**Kołowrocki Krzysztof**
ORCID ID: 0000-0002-4836-4976

**Kuligowska Ewa**
ORCID ID: 0000-0002-8274-3005

**Soszyńska-Budny Joanna**
ORCID ID: 0000-0003-1525-9392
*Maritime University, Gdynia, Poland*

# Identification methods and procedures of critical infrastructure operation process including operating environment threats

**Keywords**

critical infrastructure, operation, prediction, environment threats

**Abstract**

In the papere, there are presented the methods of identification of the critical infrastructure operation process on the basis of statistical data coming from this process realizations, in both cases, when the operation process is not related and related to the critical infrastructure operating environment threats.

## 1. Introduction

In the paper, there are presented the methods of identification of the critical infrastructure operation process on the basis of statistical data coming from this process realizations, in both cases, when the operation process is not related and related to the critical infrastructure operating environment threats. These are the methods and procedures for estimating the unknown basic parameters of the critical infrastructure operation process semi-Markov model and identifying the distributions of the critical infrastructure operation process conditional sojourn times at the particular operation states. There are given the formulae estimating the probabilities of the critical infrastructure operation process staying at the particular operation states at the initial moment, the probabilities of the critical infrastructure operation process transitions between the operation states. Moreover, there are given formulae for the estimator of unknown parameters of the distributions suitable and typical for the description of the critical infrastructure operation process conditional sojourn times at the operation states. Namely, the parameters of the uniform distribution, the triangular distribution, the double trapezium distribution, the quasi-trapezium distribution, the exponential distribution, the Weibull's distribution and the chimney distribution are estimated using the statistical methods such as the method of moments and the maximum likelihood method. The chi-square goodness-of-fit test is described and proposed to be applied to verifying the hypotheses about these distributions choice validity. The procedure of statistical data sets uniformity analysis based on Kolmogorov-Smirnov test is proposed to be applied to the empirical conditional sojourn times at the operation states coming from different realizations of the same critical infrastructure operation process.

To be able to apply this general model practically in the evaluation and prediction of the safety of real critical infrastructure it is necessary to have the statistical methods concerned with determining the unknown parameters of the proposed model [Barbu, Limnios, 2006], [Collet, 1996], [Kolowrocki, Soszynska, 2009a, 2009b, 2009d, 2009e], [Kołowrocki, Soszyńska-Budny, 2011]. Particularly, concerning the critical infrastructure operation process, the probabilities of the critical infrastructure operation process staying at the operation states at the initial moment, the probabilities of the critical infrastructure operation process transitions between the critical infrastructure operation states and the distributions of the conditional sojourn times of the critical infrastructure operation process at the particular operation states should be identified [Kolowrocki, Soszynska, 2009a, 2009b, 2009c, 2009d, 2009f], [Kolowrocki Soszynska, 2010a]. It is also necessary to have the methods of testing the hypotheses concerned with the conditional sojourn times of the critical infrastructure operation process at the operation states and the procedures of testing the uniformity of their realizations coming from different sets of empirical data [Kołowrocki, Soszyńska-Budny, 2011]. All these problems of unknown parameters identification are discussed in both cases, first for the critical infrastructure operation process not including the operating environment threats and next for the critical infrastructure operation process including the operating environment threats.

## 2. Identification of critical infrastructure operation process

We assume, as in [EU-CIRCLE Report D2.1-GMU2, 2016] and [EU-CIRCLE Report D3.3-GMU3, 2016], that acritical infrastructure during its operation at the fixed moment $t$, $t \in <0,+\infty>$, may be at one of $\nu$, $\nu \in N$, different operations states $z_b$, $b = 1,2,...,\nu$. Next, we mark by $Z(t)$, $t \in <0,+\infty>$, the critical infrastructure operation process, that is a function of a continuous variable $t$, taking discrete values in the set $\{z_1, z_2, ..., z_\nu\}$ of the critical infrastructure operation states. We assume a semi-Markov model [Kolowrocki, 2014], [Kolowrocki, Soszynska, 2009d], [Kołowrocki, Soszyńska-Budny, 2011], [Limnios, Oprisan, 2005], [Limnios et all, 2005], [Macci, 2008], [Mercier, 2008], of the critical infrastructure operation process Z(t) and we mark by $\theta_{bl}$ its random conditional sojourn times at the operation states $z_b$, when its next operation state is $z_l$, $b, l = 1,2,...,\nu$, $b \neq l$.

Under these assumption, the operation process may be described by the vector $[p_b(0)]_{1x\nu}$ of probabilities of the critical infrastructure operation process staying at the particular operations states at the initial moment $t = 0$, the matrix $[p_{bl}(t)]_{\nu x\nu}$ of the probabilities of the critical infrastructure operation process transitions between the operation states and the matrix $[H_{bl}(t)]_{\nu x\nu}$ of the distribution functions of the conditional sojourn times $\theta_{bl}$ of the critical infrastructure operation process at the operation states or equivalently by the matrix $[h_{bl}(t)]_{\nu x\nu}$ of the density functions of the conditional sojourn times $\theta_{bl}$, $b, l = 1,2,...,\nu$, $b \neq l$, of the critical infrastructure operation process at the operation states. These all parameters of the critical infrastructure operation process are unknown and before their use to the prognosis of this process characteristics have to be estimated on the basis of statistical data coming from practice.

### 2.1. Defining unknown parameters of critical infrastructure operation process and data collection

To make the estimation of the unknown parameters of the critical infrastructure operations process, the experiment delivering the necessary statistical data should be precisely planned.

First, before the experiment, we should perform the following preliminary steps:

i) to analyze the critical infrastructure operation process;

ii) to fix or to define the critical infrastructure operation process following general parameters:
  - the number of the operation states of the critical infrastructure operation process $\nu$,
  - the operation states of the critical infrastructure operation process $z_1$, $z_2$, …, $z_\nu$;

iii) to fix the possible transitions between the critical infrastructure operation states;

iv) to fix the set of the unknown parameters of the critical infrastructure operation process semi-Markov model.

