Classifying Criminal Activity: A latent class approach to longitudinal event data

BRIAN FRANCIS, Centre for Applied Statistics, Lancaster University, UK

FULVIA PENNONI, Dipartimento di Statistica, Università degli studi di Firenze, Italia
BACKGROUND

- This work arises out of a practical problem in criminology – that of classifying criminal behaviour.

- Early work in criminology in the 1970s attempted to classify a criminal – thus an offender might be judged a professional fringe violator, or a robber, or trickster (Gibbons 1972, Prentice Hall). An offender then has that label throughout their career. However, very little of this work was based on real life data, and difficult to classify new offenders to a class.

- We adopt a developmental approach – can we identify types of criminal activity in distinct age regions of an individual’s history?

- Allows the development of an offender from one crime type to another. Criminologically, follows approach of Sampson and Laub (1993, Harvard UP) of pathways through crime.
Criminal histories and Criminal careers

Need to analyse the criminal histories of offending for a collection of offenders.

There are two possible routes:

**Self report studies** – ask individuals about their offending behaviour, and track them through time in a prospective longitudinal panel survey.

Problem is that attrition causes the most active offenders to drop out of the panel survey. Also telescoping – remembering offending events at the wrong time. Offenders are also not truthful.

**Official histories**
Larger sample sizes are possible. Analysis based on arrests (US, German studies) or convictions (UK studies). However unconvicted offending is not analysed.
The data set

We use the England and Wales Offenders Index – a Home Office research data set, which is a court based record of the criminal histories of all offenders in England and Wales from 1963 to the current day.

The complete data set is rarely analysed. We analyse data from the Offenders Index Cohort study, taking all those born in 1953, and followed through to 1993.

This birth cohort is an approximate 1 in 13 sample of all offenders born in 1953, and looks at all offenders born in four selected weeks in 1953.

9,234 offenders (81.0%) are male and 2,168 (19.0%) are female.

The index stores dates of conviction, the offence code of the conviction and the disposal or sentence.

The Home Office estimate that one in three males born in England and Wales in 1953 will have a criminal record and be on the Offenders Index by the age of 40.
Problems with the data set.

♦ It does not contain information on death, or immigration, or emigration. An individual might have left the country (perhaps to Scotland), but this would be viewed as a period of not offending in the dataset.

♦ The dataset is formed by record matching, taking court records and matching them on name and data of birth to form criminal histories. Although this procedure compares well with police records (Francis and Crosland, 2002; Home Office) it can introduce inaccuracies.

♦ It does not contain all offences, but only standard list offences – minor offending such as speeding and public order offences are omitted.
A typical criminal history

A simplified criminal history of a typical male offender is shown below:

<table>
<thead>
<tr>
<th>age</th>
<th>14</th>
<th>17</th>
<th>20</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offences</td>
<td>Bicycle stealing</td>
<td>Shoplifting; Carrying offensive weapon</td>
<td>Fraud; Petty theft</td>
<td>Fraud; Petty theft; Receiving stolen property</td>
</tr>
</tbody>
</table>

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.

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 EuroJCrim).

We simplify the time axis, using age in completed years rather than working in continuous time, and analyse males and females separately.
**Methodology**

We are concerned with offending patterns in fairly short time windows. We will be defining a time window or **offence strip** of size $h$ years and examining offending within that window.

We are concerned with the **nature and variety** of offending, so within the window we define an **prevalence matrix** $O_{ij}$

$$O_{ij} = 1 \quad \text{if offence } j \text{ is convicted for offender } i \text{ within the offence strip or window}$$

$$O_{ij} = 0 \quad \text{otherwise.}$$

We exclude strips with no offences and with all $O_{ij} = 0$

Thus we have a set of 73 binary indicators for every individual. **Latent Class analysis** can then then used to determine the number of classes for any fixed time window.
So, for $h=5$, we can for example look at offending between ages 18 and 22, centred on age 20.

<table>
<thead>
<tr>
<th>age</th>
<th>14</th>
<th>17</th>
<th>20</th>
<th>22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offences</td>
<td>Bicycle stealing</td>
<td>Shoplifting; Carrying offensive weapon</td>
<td>Fraud; Petty theft</td>
<td>Fraud; Petty theft; Receiving stolen property</td>
</tr>
</tbody>
</table>
Latent Class Analysis

For fixed window size and position, we define $O_i$ to be the prevalence vector for offender $i$ over the 73 offences. Assume there are $K$ classes, with $k=1 \ldots K$.

