Journal of Polish Safety and Reliability Association
Summer Safety and Reliability Seminars, Volume 7, Number 1, 2016

Grabski Franciszek
Naval University, Gdynia, Poland

Constructing stochastic models for investigation of dangerous events and accidents number in Baltic Sea region ports

Keywords

stochastic process, Poisson process, safety characteristics

Abstract

The stochastic processes theory provides concepts and theorems that allow to build probabilistic models concerning incidents or (and) accidents. Counting processes are applied for modelling number of the dangerous events and accidents number in Baltic Sea region ports in the given time intervals. A crucial role in construction of the models plays a Poisson process and its generalizations. Three models of the incidents or (and) accidents number in the seaports are here constructed. Moreover some procedures of the model parameters identification and the computer procedures for anticipation of the dangerous events number are presented in the paper.

1. Introduction

The dangerous events and accidents number in Baltic Sea region ports in the interval [0,t] are the randomly changing quantities. The theory of stochastic (random) processes allows the modelling of the random evolution of systems through the time. We will present briefly the basic concepts of the theory of random processes, which is necessary to build the models of the dengerous incidents and accidents number in the seaports of the Baltic region.

2. Stochastic processes

Let $\mathbb{T}$ be an arbitrary subset of real numbers $\mathbb{R}$. A family of real random variables $\{X(t): t \in \mathbb{T}\}$ defined on a joint probability space $(\Omega, \mathcal{F}, P)$ taking their values in a subset $S \subset \mathbb{R}$ is called a random process or a stochastic process. A set S is called a state space or phase space of the stochastic process, while $\mathbb{T}$ is called a set of its parameters. The stochastic process is also denoted by $\{X_t: t \in \mathbb{T}\}$. From this definition it follows that for every $t \in \mathbb{T}$, $X(t)$ is a random variable taking values in S, with domain Ω. For any fixed outcome $\omega \in \Omega$, and a fixed $t \in \mathbb{T}$, a number $x(t) \in S$ is a value of the random variable $X(t)$. A function $x(\cdot) = \{x(t): t \in \mathbb{T}\}$ is said to be *a trajectory* or *realization* of a random process $\{X(t): t \in \mathbb{T}\}$. This function is also called *a sample function or a path-function.*

For a countable or finite set of parameters $\mathbb{T}$, $\{X(t): t \in \mathbb{T}\}$ is called a stochastic process with *a discrete time*. If $\mathbb{T} = \mathbb{N}_0 = \{0,1,2,\dots\}$ or $\mathbb{T} = \mathbb{N} = \{1,2,\dots\}$ or $\mathbb{T} = \mathbf{N} = \{1,2,\dots,\mathrm{n}\}$ then the stochastic process is said to be a *random sequence* or a *random chain*. If $\mathbb{T}$ is an uncountable subset of, for example $\mathbb{T} = [0, \infty)$, then the process is called a *continuous time stochastic process.*

For a fixed $t \in \mathbb{T}$, $X(t)$ is the random variable, that can have different types of distribution: continuous, discrete or singular. The distribution of this random variable can be also a mix of those three distribution types. In the mostly considered cases the distribution of $X(t)$ is continuous or discrete. The random process $\{X(t): t \in \mathbb{N}_0\}$ with discrete time and a discrete distribution of the random variable $X(t)$ for each $t \in$ is called a random process of DD type. The continuous time random process which has the discrete distribution of the random variable $X(t)$ for each $t \in$ is said to be the random process of CD type. The continuous time random process with continuous distribution in an instant t is said to be the random process of CC type.

A random process $\{X(t): t \geq 0\}$ is said to be *process with independent increments* if for all $t_1, \dots, t_n$ such that

$$0 < t_1 < t_2 < \cdots < t_n$$

the random variables

$$X(0), X(t_1) - X(0), \ldots\ X(t_n) - X(t_{n-1})$$

are mutually independent. If the increments $X(s) - X(t)$ and $X(s+h) - X(t+h)$ for all $t, s, h > 0, s > t$ have the identical probability distributions then $\{X(t): t \geq 0\}$ is called a process with the *stationary independent increments* (SII). It is proved, that for the SII processes such that $X(0) = 0$ an expectation and a variance are

$$E[X(t)] = m_1\, t,\ V[X(t)] = \sigma_1^2\, t, \qquad (1)$$

where

$$m_1 = E[X(1)] \text{ and } \sigma_1^2 = V[X(1)]. \qquad (2)$$

An example of a SII random process is a Poisson process. A stochastic process $\{X(t);\ t \geq 0\}$ taking values on $S = \{0,1,2,\ldots\}$, with the right continuous trajectories is said to be a Poisson process with parameter $\lambda > 0$ if:

