Journal of Traffic and Transportation Engineering 12 (2024) 81-91
doi: 10.17265/2328-2142/2024.02.003

Statistical Analysis of Accident Proneness of Drivers

S. V. Gerus

Kotelnikov Institute of Radioengineering and Electronics (Fryazino branch) of the RA Moscow Oblast 141190, Russia

Abstract: A mathematical model describing the risks of road accidents has been built on the basis of statistical data of drivers' accident rate. It has been revealed that drivers can be divided by the degree of their accident proneness into four categories with sharply differing probabilities of road accidents. It has been shown that there is a possibility of classification of drivers in accordance with specified categories.

Key words: Accident, probability, driver, category, classification, proneness.

1. Introduction

The work of an operator of technical means, which are potentially hazardous for human life and the environment, is increasingly supported by electronic systems. These, on the one hand, help the operator to exercise his functions, and, on the other one, ensure the required level of operability and prevent him from committing errors [1, 2]. A locomotive engineer, a car driver, a ship pilot, a dispatcher, an operator of a nuclear power plant—each of them can make a mistake, doze off, faint. To properly design electronic aids and operator's controllers, one must realize the level of risk of the subject to control, that is, the human being himself [3].

Measures suggested for reducing the accident rate which is determined by the so-called human factor are extremely diverse. These include a special psychological training for drivers and train operators, their professional selection, pre-trip control, etc. [4]. Yet, errors do occur, drivers fall asleep, accidents happen. To what extent is the operator prone to drowsiness and erroneous actions, how often does he commit them, and to what extent are people similar or different as far as emergency situations are concerned? These questions are tackled in a large number of publications (see, e.g., Refs. [5, 6]). As for motor traffic, the driver's behavior on the road is analyzed with consideration of his/her age, psychophysiological condition, peculiarities of social behavior, etc. [6]. Studies are either based on special laboratory experiments [2, 7], or on road traffic statistics [3, 8].

Here we are to analyze accident rate statistics of motor drivers. Such an approach would, with sufficiently complete and reliable data available, provide a generalized, averaged idea of the driver and his accident proneness. Such approach makes it possible to formulate the most adequate requirements to the equipment monitoring the driver's activities.

The statistical accident rate is generally established through comparison of empirically obtained distribution of RTAs (road traffic accidents) with Poisson distribution, or the negative binomial distribution [6], or by means of comparison and correlation of the number of RTA committed by a group of drivers over certain time periods [8].

Comparison of the nature of empirical distribution of RTA with Poisson distribution reveals considerable differences between them. According to Klebelsberg [6], it has turned out that empirical distributions of RTA somewhat better agree with the so-called negative binomial distribution than with the simple Poisson distribution. However, the choice of the most appropriate distribution has not been further developed so far.

Corresponding author: S. V. Gerus, research fields: numerical simulation.

Using correlation factors as a measure of linear relation between the RTA frequencies at different time periods also draw some criticism of Klebelsberg [6] and Kunkel [9] determined by non-linear relationship between accidents at various time periods. Besides, the correlation coefficients enable only to denote a problem, with specification of the number of drivers per a given number of RTA, while it remains impossible to classify drivers according to their RTA proneness. Nevertheless, according to Quenault and Fuhrmann [10], one can have a rather reliable way to classify drivers, e.g., by their road behavior and driving style, which ultimately affect the RTA rate.

2. Statistic Data

We are now to consider the problem of classification of drivers according to their proneness to accidents. To this effect, we are to use statistic data presented in a classic paper [8]. The drivers' accident frequency was studied in the state of North Carolina (USA) during two successive two-year periods. The obtained data are presented in Table 1.

Based on these results, the following conclusions have been drawn:

- 80% (2,002,577) of all the drivers were not involved in RTA either at the 1st, or at the 2nd period;
- the totality of all the RTA at the 1st and 2nd periods (629,765) accounts for 20% of all the drivers (499,663);
- 11% (267,663) of all the drivers with one or more RTAs at the first period account for 19% (63,053) of all the RTA at the 2nd period (324,017);
- 81% (260,964) of all the RTA at the 2nd period are accounted for drivers who also had an RTA at the 1st period;
- 0.2% (4 664) of all the drivers had 3 or more RTAs at the 1st period: this group accounts for 0.8% (2,483) of all the RTA at the 2nd period;
- RTA frequencies for the 2nd period increase with the number of accidents at the 1st period. This relation is non-linear.

In the paper [8], RTA frequencies of empirical data of Table 1 were simulated. Using a Monte-Carlo method, a bivariate distribution function symmetrical by lines and columns was selected, which would have marginal distributions (in Table 1, the second column and the bottom line) as close as possible to empirical ones. Correlation was found between the RTA frequencies at the first and second periods, and the RTA number was predicted for Spearman correlation coefficient ρ increasing from 0.01 to 0.76 (for coefficient $\rho = 0.11$ actually obtained for North Carolina).