Let $\pi(k)$ be the probability of membership of class $k$, and $p_{jk}$ the probability that there is at least one offence of type $j$ given that the offender belongs to class $k$.

Then the likelihood is

$$L = f(O) = \prod_i \sum_k \pi(k)p(O_i | k)$$

where

$$p(O_i | k) = \prod_j p_{ij}^{O_{ij}} (1 - p_{jk})^{1-O_{ij}}$$

We now use and modify the concept of local likelihood to analyse the data for all possible windows of size $h$. 

Brian Francis and Fulvia Pennoni  54th ISI, Berlin, 15th August, 2003
Local Likelihood estimation

Proposed by Tibshirani and Hastie (1987, JASA). Idea is to produce maximum likelihood estimation in a simple linear model

$$\beta_0 + \beta_1 x$$

in a symmetric neighbourhood of the covariate $x$ – in a window around each $x$ value.

Produces a smooth estimate $s(x)$ of the effect of $x$ rather than an estimate of $\beta_1$.

The likelihood for the model then sums over all neighbourhoods.

Can be thought of as similar to a uniform kernel smoother applied to the data. Other forms of kernel smoothing can however be applied.
Latent Class analysis with local likelihood.

We wish to fit a model where the latent classes are estimated globally over all data points, but the data points now represent local events in the neighbourhood of age \(a\). The posterior probability of cluster membership will now vary by age.

We extend the definition of the prevalence matrix to be \(O_{ija}\)

\[ O_{ija} = \begin{cases} 1 & \text{if offender } i \text{ is convicted for offence } j \text{ within the offence strip } a- \text{ the window of width } h \text{ years centred on age } a \\ 0 & \text{otherwise.} \end{cases} \]

With \(k\) classes, the latent class model then becomes:

\[
L = f(O) = \prod_{i} \prod_{a} \sum_{k} \pi_{k} p(O_{ia} | k) \]

where

\[
p(O_{ia} | k) = \prod_{j} p_{jk}^{O_{ija}} (1 - p_{jk})^{1-O_{ija}}
\]
Posterior probability of cluster membership of strip

\[
q_{ika} = \frac{\pi(k) \prod_j (p_{jk})^{O_{ija}} (1 - p_{jk})^{1-O_{ija}}}{\sum_{k=1}^{K} \pi(k) \prod_j (p_{jk})^{O_{ija}} (1 - p_{jk})^{1-O_{ija}}}
\]

We sum the \( q_{ika} \) over \( i \) to produce expected numbers of individuals and probability profiles for each cluster at each age.
Results (h=5 years)

**Males:** 11 cluster solution based on using BIC and classification error.

Some clusters are **single offence** - high $p_{jk}$ for one offence and low $p_{jk}$ for remainder.

- Cluster 2 – petty theft
- Cluster 5: Wounding/violence
- Cluster 6: shoplifting.

Other clusters are **multiple offence**:

- 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 age profiles for each cluster can be seen to vary dramatically. Cluster 7 has high activity between 12-20, whereas Cluster 9 does not reach a peak until age 28.
Results (h=5 years)

Females: 6 cluster solution

Largest cluster is shoplifting (38%). Single offence cluster with no other offences contributing.

Other large clusters:

Cluster 2: Fraud and receiving stolen goods (25%)
Cluster 3: Petty theft. (14%)
Cluster 4: Violence, assault and criminal damage. (14%)
Choice of window size.

With $h$ large, the model degenerates to a single latent class analysis for all the data.

As $h$ becomes smaller, the latent classes more fully represent local behaviour in a small age range.

With $h$ very small, the latent classes will represent individual patterns at a particular event date.

Choice of $h=5$ is reasonable. Other $h$ values have been investigated, for example $h=11$ produces fewer ‘single offence’ clusters. Choice of window size is perhaps an expert judgement.
Discussion and Conclusions

• Results from latent class analysis are useful for understanding criminal behaviour. It is possible to examine transitions between clusters, to look at the number of clusters an individual pathway encounters and much more.

• More to do – need to examine changes over time – perhaps typologies to not remain constant. Are typologies for a 1983 birth cohort the same as those for a 1953 birth cohort?

• Need to look at both quantity and quality of offending, and to consider non-uniform kernel functions.

• Generally, the method gives a way of forming typologies of other forms of event history data, for example childhood behaviour.

• There are links to latent transition analysis (Collins & Wulgalter, 1992 MultBehRes), but LTA works in discrete time (stages) rather than in continuous time.