Statistics and Classification

Clive Orton

Markov models for museums

1 Introduction

This paper is a sequel to studies of the use of sampling methods in the assessment of the condition of museum collections, carried out at the Museum of London (Keene/Orton 1992) and the British Museum (Leese/Bradley 1995). These studies looked mainly at the problem of obtaining a 'snapshot' of the condition of a collection at a point in time; this paper looks at the related problem of monitoring changes in condition over time.

The methods described below arose from a request from the Horniman Museum in south London for advice on the statistical aspects of monitoring the condition of the Museum's collections, following a 'census' of their condition (Walker/Bacon 1987). The design work was mainly done in 1991, but for various reasons (including the subsidence of the Ethnographic Gallery) has not yet been fully implemented. As the Museum is now fully engaged in preparations for its Centenary in 2001, it seems useful to publish 'the story so far' without waiting for full implementation.

2 Background

The Museum has three separate collections: ethnography, natural history, and musical instruments. In each collection, objects are stored by location code (e.g. type of object) and within that, by broad provenance. In the course of the census, information had been recorded on the condition of every object, on a four-point scale of priority: G (good), F (fair), U (urgent), and I (immediate), reflecting the need for remedial treatment. These correspond roughly to the four conservation priorities of the Museum of London survey (Keene/Orton 1992: 163) — Little, Low, High and Urgent — but the precise definitions may differ. Objects were generally recorded individually, i.e. one to each line of the census form, but the records for some types were 'bulked', e.g. recorded as 'F(X5), G(X20)' (meaning five objects in fair condition and twenty in good condition) on one line. It was believed that, for most types, the objects included in a 'bulk' record could be recognised individually in any subsequent survey. Exceptions are the eggs and fossils groups of the natural history collection; a different approach is needed for such groups (see below).

3 Aim

The aim was to design a system of sample surveys which would enable the condition of the collections to be monitored annually. Particular attention was to be paid to:

i. locating problem areas in stores ('hot spots'),
ii. identifying problem materials, with implications for specialist help,
iii. assessing long-term trends, e.g. suitability of particular stores,
iv. assessing staffing implications.

Points (i) to (iii) are inter-related, in that 'hot spots' are likely to occur where the environment is wrong for the type of material stored there, while 'problem materials' are usually only problems in terms of long-term preservation if they are in the wrong environment. The long-term suitability of particular stores will depend on material types and their needs. In museums which do not store objects by function and type, these points will be less interconnected.

It is not practical to carry out a census every year, nor is one needed in order to meet these aims. The need is for sample survey methods which will enable (a) the numbers of objects currently in each of the four priorities, and (b) the rate of movement of objects from one priority to another, to be estimated for each type of object.

4 Model

Objects must at any point in time be in one, and only one, of the four priorities. They may at any time move from a priority to a higher priority (i.e. their condition may worsen), but they cannot (without intervention) move to a lower priority. We cannot observe this process directly: all we can observe is the condition of selected objects at fixed intervals of time (in this example, the interval is one year). This situation can be modelled as a Markov chain (Cox/Miller 1965: 84), in which the probability of objects moving from one priority to another is expressed as a matrix of 'transition probabilities' p_{ij} from state i to state j over the fixed interval. The 'states' of the statistical theory correspond to the priorities described above.

For any chosen group, we say that the number of objects in the census, held at time 0, is $N(0)$, of which $N_i(0)$ (i = 1, …, 4) are in the i th priority (G, F, U, I). The number $N_j(t)$ in the j th priority at time t is given by

$$N_j(t) = \sum_i N_i(t-1)p_{ij},\ i = 1, \dots, 4;\ j = 1, \dots, 4;\ t = 1, 2, \dots,$$

or in matrix notation $\boldsymbol{N}'(t) = \boldsymbol{N}'(t-1)\boldsymbol{P}$.

The transition matrix $\boldsymbol{P}$ is initially given the form

$$\boldsymbol{P} = \begin{matrix} p_{11} & p_{12} & p_{13} & p_{14} \\ 0 & p_{22} & p_{23} & p_{24} \\ 0 & 0 & p_{33} & p_{34} \\ 0 & 0 & 0 & p_{44} \end{matrix}$$

where $\sum_j p_{ij} = 1$ for i = 1, …, 4.