The authors make a conclusion that the quality of RTA prediction is not very good, though it does improve with a growth of the value of ρ. At the same time, for coefficient $\rho = 0.19$ (which is nearly twice as much as for the one actually obtained in North Carolina), the prediction of getting into a hypothetical accident would come true only in 31% of cases.

Table 1 RTA of drivers during consecutive two-year periods.

Number of RTA in period one (m)	Distribution of drivers by the number of RTA in period one	Distribution of drivers by the number of RTA in period two						
		Number of RTA (n)						% of drivers involved in RTA
		0	1	2	3	4	5	
0	2,234,577	2,002,577	206,778	22,080	2,639	406	97	10
1	235,080	192,684	35,363	5,842	986	172	33	18
2	27,919	20,467	5,694	1,369	300	65	24	27
3	3,953	2,546	1,008	284	81	24	10	36
4	584	321	162	73	14	11	3	45
5 and more	127	65	36	13	8	2	3	49
Total	2,502,240	2,218,660	249,041	29,661	4,028	680	170	11

3. Data Analysis

Let us consider the data of paper [8]. As can be seen from Table 1:

(1) After the first two-year period, the drivers were found distributed into subgroups ranging from 0 to 5 (the first and second columns of Table 1). The number of drivers involved in accidents was equal to 267,663. Hence, the average percentage of drivers' involvement in RTA for all the groups was 267,663/2,502,240 = 10.7%. This is the average probability of getting into at least 1 accident over 2 years. It does not differ much from the probability of getting into just 1 RTA over the same period: 235,080/2,502,240 = 9.4%.

(2) Now, track the drivers after the second two-year period. Each of 6 subgroups again gives its distribution according to the number of RTA (columns 3-8 of Table 1). Yet, this time, these drivers are initially distributed into "very good", "good", etc. up to "very bad" ones. The percentage of these drivers involved in RTA is presented in the rightmost column of Table 1. It can be seen that the worse this subgroup showed itself during the 1st period, the higher the risk every driver from this subgroup has to get involved in an RTA over the same period. In essence, this is what constitutes the correlation of RTA frequencies of the first and second periods.

Based on the aforesaid, one can suppose that the overall number of drivers considered is divided into a number of categories of different intensity of being involved in RTA. Each category of drivers makes its contribution to 6 subgroups presented in column 1 of Table 1. Let us assume that the distribution of time intervals between RTA for each category is of an exponential nature, which is quite plausible due to the fact that:

(1) over a two-year interval, the processes may be considered stationary (probabilistic characteristics are time-independent);

(2) RTAs occur independently of each other (no after-effect);

(3) RTAs occur one by one, not in couples, triples, etc. (ordinariness condition). This condition is, generally speaking, not always observed as there occur accidents with a large number of participants. However, as we do not know the details of all the RTA, we will consider the events to be ordinariness.

Then each category is characterized by Poisson distribution describing the probability for a driver to get involved in RTA a number of times. For the above conditions, the risk for a randomly selected driver to be involved in m RTAs is calculated from the formula of total probability of incompatible events:

$$P_m = \sum_i C_i B_{im} \qquad (1)$$

where $B_{im} = \frac{a_i^m}{m!} e^{-a_i}$, $a_i = \lambda_i \tau$, $\sum_i C_i = 1$.

Here, i is index numbering the categories of drivers according to their psychophysiological proneness to RTA, m is index numbering subgroups of drivers by the number of RTAs they committed (m = 0, 1…5), C_i is probability of hypothesis of a driver's being referred to category i, B_{im} is probability (according to Poisson) for a driver of category i to commit m RTAs over the period τ = 2 years, λ_i is RTA intensity of drivers of category i. Theoretical RTA frequencies are found from the equation $K_m = N P_m$ and match experimental frequencies N_m stated in the 2nd column of Table 1, N = 2,502,240 is the total number of drivers.

Consider the evolution of distribution (Eq. (1)). Upon completion of the first period of the experiment, the drivers, as a result of the RTA that occurred, were found sorted into 6 subgroups. This means that, as a result of the above events, the probabilities of hypotheses C_i have changed. According to the Bayes' formula [11], for every m-th group, the new probabilities of hypotheses are:

$$\tilde{C}_{mj} = \frac{C_j B_{jm}}{\sum_i C_i B_{im}} \qquad (2)$$

Now, for a driver, provided that after the 1st period he was found in the *m*-th group, the probability to get involved in *n* RTA over the following 2 years is denoted as the formula below:

$$\tilde{P}_{mn} = \sum_i \tilde{C}_{mi} B_{in} \tag{3}$$

As the probability of the above condition is determined by Eq. (1), the total probability for the driver to be involved in *m* accidents at the first period, and in *n* accidents at the second one, and, therefore, to be found in a cell with indices *m*, *n* is as follows:

$$\tilde{\tilde{P}}_{mn} = P_m \tilde{P}_{mn} \equiv \sum_i C_i B_{in} B_{im} \tag{4}$$

As follows from Eq. (4), the obtained probability is symmetrical relative to permutation of indices; in other words, $\tilde{\tilde{P}}_{mn} = \tilde{\tilde{P}}_{nm}$, which is supported by statistical data. A part of Table 1 numbered with the number of RTA for *m* and *n* from 0 to 5 constitutes frequencies N_{mn} for drivers to be found in the above cells. It can be seen that the relation $N_{mn} = N_{nm}$ is observed with a good accuracy level, which suggests that the proposed model is consistent.

