NONHOMOGENEOUS GENERALISATIONS OF POISSON PROCESS IN THE MODELING OF RANDOM PROCESSES RELATED TO ROAD ACCIDENTS

Franciszek Grabski

Polish Naval Academy, Department of Mathematics and Physics, Śmidowicza 69 Str., 81-127 Gdynia, Poland, e-mail: f.grabski@amw.gdynia.pl, ORCID ID: 0000-0001-5772-4550

ABSTRACT

The stochastic processes theory provides concepts, and theorems, which allow to build the probabilistic models concerning accidents. "Counting process" can be applied for modelling the number of road, sea, and railway accidents in the given time intervals. A crucial role in construction of the models plays a Poisson process and its generalizations. The nonhomogeneous Poisson process, and the corresponding nonhomogeneous compound Poisson process are applied for modelling the road accidents number, and number of people injured and killed in Polish roads. To estimate model parameters were used data coming from the annual reports of the Polish police.

Keywords:
road accident, nonhomogeneous Poisson process, nonhomogeneous compound Poisson process

Research article

© 2020, Franciszek Grabski
This journal provides immediate open access to its content under the Creative Commons by 4.0 license. Authors who publish with this journal retain all copyrights and agree to the terms of the above-mentioned CC BY 4.0 license
1. INTRODUCTION

Von Bortkiewitsch (1898) calculated, using the data of the Prussian army, the number of soldiers who died, during 20 consecutive years, because of being kicked by a horse. He noticed that a random variable, say $X$, denoting the number of solders killed accidentally by the horse kick per year, has approximately Poisson distribution with parameter $\lambda = 0.61$ [1/year]. Since then Poisson’s distribution, and the corresponding stochastic Poisson process, have found use in various fields of science and technology.

A Poisson process and its extensions, are used in safety and reliability problems. They allow to construct the models denoting number of road, sea, and railway accidents in the given time intervals.

It should be mentioned, that this paper is an extension of article [3], because of the new data concerning the Polish road accidents in 2019 [11].

2. NONHOMOGENEOUS POISSON PROCESS

We start from definition of nonhomogeneous Poisson Process (NPP).

Let

$$\tau_0 = \vartheta_0 = 0, \quad \tau_n = \vartheta_1 + \vartheta_2 + \cdots + \vartheta_n, \quad n \in \mathbb{N},$$

(1)

where $\vartheta_1$, $\vartheta_2$, ..., $\vartheta_n$ are positive independent random variables.

$$\tau_\infty = \lim_{n \to \infty} \tau_n = \sup \{ \tau_n : n \in \mathbb{N}_0 \}.$$  \hspace{1cm} (2)

A stochastic process \{\(N(t)\): \(t \geq 0\)\} defined by the formula

$$N(t) = \sup \{ n \in \mathbb{N}_0 : \tau_n \leq t \}$$

(3)

is called a counting process corresponding to a random sequence \(\{\tau_n : n \in \mathbb{N}_0\}\).

Let \(\{N(t) : t \geq 0\}\) be a stochastic process, taking values on $S = \{0,1,2, \ldots\}$, value of which represents the number of events in a time interval $[0,t]$. 

Scientific Journal of PNA
A counting process \( \{ N(t): t \geq 0 \} \) is said to be a nonhomogeneous Poisson process (NPP) with an intensity function \( \lambda(t) \geq 0, \ t \geq 0 \), if

1. \( P(N(0) = 0) = 1 \) ;

2. The process \( \{ N(t): t \geq 0 \} \) is the stochastic process with independent increments, the right continuous and piecewise constant trajectories;

3. \( P(N(t + h) - N(t) = k) = \left( \frac{\int_t^{t+h} \lambda(x)dx}{k!} \right)^k e^{-\int_t^{t+h} \lambda(x)dx} \);

From the definition it follows, that the one dimensional distribution of NPP is given by the rule:

\[ P(N(t) = k) = \left( \frac{\int_0^t \lambda(x)dx}{k!} \right)^k e^{-\int_0^t \lambda(x)dx}, \ k = 0,1,2, ... \]

