

# NONHOMOGENEOUS POISSON PROCESS AND COMPOUND POISSON PROCESS IN THE MODELLING OF RANDOM PROCESSES RELATED TO ROAD ACCIDENTS

Franciszek Grabski

Polish Naval Academy, Department of Mathematics and Physics  
Śmidowicza Street 69, 81-127 Gdynia, Poland  
e-mail: f.grabski@amw.gdynia.pl

## Abstract

The stochastic processes theory provides concepts and theorems that allow building probabilistic models concerning accidents. So called counting process can be applied for modelling the number of the 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 new theoretical results regarding compound Poisson process are presented in the paper. A nonhomogeneous Poisson process and the corresponding nonhomogeneous compound Poisson process are applied for modelling the road accidents number and number of injured and killed people in the Polish road. To estimate model parameters were used data coming from the annual reports of the Polish police [9, 10]. Constructed models allowed anticipating number of accidents at any time interval with a length of  $h$  and the accident consequences. We obtained the expected value of fatalities or injured and the corresponding standard deviation in the given time interval. The statistical distribution of fatalities number in a single accident and statistical distribution of injured people number and also probability distribution of fatalities or injured number in a single accident are computed. It seems that the presented examples explain basic concepts and results discussed in the paper.

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

## 1. Introduction

In 1837, Simeon-Denis Poisson proved that the sequence of random variables  $\{X_n\}_{n=1,2,\dots}$  having binomial distribution  $p_k(k) = \binom{n}{k} p^k (1-p)^{n-k}$  when a parameter  $p$ , determining the probability of success in a single experiment is small; has approximately distribution given by the rule:

$$p(k) = \frac{\lambda}{k!} e^{-\lambda}, k \in S = \{0,1,2, \dots\}, \quad (1)$$

where  $\lambda = n p$ .

Poisson did not see applications of this theorem. Von Bortkiewitsch (1898) calculated from the data of the Prussian army, the number of soldiers who died during the 20 consecutive years because of the kick by a horse, noticed that a random variable, say  $X$ , denoting the number of soldiers killed accidentally by the horse kick per year, has approximately Poisson distribution with parameter  $\lambda = 0.61$  [1/year]. Since that time, the Poisson 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.

## 2. Nonhomogeneous Poisson process

Let:

$$\tau_0 = \vartheta_0 = 0, \tau_n = \vartheta_1 + \vartheta_2 + \dots + \vartheta_n, n \in \mathbb{N}, \quad (2)$$

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

Let

$$\tau_\infty = \lim_{n \rightarrow \infty} \tau_n = \sup\{\tau_n : n \in \mathbb{N}_0\}. \quad (3)$$

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

$$N(t) = \sup\{n \in \mathbb{N}_0 : \tau_n \leq t\} \quad (4)$$

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  $S = \{0, 1, 2, \dots\}$ , value of which represents the number of events in a time interval  $[0, t]$ .

A *counting process*  $\{N(t) : t \geq 0\}$  is said to be *nonhomogeneous Poisson process* (NPP) with an intensity function  $\lambda(t) \geq 0, t \geq 0$ , if

$$P(N(0) = 0) = 1, \quad (5)$$

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

$$P(N(t+h) - N(t) = k) = \frac{\left(\int_t^{t+h} \lambda(x) dx\right)^k}{k!} e^{-\int_t^{t+h} \lambda(x) dx}, \quad (6)$$

where  $h$  denotes an increment of time.