For model (Eq. (1)) to be substantiated, it was necessary to choose the parameters λ_i and C_i in such a way that theoretical frequencies $N\tilde{\tilde{P}}_{mn}$ and values of empirical frequencies N_{mn} of Table 1 would be as close as possible. The above parameters were calculated by the maximum likelihood method [12] based on a condition that the probability of a complete set of events, which actually occurred, must be the maximum of all the possible ones. The task assigned can be solved with such values of λ_i and C_i, when the maximum of the following likelihood function is obtained:

$$L = \sum_{m=0}^{5} \sum_{n=0}^{5} N_{mn} \ln \tilde{\tilde{P}}_{mn} \tag{5}$$

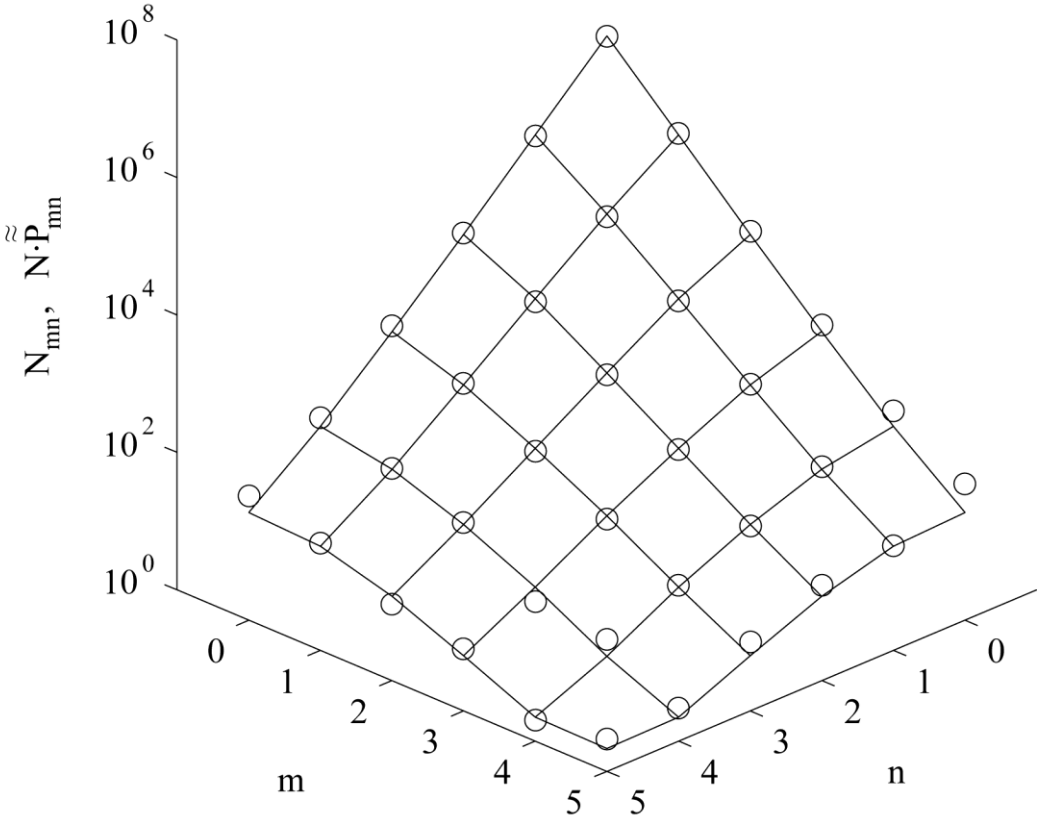

Fig. 1 Empirical N_{mn} (circles) and theoretical $N\tilde{\tilde{P}}_{mn}$ (grid nodes) drivers' accident frequencies.

Table 2 Calculated safety parameters for different categories of drivers.