The expectation and variance of NPP are the functions:

\[ \Lambda(t) = E[N(t)] = \int_0^t \lambda(x)dx \quad V(t) = V[N(t)] = \int_0^t \lambda(x)dx, \ t \geq 0. \]

The corresponding standard deviation is:

\[ D(t) = \sqrt{V[N(t)]} = \sqrt{\int_0^t \lambda(x)dx}, \ t \geq 0. \]

The expected value of the increment \( N(t + h) - N(t) \) is:

\[ \Delta(t; h) = E(N(t + h) - N(t)) = \int_t^{t+h} \lambda(x)dx. \]

The corresponding standard deviation is:

\[ D(t; h) = D(N(t + h) - N(t)) = \sqrt{\int_t^{t+h} \lambda(x)dx} \]

A nonhomogeneous Poisson process with \( \lambda(t) = \lambda, \ t \geq 0 \) for each \( t \geq 0 \), is a regular Poisson process. The increments of nonhomogeneous Poisson process are independent, but not necessarily stationary. A nonhomogeneous Poisson process is a Markov process.
3. NONHOMOGENEOUS COMPOUND POISSON PROCESS

We assume that \( \{N(t): t \geq 0\} \) is a *nonhomogeneous Poisson process* (NPP) with an intensity function \( \lambda(t) \), \( t \geq 0 \) such that \( \lambda(t) \geq 0 \) for \( t \geq 0 \), and \( X_1, X_2, \ldots \) is a sequence of the independent random variables independent of \( \{N(t): t \geq 0\} \). A stochastic process

\[
X(t) = X_1 + X_2 + \ldots + X_{N(t)}, \quad t \geq 0
\]

is said to be a *nonhomogeneous compound Poisson process* (NCPP).

**Proposition 1.**

Let \( \{X(t): t \geq 0\} \) be a nonhomogeneous compound Poisson process (NCPP). If \( E(X_1^2) < \infty \), then:

1. \( E[X(t)] = \Lambda(t) E(X_1) \)  
2. \( V[X(t)] = \Lambda(t) E(X_1^2) \),

Where:

\[
\Lambda(t) = E[N(t)] = \int_0^t \lambda(x) \, dx.
\]

**Proof [6].**

**Corollary 1**

Let \( \{X(t+h) - X(t): t \geq 0\} \) be an increment of a nonhomogeneous compound Poisson process (NCPP). If \( E(X_1^2) < \infty \), then:

\[
E[X(t+h) - X(t)] = \Delta(t; h) E(X_1), \quad (15)
\]

\[
V[X(t+h) - X(t)] = \Delta(t; h) E(X_1^2), \quad (16)
\]

where:

\[
\Delta(t; h) = \int_t^{t+h} \lambda(x) \, dx. \quad (17)
\]
4. DATA ON MOTORIZATION AND ROAD ACCIDENTS IN POLAND

Quoted data from the Central Statistical Office from 2017, 2018, and 2019 were presented in reports of the Polish Police [9], [10], [11].

4.1 GENERAL DATA ON MOTORIZATION

Since the beginning of the 90's, the number of vehicles registered in Poland has been systematically growing.

Tab.1. Number of motor vehicles in the years 2007-2018

| Years | Motor vehicles in total | Passenger cars | Trucks | Motorcycles |
|-------|-------------------------|----------------|--------|-------------|
| 2007  | 19,471,836              | 14,588,739     | 2,345,068 | 825,305     |
| 2008  | 21,336,913              | 16,079,533     | 2,511,677 | 909,144     |
| 2009  | 22,024,697              | 16,494,650     | 2,595,845 | 974,906     |
| 2010  | 23,037,149              | 17,239,800     | 2,767,035 | 1,013,014   |
| 2011  | 24,189,370              | 18,125,490     | 2,892,064 | 1,069,195   |
| 2012  | 24,875,718              | 18,744,412     | 2,920,779 | 1,107,260   |
| 2013  | 25,683,575              | 19,389,446     | 2,962,064 | 1,153,169   |
| 2014  | 26,472,274              | 20,003,863     | 3,037,427 | 1,189,527   |
| 2015  | 27,409,106              | 20,723,423     | 3,098,376 | 1,272,333   |
| 2016  | 28,601,037              | 21,675,388     | 3,179,655 | 1,355,625   |
It is easy to count, that from 2007 to 2019, the number of passenger cars increased by 38.68%, number of trucks by 30.77%, while the number of motorcycles increased by 46.68%. During this time, the total number of motor vehicles increased by 37.96%.