This model is a simplification, and in real life further factors would have to be taken into account:

1. gain of objects: at any time, new objects may be added to the collections,

2. loss of objects: at any time, objects may be removed from the collection, either by disposal, by temporary absence (e.g. for display, or for loan to another museum), or because they have decayed irretrievably. Depending on the exact meaning given to the priority 'immediate', one might say that any object in priority I in year t will have decayed irretrievably by the year $t + 1$ (or perhaps $t + 2$?). This could be modelled by introducing a fifth priority D (= dead)), with a transition probability p_{45} depending on the definitions (e.g. p_{45} = 1),

3. remedial action to individual objects: surveys of collection condition are set in a context of programmes of conservation work designed to maintain or improve the overall condition of a collection. Thus the 'below diagonal' elements of the transition matrix $\boldsymbol{P}$ will not in practice be zeros. However, there are benefits in using survey data to monitor condition 'without treatment', and to write in transition probabilities reflecting actual or planned treatment programmes.

4. remedial action to stores: as one aim of monitoring is to improve overall storage conditions, it would be surprising and disappointing if the transition probabilities did not change over time, with the aim being to increase the 'diagonal' elements and decrease the 'above diagonal' ones. This means that the transition probabilities should be re-estimated at each survey, to see whether improvement has in fact taken place.

The fourth point might seem to invalidate the use of the Markov chain model, since that model assumes that the transition probabilities are independent of time (Cox/Miller 1965: 84). However, the model can be usefully employed to predict the future condition of a collection on assumptions of (for example) no intervention, or intervention at a set level of conservation of objects, and to assess the likely impact of different programmes of intervention.

The predictive abilities of a Markov chain model arise from its independence from time. Since $\boldsymbol{N}'(t) = \boldsymbol{N}'(t-1)\boldsymbol{P}$, the matrix $\boldsymbol{P}$ can be estimated by comparing $\boldsymbol{N}(0)$ (the census) with $n(1)$ (results from the first survey). This can be used to predict, or more correctly project, $\boldsymbol{N}(t)$ as $\boldsymbol{N}'(0)\hat{\boldsymbol{P}}^t$, although it must be realised that, as t increases, errors in the estimate $\hat{\boldsymbol{P}}$ will accumulate through successive $\boldsymbol{N}$ s, which will therefore become less and less reliable. Although they are not to be believed, such projections have considerable descriptive, political and management value. They can provide a dynamic description of condition: not just the present state, but also incorporating rates of change. For example, one could use the formula to project the date by which a certain proportion (e.g. 50%) of a collection will be in priority 4 (immediate), and hence (for example) the likely half-life of the collection. Such a figure could be used to highlight a need for additional resources, and the effect on such a date of the application of extra resources could be calculated. Projections made on the basis of successive surveys could show whether the collection is 'gaining' or 'losing' ground, according to whether the expected life (or half-life) is increasing or decreasing.

It has been pointed out that in standard Risk Assessment models the risk is assessed as:

Risk = Threats + Vulnerabitities + Asset value.

In the museum context, the 'asset value' of an object is made up of its historic value, its uniqueness, and its relevance to the institution. The overall condition should in principle be weighted to take account of this, since a collection in which a few valuable objects were deteriorating rapidly, while the rest were relatively stable, would be in worse condition than the raw data would imply. This has not been attempted in this survey; it would be straightforward to take account of variations is asset value between types, but much more difficult for variation within types.

5 Sample design and implementation

5.1 Sampling: theory

It seems reasonable to make each group or location code (see above) correspond to a stratum in the statistical sense, and to use stratified random sampling methods. Results can then be obtained separately for each group (type of object) and aggregated to give an overall picture of the collection.

For any one stratum, we suppose that the population at the time of the survey is N, and that a sample of size n is selected. For the objects in this sample, we know both their priority at the time of the census and their priority at the time of the survey. The number in the i th priority at the census we call $n_{i.}$, the number in the j th priority at the survey we call $n_{.j}$, and the number in the i th priority at the census and in the j th at the survey we call n_{ij}.

Then we can estimate the transition probabilities $\{p_{ij}\}$ by

$$\hat{p}_{ij} = n_{ij} / n_{i.}$$

and the numbers N_j in each priority by

$$\hat{N}_j = \sum_i N_i \hat{p}_{ij},$$

adjusting if necessary to allow for acquisitions and disposals.

It can be shown (see below) that this approach to the estimation of the Nj, known as *ratio estimation*, will give better estimates than the simpler approach $\hat{N}_j = N (n_{.j} / n)$, at least for the sorts of values of $\{p_{ij}\}$ that are likely to be encountered.

Results can be aggregated across groups to give figures for the entire collection.

This approach is very straightforward, but it assumes that the census is followed by a single survey. Our aim is to carry out a series of surveys at regular intervals, thus leading us to the theory of *repeated sampling*. Sampling on two or more occasions is discussed in detail by Cochran (1963: 341-352), who lists three aspects that one may wish to estimate:

1. the change in $\boldsymbol{N}$ from one occasion to the next,
2. the average value of $\boldsymbol{N}$ over all occasions,
3. the value of $\boldsymbol{N}$ for the most recent occasion.