Characteristics	Categories of drivers			
Number	1	2	3	4
λ_i (1/h)	3.5×10^{-6}	1.8×10^{-5}	6.1×10^{-5}	2.6×10^{-4}
C_i (%)	77	22	0.66	3.2×10^{-4}
Number of RTA D_i (%)	37.5	56.8	5.6	0.01

Fig. 1 presents, on a logarithmic scale, comparison of calculated theoretical frequencies with the data of Table 1. The graph shows a good coincidence of empirical and theoretical data, and demonstrates the above-mentioned symmetry of distribution relative to indices *m* and *n*. Based on results of calculations, there was also controlled coefficient of correlation ρ between frequencies of RTA at the first and second periods, for which purpose the values of theoretical and empirical frequencies were used as correlation tables. Thus, for the data of Table 1, the value of coefficient is as follows: ρ = 0.11 [8]. The value of correlation coefficient found from calculations of Eqs. (1)-(5) was found equal to 0.12, which also confirms the good accuracy of the model. The found parameters of the model are presented in Table 2.

4. Discussion of Results

As can be seen from Table 2, to describe empirical results presented in Table 1, it is sufficient to be limited in Eq. (1) to four categories of drivers, with the first two categories making the largest contribution to traffic safety. Thus, coefficient C_i showing the share of drivers of *i*-th category in the total number is the highest for the 1st category (C_1 = 77%). These are the safest drivers. They have the lowest RTA probability determined by parameter $\lambda_1 = 3.5\times10^{-6}$ 1/h. The fourth category is the most hazardous one; it has the RTA intensity $\lambda_4 = 2.6\times10^{-4}$ 1/h, though the share of such drivers is negligibly small: $C_4 = 3.2\times10^{-4}$%. The second and third categories hold the intermediate positions accounting for 22% and 0.66%, respectively.

Now, consider the number of RTA committed by drivers of different categories. The share of RTA of the *i*-th category of drivers is determined based on mathematical expectation of the number of events for Poisson distribution:

$$D_i = \lambda_i C_i / \sum_j \lambda_j C_j \tag{6}$$

According to Table 2, the second category drivers, though 3.5 times less populous than the first category drivers, cause 1.5 times more RTAs. Drivers of a small (0.66%) 3rd category cause nearly 6% of all the RTAs. This is due to the fact that, though the number of drivers C_i goes down with the number of category, the RTA intensity λ_i grows at a virtually geometric series. This is such a bidirectional pattern in variation of parameters that leads to the aforesaid effect.

The average intensity of RTA $\lambda_s = 7.2\times10^{-6}$ 1/hr is found directly from Table 1 through calculation of the number of RTA per driver in 1 h. For drivers of category *i*, probability B_{im} of committing *m* accidents over time interval *t* is calculated from Poisson equation:

$$B_{im} = \frac{(\lambda_i t)^m}{m!} e^{-\lambda_i t} \tag{7}$$

Therefore, it can be seen that the entire population of drivers may, to a sufficiently high accuracy, be divided into 4 categories, each distinguished for its proneness to RTA characterized by parameter λ_i. For drivers of different categories, probabilities p_i = 1-exp(-$\lambda_i t$) of committing at least one RTA over, e.g. a year, differ much and, respectively, are:

$$p_i = 0.03;\ 0.15;\ 0.4;\ 0.9\ (i = 1, 2 \ldots 4) \tag{8}$$

Note that the results obtained noticeably differ from conclusions drawn in the cited paper [8], according to which the bulk of RTA is committed by ordinary, "good" drivers (just making mistakes), while the share of "bad" drivers makes 0.2% and they commit 0.8% of all the RTAs. According to the model Eq. (1), the number of drivers' categories is more than two; respectively, the RTA distribution by categories is totally different.

5. Classification of Drivers

Next, a question arises on whether or not this model may be used as a basis for classifying the drivers one way or another, with each particular driver to be referred to a category. Attempts of such classification based on computer simulation of RTA and correlation of these RTAs for two observation periods were made by Campbell and Levine [8]. The classification was performed based on a fact of a driver's being referred to one of the cells of Table 1 over the first period. The forecast for this driver to get involved in an accident at the second period was based on whether the driver had more or less than two accidents over the 1st period. In the authors' opinion, the results of such classification did not prove to be very successful owing to a large percentage of errors; at the same time, availability of a wide range of drivers in the benchmark data prevented the forecast from reaching an acceptable reliability level.

In this paper, simulation of the entire population of drivers is exercised based on results of the both periods. Classification of a particular driver is exercised as by Campbell and Levine [8], based on the cells of Table 1 the driver found himself in. As distinct from Campbell and Levine [8], instead of using the mere notion of a "bad" or "good" driver, a mathematical assessment of the driver's quality is introduced (parameter λ_i) formed on the analysis of all the statistical data of Campbell and Levine [8]. In this case, not a randomly selected driver's risk of getting involved in RTA is forecast, as was in paper [8], but his belonging to the category in question (*i*); and then, based on parameters of this category, specified in Table 2, the driver's accident probability is determined. Moreover, the large amount of statistical data enables to hope that, to a certain extent, it will be possible to estimate representation of drivers of different categories in their totality.