Next, to estimate the unknown parameters of the critical infrastructure operation process, based on the experiment, we should collect necessary statistical data performing the following steps:

i) to fix and to collect the following statistical data necessary to evaluating the probabilities $p_b(0)$ of the critical infrastructure operation process staying at the operation states at the initial moment $t = 0$:

- the duration time of the experiment $\Theta$,
- the number of the investigated (observed) realizations of the critical infrastructure operation process $n(0)$,
- the vector of the realizations $n_b(0)$, $b = 1,2,...,\nu$, of the numbers of staying of the operation process respectively at the operation states $z_1$, $z_2$, …, $z_\nu$, at the initial moments $t = 0$ of all $n(0)$ observed realizations of the critical infrastructure operation process

$$[n_b(0)] = [n_1(0), n_2(0), ..., n_\nu(0)],$$

where

$$n_1(0) + n_2(0) + n_\nu(0) = n(0);$$

ii) to fix and to collect the following statistical data necessary to evaluating the probabilities $p_{bl}$ of the critical infrastructure operation process transitions between the critical infrastructure operation states:

- the matrix of the realizations of the numbers $n_{bl}$, $b$, $l = 1,2,...,\nu$, $b \neq l$, of the transitions of the critical infrastructure operation process from the operation state $z_b$ into the operation state $z_l$ at all observed realizations of the critical infrastructure operation process

$$[n_{bl}] = \begin{bmatrix} n_{11} & n_{12} & \dots & n_{1\nu} \\ n_{21} & n_{22} & \dots & n_{2\nu} \\ \dots \\ n_{\nu 1} & n_{\nu 2} & \dots & n_{\nu\nu} \end{bmatrix},$$

where

$$n_{bb} = 0 \text{ for } b = 1,2,...,\nu,$$

- the vector of the realizations of the numbers $n_b$, $b = 1,2,...,\nu$, of departures of the critical infrastructure operation process from the operation states $z_b$ (the sums of the numbers of the $b$-th rows of the matrix $[n_{bl}]$)

$$[n_b] = [n_1, n_2, ..., n_\nu],$$

where

$$n_1 = n_{11} + n_{12} + ... + n_{1\nu},$$
$$n_2 = n_{21} + n_{22} + ... + n_{2\nu},$$
$$\dots$$
$$n_\nu = n_{\nu 1} + n_{\nu 2} + ... + n_{\nu\nu};$$

iii) to fix and to collect the following statistical data necessary to evaluating the unknown parameters of the distributions $H_{bl}(t)$ of the conditional sojourn times $\theta_{bl}$ of the critical infrastructure operation process at the particular operation states:

- the numbers $n_{bl}$, $b$, $l = 1,2,...,\nu$, $b \neq l$, of realizations of the conditional sojourn times $\theta_{bl}$, $b$, $l = 1,2,...,\nu$, $b \neq l$, of the critical infrastructure operation process at the operation state $z_b$ when the next transition is to the operation state $z_l$ during the observation time $\Theta$,

- the realizations $\theta_{bl}^k$, $k = 1,2, \dots, n_{bl}$, of the conditional sojourn times $\theta_{bl}$ of the critical infrastructure operation process at the operation state $z_b$ when the next transition is to the operation state $z_l$ during the observation time $\Theta$ for each $b$, $l = 1,2,...,\nu$, $b \neq l$.

## 2.2. Estimating basic parameters of critical infrastructure operation process

After collecting the statistical data, it is possible to estimate the unknown parameters of the critical infrastructure operation process performing the following steps:

i) to determine the vector

$$[p(0)] = [p_1(0), p_2(0), \dots, p_\nu(0)],$$

of the realizations of the probabilities $p_b(0)$, $b = 1,2,...,\nu$, of the critical infrastructure operation process staying at the operation states at the initial moment $t = 0$, according to the formula

$$p_b(0) = \frac{n_b(0)}{n(0)} \text{ for } b = 1,2,...,\nu, \tag{2}$$

where

$$n(0) = \sum_{b=1}^{\nu} n_b(0), \tag{3}$$

is the number of the realizations of the critical infrastructure operation process starting at the initial moment $t = 0$;

ii) to determine the matrix

$$[p_{bl}] = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1\nu} \\ p_{21} & p_{22} & \cdots & p_{2\nu} \\ \dots \\ p_{\nu 1} & p_{\nu 2} & \cdots & p_{\nu\nu} \end{bmatrix}, \tag{4}$$

of the realizations of the probabilities $p_{bl}$, $b,l = 1,2,...,\nu$, of the critical infrastructure operation process transitions from the operation state $z_b$ to the operation state $z_l$ according to the formula

$$p_{bl} = \frac{n_{bl}}{n_b} \text{ for } b,l = 1,2,...,\nu, b \neq l, \; p_{bb} = 0 \tag{5}$$

for $b = 1,2,...,\nu$,

where

$$n_b = \sum_{b \neq l}^{v} n_{bl}, \; b = 1,2,...,v, \tag{6}$$

is the realization of the total number of the critical infrastructure operation process departures from the operation state $z_b$ during the experiment time Θ.

### 2.3. Estimating parameters of distributions of critical infrastructure operation process conditional sojourn times at operation states

Prior to estimating the parameters of the distributions of the conditional sojourn times of the critical infrastructure operation process at the particular operation states, we have to determine the following empirical characteristics of the realizations of the conditional sojourn time of the critical infrastructure operation process at the particular operation states:

- the realizations of the empirical mean values $\bar{\theta}_{bl}$ of the conditional sojourn times $\theta_{bl}$ of the critical infrastructure operation process at the operation state $z_b$ when the next transition is to the operation state $z_l$, according to the formula

$$\bar{\theta}_{bl} = \frac{1}{n_{bl}} \sum_{k=1}^{n_{bl}} \theta_{bl}^k, \; b,l = 1,2,...,v, b \neq l, \tag{7}$$

- the number $\bar{r}_{bl}$ of the disjoint intervals $I_j = < a_{bl}^j, b_{bl}^j)$, $j = 1,2,...,\bar{r}_{bl}$, that include the realizations $\theta_{bl}^k$, $k = 1,2,...,n_{bl}$, of the conditional sojourn times $\theta_{bl}$ at the operation state $z_b$ when the next transition is to the operation state $z_l$, according to the formula

$$\bar{r}_{bl} \cong \sqrt{n_{bl}},$$

- the length $d_{bl}$ of the intervals $I_j = < a_{bl}^j, b_{bl}^j)$, $j = 1,2,...,\bar{r}_{bl}$, according to the formula

$$d_{bl} = \frac{\bar{R}_{bl}}{\bar{r}_{bl} - 1},$$

where

$$\bar{R}_{bl} = \max_{1 \le k \le n_{bl}} \theta_{bl}^k - \min_{1 \le k \le n_{bl}} \theta_{bl}^k,$$