Our interests are likely to lie in 1 and/or 3, but not in 2.

He gives the optimum sampling strategy for each case (*ibid.*: 342) as:

for 1, it is best to retain the same sample throughout,

for 2, it is best to obtain a new sample on each occasion,

for 3, equal precision is obtained by keeping the same sample or by replacing all of it. Replacing part of the sample may give better results than either of these.

He then goes on (*ibid.*: 345-352) to discuss sampling on more than two occasions, showing that if we are only interested in need 3, it is best to replace 50% of the sample on each occasion (*ibid.*: 347), but if we are also interested in need 1, we should increase the proportion retained to, for example, 75% (*ibid.*: 349). This increase 'produces only small increases in the variance of the current estimates and gives substantially larger reductions in the variances of the estimates of change' (*ibid.*). He suggests retaining 2/3, 3/4 or 4/5 of the sample from one survey to the next if one is interested in needs 1 *and* 3.

In the event, a retention rate of 2/3 was recommended to the Museum, i.e. one-third of the sample would be replaced at each survey, so that the selected objects would be surveyed on three occasions each (except for those 'dropping out' after the first or second survey).

I had not appreciated at that time (1991) the complexities that this would bring about in the estimation of transition probabilities after the first survey. Since the priority of each object in the current survey is known from both the census and the current survey, transition probabilities from the census to the current survey can be estimated without difficulty. But since one-third of the sample in the current survey did not participate in the previous survey, estimating transition probabilities from one survey to the next is more difficult.

The approach suggested at the time was to divide the sample into a 'matched' part (observed in the current and the previous survey) and an 'unmatched' part (observed for the first time since the census in the current survey), denoted by suffices u and m respectively. Transition probabilities between the k th and l th surveys are denoted by $\boldsymbol{P}(k, l)$, and the census is called survey 0. I suggested forming one estimate from the matched part:

$${}_m\hat{\boldsymbol{N}}'(t) = \hat{\boldsymbol{N}}'(t-1)\, \hat{\boldsymbol{P}}(t-1, t)$$

and one from the unmatched part:

$${}_u\hat{\boldsymbol{N}}'(t) = \boldsymbol{N}'(0)\, \hat{\boldsymbol{P}}(0, t)$$

These could be combined by weighting them according to the inverses of their variances (a standard variance-minimising technique). Revised estimates of $\boldsymbol{P}$ could then be obtained from the combined estimates of $\boldsymbol{N}(t)$.

The estimation of the transition probabilities from such data has been approached more thoroughly by Klotz and Sharples (1994). In a remarkably parallel study (the development of coronary disease in cardiac transplantation patients), they show that maximum-likelihood estimators of the transition probabilities can be obtained, but only by iterative methods (Newton-Raphson approximation).

A more practical problem is that the transition probabilities may well change from one survey to the next. Indeed, we hope they will change (for the better), as this indicates improvements in the management of the condition of the collection. Therefore, only the matched sample should be used in estimating current transition probabilities, since including the unmatched sample may bias the outcome. From this it follows that the matched sample should be as large as possible, say 4/5 of the total sample, rather than 2/3 as recommended above. I would be reluctant to recommend retaining the entire sample for each successive survey, unless there were plans to hold a census at regular intervals (e.g. 5- or 10-yearly).

As mentioned above, it was decided to use stratified sampling with the groups as strata. This raises the question of 'optimum allocation': should the same proportion of each stratum be chosen for the survey, or could better results be obtained by choosing different proportions?

The question of optimum allocation when sampling for proportions has been discussed by Cochran (1963: 106-109). Since transition probabilities relate to proportions of objects in a priority that change to another priority, this is a useful approach. He concludes that there is little difference in precision between optimum and proportional allocation unless the proportions are (a) very small (e.g. ≤5%) and (b) vary widely from one stratum to another (e.g. from 0.1% to 5%), and that 'the simplicity and the self-weighting feature of proportional allocation more than compensates for the slight loss in precision' (*ibid*.: 109). Elsewhere he comments that 'The simplicity and self-weighting feature of proportional allocation are probably worth a 10-to-20% increase in variance' (*ibid*.: 102).

The calculation of optimum allocation would be very difficult in our situation, as we are sampling for several proportions (not just one) which are weighted in a complicated way. Also, there is no *a priori* evidence of large systematic differences between strata (although they may be revealed as work progresses). The simple approach of proportional allocation was therefore recommended.