The location of our country on the East-West transport route generates heavy transit traffic. According to the border guards, cited in the police report [10], in 2018, 12 435 345 vehicles entered the European Union’s external borders, including 9 970 787 passenger cars.

4.2 GENERAL DATA ON ROAD ACCIDENTS

A table, containing the number of accidents, and their consequences, is presented below. The data comes from police reports [9], [10]. [11].

Tab.2. Number of accidents and their consequences in the years 2007-2019

| Years | Interval [days] | Center of accident | Number of fatalities | Number of injured | Indicator $\alpha$ | Indicator $\beta$ |
|-------|-----------------|--------------------|---------------------|------------------|-------------------|------------------|
| 2007  | [0, 365)        | 183.5              | 49 536              | 5 583            | 63 224            | 0.1127           | 1.2763           |
| 2008  | [365, 731)      | 548                | 49 054              | 5 432            | 62 097            | 0.1108           | 1.2658           |
| 2009  | [1096, 1461)    | 1278.5             | 44 196              | 4 572            | 56 046            | 0.1034           | 1.2681           |
The table does not contain the number of traffic collisions. For example, in 2018, 436,414 road collisions were reported.

### 5. MODEL OF THE ROAD ACCIDENT NUMBERS

Due to the nature of these events, pre-assumption that it is a nonhomogeneous Poisson process with some parameter $\lambda(t) > 0$, seems to be justified. The expected value of increment of this process is given by (9), while its one dimensional distribution is determined by (5). We can use practically these rules if the intensity
function $\lambda(t) > 0$ in known. To define this function one utilize information presented in table 2. The statistical analysis of the data shows that the intensity function $\lambda(t)$ can be approximated by the linear function $\lambda(t) = at + b$.

5.1. ESTIMATION OF THE MODELS PARAMETERS

Dividing the number of accidents in each year, by 365 or 366 we get the intensity in units of [1 / day].

We approximate the empirical intensity by a linear regression function $y = ax + b$ that satisfied condition

$$S(a, b) = \sum_{i=1}^{n} [y_i - (ax_i + b)]^2 \rightarrow \min$$

Recall, that solution of above optimization problem leads to finding parameters $a$ and $b$. The parameters are given by the rules:

$$a = \frac{\mu_{11}}{\mu_{20}}, \quad b = m_{01} - am_{10}, \quad \bar{x} = m_{10} = \frac{1}{n} \sum_{i=1}^{n} x_i, \quad \bar{y} = m_{01} = \frac{1}{n} \sum_{i=1}^{n} y_i, \quad (22)$$

$$m_{11} = \frac{1}{n} \sum_{i=1}^{n} x_i y_i, \quad \mu_{11} = m_{11} - m_{10} m_{01},$$

$$m_{20} = \frac{1}{n} \sum_{i=1}^{n} x_i^2, \quad \mu_{20} = m_{20} - m_{10}^2.$$  

Applying the rules (26) for the data from Table 2 and using Excel system we obtain:

$$a = -0.000003658, \quad b = 2.534128 \quad (23)$$

The linear intensity of accidents is:

$$\lambda(x) = -0.000113629 x + 3.442384 \geq 0 \quad (24)$$

This function is shown in figure 1.
From (7) we obtain:

\[ \Lambda(t) = -0.000113629t^2 + 2.534128t, \quad t \geq 0. \]  

(25)

Therefore the one dimensional distribution of NPP is:

\[ P(N(t) = k) = \left(\frac{\Lambda(t)}{k!}\right)^k e^{-\Lambda(t)} k = 0,1,2,... \]  

(26)

where \( \Lambda(t) \) is given by (25).

Finally one can say that the model of the accident number on Polish roads is the nonhomogeneous Poisson process with the parameter \( \Lambda(t), \ t \geq 0 \), determined by (25).