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

$$P(N(t) = k) = \frac{\left(\int_0^t \lambda(x) dx\right)^k}{k!} e^{-\int_0^t \lambda(x) dx}, k = 0, 1, 2, \dots \quad (7)$$

The expectation and variance of NPP are the functions

$$\Lambda(t) = E[N(t)] = \int_0^t \lambda(x) dx, \quad V(t) = V[N(t)] = \int_0^t \lambda(x) dx, t \geq 0. \quad (8)$$

The corresponding standard deviation is

$$D(t) = \sqrt{V[N(t)]} = \sqrt{\int_0^t \lambda(x) dx}, t \geq 0. \quad (9)$$

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

$$\Delta(t; h) = E(N(t+h) - N(t)) = \int_t^{t+h} \lambda(x) dx. \quad (10)$$

The corresponding standard deviation is

$$D(t; h) = D(N(t+h) - N(t)) = \sqrt{\int_t^{t+h} \lambda(x) dx}. \quad (11)$$

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  $X_1, X_2, \dots$  is a sequence of the independent random variables independent of  $\{N(t) : t \geq 0\}$ . A stochastic process

$$X(t) = X_1 + X_2 + \dots + X_{N(t)}, t \geq 0, \quad (12)$$

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

**Proposition 1**

Let  $\{X(t): t \geq 0\}$  be a *nonhomogeneous compound Poisson process* (NCP).

If  $E(X_1^2) < \infty$ , then

$$E[X(t)] = \Lambda(t) E(X_1), \tag{13}$$

$$V[X(t)] = \Lambda(t) E(X_1^2), \tag{14}$$

where

$$\Lambda(t) = E[N(t)] = \int_0^t \lambda(x) dx. \tag{15}$$

Proof [6].

**Corollary 1**

Let  $\{X(t+h) - X(t): t \geq 0\}$  be an increment of a *nonhomogeneous compound Poisson process* (NCP).

If  $E(X_1^2) < \infty$ , then:

$$E[X(t+h) - X(t)] = \Delta(t; h) E(X_1), \tag{16}$$

$$V[X(t+h) - X(t)] = \Delta(t; h) E(X_1^2), \tag{17}$$

where:

$$\Delta(t; h) = \int_t^{t+h} \lambda(x) dx. \tag{18}$$

**4. Data on motorization and road accidents in Poland**

We will quote data from the Central Statistical Office from 2017 and 2018 that were presented in reports of the Polish police [9, 10].

**4.1. General data on motorization**

Since the beginning of the nineties, 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   |
| 2017  | 29 149 178              | 22 109 572     | 3 212 690 | 1 398 609   |
| 2018  | 29 656 238              | 22 514 047     | 3 249 961 | 1 428 299   |

It is easy to count that from 2007 to 2018, the number of passenger cars increased by 54.32%, number of trucks by 38.59%, while the number of motorcycles increased by 73.06%. During this time, the total number of motor vehicles increased by 52.30%.

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 Tab. 2 contains the number of accidents and their some consequences. The data comes from police reports [9, 10].

Tab. 2. Number of accidents and their consequences in the years 2008-2018

| Years | Interval [year] | Centre of interval | Number of accident | Number of fatalities | Number of injured | Killed / Accident number | Injured / Accident number |
|-------|-----------------|--------------------|--------------------|----------------------|-------------------|--------------------------|---------------------------|
| 2007  | [0, 1)          | 0.5                | 49 536             | 5 583                | 63 224            | 0.1127                   | 1.2763                    |
| 2008  | [1, 2)          | 1.5                | 49 054             | 5 432                | 62 097            | 0.1108                   | 1.2658                    |
| 2009  | [2, 3)          | 2.5                | 44 196             | 4 572                | 56 046            | 0.1034                   | 1.2681                    |
| 2010  | [3, 4)          | 3.5                | 38 832             | 3 907                | 48 952            | 0,1006                   | 1.2606                    |
| 2011  | [4, 5)          | 4.5                | 40 065             | 4 189                | 49 501            | 0.1045                   | 1.2355                    |
| 2012  | [5, 6)          | 5.5                | 37 046             | 3 571                | 45 792            | 0.0963                   | 1.2360                    |
| 2013  | [6, 7)          | 6.5                | 35 847             | 3 357                | 44 059            | 0.0957                   | 1.2571                    |
| 2014  | [7, 8)          | 7.5                | 34 970             | 3 209                | 42 545            | 0.0915                   | 1.2166                    |
| 2015  | [8, 9)          | 8.5                | 32 967             | 2 938                | 39 778            | 0.0891                   | 1.2066                    |
| 2016  | [9, 10)         | 9.5                | 33 664             | 3 026                | 40 766            | 0.0898                   | 1.2109                    |
| 2017  | [10, 11)        | 10.5               | 32 760             | 2 831                | 39 466            | 0.0864                   | 1.2047                    |
| 2018  | [11, 12)        | 11.5               | 31 674             | 2 862                | 37 359            | 0.0903                   | 1.1794                    |