The recommendation might have been different for a museum with a predominance of ceramic and/or stone objects in its collections. Many such objects, unless in a weakened state on arrival, are unlikely to suffer deterioration other than from mechanical damage or a general storeroom disaster. They could therefore be sampled less intensively than more vulnerable objects, either by using a smaller sampling fraction or perhaps by sampling less frequently.

5.2 Sampling – practical issues

Theoretical considerations are only part of the story. The design of a sampling scheme must also take account of the fact that it will be undertaken by museum staff, or possibly temporary staff on short-term contract, who cannot be expected to have any statistical expertise. This means that any scheme should be as simple as possible, and appear straightforward and reasonable to the user. It should also be designed so that the analysis is straightforward. These points reinforce the decision to use the same sampling fraction in all strata (proportional allocation).

They also point towards a scheme of systematic sampling in each stratum, as was used in the Museum of London survey, with simple instructions for the replacement of a proportion of the sample at each survey.

The design was presented to the Museum as a 'rotating panel', selected systematically. The selected objects were to be numbered 1, 2, 3, 1, …, as they were selected, so that after the first year all the '1s' would be replaced, the next year all the '2s', and so on. Replacement would be by the next object at the same location; if the last object were to be replaced, it would be by the first. This approach would maintain the systematic nature of the same and make its implementation simple.

5.3 Bulk sampling

The strata which have been identified as having 'bulk' records (see above) have to be treated differently, both for selection and estimation. The practical problem is that it is not reasonable to expect a surveyor to remember which of a tray of (say) 200 bird eggs were in which condition at the census. The suggested solution was to treat the 'unit' (i.e. whatever grouping of objects had been entered on one line of the census form) as the unit of sampling, instead of the individual object. Systematic sampling would be used to 'select' an object, but the entire unit to which it belonged would then be sampled for the sample. This is the technique known as sampling 'with probability proportional to size' (i.e. of the unit), abbreviated to pps (*ibid*.: 308).

6 Estimation

The formulae used for estimating numbers currently in each priority, and the transition probabilities, were given above. However, they should not be presented to museum staff in this form. Ideally, specialist software covering sample design, selection, data input and analysis, should be provided, analogous to the Rothamsted General Survey Program (Anon 1989). Neither the time nor the resources were available for this task, so a spreadsheet was designed for calculating numbers in each priority, their standard deviations, and the transition probabilities. A second spreadsheet was needed to perform the calculations for the bulk samples, because they require rather different calculations.

The use of these spreadsheets has not been tested; for reasons given above I would now place more emphasis on the short-term transition probabilities and in detecting trends in them.

7 Conclusions

Statistical sampling techniques have potentially an even greater role in monitoring changes in the condition of museum collections than they do in establishing the conditions at a point in time, because the scale of resources that can be devoted to a 'one-off' census is not likely to be available on a regular (e.g. annual) basis. Modelling the varying conditions of a collection can help in the design of regular surveys, as well as suggesting novel statistics which may be of use for management or political purposes. Statistical nicety needs to be tempered with practicability to achieve a design which is reasonably efficient and which can be implemented by staff whose expertise lies elsewhere.

Acknowledgements
I am grateful to Louise Bacon of the Horniman Museum for introducing me to this problem, for discussing practical and theoretical issues with me, and for making data available. Elaine Sansom of UCL Institute of Archaeology made some constructive comments on a first draft of this paper.

references

Anon. 1989 *The Rothamsted General Survey Program (RSGP)*. Rothamsted, Herts: Lawes Agricultural Trust.

Cochran, W.G. 1963 *Sampling Techniques*. 2nd edition. New York: John Wiley.

Cox, D.R., H.D. Miller 1965 *The Theory of Stochastic Processes*. London: Methuen.

Keene, S., C. Orton 1992 Measuring the condition of museum collections. In: G. Lock/J. Moffett (eds), *Computer Applications and Quantitative Methods in Archaeology 1991*, 163-166, BAR International Series 577, Oxford: Tempus Reparatum.

Leese, M.N., S.M. Bradley 1995 Conservation condition surveys at the British Museum. In: J. Huggett/N. Ryan, (eds), *Computer Applications and Quantitative Methods in Archaeology 1994*, 81-86, BAR International Series 600, Oxford: Tempus Reparatum.

Klotz, J.H., L.D. Sharples 1994 Estimation for a markov heart transplant model, *Journal of the Royal Statistical Society D* 43 (3), 431-438.

Walker, K., L. Bacon 1987 *A Condition Survey of Specimens in the Horniman Museum*, 337-341. London: Summer Schools Press.

Clive Orton
UCL Institute of Archaeology
31-34 Gordon Square
London WC1H 0PY
United Kingdom
e-mail: tcfa002@ucl.ac.uk