MATHEMATICAL MODELLING WITHIN THE ROAD SAFETY MANAGEMENT

Roman Ondrejka
Veký Diel 3323, Výskumný ústav dopravný, a.s., ondrejka@vud.sk
Lenka Moravčíková
Veký Diel 3323, Výskumný ústav dopravný, a.s., lenka.moravcikova@vud.sk

Keywords: modelling, road, safety, management
Abstract: The precondition for effective management of road safety on the part of the road administrators is a thorough and correct identification of sites with increased frequency of accidents on individual sections of the road network. The adequacy of identification results from the application of such specific methodologies, which will allow to reveal sites with a high rate of local risk factors with the potential to cause collisions in road traffic and thus appropriately allocate adequate funds to remove them. Besides the empirical statistical records, one of reliable methods for the identification of potential black spots is also the mathematical modelling of adverse events on individual road entities and their subsequent estimation in terms of probability of occurrence in the future. The application of the method described in the article allows, following the discovery of a causal link between the frequency of accidents and their consequences and possible risk factors concerning the nature of traffic, to identify potential black spots without the necessity of reporting high accident rate in the past.

1 Identification of sites within increased accident rate

Identification, analysis and evaluation of the black spots on the road network and the intervention through the adoption of appropriate measures are one of the key tasks of road managers in relation to the road safety improvement. On one hand, it is the responsibility of the manager to create conditions for reduction of the number of potentially negative events in the form of road accidents through the elimination local risk factors and on the other hand to create conditions locally in the form of adopting effective measures preventing the occurrence of these events with consequences for the life and health of road users [1,2].

Currently, there are several approaches in the area of determination of sites with increased frequency of accidents that vary in data, time and computational complexity as well as in the identification reliability. In general, these approaches can be divided into those, whose identification is based on empirical data on accidents from the past and those, whose identification on the site under consideration is based on the expected number of traffic accidents or their consequences at the given site in the future [2].

From a theoretical perspective, for the assessment of the suitability of the applied approach it is necessary to consider, whether the used method eliminates the influence of random fluctuations in the recorded number of road accidents or their consequences on the overall assessment of the section as a site with an increased occurrence of accidents or consequences in terms of fatalities or health damage. Elimination of this systematic error within the determination of black spots is possible only through targeted inquiries and demonstration of causal relationships between traffic accidents and local conditions. They are represented through a wide range of factors related both with the traffic itself and with the construction and technical conditions of considered road entity [3].

In case that the causal relationship of accidents with some indicators is identified on a specific road segment, resulting in higher expected number of accidents or their consequences as is the case at similar sites in the future, it is necessary in the next step of the management of the identified black spots to quantify these local risk factors and to take measures to improve the safety at the given site to an acceptable level expressed by a reasonable level of accidents or their consequences [2,3].

2 Application of mathematical modellingi n the identification process

Within the evaluation of the current safety level of individual road entities, i.e. road sections, intersections and level railway crossings, it is necessary to not only consider the recorded number of accidents, thus not only what happened on the entity in the past, but also of what will happen, i.e. the expected number of accidents or their consequences in the future. It is necessary to build upon this basic premise when deciding, what methods to apply in order to reliably identify sites with increased frequency of accidents.

Safety defined in this manner should be explicitly modelled mathematically, allowing the application of theoretical knowledge about the so-called Bayesian methods. These represent advanced mathematical - statistical evaluation methods, which are based on the basic premise that empirical data in the form of e.g. total number of accidents at a particular road section, or a number of accidents with certain consequences, or the number of persons with specific consequences for life and health are a discrete random variable with a value of an unknown parameter expressed by means of a probability distribution.

The value of an unknown parameter, a random variable, is actually based on the statistics of traffic accidents managed by the Presidium of the Police Force of the Slovak republic - Department of Traffic Police, presently through the Road Accidents Information System (ISDN). The main difference with the simple procedure of determining the critical value based on the cumulative probability of a certain level of significance, utilized presently for the identification of black spots by the Slovak Road Administration, is that the mentioned variable is acting as a dependent variable in a certain mathematical formulation of a set of independent variables.

The principle of modelling of safety on road entities, i.e. road sections with length of 0.5 or 1 km, or three-way or multi-way intersections, or level crossings of the road with the railways, should be based on two basic input assumptions. The first one is the frequency of accidents expected on the entities with the same characteristics, i.e. attributes that act as exogenous in developed models, i.e. explanatory variables. In case of road sections it is concerning for example the daily traffic volume, type of area (rural area, urban area), percentage of heavy vehicles, the vertical alignment slope, radius of horizontal curves, etc.