1. $X(0) = 0$,
2. $\{X(t): t\ \geq 0\}$ is the process with the stationary independent increments,
3. For all $t > 0, h \geq 0$,

$$P(X(t+h) - X(t) = k) = \frac{(\lambda\, h)^k}{k!}\ e^{-\lambda\, t}, k \in S \quad (3)$$

For $t = 0$ we get a first order distribution of the Poisson process:

$$p_k(h) = P(X(h) = k) = \frac{(\lambda\, h)}{k\,!}\ e^{-\lambda\, h}, k \in S \qquad (4)$$

For $h = 1$ we obtain the Poisson distribution with parameter λ. Hence $E[X(1)] = \lambda$ and $V[X(1)] = \lambda$. Therefore, from (1) and (2), we obtain the expectation and the variance of the Poisson process:

$$E[X(t)] = \lambda\, t,\ V[X(t)] = \lambda\, t, t \geq 0. \qquad (5)$$

For a fixed t this formula determines the Poisson distribution with parameter $\Lambda = \lambda\, t$:

$$p(k) = P(X = k) = \frac{(\Lambda)^k}{k\,!}\ e^{-\Lambda}, k \in S. \qquad (6)$$

Using this formula we have written a short procedure in a MATHEMATICA computer program which allows to calculate these probabilities.

The procedure of calculation of the Poisson distribution

$a = \Lambda$
Print ["*n*=", *n*=4, " *a*=", *a*=3]
For [*k* = 0, *k* ≤ *n*, *k*++,
Print [" *p* [",*k*,"]", *p*[*k*] = $\frac{a^k}{k!} e^{-a}$ // N]]
Print [" *P*(*X* ≤ *n*) = ", " *P*(*X* > *n*) = " $\sum_{k=0}^{n}$ *p*[*k*], "1-$\sum_{k=0}^{n}$ *p*[*k*]]

Let $0 < \tau_1, < \tau_2, \ldots$ represent the consecutive instants of the state changes (jumps) in the Poisson process or another process with the right continuous, nondegreasing and picewise constant trajectories. The random variables $\vartheta_1 = \tau_1,\ \ \vartheta_2 = \tau_2 - \tau_1,\ \ \ldots$ denote the sojourn times of the states $0, 1, \ldots$. Let us notice that

$$\tau_0 = \vartheta_0 = 0, \tau_n = \vartheta_1 + \vartheta_2 + \cdots + \vartheta_n, n \in \mathbb{N},$$
$$\tau_\infty = \lim_{n\to\infty} \tau_n = \sup\{\tau_n : n \in \mathbb{N}_0\}.$$

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

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

is called a *counting process* corresponding to a random sequence $\{\tau_n : \in \mathbb{N}_0\}$.
For the Poisson process with parameter λ the random variables $\vartheta_1, \vartheta_2, \ldots, \vartheta_n, n = 2,3, \ldots$ are mutually independent and *exponentially distributed with the identical parameter* λ.
The Poisson process is a *counting process* which is genereted by the random sequence $\{\tau_n : \in \mathbb{N}_0\}$, where $\tau_n = \vartheta_1 + \vartheta_2 + \cdots + \vartheta_n, n \in \mathbb{N}$.

3. Poisson process as a stochastic model of dangerous events and accidents number in Baltic Sea region ports

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 accidents or dangerous incidents at a particular port of the Baltic Sea in a time interval $[0, t]$. Due to the nature of these events, pre-assumption that it is a Poisson process with some parameter $\lambda > 0$, seems to be justified.