This 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 (10) while its one-dimensional distribution is determined by (6). We can use practically these rules if will know the intensity function  $\lambda(t) > 0$ . To define this function we utilize information presented in Tab. 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

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

$$S(a, b) = \sum_{i=1}^n [y_i - (ax_i + b)]^2 \rightarrow \min. \tag{19}$$

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

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

$$m_{11} = \frac{1}{n} \sum_{i=1}^n x_i y_i, \quad \mu_{11} = m_{11} - m_{10} m_{01}, \quad m_{20} = \frac{1}{n} \sum_{i=1}^n x_i^2, \quad \mu_{20} = m_{20} - m_{10}^2.$$

Applying the rules (20) for the data from Tab. 2 and using Excel system we obtain:

$$a = -1765.9331, \quad b = 48979.849. \quad (21)$$

The linear intensity of accidents is:

$$\lambda(t) = -1765.9331 t + 48979.849 \quad t \geq 0. \quad (22)$$

From (7) we have

$$\Lambda(t) = \int_0^t (1765.9331 x + 48979.849) dx. \quad (23)$$

Hence, we obtain

$$\Lambda(t) = -882.96655 t^2 + 48979.849 t, \quad t \geq 0. \quad (24)$$

Therefore the one-dimensional distribution of NPP is:

$$P(N(t) = k) = \frac{(\Lambda(t))^k}{k!} e^{-\Lambda(t)}, \quad k = 0, 1, 2, \dots, \quad (25)$$

where  $\Lambda(t)$  is given by (24).

Finally, we can say that the model of the accident number in Polish road is the nonhomogeneous Poisson process with the parameter  $\Lambda(t), t \geq 0$  determined by (24).

## 6. Anticipation of accident number

From (6) and (11) get

$$P(N(t+h) - N(t) = k) = \frac{\Delta(t;h)^k}{k!} e^{-\Delta(t;h)}. \quad (26)$$

It means that we 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 (11). For the function:

$$\Lambda(t) = a \frac{t^2}{2} + b t, \quad (27)$$

we obtain the expected value of the accidents number at time interval  $[t, t+h)$

$$\Delta(t; h) = h \left( \frac{a h}{2} + b + a t \right). \quad (28)$$

The corresponding standard deviation is

$$\sigma(t; h) = \sqrt{h \left( \frac{a h}{2} + b + a t \right)}. \quad (29)$$

### 6.1. Example

We want to predict the number of accidents from June 1 of 2019 to August 30 of 2019. We also want to calculate the probability of a given number of accidents. First have to determine parameters  $t$  and  $h$ . As extension of Tab. 2 on year 2019, we obtain an interval  $[12, 13)$ . From January 1 of 2019 to June 1 of 2019 have passed 151 days = 0.4137 year. Hence  $t = 12.4137$ . From June 1 to August 31 have passed  $h = 92$  days = 0.2520 year. Hence the interval  $[t, t+h) =$

= [12.4137, 12.665). For these parameters using (28) and (29) we obtain  $\Delta(t; h) = 6765.69$ ,  $\sigma(t; h) = 82.2538$ .

