Accident Analysis and Prevention 111 (2018) 147–154

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Identifying traffic accident black spots with Poisson-Tweedie models

Birgit Debrabant[a], Ulrich Halekoh[a,*], Wagner Hugo Bonat[b], Dennis L. Hansen[c,d], Jacob Hjelmborg[a], Jens Lauritsen[c,d]

[a] Department of Public Health, University of Southern Denmark, Odense, Denmark
[b] Department of Statistics, Paraná Federal University, Curitiba, Brazil
[c] Accident Analysis Group, Department of Ortopedics, Odense University Hospital, Odense, Denmark
[d] Department of Clinical Medicine, University of Southern Denmark, Odense, Denmark

ARTICLE INFO

Keywords:
Black spot detection
Traffic accidents
Hospital admission data
Poisson–Tweedie distribution

ABSTRACT

This paper aims at the identification of black spots for traffic accidents, i.e. locations with accident counts beyond what is usual for similar locations, using spatially and temporally aggregated hospital records from Funen, Denmark. Specifically, we apply an autoregressive Poisson–Tweedie model, which covers a wide range of discrete distributions and handles zero-inflation as well as overdispersion. The estimated power parameter of the model was 1.6 ($SE = 0.06$) suggesting a distribution close to the Pólya-Aeppli distribution. We identified nine black spots consistently standing out in all six considered calendar years and calculated by simulations a probability of $p = 0.03$ for these to be chance findings. Altogether, our results recommend these sites for further investigation and suggest that our simple approach could play a role in future area based traffic accident prevention planning.

1. Introduction

We present a case study of black spot detection for traffic accidents, based on six years of hospital admissions data for traffic accidents on the island of Funen, Denmark. The main goal of black spot detection is to identify specific sites, e.g. intersections or road segments, as candidates for traffic safety improvements.

This is an active area of research, see e.g. Thomas and DeRobertis (2013), De Pauw et al. (2014), Vandenbulcke et al. (2014). The concern for traffic accident prevention stems from the fact that traffic accidents are estimated to be the eighth leading cause of death at the moment and are predicted to be the third leading cause of death by 2030 (WHO, 2013).

The data for the present study originated from records of all traffic-related injuries in the Funen region for the period 2002–2007, using hospital admissions data from all three hospitals. No study has yet been done in Denmark using this kind of data, and previous decisions regarding traffic safety improvements have been based on accident records by the police. Although hospital data do not contain those accidents where only material damage occurred, police records, on the other hand, tend to substantially under-represent vulnerable road users such as pedestrians and cyclists.

It has also been documented (Lauritsen et al., 2002) that for the region covered (Funen) more than 90% of treatment costs as well as societal costs after person injury is covered by those patients seeking treatment at the hospital. In general police records only cover 15–18% as seen since the mid-1980s after traffic accidents based on direct coupling at person level of police and hospital records (see www.ouh.dk/uag). This suggests that hospital records give a fuller picture of the health care-related consequences of traffic injuries.

The purpose of this article was to develop a simple yet sufficiently flexible statistical method suited to our dataset for the identification of black spots, the latter being locations with higher accident rates than expected given characteristics of the location and its neighbourhood.

A wide variety of statistical distributions and methods has been proposed for analysing traffic accident count data. Common distributions include Poisson and negative binomial distributions, Poisson-lognormal distributions as well as zero-inflated Poisson and negative binomial distributions, which have been adopted by, e.g. Jovanis and Li Chang (1986), Joshua and Garber (1990), Miaou and Lum (1993), Miaou (1994,1994), Maycock and Hall (1984), Turner and Nicholson (1998), Amoros et al. (2003), Cafiso et al. (2010), Miaou et al. (2005), Lord and Miranda-Moreno (2008), Aguero-Valverde and Jovanis (2008), Lord et al. (2005) and Lord et al. (2007). Random effects can be used to take correlations among observations as well as unobserved heterogeneity into account, see, e.g. Shankar et al. (1998), Miaou et al. (2003), El-Basyouny and Sayed (2009), Venkataraman et al. (2013) and Barua et al. (2015). Further modern modeling strategies, which have

* Corresponding author.
E-mail address: uhalekoh@health.sdu.dk (U. Halekoh).

https://doi.org/10.1016/j.aap.2017.11.021
Received 10 April 2017; Received in revised form 20 September 2017; Accepted 16 November 2017

been applied to accident data, include latent-class (finite mixture) models (e.g. Park and Lord, 2009; Buddhavarapu et al., 2016; Heydari et al., 2017), Markov switching count models (e.g. Malyshkina and Mannering, 2009, 2010), hierarchical models (e.g. Jones and Jørgensen, 2003; Kim et al., 2007; Dupont et al., 2013), multivariate models (e.g. Miaou and Lord, 2003; Depaire et al., 2008; Dong et al., 2014; Heydari et al., 2017), Bayesian methods (e.g. Li et al., 2007; Elvik, 2008; Pei et al., 2011) and neural networks (Zeng et al., 2016). For a more complete overview of models used in accident research and application studies we refer the reader to Lord and Mannering (2010) and Mannering et al. (2016).

In this article, we develop a spatial autoregressive model for accident counts aggregated to squares of size 1 km^2. We use the family of extended Poisson–Tweedie distributions (Bonat et al., 2017), which provides a flexible class of models to deal with under-, equi- and overdispersed count data as well as highly skewed count data with excessive zeros as usual in traffic accidents applications. Poisson–Tweedie distributions include the Neyman Type A, Pólya-Aeppli, negative binomial and Poisson inverse-Gaussian distributions as special cases.

The dataset is presented in more detail in Section 2, Section 3 introduces Poisson–Tweedie distributions, the statistical model and elaborates on our definition of black spots. Results are given in Section 4 followed by a discussion in Section 5. In Appendix A we give a detailed description of our simulations and in Appendix B a computer code for fitting our proposed model is given.

