Analytic Methods in Accident Research 12 (2016) 48–65

Contents lists available at ScienceDirect

Analytic Methods in Accident Research

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

The Palm distribution of traffic conditions and its application to accident risk assessment

Ilkka Norros[a,*], Pirkko Kuusela[a], Satu Innamaa[a], Eetu Pilli-Sihvola[b], Riikka Rajamäki[b]

[a] VTT Technical Research Centre of Finland Ltd, Finland
[b] Finnish Transport Safety Agency Trafi, Helsinki, Finland

ARTICLE INFO

Keywords:
Palm distribution
Traffic accident risk
Traffic condition
Weather condition
Statistical testing

ABSTRACT

We introduce a method for assessing the influence of various road, weather and traffic conditions on traffic accidents. The idea is to contrast the distribution of conditions as seen by the driver involved in an accident with their distribution as seen by an arbitrary driver. The latter is considered as a variant of the notion of Palm probability of a point process, and it is easy to compute when road, weather and traffic measurement data are available. The method includes straightforward assessment of the statistical significance of the findings. We then study a single large example case, Ring-road I in Helsinki observed over five years, and present a comprehensive analysis of the influence of traffic, road and weather conditions on traffic accidents. Our results are in line with existing knowledge; for example, the traffic volume as such has hardly any influence on accidents, whereas the afternoon rush hours are considerably more risky than the morning ones, and heavy rain and snowfall as well as reduced visibility in general increase the accident risk substantially. The notion of Palm probability offers a transparent and uniform approach to such questions, and the proposed approach can be applied as a semi-automatic risk assessment tool prior to deeper analyses.

1. Introduction

Traffic accident records provide usually a long list of conditions and characteristics prevailing at the time of the accident: time, place and its features, weather, road surface etc. Another group of attributes describes the accident itself, including the casualties. From a data set consisting of accident records, one obtains an empirical distribution of conditions 'seen by the accidents', which we shall in this paper call the *accident distribution*. The significance of a statistical analysis of the accident distribution alone is limited, because as such it gives no information about the normal state of the world outside the accident times.

It has of course been always understood that traffic accidents only happen when there is traffic, and therefore a major factor of traffic accident frequency is their *exposure*, i.e. the traffic itself. As an example already twenty years old, Fridstrøm et al. (1995) found that the exposure and sheer randomness together accounted for 80–90% of the accidents in the Nordic countries. In their times, the exposure could be measured only by indirect means, and they inferred it from the fuel consumption that was recorded at county level. The notion of exposure is fundamental also, e.g., in Nilsson (2004).

With the emerging Internet of Things, roads will be more and more accurately monitored for traffic, weather and surface

* Corresponding author.
E-mail addresses: ilkka.norros@vtt.fi (I. Norros), pirkko.kuusela@vtt.fi (P. Kuusela), satu.innamaa@vtt.fi (S. Innamaa), eetu.pilli-sihvola@trafi.fi (E. Pilli-Sihvola), riikka.rajamaki@trafi.fi (R. Rajamäki).

http://dx.doi.org/10.1016/j.amar.2016.10.002
Received 30 March 2016; Received in revised form 4 October 2016; Accepted 4 October 2016
2213-6657/

conditions. The high-dimensional variability of all these conditions can be described in terms of an empirical distribution computed from the Big Data provided by the monitoring systems. For the needs of accident analysis however, an empirical distribution computed as a time average is not the best one to describe the wanted 'normal world outside the accidents'. We propose that the most relevant description is obtained by weighing the time series data by traffic density, so that the distribution of conditions corresponds to that obtained by picking a vehicle uniformly randomly from the spatio-temporal observation domain and recording the conditions seen by that driver in that moment. We adopt for this distribution the term *Palm distribution* of conditions.

Palm probability plays an important role in the theory of random point processes, where it expresses the distribution of the world from the viewpoint of an arbitrary point of the point process. Our adaptation of this concept is the distribution of road and traffic conditions seen by an arbitrary driver. Indeed, if each driver had a low constant intensity of causing an accident, then the accident distribution would, in long run, equal the Palm distribution.

The idea to apply a variant of Palm probability in analysing different traffic and road weather conditions under which the accident risk is increased was created and piloted in a recent research project (Innamaa et al., 2013). The recognition that weighing the conditions by traffic density presents an instance of a Palm probability seems to be new, and we believe that the this theoretical notion helps significantly in the formulation of statistical problems and results.

The main contributions of this paper are (i) the explicit conceptualization of the Palm distribution in the vehicular traffic context, and (ii) that a lot of information on the influence of various conditions on accidents can be obtained simply by comparing the accident distribution of conditions with their Palm distribution. Further, we realise that in order to make a single high-dimensional Palm distribution manageable and comparable with the accident distribution, (iii) each variable has to be quantized to a sufficiently coarse granularity. Finally, (iv) we demonstrate our approach by a single but extensive case study of the Ring-road I in Helsinki observed over five years.

This paper is structured as follows. Section 2 first defines our notion of Palm probability in an abstract fashion and then shows how it can be instrumentalized for practical computations based on extensive traffic, road and weather data. Section 3 introduces our accident data and presents the statistical methods we apply. The results on accident risk on Ring-road I are presented in Section 4. We discuss the results along their presentation and relate them to the literature. A general discussion of the method, including its extension perspectives as well as some already recognized defects and challenges, is given in the concluding Section 5.

2. Palm distribution of road traffic conditions

2.1. Palm probability

Consider a road section R of length L during a time interval $[0, T]$. We focus on the simplest case where we abstract from all internal structure of R. The rationale of this assumption is that the traffic on R is observed at a single station, typically through inductive loops. Let vehicles arrive to R according to a point process with time points A_n and counting process N_t such that

$$0 \le A_1 < A_2 < \cdots < A_{N_T} \le T. \tag{1}$$

Denote the speed of vehicle n by V_n, and assume it to be constant through the vehicle's stay in R, so that vehicle n leaves R at time

$$B_n = A_n + \frac{L}{V_n}. \tag{2}$$

The number M_t of vehicles in R at time $t \in [0, T]$ can then be expressed as

$$M_t = \sum_{n=1}^{N_t} 1_{\{B_n > t\}}. \tag{3}$$

In addition to speed, each vehicle n possesses another characteristic W_n, with values in some measurable space $\mathcal{W}$.

The last element of our framework is the process $(X_t)_{t\in[0,T]}$ of road conditions taking values in a measurable space $\mathcal{Z}$. Note that this abstract setup allows both W_n and X_t to be vector-valued.

Let us now define the *state* of the road section R at time t as

$$S_t = (\mu_t, X_t) = \left(\sum_{n=1}^{N_t} 1_{\{A_n \le t \le B_n\}}\,\delta_{(V_n, W_n)},\, X_t\right), \tag{4}$$

where $\delta_{(v,w)}$ denotes a Dirac measure at point $(v, w) \in [0, \infty) \times \mathcal{W}$, so that μ_t is an integer-valued measure on $[0, \infty) \times \mathcal{W}$. Note that

$$M_t = \mu_t([0, \infty) \times \mathcal{W}), \tag{5}$$

and the traffic density (vehicles per length unit) at time t is

$$D_t = \frac{M_t}{L}. \tag{6}$$

Thus, the state consists of the momentary traffic constitution, coded in μ_t, and the momentary road conditions, coded in X_t. We denote the space of all possible states as $\mathcal{S}$, and equip it with an appropriate σ-algebra.

In this modeling framework, the objects W_n and X_t can contain observable as well as unobservable elements. In the present