3.1. Procedure of parameter λ identification

Let $0 < \tau_1, < \tau_2, \ldots$ represent the consecutive moments of dangerous incidents. It was above mentioned, the random variables $\vartheta_1 = \tau_1$,

$\vartheta_2 = \tau_2 - \tau_1, \ldots, \; n = 1,2,3, \ldots$ are mutually independent and *exponentially distributed with the identical parameter* λ. Because

$$E[\vartheta] = \frac{1}{\lambda} \tag{7}$$

then

$$\lambda = \frac{1}{E[\vartheta]}, \tag{8}$$

If the numbers $x_1, x_2, \ldots, x_n$ are values of random variables $\vartheta_1, \vartheta_2, \ldots, \vartheta_n$, than an estimate of $E[\vartheta]$ is mean

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}. \tag{9}$$

Finally the parameter λ can be calculated by the rule

$$\lambda = \frac{n}{x_1 + x_2 + \cdots + x_n} = \frac{1}{\bar{x}}. \tag{10}$$

3.2. Illustrative example

In *Table 1* are presented "observations" of the times between incidents (accidents).

Table 1. Observations of the times between incidents

Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Time between incidents [days]	1.2	27.9	4.7	9.3	15.0	13.2	19.8	1.3
Number	9	10	11	12	13	14	15	16
Time between incidents [days]	8.5	10.6	28.4	3.1	3.8	17.7	9.6	12.3

Using (9) we obtain

$$\lambda = \frac{1}{\bar{x}} = \frac{1}{11.65} = 0.0858 \left[\frac{1}{day}\right] \tag{11}$$

A first order distribution of the Poisson process describing number dangerous incidents at a particular port in a time interval $[0, t]$ is

$$P(N(t) = k) = \frac{(0.0858\, t)^k}{k\,!}\, e^{-0.0858\, t}, \quad k = 0, 1, 2, \ldots \tag{12}$$

3.3. Anticipattion of the dangerous incidents number

From (3) we get

$$P(N(t+h) - N(t) = k) = \frac{(\lambda\, h)^k}{k!}\, e^{-\lambda\, h}, \tag{13}$$

It means that we can anticipate number of dangerous incidents at any time interval with a length of *h.*

3.4. Numerical Example

Under assumption that the parameter $\lambda = 0.0858 \left[\frac{1}{day}\right]$, the distribution of dangerous incidents number at the time interval $[146, 206]$ has the Poisson distribution with parameter

$$\Lambda = \lambda\, h = 0{,}0858 \cdot 60 = 5.148.$$

Therefore

$$p_k(60) = P(N(206) - N(146) = k) = \frac{(5.148)^k}{k!}\, e^{-5.148}, k = 0,1,2, \ldots$$

The procedure of calculation of the number dangerous incidents distribution in this case takes the form

```
a= 5.148
Print ["n=", n=6, " a=", a=5.148]
For [k = 0, k ≤ n, k++,
Print [" p [",k,"]", p[k] = a^k/k! e^(-a) // N] ]
Print [" P(X ≤ n)= ", " P(X > n)= "
      ∑_{k=0}^{n} p[k], "1-∑_{k=0}^{n} p[k] ]
```

Finnally we obtain:

n= 6 , *a*= 5.148
p[0]= 0.00581102
p[1]= 0.0299151
p[2]= 0.0770015
p[3]= 0.132135
p[4]= 0.170057
p[5]= 0.175091
p[6]= 0.150228
$P(X \le n) = 0.740238$, $P(X > n) = 0.259762$

We can see that the most probable number of incidents in a time interval lengh of 60 days is 5, but the probability of this event is only 0.175091. The number of incidents is not greater than 6 with probability 0.740238.

4. Model describing total sum of different kind of accidents in seaport

Models describing total sum of accidents in the different seaports and the in the Baltic see region. The procedures of assessment of the models parameters.

The procedures for calculating the probabilities of expected number of accidents.

Let $\{N_k(t);\ t \geq 0\}, k = 1,2, \dots, n$ denote numbers of different kind of accidents in a seaport. We suppose that $\{N_k(t);\ t \geq 0\}, k = 1,2, \dots, n$ are independent Poisson processes with parameters $\lambda_1, \lambda_2, \dots, \lambda_n$. The parameter λ_k may be estimated based on the length of the intervals between successive accidents of a type *k* the same way like for a parameter λ,(see (10)). From the so-called theorem on adding of the random variables with Poisson distributions it follows that *the sum of* n independent Poisson processes with parameters $\lambda_1, \lambda_2, \dots, \lambda_n$, *is the Poisson process with parameter* $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$. It means that the process $\{N(t);\ t \geq 0\},\ N(t) = N_1(t) + N_2(t) + \dots + N_n(t)$ is the Poisson process with parameter $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$. This process counts total number of accidents in a seaport in a time interval $[0, t]$. A first order distribution of this process is given by the rule

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

where $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$.