- the ends $a_{bl}^j$, $b_{bl}^j$, of the intervals $I_j = < a_{bl}^j, b_{bl}^j)$, $j = 1,2,...,\bar{r}_{bl}$, according to the formulae

$$a_{bl}^1 = \max\{\min_{1 \le k \le n_{bl}} \theta_{bl}^k - \frac{d_{bl}}{2}, 0\},$$
$$b_{bl}^j = a_{bl}^1 + j d_{bl}, \; j = 1,2,...,\bar{r}_{bl},$$
$$a_{bl}^j = b_{bl}^{j-1}, \; j = 2,3...,\bar{r}_{bl},$$

in such a way that

$$I_1 \cup I_2 \cup ... \cup I_{\bar{r}_{bl}} = < a_{bl}^1, b_{bl}^{\bar{r}_{bl}})$$

and

$$I_i \cap I_j = \emptyset \text{ for all } i \neq j, \; i, j \in \{1,2,...,\bar{r}_{bl}\},$$

- the numbers $n_{bl}^j$ of the realizations $\theta_{bl}^k$ in the intervals $I_j = < a_{bl}^j, b_{bl}^j)$, $j = 1,2,...,\bar{r}_{bl}$, according to the formula

$$n_{bl}^j = \#\{k : \theta_{bl}^k \in I_j, k \in \{1,2,...,n_{bl}\}\}, \; j = 1,2,...,\bar{r}_{bl},$$

where

$$\sum_{j=1}^{\bar{r}_{bl}} n_{bl}^j = n_{bl},$$

whereas the symbol # means the number of elements of the set;

To estimate the parameters of the distributions of the conditional sojourn times of the critical infrastructure operation process at the particular operation states distinguished in [EU-CIRCLE Report D2.1-GMU2, 2016], we proceed respectively in the following way:

- for the uniform distribution with the density function given by (2.5) in [EU-CIRCLE Report D2.1-GMU2, 2016], the estimates of the unknown parameters are:

$$x_{bl} = a_{bl}^1 \; y_{bl} = x_{bl} + \bar{r}_{bl} d_{bl}; \tag{8}$$

- for the triangular distribution with the density function given by (2.6) in [EU-CIRCLE Report D2.1-GMU2, 2016], the estimates of the unknown parameters are:

$$x_{bl} = a_{bl}^1, \; y_{bl} = x_{bl} + \bar{r}_{bl} d_{bl}, \; z_{bl} = \bar{\theta}_{bl};$$

- for the double trapezium distribution with the density function given by (2.7) in [EU-CIRCLE Report D2.1-GMU2, 2016], the estimates of the unknown parameters are:

$$x_{bl} = a_{bl}^{1},\ y_{bl} = x_{bl} + \bar{r}_{bl} d_{bl},\ q_{bl} = \frac{n_{bl}^{1}}{n_{bl} d_{bl}},$$
$$w_{bl} = \frac{n_{bl}^{\bar{r}_{bl}}}{n_{bl} d_{bl}},\ z_{bl} = \bar{\theta}_{bl}; \tag{10}$$

- for the quasi-trapezium distribution with the density function given by (2.8) in [EU-CIRCLE Report D2.1-GMU2, 2016], the estimates of the unknown parameters are:

$$x_{bl} = a_{bl}^{1},\ y_{bl} = x_{bl} + \bar{r}_{bl} d_{bl},\ q_{bl} = \frac{n_{bl}^{1}}{n_{bl} d_{bl}},$$
$$w_{bl} = \frac{n_{bl}^{\bar{r}_{bl}}}{n_{bl} d_{bl}},\ z_{bl}^{1} = \bar{\theta}_{bl}^{1},\ z_{bl}^{2} = \bar{\theta}_{bl}^{2}, \tag{11}$$

where

$$\bar{\theta}_{bl}^{1} = \frac{1}{n_{(me)}} \sum_{k=1}^{n(me)} \theta_{bl}^{k},\ \bar{\theta}_{bl}^{2} = \frac{1}{n_{bl} - n_{(me)}} \sum_{k=n(me)+1}^{n_{bl}} \theta_{bl}^{k},$$
$$n_{(me)} = [\frac{n_{bl}+1}{2}], \tag{12}$$

and $[x]$ denotes the entire part of *x*;

- for the exponential distribution with the density function given by (2.9) in [EU-CIRCLE Report D2.1-GMU2, 2016], the estimates of the unknown parameters are:

$$x_{bl} = a_{bl}^{1},\ \alpha_{bl} = \frac{1}{\bar{\theta}_{bl} - x_{bl}}; \tag{13}$$

- for the Weibull's distribution with the density function given by (2.10) in [EU-CIRCLE Report D2.1-GMU2, 2016], the estimates of the unknown parameters are (the expressions for estimates of parameters $\alpha_{bl}$ and $\beta_{bl}$ are not explicit):

$$x_{bl} = a_{bl}^{1},\ \alpha_{bl} = \frac{n_{bl}}{\sum_{k=1}^{n_{bl}} (\theta_{bl}^{k})^{\beta_{bl}}},$$
$$\alpha_{bl} = \frac{\frac{n_{bl}}{\beta_{bl}} + \sum_{k=1}^{n_{bl}} \ln(\theta_{bl}^{k} - x_{bl})}{\sum_{k=1}^{n_{bl}} (\theta_{bl}^{k})^{\beta_{bl}} \ln(\theta_{bl}^{k} - x_{bl})}; \tag{14}$$

- for the chimney distribution with the density function given by (2.11) in [EU-CIRCLE Report D2.1-GMU2, 2016], the estimates of the unknown parameters are:

$$x_{bl} = a_{bl}^{1},\ y_{bl} = x_{bl} + \bar{r}_{bl} d_{bl}, \tag{15}$$

and moreover, if

$$\hat{n}_{bl} = \max_{1 \le j \le \bar{r}_{bl}} \{n_{bl}^{j}\} \tag{16}$$

and $i$, $i \in \{1,2,...,\bar{r}_{bl}\}$, is the number of the interval including the largest number of realizations i.e. such as that

$$n_{bl}^{i} = \hat{n}_{bl}, \tag{17}$$

then:

- for $i = 1$

either

$$z_{bl}^{1} = x_{bl} + (i-1)d_{bl},\ z_{bl}^{2} = x_{bl} + i d_{bl},\ A_{bl} = 0,$$
$$C_{bl} = \frac{n_{bl}^{i}}{n_{bl}},\ D_{bl} = \frac{n_{bl}^{i+1} + ... + n_{bl}^{\bar{r}_{bl}}}{n_{bl}}, \tag{18}$$

while

$$n_{bl}^{i+1} = 0 \text{ or } n_{bl}^{i+1} \ne 0 \text{ and } \frac{n_{bl}^{i}}{n_{bl}^{i+1}} \ge 3, \tag{19}$$

or

$$z_{bl}^{1} = x_{bl} + (i-1)d_{bl},$$
$$z_{bl}^{2} = x_{bl} + (i+1)d_{bl},\ A_{bl} = 0,$$
(20)

$$C_{bl} = \frac{n_{bl}^{i} + n_{bl}^{i+1}}{n_{bl}},\ D_{bl} = \frac{n_{bl}^{i+2} + ... + n_{bl}^{\bar{r}_{bl}}}{n_{bl}}, \tag{21}$$

while

$$n_{bl}^{i+1} \ne 0 \text{ and } \frac{n_{bl}^{i}}{n_{bl}^{i+1}} < 3; \tag{22}$$