This means that the average predicted number of accidents between June 1, 2019 and August 31, 2019 is about 7074 with a standard deviation of about 84.

$$P_{c \leq X \leq d} = P(c \leq N(t+h) - N(t) \leq d) = \sum_{x=c}^{x=d} \frac{6765.69^x}{x!} e^{-6765.69}; x = 0, 1, 2, \dots \quad (30)$$

Applying approximation by normal distribution, we get:

$$P_{c \leq X \leq d} = \Phi\left(\frac{d-6765.69}{82.2538}\right) - \Phi\left(\frac{c-6765.69}{82.2538}\right) \quad (31)$$

For  $d = k \sigma, c = -k \sigma, k = 1, 2, 3$  we obtain  $k$ -sigma formula:

$$P_{-k \sigma \leq X \leq k \sigma} = 2\Phi(k) - 1 = \begin{cases} 0.6827 & \text{for } k = 1 \\ 0.9545 & \text{for } k = 2 \\ 0.9973 & \text{for } k = 3 \end{cases} \quad (32)$$

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

$$\begin{aligned} P(X \in [6683.436, 6847.944]) &= 0.6827, \\ P([6601.183, 6930.198]) &= 0.9545, \\ P(X \in [6518.929, 7012.451]) &= 0.9973. \end{aligned} \quad (33)$$

## 7. Anticipation of the accident consequences

Let  $X = X_i, i = 1, 2, \dots, N(t)$  denotes number of fatal events in a single accident. We suppose that the random variables  $X_i, i = 1, 2, \dots$  have the identical Poisson distribution with parameters  $E(X_i) = V(X_i) = \mu, i = 1, 2, \dots, N(t)$ . From the data analysis regarding road fatalities, it follows that the parameter  $\mu$  is not constant. This parameter depends on time:  $\mu = \mu(t)$ . The function we approximate by linear function using data from Tab. 2. Utilizing the EXCEL system, we calculated the linear regression.

$$\mu(t) = -0.0023 t + 0.111402, t \geq 0. \quad (34)$$

Note that the function in the time interval  $[t, t+h]$  is almost constant if the parameter  $h$  is not large. Therefore, to simplify the model we can assume that in the interval of prediction  $[t, t+h]$  for given  $t$  and  $h$ , the random variables  $X_i$  are i. i. d. with the constant expectation and variance given by the rule:

$$E(X_i) = V(X_i) = \tilde{\mu} = \frac{\mu(t) + \mu(t+h)}{2}. \quad (35)$$

For the data from Example (sub-chapter 6.1), we have:

$$\tilde{\mu} = 12 \frac{151+46}{365} = -0.0023 * 12.5397 + 0.111402 = 0.0825. \quad (36)$$

The predicted number of fatal events in the time interval  $[t, t+h]$  is described by the expectation of the increment  $X(t+h) - X(t)$ .

For the data from Example (sub-chapter 6.1) using (32), (33), (34) and (35) we can obtain the expected value of fatalities number **EFN** and the corresponding standard deviation **DFN** in the time interval  $[t, t+h]$ . From (16), (17) and (18) we obtain the rules:

$$\mathbf{EFN} = \Delta(t; h) \times \tilde{\mu}, \mathbf{DFN} = \sqrt{\Delta(t; h) \times (\tilde{\mu} + \tilde{\mu}^2)}. \quad (37)$$

For  $\Delta(t; h) = 6765.69, \tilde{\mu} = 0.0825$  and  $[t, t+h] = [12.4137, 12.665)$ , finally, we obtain **EFN = 557.926** and **DFN = 24.5755**.

Table 3 shows the Poisson distribution of fatalities number in a single accident.

*Tab. 3. Probability distribution of fatalities number in a single accident*

|          |          |           |           |           |             |                        |                        |
|----------|----------|-----------|-----------|-----------|-------------|------------------------|------------------------|
| $x$      | 0        | 1         | 2         | 3         | 4           | 5                      | 6                      |
| $P(X=x)$ | 0.920811 | 0.0759669 | 0.0031336 | 0.0000017 | 0.000001774 | $2.936 \times 10^{-8}$ | $4.03 \times 10^{-10}$ |