Probability of the *k* different kind of accidents at any time interval with a length of *h* is

$$P(N(t+h) - N(t) = k) = \frac{(\lambda\, h)^k}{k!} e^{-\lambda\, h}, \quad k = 0, 1, 2, \dots . \tag{14}$$

5. Random parameter in Poisson model

The expected number of accidents often depends on changing randomly external conditions. Thus it can be assumed that the parameter λ is a random variable. We assume that this random variable has a gamma distribution with a density

$$f(u) = \begin{cases} \frac{\alpha^\nu}{\Gamma(\nu)} u^{\nu-1} e^{-\alpha\, u} & \text{for } u > 0 \\ 0 & \text{for } u \leq 0 \end{cases} \tag{15}$$

where $\alpha > 0,\ \ \nu > 1$.

Suppose that a condition distribution of the accidents number given λ has a Poisson distribution

$$P(N(t) = k|\ \lambda) = \frac{(\lambda\, t)^k}{k\,!}\ e^{-\lambda\, t}, \quad k = 0, 1, 2, \dots, \tag{16}$$

Using the formula for the total probability is calculated unconditional one-dimensional distribution of the process.

For $k = 0$ we obtain

$$P(N(t) = 0) = P(\vartheta_1 > t) = \left(\frac{\alpha}{\alpha+t}\right)^\nu. \tag{17}$$

For $k = 1, 2, \dots$ we have

$$P(N(t) = k) = \int_0^\infty \frac{(ut)^k e^{-ut}}{k!}\ \frac{\alpha^\nu}{\Gamma(\nu)} u^{\nu-1} e^{-\alpha\, u} du \tag{18}$$

Finally we obtain [4]:

$$P(N(t) = k) = \\ = \frac{\nu(\nu+1)\dots(\nu+k-1)}{k!} \left(\frac{t}{t+\alpha}\right)^k \left(\frac{\alpha}{t+\alpha}\right)^\nu, \quad k = 1, 2, \dots, \nu > 1, \lambda > 0. \tag{19}$$

The random variable $T = \vartheta_n,\ n = 1, 2, \dots$ is the time which elapses between successive accidents.

The function

$$R(t) = P(T > t) = \left(\frac{\alpha}{\alpha+t}\right)^\nu, t \geq 0 \tag{20}$$

is called *survival function*. Cumulative distribution function (CDF) of the random variable *T* has the form

$$F(t) = 1 - \left(\frac{\alpha}{\alpha+t}\right)^\nu, t\ \geq 0. \tag{21}$$

and corresponding to it a probability density function is

$$f(t) = \frac{\nu\, \alpha^\nu}{(\alpha+t)^{\nu+1}}\ \ , t\ \geq 0 \tag{22}$$

The expected value of this random variable is

$$E(T) = \int_0^\infty \left(\frac{\alpha}{\alpha+t}\right)^\nu dt = \frac{\alpha}{\nu-1}. \tag{23}$$

The second moment is

$$E(T^2) = 2\int_0^\infty t \left(\frac{\alpha}{\alpha+t}\right)^\nu dt = \frac{2\alpha^2}{(\nu-1)(\nu-2)}. \tag{24}$$

The variance is

It should be mentioned that the variance there exists if $\nu > 2$.

$$V(T) = \frac{\alpha^2 \nu}{(\nu-1)^2(\nu-2)}. \tag{25}$$

The standard deviation takes the form

$$\sigma(T) = E(T)\sqrt{\frac{\nu}{\nu-2}} \tag{26}$$

5.1. Procedure of parameters identification

Notice that these values depend on the two parameters: both α and ν. There is a natural question, how to determine these parameters. One method of estimating the unknown parameters is the so called the method of moments. In this method the unknown parameters are replaced by their statistical estimates derived from the results of observation. In this case, the expected value is replaced by the average of the sample and the second moment of the random variable is replaced by the second moment of the sample. Solving the corresponding system of equations we obtain the unknown parameters of the distribution.
Let the numbers $x_1, x_2, \dots, x_n$ are the values of random variables $\vartheta_1, \vartheta_2, \dots, \vartheta_n$, denoting the time between successive accidents. An estimate of the expectation $\mathrm{E(T)}$ is mean:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \tag{27}$$

and the estimate of is the the second moment from the sample:

$$\overline{x^2} = \frac{{x_1}^2 + \cdots + {x_n}^2}{n}. \tag{28}$$

Applying the method of moments we obtain the system of equations

$$\bar{x} = \frac{\alpha}{\nu - 1}, \ \ \overline{x^2} = \frac{2\alpha^2}{(\nu-1)(\nu-2)}. \tag{29}$$