- for $i = 2,3,...,\bar{r}_{bl} - 1$

either

$$z_{bl}^{1} = x_{bl} + (i-1)d_{bl},\ z_{bl}^{2} = x_{bl} + i d_{bl},$$
$$A_{bl} = \frac{n_{bl}^{1} + ... + n_{bl}^{i-1}}{n_{bl}},\ C_{bl} = \frac{n_{bl}^{i}}{n_{bl}}, \tag{23}$$

$$D_{bl} = \frac{n_{bl}^{i+1} + \ldots + n_{bl}^{\bar{r}_{bl}}}{n_{bl}}, \quad (24)$$

while

$$n_{bl}^{i-1} = 0 \text{ or } n_{bl}^{i-1} \neq 0 \text{ and } \frac{n_{bl}^{i}}{n_{bl}^{i-1}} \geq 3 \quad (25)$$

and while

$$n_{bl}^{i+1} = 0 \text{ or } n_{bl}^{i+1} \neq 0 \text{ and } \frac{n_{bl}^{i}}{n_{bl}^{i+1}} \geq 3, \quad (26)$$

or

$$z_{bl}^{1} = x_{bl} + (i-1)d_{bl},\ z_{bl}^{2} = x_{bl} + (i+1)d_{bl},$$
$$A_{bl} = \frac{n_{bl}^{1} + \ldots + n_{bl}^{i-1}}{n_{bl}},$$

(27)

$$C_{bl} = \frac{n_{bl}^{i} + n_{bl}^{i+1}}{n_{bl}},\ D_{bl} = \frac{n_{bl}^{i+2} + \ldots + n_{bl}^{\bar{r}_{bl}}}{n_{bl}},$$

(28)

while

$$n_{bl}^{i-1} = 0 \text{ or } n_{bl}^{i-1} \neq 0 \text{ and } \frac{n_{bl}^{i}}{n_{bl}^{i-1}} \geq 3 \quad (29)$$

and while

$$n_{bl}^{i+1} \neq 0 \text{ and } \frac{n_{bl}^{i}}{n_{bl}^{i+1}} < 3, \quad (30)$$

or

$$z_{bl}^{1} = x_{bl} + (i-2)d_{bl},\ z_{bl}^{2} = x_{bl} + id_{bl},$$
$$A_{bl} = \frac{n_{bl}^{1} + \ldots + n_{bl}^{i-2}}{n_{bl}},\ C_{bl} = \frac{n_{bl}^{i-1} + n_{bl}^{i}}{n_{bl}}, \quad (31)$$
$$D_{bl} = \frac{n_{bl}^{i+1} + \ldots + n_{bl}^{\bar{r}_{bl}}}{n_{bl}}, \quad (32)$$

while

$$n_{bl}^{i-1} \neq 0 \text{ and } \frac{n_{bl}^{i}}{n_{bl}^{i-1}} < 3 \quad (33)$$

and while

$$n_{bl}^{i+1} = 0 \text{ or } n_{bl}^{i+1} \neq 0 \text{ and } \frac{n_{bl}^{i}}{n_{bl}^{i+1}} \geq 3, \quad (34)$$

or

$$z_{bl}^{1} = x_{bl} + (i-2)d_{bl},\ z_{bl}^{2} = x_{bl} + (i+1)d_{bl},$$
$$A_{bl} = \frac{n_{bl}^{1} + \ldots + n_{bl}^{i-2}}{n_{bl}}, \quad (35)$$
$$C_{bl} = \frac{n_{bl}^{i-1} + n_{bl}^{i} + n_{bl}^{i+1}}{n_{bl}},\ D_{bl} = \frac{n_{bl}^{i+2} + \ldots + n_{bl}^{\bar{r}_{bl}}}{n_{bl}}, \quad (36)$$

while

$$n_{bl}^{i-1} \neq 0 \text{ and } \frac{n_{bl}^{i}}{n_{bl}^{i-1}} < 3 \quad (37)$$

and while

$$n_{bl}^{i+1} \neq 0 \text{ and } \frac{n_{bl}^{i}}{n_{bl}^{i+1}} < 3; \quad (38)$$

- for $i = \bar{r}_{bl}$

either

$$z_{bl}^{1} = x_{bl} + (i-1)d_{bl},\ z_{bl}^{2} = x_{bl} + id_{bl},$$
$$A_{bl} = \frac{n_{bl}^{1} + \ldots + n_{bl}^{i-1}}{n_{bl}}, \quad (39)$$
$$C_{bl} = \frac{n_{bl}^{i}}{n_{bl}},\ D_{bl} = 0, \quad (40)$$

while

$$n_{bl}^{i-1} = 0 \text{ or } n_{bl}^{i-1} \neq 0 \text{ and } \frac{n_{bl}^{i}}{n_{bl}^{i-1}} \geq 3, \quad (41)$$

or

$$z_{bl}^{1} = x_{bl} + (i-2)d_{bl},\ z_{bl}^{2} = x_{bl} + id_{bl},$$
$$A_{bl} = \frac{n_{bl}^{1} + \ldots + n_{bl}^{i-2}}{n_{bl}}, \quad (42)$$
$$C_{bl} = \frac{n_{bl}^{i-1} + n_{bl}^{i}}{n_{bl}},\ D_{bl} = 0, \quad (43)$$

while

$$n_{bl}^{i-1} \neq 0 \text{ and } \frac{n_{bl}^{i}}{n_{bl}^{i-1}} < 3. \quad (44)$$

## 2.4. Identification of distribution functions of critical infrastructure operation process conditional sojourn times at operation states

To formulate and next to verify the non-parametric hypothesis concerning the form of the distribution of the critical infrastructure operation process conditional sojourn time $\theta_{bl}$ at the operation state $z_b$ when the next transition is to the operation state $z_l$, on the basis of at least 30 its realizations $\theta_{bl}^{k}$, $k = 1,2,...,n_{bl}$, it is due to proceed according to the following scheme:

- to construct and to plot the realization of the histogram of the critical infrastructure operation process conditional sojourn time $\theta_{bl}$ at the operation state $z_b$, defined by the following formula

$$\bar{h}_{n_{bl}}(t) = \frac{n_{bl}^{j}}{n_{bl}} \text{ for } t \in I_j, \quad (45)$$

