the capture of left and right paired chicken bones in a scavenged assemblage to be statistically independent events. The independence assumptions are reasonable for animals like sheep or cattle that are too large to be dismembered, consumed, dumped and scavenged in a single sitting; they are also reasonable with small animals whenever the material is not heavily scavenged. In general, the impact of slight statistical dependence will be less severe when estimates are made from large assemblages than when estimates are made from small ones.

The second problem is that of reliably recognising pairs. It is hard enough to tell one sheep from another in the flesh, but distinguishing animals of similar stature from bone fragments is even harder. Inefficient estimates of the number of left-right pairs recovered can vitiate PNI studies. An analogous problem to this has also been faced by population biologists who may be obliged to overestimate the number of animal 'recaptured' and a further estimator derived with this problem in mind is applicable in archaeozoology. The removal estimator has been used recently to estimate the size of populations of migrating bowhead whales from sightings at two points a few kilometres apart. Because the biologists can not always be certain whether they have actually sighted the same whale twice, they take a conservative approach, assuming that any whale at the second station that might have been among those sighted at the first is not a 'new' whale. The analogy with archaeozoological pairing studies is clear since any attempt to use morphology, age and size to recognise pairs may form 'pseudo-pairs' by matching left and right elements from animals of similar stature. Since true pairs are unlikely to be missed by well executed pairing exercises, archaeozoologists take a similarly conservative approach to recapture. The removal estimator

is a MLE of target population size subject to certain assumptions (Winder 1992a and references therein). Archaeozoological formulations of the removal estimator using conservative pairing data can be had from one of two formulae, one for right:

$$PNI_r = \frac{Rt^2}{(Rt-Lt+Pr)} \dots\dots\dots (6)$$

and one for left elements:

$$PNI_l = \frac{Lt^2}{(Lt-Rt+Pr)} \dots\dots\dots (7).$$

The removal estimate is only a MLE if its divisor is positive. If both versions have positive divisors the two can be averaged. Although inefficient estimates of *Pr* will have adverse effects of the overall efficiency of the removal estimate, the problem of estimating abundance with conservative recapture (i.e. pairing) data is obviously not insurmountable. Another way of using release recapture statistics with bones that can not be paired perfectly is described in the next section.

Turning MNI statistics into PNI statistics

Minimum Number of Individuals (MNI) is a descriptive statistic which pivots on the central notion of the reconstructed individual (the animal, the pottery vessel, the flint core). MNI is the minimum number of individuals required to recreate an assemblage and can be used as a measure of the number of individuals captured in a body of sediments. MNI statistics were introduced to archaeozoology by White in 1953. More refined methods for bones were proposed by Bökönyi (1970) and Chaplin (1971) among others. Interestingly, MNI statistics for potsherds are described by Vince (1977) and similar methods can be used to obtain MNI statistics from refit assemblages of chipped stone artefacts.

To produce an MNI statistic, we sort and identify the fragments, refitting broken pieces as far as possible. We then reconstruct recognisable individuals from the refitted fragments. Fragments which might have come from the same individual are treated as if they had done so, even when it is not possible to show that they did. If we use MNI as a measure of the number of individuals captured by a body of sediments, this conservative approach to counting implies that

an individual which can not be distinguished from another individual has, in effect, 'escaped'.

Many archaeologists are unhappy about MNI methods. They point out that it is hard work getting MNI statistics and unwise to use them as estimates of the relative size of target populations because of consistent variation in capture probability. Some are also worried by their non-additivity when assemblages from two sedimentary units are aggregated. Consider a simple archaeological site split into two 'contexts', 1 and 2. We can obtain three separate MNI statistics from this site: MNI_1 is the MNI statistic from Context 1 alone, MNI_2 is that from Context 2 alone and MNI_{tot} is that obtained when objects from both contexts are combined. These MNI statistics satisfy the inequality described by Grayson (1973; 1984):

$$MNI_{tot} \leq MNI_1 + MNI_2 \dots\dots\dots (8).$$

MNI statistics were originally used to infer the percentage contribution of different animals to the diet and it was felt that non-additivity was not the sort of property one would expect such a statistic to possess. Since other descriptive statistics (fragment counts and weights, for example) are additive across aggregated contexts, many archaeologists felt there must be something odd about the MNI statistic.