In the case of intersections it as also concerning the daily traffic volume, type of area, the number and hierarchy of gateways, share of vehicle types, type of traffic control (traffic rules, traffic signal system) and other. For railway crossings, such attributes include mainly the intensity of train traffic, traffic volume on intersecting road, security level, i.e. installed level crossing protection system (mechanical, lighting), number of tracks, etc.

The second, equally important prerequisite for modelling is the frequency of accidents on the assessed road section or intersection for a period of one, two or three years. The Decree of the Ministry of Transport, Construction and Regional Development No. 251 of 2011 also suggests considering all traffic accidents occurring on the relevant entity for the last three years. It also indicates the method of classification of identified black spots by:

- ✓ quantified social costs associated with accidents,
- ✓ severity of accidents,
- ✓ functional class and category of road,
- ✓ capacity of road and traffic volume,
- ✓ conclusions of Cost-Benefit Analysis of the countermeasure implementation.

The resulting estimate of safety is subsequently determined by the quality of the estimated mathematical model of independent variables in the form of a safety function, which is then used for calculation of the share of the modelled safety component in the total amount of the safety on a section, intersection or level railway crossing in the future, usually for the next three years.

3 Calibration of safety regression models

The starting point for determination of the current level of safety is the calibration of safety functions in a form of mathematical formulations of behavioural patterns of the dependent variable, in this case, the number of accidents on individual entities, the frequency distribution of which we assume has the character of a negative binomial distribution. This assumption can be verified by the calculation of a so-called empirical distribution compliance testing with estimated theoretical distribution at the selected level of statistical error. Since this variable will be modelled using regression, the validity of this assumption will be verified by the significance of the developed model, which requires the random variable represented by the number of traffic accidents to be distributed by the negative binomial distribution.

The general form of prediction models of accident rate can be expressed as a function of:

$$E(Y_i) = f\{x_i, \beta_j\} \qquad (1)$$

In our case, we consider the function in a form of a correlation of relative change in the number of road accidents from absolute change in explanatory risk factors. The form of the negative binomial regression model can then be written in the following matrix form:

$$\vec{y} = X\vec{\beta} + \vec{\varepsilon} \qquad (2)$$

where $\vec{\mathbf{y}}$ is a n-dimensional column vector of observations of endogenous variable, i.e. number of road accidents in the form of a logarithm on the entities i = 1,2,3, .. n, $\mathbf{X}$ is the matrix of observations of explanatory variables, i.e. individual characteristics of entities, the vector $\vec{\boldsymbol{\beta}}$ is a column vector of unknown parameters of the model and $\vec{\boldsymbol{\varepsilon}}$ is the n-dimensional vector of random disturbances that meet the assumption of distribution normality. Then the Log-linear model for the mean value η_i is as follows:

$$\ln(\eta_i) = x_i^T \beta + \varepsilon_i \qquad (3)$$

Random failure ε_i has an expected value of zero, constant variance and each is generated independently, which means that the residual variations are not mutually autocorrelated.

4 Estimation of model parameters by means of log-likelihood function

In the following interpretation we are going to omit the stochastic errors and work only with the deterministic form of the model in the following form:

$$E(Y_i \mid x_i) = \eta_i = \exp(\beta_0 + \sum_{j=1}^{n} x_{ij}\beta_j) \quad (4)$$

where $E(Y_i \mid x_i)$ is modelled mean value of the number of accidents expected on the entity i (concerning road sections per unit length) over a time period of model calibration, x_{ij} in a form $x_{i1}\ldots x_{in}$ are the values of characteristics, independent variables of individual entities within the same time period, β_j is the intercept or else a level constant, which may not be explicitly estimated in each model coefficients and coefficients $\beta_1\ldots\beta_n$ are parameters of explanatory variables estimated by the model.

For the negative binomial distribution with a mean value η_i and the dispersion parameter ϕ is the vector of unknown parameters β, mean values η_i, also known as $\hat{y}_i$, i.e. estimated values of the number of road accidents to the actual number of accidents y_i, together with dispersion parameter $\hat{\theta}$ obtained by maximizing the so-called log-likelihood function $\ell(\boldsymbol{\beta},\theta)$ in the form:

$$\ell(\boldsymbol{\beta},\theta) = \sum_i \left\{ \sum_{r=1}^{y_i-1} \log(1+\theta r) \right\} - y_i \log(a) - \log(y_i!) + y_i \log(\theta\eta_i) - (y_i + \theta^{-1}) \log(1+\theta\eta_i) \quad (5)$$