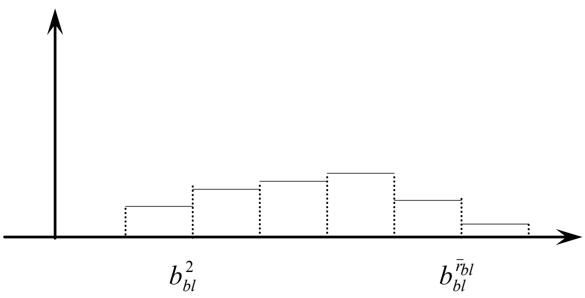


*Figure 1.* The graph of the realization of the histogram of the system operation process conditional sojourn time $\theta_{bl}$ at the operation state $z_b$

- to analyze the realization of the histogram $\bar{h}_{n_{bl}}(t)$, comparing it with the graphs of the density functions $h_{bl}(t)$ of the previously distinguished in [EU-CIRCLE Report D2.1-GMU2, 2016] distributions, to select one of them and to formulate the null hypothesis $H_0$, concerning the unknown form of the distribution of the conditional sojourn time $\theta_{bl}$ in the following form:

$H_0$: The critical infrastructure operation process conditional sojourn time $\theta_{bl}$ at the operation state $z_b$ when the next transition is to the operation state $z_l$, has the distribution with the density function $h_{bl}(t)$;

- to join each of the intervals $I_j$ that has the number $n_{bl}^{j}$ of realizations less than 4 either with the neighbour interval $I_{j+1}$ or with the neighbour interval $I_{j-1}$ this way that the numbers of realizations in all intervals are not less than 4;
- to fix a new number of intervals $\bar{\bar{r}}_{bl}$;
- to determine new intervals

$$\bar{I}_j = <\bar{a}_{bl}^{j}, \bar{b}_{bl}^{j}), \ j = 1,2,..,\bar{\bar{r}}_{bl};$$

- to fix the numbers $\bar{n}_{bl}^{j}$ of realizations in new intervals $\bar{I}_j$, $j = 1,2,..,\bar{\bar{r}}_{bl}$;
- to calculate the hypothetical probabilities that the variable $\theta_{bl}$ takes values from the interval $\bar{I}_j$, under the assumption that the hypothesis $H_0$ is true, i.e. the probabilities

$$p_j = P(\theta_{bl} \in \bar{I}_j) = P(\bar{a}_{bl}^{j} \leq \theta_{bl} < \bar{b}_{bl}^{j})$$

$$= H_{bl}(\bar{b}_{bl}^{j}) - H_{bl}(\bar{a}_{bl}^{j}), \ j = 1,2,..,\bar{\bar{r}}_{bl},$$

where $H_{bl}(\bar{b}_{bl}^{j})$ and $H_{bl}(\bar{a}_{bl}^{j})$ are the values of the distribution function $H_{bl}(t)$ of the random variable $\theta_{bl}$ corresponding to the density function $h_{bl}(t)$ assumed in the null hypothesis $H_0$;

- to calculate the realization of the $\chi^2$ (chi-square)-Pearson's statistics $U_{n_{bl}}$, according to the formula

$$u_{n_{bl}} = \sum_{j=1}^{\bar{\bar{r}}_{bl}} \frac{(\bar{n}_{bl}^{j} - n_{bl} p_j)^2}{n_{bl} p_j}; \quad (47)$$

- to assume the significance level $\alpha$ ($\alpha = 0.01$, $\alpha = 0.02$, $\alpha = 0.05$ or $\alpha = 0.10$) of the test;
- to fix the number $\bar{\bar{r}}_{bl} - l - 1$ of degrees of freedom, substituting for $l$ dependent on the distinguished in [EU-CIRCLE Report D2.1-GMU2, 2016] distributions respectively the following values: $l = 0$ for the uniform, triangular, double trapezium, quasi-trapezium and chimney distributions, $l = 1$ for the exponential distribution and $l = 2$ for the Weibull's distribution;
- to read from the Tables of the $\chi^2$ – Pearson's distribution the value $u_\alpha$ for the fixed values of the significance level $\alpha$ and the number of degrees of

freedom $\bar{\bar{r}}_{bl} - l - 1$ such that the following equality holds

$$P(U_{n\,bl} > u_{\alpha}) = \alpha, \tag{48}$$

and next to determine the critical domain in the form of the interval $(u_{\alpha}, +\infty)$ and the acceptance domain in the form of the interval $< 0, u_{\alpha} >$,

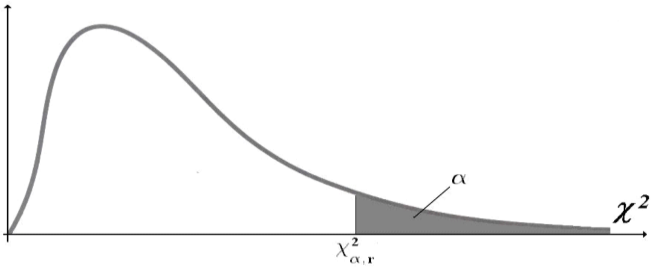

*Figure 2.* The graphical interpretation of the critical interval and the acceptance interval for the chi-square goodness-of-fit test

- to compare the obtained value $u_{n_{bl}}$ of the realization of the statistics $U_{n_{bl}}$ with the read from the Tables critical value $u_{\alpha}$ of the chi-square random variable and to decide on the previously formulated null hypothesis $H_0$ in the following way: if the value $u_{n_{bl}}$ does not belong these to the critical domain, i.e. when $u_{n_{bl}} \le u_{\alpha}$, then we do not reject the hypothesis $H_0$, otherwise if the value $u_{n_{bl}}$ belongs to the critical domain, i.e. when $u_{n_{bl}} > u_{\alpha}$, then we reject the hypothesis $H_0$.

## 2.5. Testing uniformity of statistical data of critical infrastructure operation processes

The statistical data that are needed in Section 7.2.4 for estimating the unknown parameters of the critical infrastructure operation process very often are coming from different experiments of the same operation process and they are collected into separate data sets. Before joining them into one set of data in order to do the unknown parameters evaluation with the methods and procedures described in Section 2.4, we have to make the uniformity testing these statistical data sets.