Using data from the Table 1 and Figure 2 one can compute the indicator of fatalities intensity in road accidents in relation to the number of vehicle crossing

\[ \alpha = \frac{NF}{NVC} \]  

(27)

where \( NF \) denotes number of fatalities, \( NVC \) designates number of vehicles crossing.
Fig. 2. Indicators of fatalities intensity in road accidents in relation to the crossing number

\[ \beta = \frac{NI}{NVC} \]
6. **ANTICIPATION OF ACCIDENT NUMBER**

From (5) and (10) we get:

\[ P(N(t + h) - N(t) = k) = \frac{\Delta(t; h)}{k!} e^{-[\Delta(t; h)]}. \]  

(28)

It means that one can anticipate number of accidents at any time interval, with a length of \( h \). The expected value of the increment \( N(t + h) - N(t) \) is defined by (10). For the function:

\[ \Lambda(t) = \frac{a t^2}{2} + b t \]  

(29)

we obtain the expected value of the accidents at time interval \([t, t + h] \)

\[ \Delta(t; h) = h\left(\frac{a h}{2} + b + a t \right), \]  

(30)

The corresponding standard deviation is:

\[ \sigma(t; h) = \sqrt{h\left(\frac{a h}{2} + b + a t \right)}. \]  

(31)

**Example 1.**

We want to predict the number of accidents from June 1\( ^{st} \) of 2020 to August 30\( ^{th} \) of 2020. We also want to calculate the probability of a given number of accidents. First we have parameters \( t \) and \( h \). As extension of table 2 on year 2019 we can obtain an interval [4749, 5114). From January 1\( ^{st} \) of 2020 to June 1\( ^{st} \) of 2020 152 days have passed. Hence \( t = 4749 + 152 = 4901 \). From June 1\( ^{st} \) to August 31\( ^{st} \) \( h = 92 \) days have passed. For these parameters using (29) and (30) we obtain \( \Delta(t; h) = 7074.406 \), \( \sigma(t; h) = 84.109 \).

This means, that the average predicted number of accidents between June 1\( ^{st} \), 2019 and August 31\( ^{st} \), 2019 is about 7074, with a standard deviation of about 84.

\[ P_{c \leq X \leq d} = P(c \leq N(t + h) - N(t) \leq d) = \sum_{x=c}^{x=d} \frac{7074.406^x}{x!} e^{-7074.406} ; x = 0,1,2,\ldots \]

Applying approximation by normal distribution we get:

\[ P_{c \leq X \leq d} = \Phi \left( \frac{d - 7074.406}{84.109} \right) - \Phi \left( \frac{c - 7074.406}{84.109} \right) \]
For $d = k\sigma$, $c = -k\sigma$, $k = 1,2,3$ we obtain $k$-sigma formula:

$$P_{-k\sigma \leq X \leq k\sigma} = 2\Phi(k) - 1 = \begin{cases} 0.6827 & \text{for } k = 1 \\ 0.9545 & \text{for } k = 2 \\ 0.9973 & \text{for } k = 3 \end{cases}$$

Therefore, for the predicted number of accidents between June 1st, 2019 and August 31st, 2019:

$$P(X \in [6990.297, 7158.515]) = 0.6827,$$
$$P(X \in [6906.188, 7242.624]) = 0.9545,$$
$$P(X \in [6822.079, 7326.733]) = 0.9973.$$ 

6.1. ANTICIPATION OF THE ACCIDENTS CONSEQUENCES

Let $X = X_i$, $i = 1,2, ..., N(t)$ denotes number of fatal events in a single accident. We suppose that the random variables $X_i$, $i = 1,2, ...$ have the identical Poisson distribution with parameters $E(X_i) = V(X_i) = \mu$, $i = 1,2, ..., N(t)$.