Now, we are to look at what goes on with distribution of drivers by categories, as the drivers are systematized by subgroups following the first and second observation periods. In Table 3 and in Fig. 2 it is shown, how, after the first period, the share $\tilde{C}_{mj}$ of a category j changes depending on group m, where the drivers found themselves according to the number of RTAs committed by them.

In the group with the number of RTA equal to zero, distribution $\tilde{C}_{0j}$ changed insignificantly compared to initial C_i (Table 2)—the value of $\tilde{C}_{01}$ increased by 5%, whereas the share of others somewhat decreased. In other words, the outflow of "bad" drivers resulted in an increase in the share of "good" ones. In subsequent groups (m = 1, 2 … 5), coefficient $\tilde{C}_{m1}$ falls, coefficients $\tilde{C}_{m3}$ and $\tilde{C}_{m4}$ of "bad" drivers grow, and the share $\tilde{C}_{m2}$ of "medium" drivers first rises to 76% for index m = 2, and then declines. With the same index m = 2, coefficients $\tilde{C}_{m1}$ and $\tilde{C}_{m3}$ reach the value of 12%. It is worth noting that the overall number of participants of the m-th group, according to Eq. (1), would fall with the growth of m as $N \cdot P_m$, with coefficients $\tilde{C}_{mj}$ showing the relation between categories j for each subgroup of m only. The presented results enable to draw a conclusion that in subgroups with $m \geq 2$ the share of drivers of the first category constitutes a very small part, in subgroups with m = 2, 3 drivers of the second category prevail, while the third category drivers make a majority in subgroups with m = 4, 5. Therefore, right after the first observation period, the presented model makes it possible to divide drivers into categories according to their accident proneness.

Table 3 Distribution of drivers' categories (coefficient $\tilde{C}_{mj}$ %) after the first period.

	Category of driver			
Number of RTAs at the first period	1	2	3	4
0	82	18	0.2	0
1	45	52	2	0
2	12	76	12	0
3	2	65	33	0.04
4	0.2	37	63	0.3
5	0	13	83	4

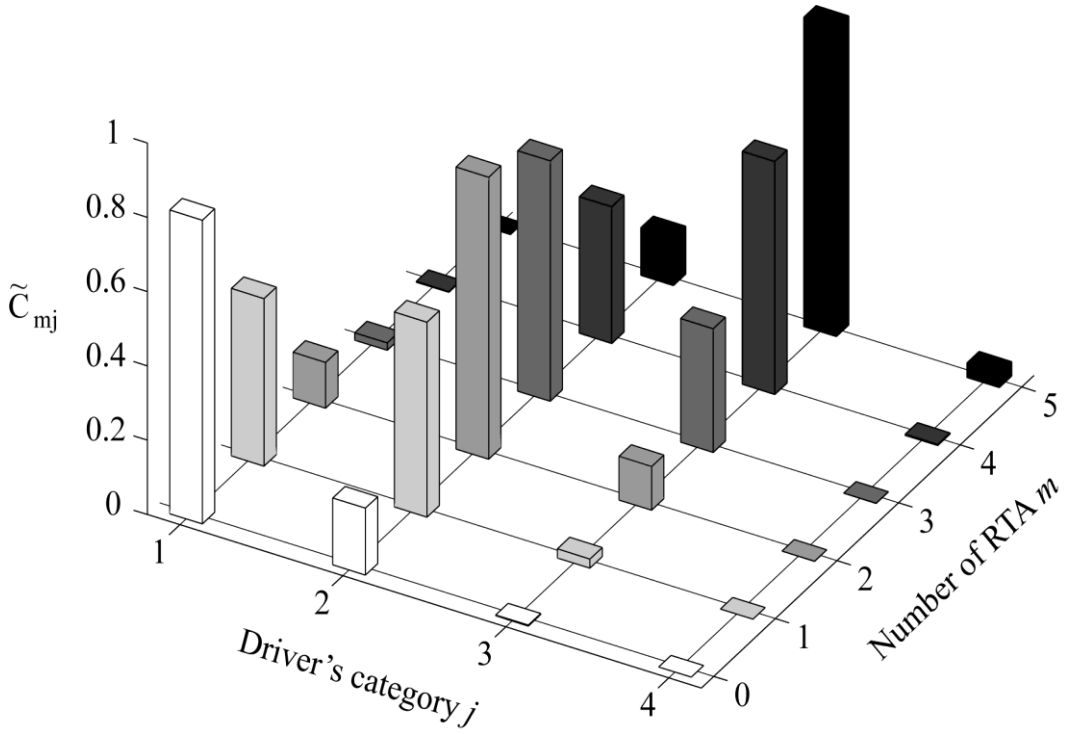

Fig. 2 Distribution of drivers' categories after the first period.