### *2.5.1. Procedure of critical infrastructure operation process data collection*

To make the uniformity testing of the statistical data collected in two separate data sets coming from the same system operation process realizations in two different experiments, we should collect necessary statistical data performing the following steps:

i) to fix two independent experiments of the critical infrastructure operation process data collection and their following basic parameters:

- the duration times of the experiments $\Theta_1$, $\Theta_2$,
- the critical infrastructure operation processes single realizations,
- the numbers of the investigated (observed) realizations of the critical infrastructure operation process $n_1(0)$, $n_2(0)$;

ii) to fix and to collect the following statistical data concerned with the empirical distributions of the conditional sojourn times $\theta^1_{bl}$ and $\theta^2_{bl}$, $b, l \in \{1,2,\dots,\nu\}$, $b \ne l$, of the critical infrastructure operation process at the particular operation states, respectively in the first experiment and in the second experiment:

- the number of realizations

$$n^1_{bl},\ b,l \in \{1,2,\dots,\nu\},\ b \ne l,$$

of the sojourn time $\theta^1_{bl}$, $b,l \in \{1,2,\dots,\nu\}$, in the first experiment,

- the sample of non-decreasing ordered realizations

$$\theta^{1k}_{bl},\ k = 1,2,\dots,n^1_{bl},\ b \ne l, \tag{49}$$

of the sojourn time $\theta^1_{bl}$, $b,l \in \{1,2,\dots,\nu\}$, in the first experiment,

- the number of realizations

$$n^2_{bl},\ b,l \in \{1,2,\dots,\nu\},\ b \ne l,$$

of the sojourn time $\theta^2_{bl}$, $b,l \in \{1,2,\dots,\nu\}$, in the second experiment,

- the sample of non-decreasing ordered realizations

$$\theta^{2k}_{bl},\ k = 1,2,\dots,n^2_{bl},\ b \ne l, \tag{50}$$

of the sojourn time $\theta^2_{bl}$, $b,l \in \{1,2,\dots,\nu\}$, in the second experiment.

### *2.5.2. Procedure of testing uniformity of distributions of critical infrastructure operation process conditional sojourn times at operation states*

We consider test $\lambda$ based on Kolmogorov-Smirnov theorem [Kołowrocki, Soszyńska-Budny, 2011] that can be used for testing whether two independent samples of realizations of the conditional sojourn time $\theta_{bl}$, $b,l \in \{1,2,...,\nu\}$, $b \neq l$, at the particular operation states of the critical infrastructure operation process are drawn from the population with the same distribution.

We assume that we have defined in previous section two independent samples of non-decreasing ordered realizations (49) and (50) of the sojourn times $\theta_{bl}^1$ and $\theta_{bl}^2, b,l \in \{1,2,...,\nu\}$, $b \neq l$, coming from two different experiments, respectively composed of $n_{bl}^1$ and $n_{bl}^2$ realizations and we define their corresponding empirical distribution functions

$$H_{bl}^1(t) = \frac{1}{n_{bl}^1} \# \{k : \theta_{bl}^{1k} < t, k \in \{1,2,...,n_{bl}^1\}\},$$

(51)

$t \geq 0$, $b,l \in \{1,2,...,\nu\}$, $b \neq l$,

and

$$H_{bl}^2(t) = \frac{1}{n_{bl}^2} \# \{k : \theta_{bl}^{2k} < t, k \in \{1,2,...,n_{bl}^2\}\}, \quad (52)$$

$t \geq 0$, $b,l \in \{1,2,...,\nu\}$, $b \neq l$.

Then, according to Kolmogorov-Smirnov theorem [Kołowrocki, Soszyńska-Budny, 2011], the sequence of distribution functions given by the equation

$$Q_{n_1 n_2}(\lambda) = P(D_{n_1 n_2} < \frac{\lambda}{\sqrt{n}}) \quad (53)$$

defined for $\lambda > 0$, where

$$n_1 = n_{bl}^1,\ n_2 = n_{bl}^2,\ n = \frac{n_1 n_2}{n_1 + n_2}, \quad (54)$$

and

$$D_{n_1 n_2} = \max_{-\infty < t < +\infty} |H_{bl}^1(t) - H_{bl}^2(t)|, \quad (55)$$

is convergent, as $n \to \infty$, to the limit distribution function

$$Q(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2},\ \lambda > 0. \quad (56)$$

The distribution function $Q(\lambda)$ given by (56) is called $\lambda$ distribution and its Tables of values are available.

The convergence of the sequence $Q_{n_1 n_2}(\lambda)$ to the $\lambda$ distribution $Q(\lambda)$ means that for sufficiently large $n_1$ and $n_2$ we may use the following approximate formula

$$Q_{n_1 n_2}(\lambda) \cong Q(\lambda). \quad (57)$$

Hence, it follows that if we define the statistic

$$U_n = D_{n_1 n_2} \sqrt{n}, \quad (58)$$

where $D_{n_1 n_2}$ is defined by (9.55), then by (9.53) and (9.57), we have

$$P(U_n < \lambda) = P(D_{n_1 n_2} \sqrt{n} < \lambda)$$
$$= P(D_{n_1 n_2} < \frac{\lambda}{\sqrt{n}}) = Q_{n_1 n_2}(\lambda) \cong Q(\lambda) \quad (59)$$

for $\lambda > 0$.

This result means that in order to formulate and next to verify the hypothesis that the two independent samples of the realizations of the critical infrastructure operation process conditional sojourn times $\theta_{bl}^1$ and $\theta_{bl}^2$, $b,l \in \{1,2,...,\nu\}$, $b \neq l$, at the operation state $z_b$ when the next transition is to the operation state $z_l$ are coming from the population with the same distribution, it is necessary to proceed according to the following scheme:

- to fix the numbers of realizations $n_{bl}^1$ and $n_{bl}^2$ in the samples,

- to collect the realizations (49) and (50) of the conditional sojourn times $\theta_{bl}^1$ and $\theta_{bl}^2$ of the critical infrastructure operation process in the samples,