2. Description of data

The data were collected by the Accident Analysis Group (Hansen and Lauritsen, 2008) at hospitals located on Funen, Denmark, in the period from 2002 to 2007 (Fig. 1).

Each patient reporting at a hospital as having been involved in a traffic accident was asked several questions regarding the accident location and other relevant information. For the analysis we used only accidents for which a location could be related to a house number or an intersection and we confined us to traffic accidents which occurred on public roads.

We covered Funen with a grid of 1 km^2 squares defined by the UTM coordinates (UTM zone 31N, WGS84). The injury data was quality assured and aggregated to the grid as described in Hansen and Lauritsen (2008). The quality assurance excluded hospital contacts with uncertain location (e.g. not on a regular road), imprecise geocoding (e.g. "somewhere on a 20 km long road") or at locations only occurring partially over the years. Among a total of 27,957 verified traffic accidents on public roads 13,924 (50%) could be located with a precision which allowed allocation to a given 1x1 km square and therefore inclusion in the analysis. The traffic accidents in the following analysis are the sum of the accidents in each square for each year. Using data essentially accumulated in grid cells allowed to circumvent the difficult task to relate single accident locations to specific intersections. For other aspects related to the use of grids we refer to Xie et al. (2017).

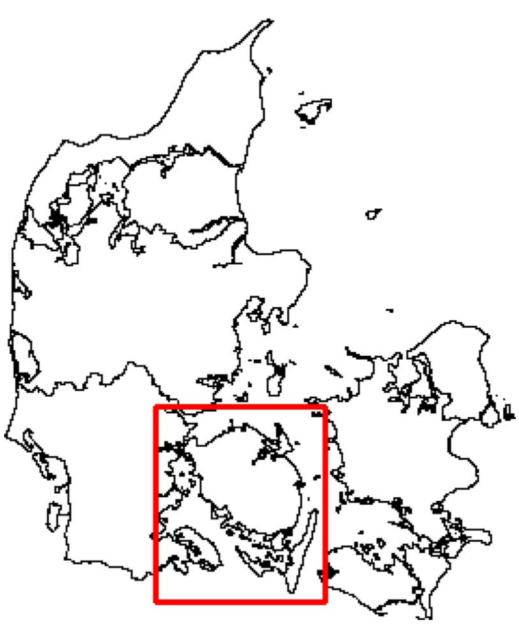

Fig. 1. Denmark with the island Funen.

Fig. 2 shows the number of accidents for the first year (2002) and the average of the year totals over the six years 2002–2007. In 2002 there was a total of 2145 traffic accidents and the average of the year totals was 2321. In 2002 in 3335 (85.3%) of the 3911 squares no accident was reported and over all six years 2427 (62.3%) locations had no reported accident.

The number of intersections and the street-length in a square are used as risk indicators for traffic accidents at locations. These values are shown in Fig. 3 and are assumed to be constant over the six years. Our data did not contain more detailed exposure data, such as accurate traffic intensity records, nor more precise information about local risk factors such as the geometry and capacity of intersections.

3. Statistical model and definition of black spots

Accident counts are known for exhibiting overdispersion and zero-inflation relative to the Poisson distribution (Lord and Mannering, 2010). However, given the wealth of discrete distributions, it is difficult to commit oneself to a single distributional model as being the most appropriate one. Therefore, and in order to take the mentioned features into account in a more unified manner, we consider the broader class of Poisson–Tweedie mixture distributions. A distribution from this family of discrete distributions (see Jørgensen and Kokonendji, 2016 for a formal definition) is specified by three parameters μ, τ and p. Here, $\mu > 0$ denotes the mean, $\tau > 0$ the dispersion and $p \geq 1$ the shape/power parameter. The variance is given by $\mu + \tau \cdot \mu^p$ and τ larger than zero indicates overdispersion. The family of Poisson–Tweedie distributions allows for zero-inflation and can further be extended to incorporate underdispersed count data with nonnegative dispersion τ, see Bonat et al. (2017) and Bonat (2016, 2017). For $p = 1$, $p = 1.5$, $p = 2$ and $p = 3$ the Poisson–Tweedie distribution respectively corresponds to Neyman type A, Pólya-Aeppli, negative binomial and Poisson-inverse Gaussian/Sichel distribution, see Kokonendji et al. (2004), all of which are well-known distributions in accident modelling (Kemp, 1967; Minkova and Balakrishnan, 2014; Özel and İnal, 2010; Lord and Mannering, 2010; Zha et al., 2016). Since the class of extended Poisson–Tweedie distributions comprises major families of distributions used for traffic modelling and is additionally richer than each single of these families alone, we consider it well-suited for our purposes.

In the sequel we denote by Y_{it}, $i = 1, \ldots, 3911$, $t = 2002, \ldots, 2007$, the number of accident counts at location i in year t and consider the following auto-regressive model containing the number of accidents from neighbouring locations, the calendar year, the street length and number of intersections as covariates:

Y_{it} is Poisson–Tweedie distributed with dispersion τ and power p.
Its mean value μ_{it} is given by

$$\log(\mu_{\mathrm{it}}) = m_t + \sum_{d=1}^{D} a_d \log(\overline{Y}_{i,t}^{(d)} + 0.02) + b\log(S_i + 0.5) + c\log(L_i), \tag{1}$$

where S_i and L_i are the number of intersections and street length in location i, $\overline{Y}_{i,t}^{(d)}$ denotes the average accident count at time t over all neighbouring locations of cell i at distance d, and $D \in \{1, 2, \ldots\}$ is the maximum distance considered. As distance measure we use the supremum norm between the square centres. The set of all neighbouring