After the second observation period, the drivers underwent further segregation. Each driver found himself in one or another cell of Table 1, for which, according to Eqs. (3) and (4), there exists its own distribution function; i.e., for each cell, there are its own coefficients $\tilde{\tilde{C}}_{mnj}$, which determine the relations between the number of drivers of different categories. These coefficients are calculated from an equation similar to Eq. (2):

$$\tilde{\tilde{C}}_{mnj} = \frac{\tilde{C}_{mj} B_{jn}}{\sum_i \tilde{C}_{mi} B_{in}}, \qquad \tilde{\tilde{C}}_{mnj} = \tilde{\tilde{C}}_{nmj} \qquad (9)$$

These distributions are presented in Fig. 3. It is appropriate to recall than it is only possible to compare coefficients $\tilde{\tilde{C}}_{mnj}$ with different j, but the same values of m and n, as the cells differ in the number of drivers $N \cdot \tilde{\tilde{P}}_{mn}$ (also see Table 1). The drivers who got involved in no RTA at this period ($m = n = 0$) were distributed by categories as follows: $\tilde{\tilde{C}}_{00\,j}$ = 85%, 14%, 0.1% (j = 1, 2, 3). Compared to the first period, the share of the first category drivers increased by 3%, mostly at the expense of the second category drivers. The concentration of the second

category drivers reaches its maximum of 78% at two diagonals: $m+n = 2$ and $m+n = 3$. This is just 2% higher than the maximum of this category after the first period. The third category drivers prevail on diagonals with $m+n = 5, 6…9$, making 68% to 96% there, which is considerably higher than the maximum of the first period. Finally, the drivers of the fourth (the smallest) category were revealed very clearly (according to Table 1, there are just 3 of them) who, with a probability of 82%, would take a place in a corner cell ($m = n = 5$). The remaining 18% suppose that the drivers of this cell may be referred to the third category.

As filling of the cells would fall with a growth of indices m and n, to identify the maximum number of drivers, one must exercise classification for one category or another at the minimum possible values of m for the first period, and $m + n$ for the second one. To this effect, it is necessary to determine an acceptable classification error level R_0 and, based thereon, to select the cells of m and n, for which the probability of $\tilde{C}_{mj}$ or $\tilde{\tilde{C}}_{mnj}$ would exceed the value $1 - R_0$. The aforesaid applies to results of the both observation periods.