The predicted number of fatal events in the time interval $[t, t+h)$ is described by the expectation of the increment $X(t+h) - X(t)$. Recall that the expected value, and standard deviation of the accidents number in the time interval $[t, t+h)$ are given by (10) and (11). For the data from Example 1 using (29) and (30), we obtain the expected value of fatalities number ($EFN$), and the corresponding standard deviation ($DFN$) in the time interval $[t, t+h)=[4900, 4992)$:

$$\Delta(t;h)=\Delta(t;h) \times \mu, \text{ and } DFN = \sqrt{\Delta(t;h) \times (\mu + \mu^2)}. \text{ We assume that the NCPP is homogenous in this time interval, and a mean is calculated in center of interval.}$$

Finally we obtain $EFN = 315.034$ and $DFN=17.7492$.

For the same data we obtain the expected value of injured number, denoted by $EIN$, and corresponding standard deviation ($DIN$) in the time interval $[4900, 4992)$. We assume that the NCPP is homogenous in this time interval, and a mean is calculated in center of that one. In this case $\Delta(t;h) = 626.62$. Using the same formulas we get the expectation $EIN$, and the standard deviation $DGN = 25.03$ of injured people number.
7. CONCLUSIONS

The nonhomogeneous Poisson process, and the corresponding nonhomogeneous compound Poisson process are applied for modelling the road accidents number, and the number of injured and fatalities on Polish roads. To estimate model parameters one used data coming from the annual reports of the Polish police. Constructed models allowed to anticipate number of accidents at any time interval, with a length of \( h \) and the accident consequences. One obtained the expected value of fatalities or injured, and the corresponding standard deviation in the time interval \( [t, t+h) \).

The statistical distribution of fatalities number in a single accident, and statistical distribution of injured people number, and also statistical distribution of fatalities or injured number in a single accident, are computed.

REFERENCES

[1] Di Crescenzo A., Martinucci B., Zacks S. Compound Poisson process with Poisson subordinator *Journal of Applied Probability* Vol. 52, No. 2, p. 360 - 374, 2015.

[2] Fisz M. Probability and Mathematical Statistics, Warsaw PWN; (in Polish) 1969

[3] Grabski F. Nonhomogeneous Poisson process and compound Poisson process in modeling of random processes related to road accidents. *Journal of KONES Powertrain and Transport*, Vol.26, No.1 2019. Pp 39-46.

[4] Grabski F. *Semi-Markov Processes: Applications in Systems Reliability and Maintenance* Elsevier; Amsterdam, Boston, Heidelberg, London, NewYork, Oxford, Paris, San Diego, San Francisco, Sydney, 2015.

[5] Grabski F. Nonhomogeneous Poisson process application to modelling accidents number at Baltic waters and ports. *Journal of Polish Safety and Reliability Association*, volume 8, number 1, p. 39-46, 2017.

[6] Grabski F. Nonhomogeneous stochastic processes connected to Poisson process. *Scientific Journal of Polish Naval Academy*, 2 (213), pp. 5-15, Gdynia 2018 (LIX).

[7] Limnios N, Oprisan G., *Semi-Markov Processes and Reliability*; Birkhauser, Boston 2001.

[8] Shiryayev A. N., Probability. Springer-Verlag, New York, Berlin, Heidelberg, Tokyo 1984.

[9] Symon E. *Wypadki drogowe w Polsce w roku 2017*, Wydział Opiniodawczo-Analityczny Biura Ruchu Drogowego Komendy Głównej Policji, Warszawa 2018.
Niejednorodne uogólnienie procesu Poissona w modelowaniu losowych procesów związanych z wypadkami drogowymi

Streszczenie

W pracy przedstawiono niektóre uogólnienia procesu Poissona i ich własności. Skupiono się na dwóch uogólnieniach – niejednorodnym procesie Poissona i niejednorodnym złożonym procesie Poissona. Niejednorodny proces Poissona pozwala na skonstruowanie modelu probabilistycznego, opisującego liczbę różnych rodzajów wypadków. Niejednorodny złożony proces Poissona pozwala matematycznie opisywać konsekwencje tych wypadków. Przedstawione tu wyniki teoretyczne dają możliwość przewidywania liczby wypadków i ich konsekwencji. Estymacja parametrów modelu została wykonana na podstawie danych zamieszczonych w rocznych raportach Policji [9], [10], [11].

Słowa kluczowe:

wypadek drogowy, niejednorodny proces Poissona, niejednorodny złożony proces Poissona