

Classifying criminal activity: A latent class approach to longitudinal event data

Brian Francis

Centre for Applied Statistics, Fylde College, Lancaster University, Lancaster LA1 4YF, UK
B.Francis@Lancaster.ac.uk

Fulvia Pennoni

Dipartimento di Statistica Applicata G. Parenti, Università degli Studi di Firenze, Firenze, Italy,
Pennoni@ds.unifi.it

This paper is concerned with classifying longitudinal event data for a set of individuals. However, rather than classifying *individuals*, we are interested in identifying homogeneous periods of activity within such a longitudinal time series, and in classifying these periods of activity. Such data can arise in many circumstances. For example, a psychological classroom-based observational study of children may identify types of core behaviour (staring out of window, tapping foot, listening intently) and the interest would be in which of these co-occur at the same time. In such a study we would wish to build behavioural typologies, and identify sequences of behaviour and the amount of time the child spends in each state.

In this study we are concerned instead with criminal behaviour for *males* measured through official criminal histories. We use the Offenders Index, which is a large official database of criminal convictions in England and Wales of all offenders since 1963. An anonymised subset of this dataset is publicly available, and we analyse a fixed birth cohort of a one in thirteen sample of all offenders born in 1953. The database contains information on criminal history of each of the 10,000 male offenders, including the dates of conviction, the offence code of the conviction and the disposal or sentence. We simplify the data, reducing the offence codes to 73 major offences, after combining categories and eliminating offences with less than ten occurrences in the whole cohort (Francis et al. 2004). The problem is then to identify which offences co-occur at the same period or age in the criminal history of an offender, to identify the number and nature of these activity groups, and to examine transition from one activity group to another. For example, a simplified criminal history of a typical offender is shown below:

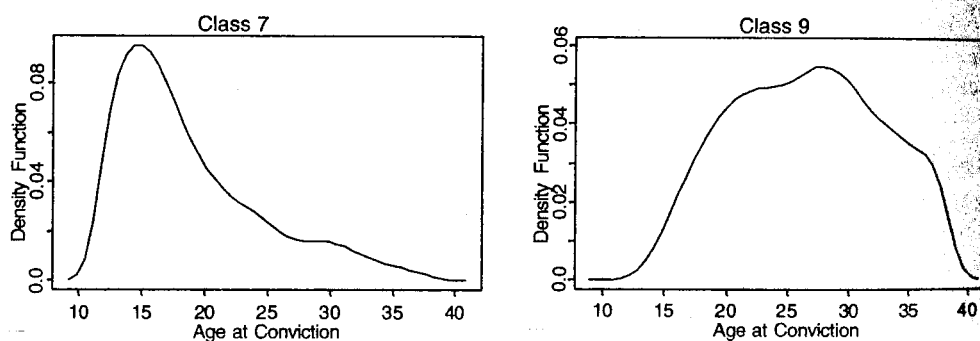
| age | 14 | 17 | 20 | 22 |
|----------|------------------|----------------------------------|-----------------------|--|
| Offences | Bicycle stealing | Shoplifting; Offensive weapon | Fraud; Petty theft | Fraud; Petty theft; Receiving stolen property |

We would like, for example, to determine whether bicycle stealing and shoplifting tend to co-occur in this cohort, whether fraud and receiving stolen property co-occur, and at what ages these offences are most prevalent. This study is primarily concerned with the nature and variety of offending, and so we form a prevalence vector to represent whether a conviction occurred for each of the 73 offence codes in a fixed period.

Our aim is to produce a set of classes which represent distinct types of co-offending activity. One possible way of proceeding is to examine the offence prevalence vector at each age, and to carry out a clustering procedure on all active age profiles for all offenders. However, it is unlikely that a true picture of an offending profile can be obtained purely from a single year, as there is substantial year to year variation. We therefore widen the period of examination, and take a fixed window width of h years, centring the windows on age a , and then repeat over a grid of values of a . We construct prevalence vectors for all windows with some offending activity. Every conviction

will thus contribute to h prevalence vectors. The choice of h is crucial, and h can be viewed as a smoothing parameter on the analysis. Increasing the value of h will decrease the ability of the model to detect changes over time, until, when h approaches the age range, we are analysing, we are analysing individuals rather than periods within individuals. Too small a value of h will introduce substantial noise into the transitions of an offender from one class to another. The window can be viewed as a uniform kernel density; other kernels can be used but make little sense in this context as we are constructing prevalence vectors. We use latent class analysis, a model-based clustering procedure (Hagenaars et al, 2002) as this procedure assigns windows to classes with estimated probabilities and changes in these probabilities over time will be of interest when assessing transitions of offenders from one offending class to another.

Initial results are encouraging. Using a window size of $h = 5$ years, we identify 11 offending classes. For every age a , we can then identify the number of offenders classified into that offending class, and graphing this as gives us a picture of changing activity over time. The graphs below show the results for two of the classes. Class 7 is characterised by a very high probability (0.9999) of breaking into shops and commercial property, with smaller probabilities for petty theft (0.19) and burglary (0.08). Class 9 is characterised by a very high probability of fraud offences (0.9997) with petty theft (0.27) and receiving stolen goods (0.14) also contributing. The figures indicate the age profiles of these two classes in the form of densities. Class 7 is predominately a young person's activity, with high numbers of offenders between ages 12-20. Class 9, in contrast, is for older males, with peak activity not reached until age 28. Additional results will be presented at the meeting.



REFERENCES

- Francis, B., Soothill, K. and Fligelstone, R. (2004) Identifying patterns of offending behaviour: A new approach to typologies of crime. Accepted for *European Journal of Criminology* 1 (1).
- Hagenaars, J. A. and McCutcheon, A. J. (2002). *Applied Latent Class Analysis*. Cambridge University Press: Cambridge.

RESUME

Dans cette article ont été étudiées les histoires complètes des transgressions commises par les membres d'une cohorte de dix mille contrevenants en Angleterre et dans le Pays de Galles et le problème de classifier leurs périodes de contravention active a été adressé. Pour tout individu parmi les contrevenants on peut prévoir une progression d'étapes à travers plusieurs types d'activités criminelles au cours de leur carrière. Pour les contrevenants de chaque âge nous construisons une fenêtre d'observation temporelle ayant h années chacune de long et au centre de laquelle se place l'âge de l'individu. L'activité criminelle qui se présente à l'intérieur d'une telle fenêtre se résume au moyen d'une rangée de variables binaires indicatrices, chacune liée à une des nombreuses types de contravention. Une analyse par classes latentes de toute la rangée de fenêtres nous fournit une classification des types de délits qui ont lieu simultanément.