Event History Analysis In Demography

Daniel Courgeau
and
Éva Lelièvre

CLARENDON PRESS · OXFORD
1992

Contents

Introduction

Up to the present, longitudinal analysis has basically been developed through an approach that takes each demographic phenomenon into consideration separately. Its main objective has been to isolate each phenomenon in its 'pure state'. In particular, it was considered necessary to separate the effects of each demographic variable from those of the others, such as mortality and migrations. To this end, a certain number of hypotheses were required, but they could not be tested against existing sources (Henry, 1959, 1966).

Longitudinal analysis was developed on the basis of aggregate data such as registration statistics, or, whenever possible, data from population registers. Even though these sources allow each event to be dealt with separately, since they eliminate disturbing effects, they hardly make it possible to analyse interactions between different phenomena. This explains why, in traditional demographic manuals, we find isolated phenomena in their 'pure state' presented as the subject matter of separate chapters: nuptiality, fertility, mortality, moves and migrations (Pressat, 1961; Henry, 1972).

None the less, certain demographers have pointed out the usefulness of analysing interactions between demographic phenomena. Pressat (1966) emphasized that 'the search for correlations between demographic phenomena, even though this domain remains unexplored, should enable us to deepen our knowledge considerably'. He did not, however, give any indication of what method to follow. Similarly, Henry (1972), in his analysis of nuptiality, stated that, 'in the case of out-migrants, one might be tempted to substitute their nuptiality abroad for that which they would have experienced had they stayed in their original country; yet nuptiality in a foreign country depends on conditions that may differ considerably'. This is a definite recognition of the interaction between the two phenomena and the change in an individual's nuptial behaviour following his out-migration. As data were lacking, however, Henry did not pursue this analysis any further.

Another important problem is that of heterogeneity. This point was also touched on by Henry (1959). Although, 'in the case of a homogeneous cohort, the statistical history of the component individuals is the same as the cohort's statistical history', this result no longer holds true when a heterogeneous cohort is being studied. Thus, for example, in the simplest of cases, where two sub-populations each have a constant though different probability of occurrence of the studied event over the period, the overall population will no longer have a constant rate, taken as an average of the two sub-populations' rates. It may well be the case at the outset, but with time the sub-population with the highest hazard rate will be eliminated from the population at risk by a selection process. This means that after a while the hazard rate of the observed population will converge with that of the sub-population with the lowest hazard rate. This heterogeneity may, of course, be of a much more complex nature and would need to be studied further.

> A precise knowledge of the practical implications of heterogeneity in human groups would necessitate further differential demographic research into the individual's physical and psychological characteristics, in order to study both the dispersion and correlation of the intra-group demographic indices which have so far been studied in a rather cursory fashion. (Henry, 1959)

As long as demographers use statistics such as those published in registration records or population registers, they have no way of dealing with the two basic problems: the analysis of interactions between demographic phenomena, and the analysis of heterogeneity in human groups. Other sources must now be used in order to observe a group of individuals over their entire lifetime, or at least part of it, as well as to collect a greater number of characteristics for each respondent.

It can thus be seen that the unit of analysis will no longer be the event (death, marriage, birth, migration, etc.): instead, each individual biography will be viewed as a more complex process. The question is no longer one of trying to isolate each phenomenon in its 'pure state': on the contrary, we must try and see how one event in an individual's existence can influence his life-course, and how certain characteristics can induce an individual to adopt behaviour patterns that are different from those of another individual.

This change of view leads us to reformulate the fundamental notions of demographic analysis in terms of complex stochastic processes. Let us take a more detailed look at how this is to be done.

Demographic processes do not occur in an abstract space–time, but originate within a given social structure. Someone born into a Lobi tribe in the early twentieth century will have quite a different biography from someone born into rural France of the same period, or someone born into present-day urban France. In each of these social structures, however, it is possible to distinguish relational systems that have developed to a greater or lesser extent depending on the group or the society concerned: 'family, economic, political, religious, educational, associational and informal systems' (Kimball and Pearsall, 1954). There is, of course, nothing to prevent new types of relational systems from appearing in the future. Our approach does not consider a society as a closed entity, but rather as one in constant evolution.

Each member of a given society is simultaneously involved in the various systems. For example, someone living in present-day France may be part of the family system as a member of an unmarried couple and father of a child; in the economic system as an engineer in the car industry; in the political system as a town councillor; in the religious system as a non-practising Catholic; in the educational system as a recipient of professional training; in the associational system as an amateur footballer; and, finally, in the informal system through his occasional attendance at parent–teacher meetings to solve his child's educational problems.

It is this interaction between the different types of involvement that creates a space and time specific to each situation. The geographical or occupational mobility of a single person may be much more frequent and may occur over greater distances than that of a married person, especially if the latter has one or more children. The married person is naturally tied to his place of residence and work by constraints related to his spouse's work-place, the location of his children's school, etc.

Event history analysis will thus attempt to place these changes in the time and space of an individual's life, in his social context. The point is to see how an event of a family, economic or other nature experienced by the individual will change the probability of other events happening to him over his lifetime. We shall, for instance, try to discover how his marriage can influence his professional career, his spatial mobility and other occurrences, such as the birth of a child or a break with his original family ties.