- to find the realization of the empirical distribution functions $H_{bl}^1(t)$ and $H_{bl}^2(t)$ defined by (51) and (52) respectively, in the following forms:

$$H_{bl}^{1}(t)=\begin{cases}\frac{n_{bl}^{11}}{n_{bl}^{1}}=0, & t\le\theta_{bl}^{11}\\ \frac{n_{bl}^{12}}{n_{bl}^{1}}, & \theta_{bl}^{11}<t\le\theta_{bl}^{12}\\ \frac{n_{bl}^{13}}{n_{bl}^{1}}, & \theta_{bl}^{12}<t\le\theta_{bl}^{13}\\ \cdots & \\ \frac{n_{bl}^{1k}}{n_{bl}^{1}}, & \theta_{bl}^{1k-1}<t\le\theta_{bl}^{1k}\\ \cdots & \\ \frac{n_{bl}^{1n_{bl}^{1}}}{n_{bl}^{1}}, & \theta_{bl}^{1n_{bl}^{1}-1}<t\le\theta_{bl}^{1n_{bl}^{1}}\\ \frac{n_{bl}^{1n_{bl}^{1}+1}}{n_{bl}^{1}}=1, & t\ge\theta_{bl}^{1n_{bl}^{1}}\end{cases}, \quad (60)$$

$$H_{bl}^{2}(t)=\begin{cases}\frac{n_{bl}^{21}}{n_{bl}^{2}}=0, & t\le\theta_{bl}^{21}\\ \frac{n_{bl}^{22}}{n_{bl}^{2}}, & \theta_{bl}^{21}<t\le\theta_{bl}^{22}\\ \frac{n_{bl}^{23}}{n_{bl}^{2}}, & \theta_{bl}^{22}<t\le\theta_{bl}^{23}\\ \cdots & \\ \frac{n_{bl}^{2k}}{n_{bl}^{2}}, & \theta_{bl}^{2k-1}<t\le\theta_{bl}^{2k}\\ \cdots & \\ \frac{n_{bl}^{2n_{bl}^{2}}}{n_{bl}^{2}}, & \theta_{bl}^{2n_{bl}^{2}-1}<t\le\theta_{bl}^{2n_{bl}^{1}}\\ \frac{n_{bl}^{2n_{bl}^{2}+1}}{n_{bl}^{2}}=1, & t\ge\theta_{bl}^{2n_{bl}^{2}}\end{cases}, \quad (61)$$

where

$$n_{bl}^{11}=0,\ n_{bl}^{1\,n_{bl}^{1}+1}=n_{bl}^{1}, \quad (62)$$

and

$$n_{bl}^{1k}=\#\{j:\theta_{bl}^{1j}<\theta_{bl}^{1k},\ j\in\{1,2,\ldots,n_{bl}^{1}\}\},$$
(63)
$$k=2,3,\ldots,n_{bl}^{1},$$

is the number of the sojourn time $\theta_{bl}^{1}$ realizations less than its realization $\theta_{bl}^{1k}$, $k=2,3,\ldots,n_{bl}^{1}$, and respectively

$$n_{bl}^{21}=0,\ n_{bl}^{2\,n_{bl}^{2}+1}=n_{bl}^{2}, \quad (64)$$

And

$$n_{bl}^{2k}=\#\{j:\theta_{bl}^{2j}<\theta_{bl}^{2k},\ j\in\{1,2,\ldots,n_{bl}^{2}\}\}, \quad (65)$$
$$k=2,3,\ldots,n_{bl}^{2},$$

is the number of the sojourn time $\theta_{bl}^{2}$ realizations less than its realization $\theta_{bl}^{2k}$, $k=2,3,\ldots,n_{bl}^{2}$,

- to calculate the realization of the statistic $u_n$ defined by (58) according to the formula

$$u_n=d_{n_{bl}^{1}n_{bl}^{2}}\sqrt{n}, \quad (66)$$

where

$$d_{n_{bl}^{1}n_{bl}^{2}}=\max\{d_{n_{bl}^{1}n_{bl}^{2}}^{1},\ d_{n_{bl}^{1}n_{bl}^{2}}^{2}\},$$

$$d_{n_{bl}^{1}n_{bl}^{2}}^{1}=\max\{|H_{bl}^{1}(\theta_{bl}^{1k})-H_{bl}^{2}(\theta_{bl}^{1k})|,\ k\in\{1,2,\ldots,n_{bl}^{1}\}\}, \quad (68)$$

$$d_{n_{bl}^{1}n_{bl}^{2}}^{2}=\max\{|H_{bl}^{1}(\theta_{bl}^{2k})-H_{bl}^{2}(\theta_{bl}^{2k})|,\ k\in\{1,2,\ldots,n_{bl}^{2}\}\}, \quad (69)$$

$$n=\frac{n_{bl}^{1}n_{bl}^{2}}{n_{bl}^{1}+n_{bl}^{2}}, \quad (70)$$

- to formulate the null hypothesis $H_0$ in the following form:

$H_0$ : The samples of realizations (49) and (50) are coming from the populations with the same distributions,

- to fix the significance level $\alpha$ ($\alpha=0.01$, $\alpha=0.02$, $\alpha=0.05$ or $\alpha=0.10$) of the test,

- to read from the Tables of $\lambda$ distribution, corresponding to $1-\alpha$, the value $\lambda_0$ such that the following equality holds

$$P(U_n<\lambda_0)=Q(\lambda_0)=1-\alpha, \quad (71)$$

- to determine the critical domain in the form of the interval $(\lambda_0, +\infty)$ and the acceptance domain in the form of the interval $(0, \lambda_0 >$,

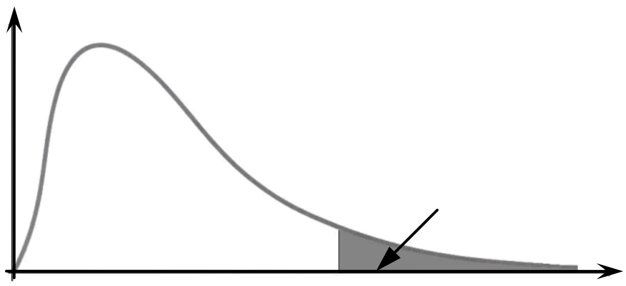

*Figure 3.* The graphical interpretation of the critical domain and the acceptance domain for the two-sample Smirnov-Kolmogorov test

- to compare the obtained value $u_n$ of the realization of the statistics $U_n$ with the read from the Tables critical value $\lambda_0$,
- to decide on the previously formulated null hypothesis $H_0$ in the following way:

if the value $u_n$ does not belong to the critical domain, i.e. when $u_n \le \lambda_0$ then we do not reject the hypothesis $H_0$, otherwise if the value $u_n$ belongs to the critical domain, i.e. when $u_n > \lambda_0$, then we reject the hypothesis $H_0$.

In the case when the null hypothesis $H_0$ is not rejected we may join the statistical data from the considered two separate sets into one new set of data and if there are no other sets of statistical data including the realizations of the sojourn time $\theta_{bl}$, we proceed with the data of this new set in the way described in Sections 2.1-2.4. Otherwise, if there are other sets of statistical data including the realizations of the sojourn time $\theta_{bl}$, we select the next one of them and perform the procedure of this section for data from this set and data from the previously formed new set. We continue this procedure up to the moment when the store of the statistical data sets including the realizations of the sojourn time $\theta_{bl}$, is exhausted.