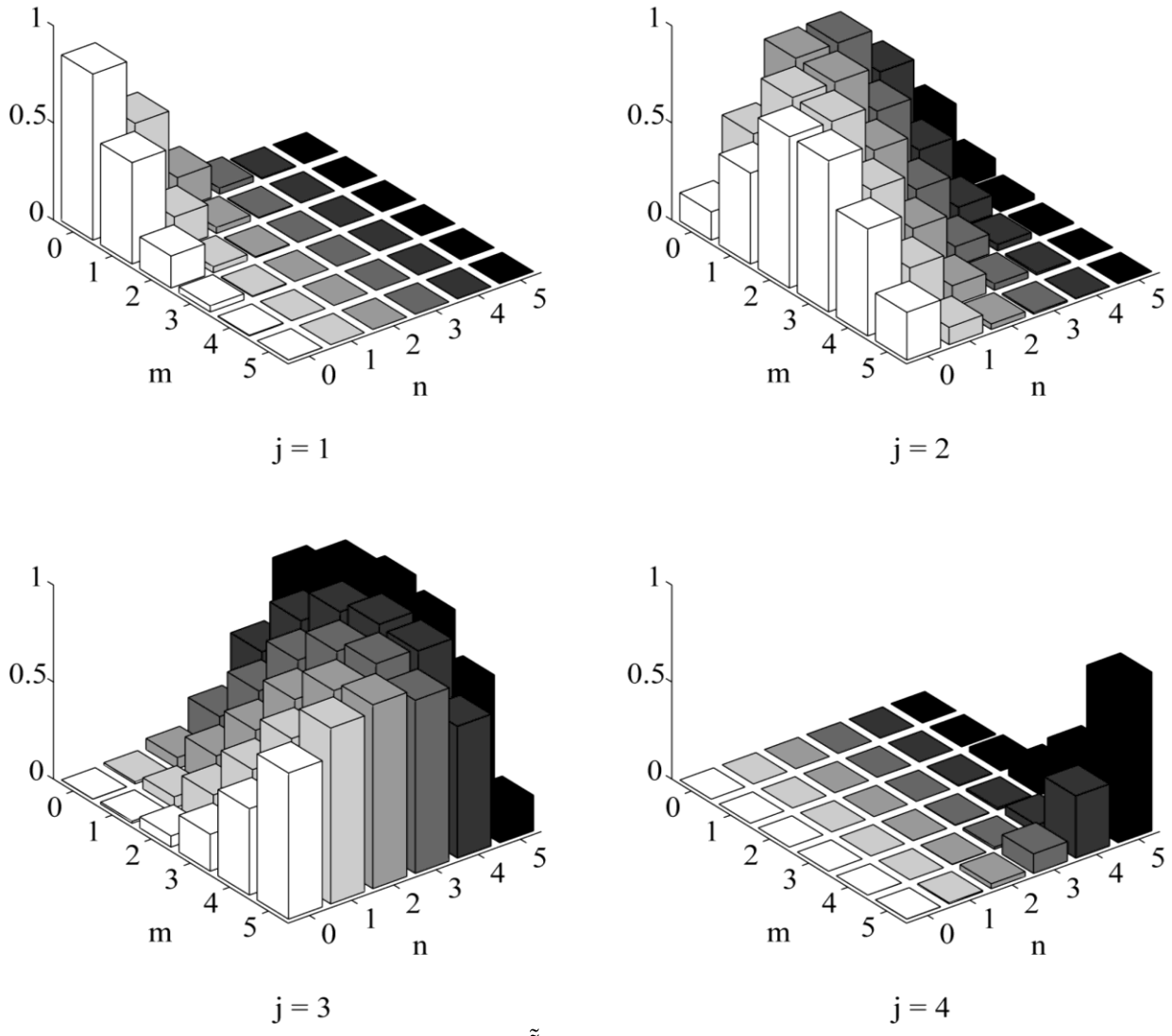

Fig. 3 Distribution of categories of drivers (coefficient $\tilde{\tilde{C}}_{mnj}$) after the second period. The values of indices *m* and *n* correspond to numbered line and column cells of Table 1. Index *j* specifies the categories of drivers.

E.g., choose the limiting level for classification error equal to R_0 = 27%, which, according to Campbell and Levine [8], is close to the acceptable level of 22%. Then, after the first period (Table 1, columns 1 and 2, and Table 3), we have the following classification results:

- 2,234,577 drivers from cell m = 0, with a probability of 82%, are referred to the first category;
- 27,919 drivers from cell m = 2, with a probability of 76%, are referred to the second category;
- 127 drivers from cell m = 5, with a probability of 83%, are referred to the third category;
- the fourth category is not identified following the first period results;
- unidentified are 239,617 people from cells with m = 1, 3, 4, which makes 9.6% of drivers.

Following the results of the second period (Table 1, columns 3-8, Fig. 3), the classification is as follows:

- 2,002,577 drivers from cell m = n = 0, with a probability of 85%, are referred to the first category;
- 94,631 drivers of diagonals m+n = 2, 3, with a probability of 78%, are referred to the second category;
- 1,472 drivers of diagonals m+n = 5, 6,…8, with probabilities of 73%-96%, are referred to the third category;
- 3 drivers from cell m = n = 5, with a probability of 82%, are referred to the fourth category;
- unidentified are 403,557 people from diagonals m+n = 1, 4, 9, which makes 16% of drivers. Virtually all the first category drivers, 17% of the second category drivers, 9% of the third category drivers and 38% of the fourth category drivers are identified with a specified margin of error.

It can be seen that after the second period, all the categories are determined with a higher probability, and the drivers of categories 2 to 4 are identified in larger numbers than over the first two-year period. After the second period, we got a seemingly paradoxical result: the number of unrecognized drivers compared with the first period was somewhat higher. However, precisely the more rigorous identification of drivers has led to this: previously not quite recognized drivers have been dropped out. The bulk of those unidentified is constituted by drivers with a single RTA, which is quite understandable as it is very difficult to provide identification based on a single episode. A relatively low percentage of identified drivers of the second and third categories is not due to shortcomings of this model; the point is that for such drivers, for a two-year period according to Eq. (8), the probabilities of committing RTAs are just 0.3 and 0.7, respectively. Most of these drivers left no information on them in the form of a number of RTAs larger than one and, therefore, were left unidentified. With the number or duration of periods increasing, the percentage of their identification must rise.

The following is to be said based on the second period. As distinct from calculation of the model parameters, classification of drivers does not require analysis of the driver's affiliation to this or that category in two periods. It is sufficient to classify the driver the way it was done above for the first period, but based on the number of RTAs committed over an entire time period. The results turn out to be virtually the same. This follows from the fact that the largest contribution to the classification at the second period was made by values of matrices $\tilde{\tilde{C}}_{mnj}$ lying on diagonals $m+n = const$. Therefore, affiliation of a driver to this category is determined based on Eq. (2) according to the number of RTAs committed by him over a sufficiently long time.

After the classification was performed, the forecast for each identified driver to get involved in RTA is provided by Eqs. (7) or (8).

6. Errors of Model

We are now to consider some inaccuracies of the model and the causes of the same. A discrepancy between the theoretical and empirical data is suggested by an approximately 10% deviation of correlation coefficients of the model (ρ = 0.12) and Table 1 (ρ = 0.11). More specifically, this inaccuracy is of about the same order as the asymmetry of initial distribution

noticeable in Fig. 1. This is caused by a number of reasons. First, a one-sided screening of samples of drivers is possible, with some drivers failing to proceed from the first period to the second one (lethal RTA, revocation of driving license, etc.), and this may lead to asymmetry of statistical data. Second, due to the lack of data, such a phenomenon was not taken into account as accidents involving more than one driver simultaneously, as well as the fact that one accident may entail a series of others. Third, it was supposed that there were just a few categories of drivers with fixed values of parameter λ_i, though, obviously, this parameter can, in its turn have a certain distribution function. However, in spite of the aforesaid inaccuracies, in our view, the model in question describes the available totality of statistical data well enough.