For the same data in a similar way, we compute the expected value **EIN** of a random variable a value of each denotes injured people number. We also calculate a corresponding standard deviation **DIN** in the same time interval. In this case, we have

$$\mu(t) = -0.008862 t + 1.28797, t \geq 0. \tag{38}$$

From (35) we get

$$\tilde{\mu} = -0.008862 * 12.5397 + 1.2879 = 1.1768. \tag{39}$$

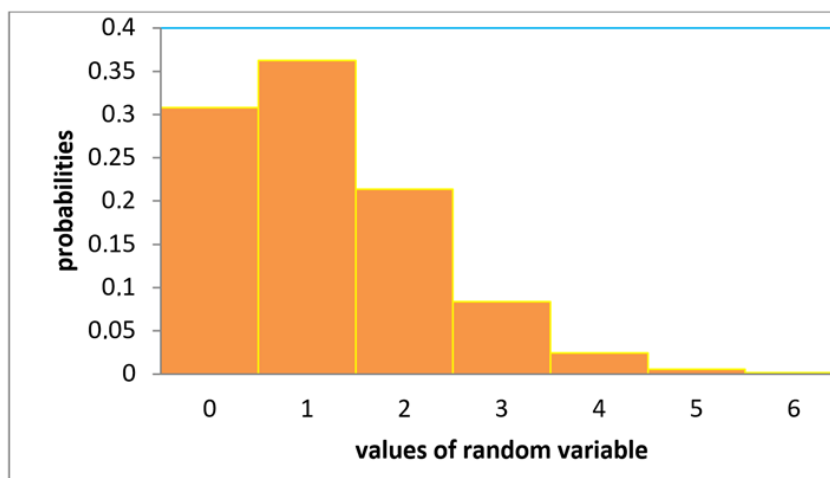
In this case  $\Delta(t;h) = 6765.69$ . Using the same formulas we get the expectation of injured people **EIN = 7961.052** and the corresponding standard deviation **DIN = 131.62**.

Table 4 shows the Poisson distribution of injured people number in a single accident.

*Tab. 4. Probability distribution of injured people number in a single accident*

|          |          |          |          |          |          |          |          |
|----------|----------|----------|----------|----------|----------|----------|----------|
| $x$      | 0        | 1        | 2        | 3        | 4        | 5        | 6        |
| $P(X=x)$ | 0.308264 | 0.362765 | 0.213451 | 0.083796 | 0.024633 | 0.005797 | 0.001137 |

Figure 1 shows the Poisson distribution of injured people number in a single accident.



*Fig. 1. Distribution of random variable denoting number of injured people in a single accident*

## 8. 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 in Polish road. To estimate model parameters were used data coming from the annual reports of the Polish police. Constructed models allowed anticipating number of accidents at any time interval with a length of  $h$  and the accident consequences. We obtained the expected value of fatalities and injured and the corresponding standard deviation in the time interval  $[t, t + h)$ .

The probability distribution of fatalities number in a single accident and probability distribution of injured people number are computed. Constructed models allowed anticipating

number of accidents and the accident consequences in relatively short time interval. The expected value of fatalities, expected value of injured and the corresponding standard deviations in the given time interval are anticipated in the paper.

## References

- [1] Di Crescenzo, A., Martinucci, B., Zacks, S., *Compound Poisson process with Poisson subordinator*, Journal of Applied Probability, Vol. 52, No. 2, pp. 360-374, 2015.
- [2] Fisz, M., *Probability and Mathematical Statistics*, PWN (in Polish), Warsaw 1969.
- [3] Grabski, F., *Semi-Markov models o reliability and operation*, IBS PAN (in Polish), Warsaw 2002.
- [4] Grabski, F., *Semi-Markov Processes: Applications in Systems Reliability and Maintenance*, Elsevier, Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Sydney 2015.
- [5] Grabski, F., *Nonhomogeneous Poisson process application to modelling accidents numberat Baltic waters and ports*, Journal of Polish Safety and Reliability Association, Vol. 8, No. 1, pp. 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] Shiriyayev, 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.
- [10] Symon, E., *Wypadki drogowe w Polsce w roku 2018*, Wydział Opiniodawczo-Analityczny Biura Ruchu Drogowego Komendy Głównej Policji, Warszawa 2019.

*Manuscript received 14 January 2019; approved for printing 25 March 2019*