There is, of course, nothing odd about non-additivity. If the catchments of Contexts 1 and 2 overlap to some degree, some of the individuals captured at Context 1 may also have been captured at Context 2. These individuals should not be counted twice on amalgamation of contexts. The Grayson inequality is evidence that MNI behaves precisely as it should, given that the same individual may be captured more than once.

We can make Lincoln Indices or Removal Estimates from MNI statistics whenever it is reasonable to suppose that two sedimentary units may have had identical catchments with similar capture probabilities by treating the MNI statistics as independent 'trappings' from a closed population of individuals. Thus for the Lincoln Index we take:

$$S_1 = MNI_1$$

and

$$S_2 = MNI_2.$$

The number of individuals trapped twice will, by the Grayson inequality, be

$$M = MNI_1 + MNI_2 - MNI_{tot}$$

From which we obtain a generalised PNI statistic (PNI_G) based on the Lincoln Index:

$$PNI_G = \frac{MNI_1 * MNI_2}{(MNI_1 + MNI_2 - MNI_{tot})} \dots \dots (9)$$

An alternative, using Bailey's modification and averaging gives:

$$PNI_G = \frac{(MNI_1+1) * MNI_2 + MNI_1 * (MNI_2+1)}{2(MNI_1 + MNI_2 - MNI_{tot} + 1)} \dots \dots (10)$$

Equation (10) can be calculated from any assemblage capable of yielding MNI statistics. Equation (9) requires that an extreme form of the Grayson inequality holds:

$$MNI_{tot} < MNI_1 + MNI_2 \dots \dots (11)$$

This can best be arranged by using large sample units and one of the more refined MNI methods. Such MNI procedures facilitate the 'capture' of poorly preserved individuals that would have 'escaped' from more primitive methods and enhance the efficiency of the PNI estimate by boosting assemblage size.

For the purposes of estimation, an individual that can not be reliably distinguished from another in the assemblage is deemed to have 'escaped'. The probability that an individual will escape from the MNI method chosen must be constant both for subsets 1, 2 and the aggregated assemblage. Systematic errors caused by variation between the diagnostic criteria used by collaborating archaeozoologists could cause problems here as could the possibility that escape probabilities are dependent on assemblage size. Aggregating assemblages 1 and 2 may increase the relative frequency of spurious pairs. Pairing exercises based on assemblages whose catchments do not overlap at all (like distinct time phases or populations of bones of the right side, for example) could be used to test for density-dependent escape probabilities and to estimate suitable correction factors.

The capture probability of an individual at Context 1 must not be affected by its having been captured in Context 2. This is not so restrictive an assumption as that which underpins 'pairing' methods, but should not

be ignored. Essentially, it is up to the archaeologist to employ a sampling stratagem for which the assumptions are supportable, at least as a working hypothesis. On the rare occasions when distributions seem 'clumped' with whole skeletons dumped in one location, the method should not be used.

Although the size of the two captured populations and the efficiency of PNI estimates are positively correlated, this does not necessarily mean that a small site can produce so few reliable PNI estimates that we will be forced to face the sampling problems caused by basing inferences on too few data points. Suppose, for example, we have ten quadrats, Q_1, Q_2, \dots, Q_{10} . These can be split into two, arbitrary subsets of 5 quadrats each to obtain a single PNI statistic. There are 252 different ways of selecting 5 quadrats from a set of ten and so 252 PNI estimates can be obtained by splitting the assemblage into two halves. These could be averaged to produce an overall estimate for the 10-quadrat 'site'. Of course, the computational problems raised by generating MNI statistics on such a grand scale are considerable but they are by no means insurmountable; suitable computer programs have been available in the literature since the late 1970s (Creak 1978).

We could use the frequency distribution of PNI estimates from different partitionings to test the PNI assumptions. If the distribution is unimodal with a small variance, we have a stable mean PNI estimate. If large variances are obtained, the estimates are unstable. Multiple modes would imply some 'clumping' in patterns of deposition. This method of testing the PNI assumptions could be a valuable analytical tool in its own right. Any analysis that could provide an empirical refutation of the PNI assumptions would allow us to search for spatial discontinuities in target populations within a given time phase. Classical spatial analysis is restricted to characterising discontinuities in the surviving assemblage.