7. Summary

The model elaborated herein is based on statistical data obtained in the state of North Carolina (USA) in the 1970s. Strictly speaking, this is true for drivers of the above region and for the above time period. Yet, owing to a tremendous representativeness and a high quality of the data obtained, one may suppose that the model is, to some extent, of international nature. Of course, for other regions, the parameters of this model may vary. After all, these parameters reflect the general situation related to the road traffic safety for a more or less closed population of drivers. Categories and their parameters identified within the model are related to a large number of mutually affecting factors of personal and social life of drivers, the quality of their training, the driving style and traditions established in the area in question, operation of road traffic safety inspection, the mentality, psychophysiological and other factors. Yet, to undertake comparative analysis of influence produced by the above factors on the drivers' accident proneness, statistical data from other regions are required. Unfortunately, there is no basis for comparison as the authors are unaware of any database of the same volume as the one studied by Campbell and Levine [8]. Nevertheless, we are to note that the intensity of RTA for freight traffic in the Russian Federation is, according to Dementienko et al. [3], equal to 4.8×10^{-6} 1/h, which is just 37% higher than the one calculated for the 1st category, and almost 4 times smaller than for the 2nd category. Moreover, the aforesaid intensity is even lower than the average value across the categories $\lambda_s = 7.2 \times 10^{-6}$ 1/h, which is understandable as Dementienko et al. [3] consider professional drivers, whose driving safety level must indeed be higher than average.

The aforesaid testifies, first, to a good agreement between the statistical data and calculations presented in papers [3, 8] and the given paper, and, second, statistical data from different regions agree with each other quite well, and correction of parameters of the model for different territories should not be significant.

8. Conclusions

Based on statistical data of Campbell and Levine [8], it has become possible to develop a mathematical model representing the entire population of drivers as a totality of four categories, each featuring its own accident proneness characterized as RTA intensity λ_i.

These parameters differ very considerable for different categories. The probability of an RTA event throughout a year depends on the driver's category and varies from 0.03 to 0.9. The largest number of RTA is committed by the category of drivers, which is the second in terms of the number of drivers and the value of parameter λ_i.

The possibility is shown to classify a driver as being referred to this or that category depending on the number of RTAs he was involved in over a certain time period.

References

[1] Rail Safety and Standards Board. 2002. *Driver Vigilance Devices: Systems Review*. London: Railway Safety, p. 105.

[2] Dementienko, V. V., Markov, A. G., Koreneva, L. G., and Shakhnarovich, V. M. 2000. "Driver Vigilance Automated

Control." *Biomedical Radioelectronics* 8: 38-47.

[3] Dementienko, V. V., Dorokhov, V. B., Gerus, S. V., Markov, A. G., and Shakhnarovich, V. M. 2007. "On the Efficiency of Driver State Monitoring Systems." *Technical Physics* 52 (6): 781-6.

[4] Transport Canada. 1997. *Commercial Motor Vehicle Driver Fatigue and Alertness Study. Technical Summary.* Vancouver: Transport Canada.

[5] Haga, S. 1984. "An Experimental Study of Signal Vigilance Errors in Train Driving." *Ergonomics* 27 (7): 755-65.

[6] Klebelsberg, D. 1982. *Verkehrspsychologie*. Berlin: Springer. (in German)

[7] Dementienko, V. V., Dorokhov, V. B., Gerus, S. V., Koreneva, L. G., Markov, A. G., and Shakhnarovich, V. M. 2008. "A Biomathematical Model of a Human Operator's Falling Asleep." *Human Physiology* 34 (5): 592-600.

[8] Campbell, B. J., and Levine, D. 1973. "Accident Proneness and Driver License Programs." In *Proceedings of the First International Conference on Driver Behavior*, Zurich, Switzerland, pp. 1-12.

[9] Kunkel, E. 1975. *Kriminalität und Fahreignung.* Köln: TÜV Rheinland. (in German)

[10] Quenault, S. W., and Fuhrmann, J. 1969. "Fahrverhaltensbeobachtung und Fahrerklassifikation im Straßenverkehr." *Zeitschrift für Verkehrssicherheit* 15: 248-61. (in German)

[11] Gmurman V. E. 1968. *Fundamentals of Probability Theory and Mathematical Statistics.* New York: American Elsevier Pub. Co.

[12] Kendall, M. G., and Stuart, A. S. 1979. *The advanced Theory of Statistics* (V. 2, 4th ed.). London: Charles Griffin.