The solution is given by the rules

$$\nu = 1 - \frac{\overline{x^2}}{(\bar{x})^2 - \overline{x^2}}, \ \ \alpha = \bar{x}\,(\nu - 1). \tag{30}$$

5.2. Illustrative example

We compute the values of parameters α and ν for the data from the *Table 1*. Using the equation (30) we obtain

$$\nu = 4 \ \ \alpha = 35{,}03. \tag{31}$$

To anticipate a number of dangerous accidents at any time interval with a length of t we can use the rules (17) and (19).
Applying (17) and (19) was written brief procedure in a MATHEMATICA computer program that for calculating probabilities in the considered distribution.
The procedure of calculation of the number dangerous accidents distribution in this case for $t = 60\ [day]$ is of the form:

```
Print ["n=", n=6, ", α=", α=35,", t=", t=60,
      ", ν=", ν=4]
Print["  p[",0,"]=", p[0]= (α/(α+t))^ν //N]
For[k=1, k ≤ n,  k++, Print ["  p[",k,"]=",
     p[k] = ∏_{i=1}^{k} ν(ν+1)…(ν+k−1)/k! (t/(t+α))^k (α/(t+α))^ν
     //N]]
Print["P(X ≤ n)=", ∑_{k=0}^{n} p[k], "
      P(X > n)=",1 − ∑_{k=0}^{n} p[k]]
```

As a result of these calculations, we get

n= 6 , α= 35 , t= 60 , ν= 4
p[0]= 0.0184237
p[1]= 0.0465442
p[2]= 0.0734908
p[3]= 0.0928305
p[4]= 0.102602
p[5]= 0.103682
p[6]= 0.0982252
$P(X \le n) = 0.5357,\ P(X > n) = 0.4643$.

We can notice that the most probable number of accidents in a time interval lengh of 60 days is 5, but the probability of this event is 0.1037. The number of accidents is not greater than 6 with probability 0.5357 Probability that there will be no accidents is equal to 0.0184 whereas in the previous model, it is 0.0058. It is a three times more than the Poisson model.

6. Conclusions

The random processes theory deliver concepts and theorems that enable to construct stochastic models concerning the incidents or (and) accidents. Counting processes and processes with independent increaments are the most appropriate for modelling number of the dangerous events and accidents number in Baltic Sea region ports in specified period of time. A crucial role in the models construction plays a Poisson process and its generalizations. Three models of the incidents and accidents number in the seaports are here constructed. Moreover some procedures of the model parameters identification and the computer procedures for anticipation of the dangerous events number are presented in the paper. To select the most appropriate model one should verify the models using appropriate statistical tests. But it requires a large number of data.

References

[1] Barlow, R. E. & Proshan, F. (1975). *Statistical theory of reliability and life testing*. New York:

Holt, Rinchart and Winston, Inc.

[2] Cinlar, E. (1969). Markov renewal theory. *Adv. Appl. Probab* 2, 1, 123-187.

[3] Feller, W. (1964). On semi-Markov processes. *Proc. Nat. Acad. Sci*. USA, 51, 4, 653-659.

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

[5] Grabski, F. (2002). *Semi-Markov models of reliability and operation*. IBS PAN. Warsaw. (in Polish).

[6] Grabski, F. (2014). *Semi-Markov Processes: Applications in Systems Reliability and Maintenance*. Elsevier; Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Sydney.

[7] Korolyuk, V. S. & Turbin, A. F. (1976). *Semi-Markov processes and their applications*. Naukova Dumka, Kiev. (in Russian).

[8] Limnios, N. & Oprisan, G. (2001). *Semi-Markov Processes and Reliability*. Birkhauser, Boston.

[9] Shiryayev, A. N. (1984). *Probability*. Springer-Verlag, New York, Berlin, Heidelberg, Tokyo.