**Estimation of population sizes using truncated Poisson models**

By Peter G.M. van der Heijden and Maarten Cruijff

Registration files can be used to generate a list of individuals from some population. If each time that an observation of a population member occurs is registered but, for one reason or another, some population members are not observed at all, the list will be incomplete and will show only part of the population. In this project we work on methods to estimate the size of a population, and its characteristics in terms of a number of covariates, from incomplete registration lists.

We give a few examples. Consider the estimation of the number of illegal immigrants in the Netherlands from police records. These records contain information on the number of times each illegal immigrant was apprehended by the police and they are incomplete since the illegal immigrants who were never apprehended do not appear in them (Van der Heijden et al., 2003a). Two other examples from the field of criminology that our method was applied to are the estimation of the size of the population of drunken drivers and illegal gun owners in the Netherlands based on police records on the number of apprehended individuals (Van der Heijden et al., 2003b). One could also think of non-criminal applications, such as the estimation of the number of individuals with a certain illness from a registration of doctor visits, the number of researchers working in a particular research area from a registration of published papers, the number of research groups working in a particular area from a registration of patents, or the number of potential clients in a hotel chain from a registration of visits of clients. All these applications have in common that, due to the nature of the registration data, a zero count cannot be observed and the data are truncated.

The truncated Poisson regression model is used to arrive at point and interval estimates of the size of a populations. The dependent variable is a count variable describing how often someone appears in the registration. The population size estimates are derived assuming that each count is a realization of a Poisson distribution, and that the Poisson parameters are related to covariates through the truncated Poisson regression model. These assumptions are discussed in detail in (Van der Heijden et al., 2003b).

The approach taken to estimate the population size is this. Say someone’s Poisson parameter specifies his probability of appearing in the registration at least once at .25, nonetheless he is appears in the registration. This implies that, for this individual, we estimate that there are three other individuals who do not appear in the registration. By performing this trick for every individual in the registration, and adding up all individual estimates, we obtain an estimate of the total number of individuals who are member of the population of interest, but are not in the registration.

Currently we study the Zelterman estimator, that only makes use of the individuals seen once or twice. Most applications we have worked on are in the field of criminology (illegal fire arms, drunk driving), illegal immigrants and public health (opiate users). This project has received funding from the Home Office and the Ministry of Justice.

**Publications in Dutch**

(you can contact Peter van der Heijden for pdf-files):


Publications in English
(you can contact Peter van der Heijden for pdf-files):