Maximum likelihood estimation $(\hat{\boldsymbol{\beta}},\hat{\theta})$ can be achieved by maximizing the function $\ell(\boldsymbol{\beta},\theta)$ with regard to β and θ. Then the associated relations are:

$$\frac{\partial \ell(\boldsymbol{\beta},\theta)}{\partial \beta_j} = \sum_i \frac{(y_i - \eta_i)x_{ij}}{1+\theta\eta_i} = 0 \quad (6)$$

for $j = 1,2,\ldots n$ and

$$\frac{\partial \ell(\boldsymbol{\beta},\theta)}{\partial \theta} = \sum_i \left\{ \sum_{r=1}^{y_i-1} \left(\frac{r}{1+\theta r} \right) \right\} + \theta^{-2} \log(1+\theta\eta_i) - \frac{(y_i + \theta^{-1})\eta_i}{1+\theta\eta_i} = 0 \quad (7)$$

By means of the maximum likelihood method the estimates $(\hat{\boldsymbol{\beta}},\hat{\theta})$ can be obtained simultaneously, whereby the mentioned computationally demanding process includes sequential iterations. In the first sequence at the initiation value θ i.e. $\theta_{(0)}$ through maximizing $\ell(\boldsymbol{\beta},\theta)$ with regard to β we get the estimate of β_1 In the second sequence at β fixed at value of β_1 the maximization of $\ell(\boldsymbol{\beta},\theta)$ with regard to θ we obtain the value $\theta_{(1)}$. Through the iterative procedure and cycling between the used fixed value θ and the used fixed value β we get by maximum likelihood estimates $(\hat{\boldsymbol{\beta}},\hat{\theta})$. Currently, this calculation is facilitated by computer equipment using appropriate mathematical software.

Obtained estimates of the parameter θ and of the parameter vector β need to be tested for the statistical significance at the respective selected level of significance. It is also necessary to verify the significance of the model as a whole through the Pearson χ^2 test at respective degrees of freedom for the selected statistical error. In case of compilation of multiple models, it is possible to assess obtained maximized values of credibility functions as well as auxiliary assessment criteria such as Akaike or Bayesian information criterion. We are not specifying the principle of the test of significance of variable parameters or the model as a whole, as it is often a part of the output sets of computer software products. The estimated equation, as a function of characteristic variables and some regression parameters then indicates the average frequency of accidents on entities of given type per calibration period.

The resulting safety function can be applied for the calculation of safety for a given set of entities with the same or similar characteristics. The calculated value in the form of the number of road accidents per one year is the mean value of the frequency on a given road section, intersection or railway crossing at individual values of input variables for each entity.

In this way we obtain deviation of the actual number of accidents from the modelled mean values, i.e. from the expected level of safety for entities with the same or similar characteristics. In the Bayesian estimation by the function of the frequency of accidents, the obtained difference represents the proportion of the local risk factors on the accident in the locality. By means of relevant mathematical and statistical formulas it is possible from the estimated frequency of accidents to express the expected proportion of accidents with consequences, and based on the estimated number of fatal, serious and minor injuries to express the amount of the expected social losses resulting from road accidents on individual road entities. After subsequent implementation of adequate measures from the side of the manager of the respective road aimed at the reduction of accidents, it is possible, after calculation of implementation costs, to express the effectiveness of invested public funds through the cost-benefit analysis.

5 Conclusions

The proposition of sophisticated methods and procedures applied in the area of road safety management increases the demands on the work procedures of road managers in the process of credible identification of sites with an increased occurrence of traffic accidents. The need for systemic change in performance of this process is highlighted not only in the mentioned Decree of the Ministry of Transport, Construction and Regional Development of the Slovak Republic No. 251/2011 on details concerning safety management on roads, but also in the EU Directive 2008/96/EC on road infrastructure safety management, regulating procedures within the implementation of safety inspections, where it recommends to take into account the attributes of roads as well. Currently, Výskumný ústav dopravný (Transport Research Institute, Inc.) develops a software solution for decision support for road managers in the process of identification of black spots, as well as in the process of implementation of the safety countermeasures. The purpose of the software tool is not only an address but mainly an effective use of funds allocated for the improvement of safety on road infrastructure in Slovakia.

References

[1] Machciník, Š. – Ondrejka, R. Návrh metodického pokynu pre riadenie a kontrolu bezpečnosti na pozemných komunikáciách (Draft methodological guideline for the road safety management and inspection) research project, Ministry of Transport, Construction and Regional Development of the Slovak Republic, final report 2012. (Original in Slovak)
[2] Decree of the Ministry of Transport, Construction and Regional Development of the Slovak Republic No. 251/2011, laying down details of the road safety management. (Original in Slovak)
[3] Directive 2008/96/EC of the European Parliament and of the Council on road infrastructure safety management (Original in English)
[4] Act No. 249/2011 Coll. on road infrastructure safety management and on amendments to some acts (Original in English)

Acknowledgement
The article was elaborated during the implementation of projects University science park ITMS 26220220184, funded within the Operational Programme Research and Development.

Review process
Single-blind peer reviewed process by two reviewers.