Here we are directly concerned with the analysis of interactions between demographic phenomena, the utility of which we have

already pointed out. This method of analysis has its place in the study of event histories.

Similarly, when trying to understand an individual's behaviour, one must take into account his social origins and his entire past history. In this case we are supposing that behaviour patterns are not innate but rather that they can change over an individual's lifetime as a result of what he experiences and acquires with time. Thus, two individuals from the same social background, but who have taken entirely different paths in life, can have attitudes to marriage, forming a family, career, etc., that diverge increasingly as time goes on.

We, thus, arrive at a method of analysis of population heterogeneity which uses a dynamic rather than static approach and which, accordingly, has its place in the study of individual event histories.

It is important to note that this analysis of population heterogeneity is not deterministic, but basically probabilistic.

Consequently, a large number of individuals who find themselves in the same conflictual situation at the outset will have different probabilities of finding solutions to the situation before a given date. Some will never find solutions; others may invent for themselves an entirely new pattern of behaviour which can provide a better solution to their conflictual situation. We are, thus, allowing the individual a margin of freedom which may lead to entirely new situations. Such a margin of freedom is of course essential, as no chain of events is predestined, but evolves with the course of time.

On this point, we are very close to Prigogine and Stengers (1988), who recognize that 'the event creates a difference between the past and the future . . . It is the intelligible outcome of a past, from which, however, it could not be deduced. It opens out an historic future where the insignificance or meaning of its consequences will be decided.' Here, one finds that same margin of freedom which can lead to entirely new situations.

After this informal introduction to event history analysis, can we now proceed to describe it in more formal terms?

When an individual is born, his life can follow a wide variety of paths. These different life-courses, however, are far from being equally probable. An individual's event history can therefore be defined as the result of a complex stochastic process, which develops over time, yet is situated within given historical, geographical, economic, and social conditions.

Let Ω_θ represent the set of event histories or partial event histories which can be observed up to a time θ. As already stated, our observation must be limited to the past, since the future brings into play new situations that cannot be deduced from the past. For example, the career of garage-owner could not have been envisaged before the appearance of motor cars, and in the near future genetic discoveries may stretch the span of human life to 150 years or more. The analysis carried out at time t only takes into account past behaviour patterns and projects them into the future without introducing elements affecting their evolution. This evolution, however, takes place at a slow enough rhythm to enable the analysis of the past to enlighten us as to the probability of various events occurring in the near future.

Let ω_n denote an entirely observed event history, where n precedes θ, and ω_θ an event history observed up to the moment θ, which is not finished. It can be said that the events in either of these histories are variables defined in the general space of Ω_θ. For example, the age at which one of these individuals marries in an *application* of Ω_θ on $[0, +\infty]$.

Now, let us give Ω_θ with a sigma-algebra[1] $\mathcal{B}_\theta$ of events that are to be analysed within a given population, together with a probability measure P_θ of $\mathcal{B}_\theta$ on [0,1], which assigns the different probabilities to the different events occurring within the observed population. In this case, $(\Omega_\theta, \mathcal{B}_\theta, P_\theta)$ does indeed define a probability space.

A random moment T will thus be a function of time on the Ω_θ probability space, which can in this case be extended beyond θ, supposing that the probabilities defined before θ remain the same over time. Thus, for example, if an individual is not married in θ, we may suppose that his age at marriage is a random variable T following the same distribution function as that observed for the individuals already married who have similar characteristics. Besides, this hypothesis is sometimes unnecessary, and one can simply work on event histories that are completely terminated, as is done in historical demography.

The method of event history analysis we present here will thus involve estimating the probability distribution of the life-courses followed by a given population. This distribution may vary from one sub-population to another and may depend on certain characteristics of the individuals in the sub-population (social and economic characteristics of parents or grandparents, for example). These life-courses

[1] The set of the Ω_θ parts.

feature random variables $T_1, T_2, \ldots, T_n$ which represent the duration of stay in the different states that constitute them. Of course, these variables are not independent, and the distribution of life-courses to be estimated is the result of their joint distribution.

Our approach, therefore, supposes that individual behaviour can be described as a complex stochastic process. Having once assumed this model, we shall begin with a statistical estimation of the distribution of the variables previously defined which we develop in the present work. Once known, these distributions make it possible to deal with the more complex distribution of the overall life-course.

We shall first examine methods for collecting these event histories. Generally, it is not possible to have an exhaustive collection of biographies. More often, one works with data on partially observed event histories collected from a sample of individuals. It is therefore necessary to examine the different problems raised by such incomplete observation.

We shall then proceed to formalize the methods of analysis and estimation, starting with the most simple case and gradually introducing an increasing degree of complexity. After studying an event, we shall move on to the reciprocal study of the interactions between two events, before extending these non-parametric models to cover more complex situations. Next, we shall examine parametric models which make it possible to introduce the effect of a large number of characteristics on the duration of stay in a given state. Finally, we shall deal with semi-parametric models which combine the two preceding approaches.

These various methods will be illustrated by applications to very different situations, so as to show the possibility for their generalization. In the Appendix, we shall indicate the programmes for carrying out these analyses, so that the reader may actually make use of them.

This book provides both a detailed theoretical presentation of methods for event history analysis and a practical application of these methods to files that exist already or that are to be created on the basis of event history surveys.

Subject Index

Author Index