Any given catchment may contain many target populations. Within limits it is possible to re-classify the assemblage so as to select which of the target populations we will investigate. Separate PNI statistics could be obtained element by element to look at differences in the catchments of sub-assemblages of particular interest on ethnographic or taphonomic grounds. Taxa can be aggregated to compare the target populations of specific

sets of the elements of taxa and compound taxa within and between phases. We could compare the target population of ovicaprid tibia with pig tibia, small artiodactyl long bones with large artiodactyl long bones and so on. The key to exploiting the analytical leverage PNI methods provide is to work creatively with the target populations we can actually access, rather than to agonize about the inaccessibility of populations on some theoretician's wish-list.

Conclusions

Even non-archaeozoologists know that one of the reasons animal bones are collected is to allow us to study palaeoeconomies. Information about the relative contributions of different species to the diet is, of course, an important aspect of this work. Rather fewer outsiders realise that the differential preservation we can infer by comparing bone assemblages with whole skeletons and can observe in taphonomic and ethnographic data means that almost everything written on the subject is based on hopeful guesswork. It is hard to see why so little effort has gone in to overcoming the theoretical difficulties raised by the early PNI methods when, for all their imperfections, they provided the best independent check on traditional methods.

Although statistics have been derived for pottery assemblages by Orton (1982) which some think are similar to the faunal estimators, all the pottery methods hinge on the assumption that artefacts may be broken but will never be destroyed. Orton's pottery statistics can be used to get cost-effective figures for relative frequencies in the modern assemblage but offer no analytical leverage when trying to infer the relative sizes of ancient populations. Archaeozoology is alone among the post-excavational specialisms in having methods that may compensate for some of the effects on assemblage composition of different recovery probabilities. The potential pay-off to be won by applying these palaeoeconomic estimators both to inter- and to intra-stratum studies and, indeed, of extending their use to populations of chipped stone and potsherds, is considerable.

An initial case study has been undertaken in which a few dozen PNI statistics have been calculated from a faunal dataset describing about 10,000 fragments. These suggest capture probabilities between 0.32 and 0.36 for the

mandibular teeth of ovicaprids and 0.41-0.49 for those of cattle in the plough-damaged, abandonment fills of a Romano-British Farmstead (Winder 1992b). However, much more work will be required to refine these methods further and to prepare data for their serious evaluation. A custom-built, computerised database and computer programs capable of obtaining MNI statistics from any context or group of contexts both for bones and potsherds or bones and chipped stones would be a prerequisite of such a study. Statistical software to analyse the frequency distributions of replicate PNI estimates and graphical software to superimpose colour-coded target population densities on three-dimensional representations of the excavated sediments would also be needed for the preparation of reports.

References

- Bailey, N. (1951). On estimating the size of mobile populations from capture-recapture data. *Biometrika* 38, 293-306.
- Bökönyi, S. (1970). A new method for the determination of the minimum number of individuals in animal bone material. *American Journal of Archaeology* 74, 291-2.
- Chaplin, R. (1971). *The study of animal bones from archaeological sites*. New York: Seminar Press.
- Creak, G. A. (1978). Matching routines. *Auckland University Computer Centre Technical Report* 11.
- Fieller, N. and Turner, A. (1982). Number estimation in vertebrate samples. *Journal of Archaeological Science* 9, 49-62.
- Grayson, D. (1973). On the methodology of faunal analysis. *American Antiquity* 39, 432-9.
- Grayson, D. (1984). *Quantitative zooarchaeology: topics in the analysis of archaeological faunas*. New York: Academic Press.
- Krantz, G. (1966). A new method of counting animal bones. *American Journal of Archaeology* 72, 286-8.
- Orton, C. (1982). Computer simulation experiments to assess the performance of measures of quantity of pottery. *World Archaeology* 39, 1-20.

Seber, G. (1973). *The estimation of animal abundance*. London: Griffin.

Vince, A. (1977). Some aspects of pottery quantification. *Medieval Ceramics* 1, 63-74.

White, T. (1953). A method of calculating the dietary percentages of various food animals utilized by aboriginal peoples. *American Antiquity* 18, 396-8.

Winder, N. (1992a). The removal estimator; a 'probable numbers' statistic that requires no matching. *International Journal of Osteoarchaeology* 2, 15-18.

Winder, N. (1992b). The animal bones from Boxfield Farm, Chells. *Cambridge Faunal Remains Unit Archive Report*.

Disk copy received August 1992