## 3. Identification of critical infrastructure operation process including operating environment threats based on statistical data

critical infrastructure during its operation including environment threats, at the fixed moment $t$, may be at one of $\nu'$, $\nu' \in N$, different operations states $z'_b$, $b = 1,2,...,\nu'$. Next, we mark by $Z'(t)$, $t \in < 0,+\infty >$, the critical infrastructure operation process related to its operating environment threats, that is a function of a continuous variable $t$, taking discrete values in the set $\{z'_1, z'_2, ..., z'_\nu\}$ of the critical infrastructure operation states defined in [EU-CIRCLE Report D2.1-GMU2, 2016]. We assume a semi-Markov model [Kolowrocki, 2014], [Kolowrocki, Soszynska, 2009d], [Kołowrocki, Soszyńska-Budny, 2011], [Limnios, Oprisan, 2005], [Limnios et all, 2005], [Macci, 2008], [Mercier, 2008], of the critical infrastructure operation process Z'(t) and we mark by $\theta'_{bl}$ its random conditional sojourn times at the operation states $z'_b$ when its next operation state is $z'_l$, $b, l = 1,2,...,\nu'$, $b \ne l$.

Under these assumption, the operation process may be described by the vector $[p'_b(0)]_{1\times\nu'}$ of probabilities of the critical infrastructure operation process including environment threats staying at the particular operations states at the initial moment $t = 0$, the matrix $[p'_{bl}(t)]_{\nu'\times\nu'}$ of the probabilities of the critical infrastructure operation process transitions between the operation states and the matrix $[H'_{bl}(t)]_{\nu'\times\nu'}$ of the distribution functions of the conditional sojourn times $\theta'_{bl}$ of the critical infrastructure operation process at the operation states or equivalently by the matrix $[h'_{bl}(t)]_{\nu'\times\nu'}$ of the density functions of the conditional sojourn times $\theta'_{bl}$, $b, l = 1,2,...,\nu'$, $b \ne l$, of the critical infrastructure operation process at the operation states. These all parameters of the critical infrastructure operation process are unknown and before their use to the prognosis of this process characteristics have to be estimated on the basis of statistical data coming from this operation process realizations.

## 3.1. Defining unknown parameters of critical infrastructure operation process and data collection including operating environment threats

To make the estimation of the unknown parameters of the critical infrastructure operations process, the experiment delivering the necessary statistical data should be precisely planned.
First, before the experiment, we should perform the following preliminary steps:
i) to analyze the critical infrastructure operation process;
ii) to fix or to define the critical infrastructure operation process following general parameters:

- the number of the operation states of the critical infrastructure operation process $\nu'$,
- the operation states of the critical infrastructure operation process $z'_1$, $z'_2$, …, $z'_{\nu'}$;

iii) to fix the possible transitions between the critical infrastructure operation states;
iv) to fix the set of the unknown parameters of the critical infrastructure operation process semi-Markov model.
Next, to estimate the unknown parameters of the critical infrastructure operation process, based on the experiment, we should collect necessary statistical data performing the following steps:
i) to fix and to collect the following statistical data necessary to evaluating the probabilities $p'_b(0)$ of the critical infrastructure operation process staying at the operation states at the initial moment $t = 0$:
- the duration time of the experiment $\Theta'$,
- the number of the investigated (observed) realizations of the critical infrastructure operation process $n'(0)$,
- the vector of the realizations $n'_b(0)$, $b = 1,2,...,\nu'$, of the numbers of staying of the operation process respectively at the operation states $z'_1$, $z'_2$, …, $z'_{\nu'}$, at the initial moments $t = 0$ of all $n'(0)$ observed realizations of the critical infrastructure operation process

$$[n'_b(0)] = [n'_1(0), n'_2(0), ..., n'_{\nu'}(0)],$$

where

$$n'_1(0) + n'_2(0) + n'_{\nu'}(0) = n'(0);$$

ii) to fix and to collect the following statistical data necessary to evaluating the probabilities $p'_{bl}$ of the critical infrastructure operation process transitions between the critical infrastructure operation states:
- the matrix of the realizations of the numbers $n'_{bl}$, *b, l* = 1,2,...,*ν*', *b* ≠ *l*, of the transitions of the critical infrastructure operation process from the operation state $z'_b$ into the operation state $z'_l$ at all observed realizations of the critical infrastructure operation process

$$[n'_{bl}] = \begin{bmatrix} n'_{11} \; n'_{12} \dots n'_{1\nu} \\ n'_{21} \; n'_{22} \dots n'_{2\nu} \\ \dots \\ n'_{\nu'1} \; n'_{\nu'2} \dots n'_{\nu'\nu'} \end{bmatrix},$$

where

$$n'_{bb} = 0 \text{ for } b = 1,2,...,\nu',$$

- the vector of the realizations of the numbers $n'_b$, *b* = 1,2,...,*ν*', of departures of the critical infrastructure operation process from the operation states $z'_b$ (the sum of the numbers of the *b*-th row of matrix $[n'_{bl}]$)

$$[n'_b] = [n'_1, n'_2, ..., n'_\nu],$$

where

$$n'_1 = n'_{11} + n'_{12} + ... + n'_{1\nu'},$$
$$n'_2 = n'_{21} + n'_{22} + ... + n'_{2\nu'},$$
$$\dots$$
$$n_{\nu'} = n_{\nu'1} + n_{\nu'2} + ... + n_{\nu'\nu'};$$

iii) to fix and to collect the following statistical data necessary to evaluating the unknown parameters of the distributions $H'_{bl}(t)$ of the conditional sojourn times $\theta'_{bl}$ of the critical infrastructure operation process at the particular operation states:
- the numbers $n'_{bl}$, *b, l* = 1,2,...,*ν*', $b \neq l$, of realizations of the conditional sojourn times $\theta'_{bl}$, *b, l* = 1,2,...,*ν*', $b \neq l$, of the critical infrastructure operation process at the operation state $z'_b$ when the next transition is to the operation state $z'_l$ during the observation time $\Theta$,

- the realizations $\theta'^{k}_{bl}, k = 1,2, \ldots, n'_{bl}$, of the conditional sojourn times $\theta'_{bl}$ of the critical infrastructure operation process at the operation state $z'_b$ when the next transition is to the operation state $z'_l$ during the observation time $\Theta'$ for each $b, l = 1,2,...,\nu'$, $b \neq l$.

### 3.2. Estimating basic parameters of critical infrastructure operation process including operating environment threats

After collecting the statistical data, it is possible to estimate the unknown parameters of the critical infrastructure operation process performing the following steps:

i) to determine the vector

$$[p'(0)] = [p'_1(0), p'_2(0), \ldots, p'_{\nu'}(0)], \quad (72)$$

of the realizations of the probabilities $p'_b(0)$, $b = 1,2,...,\nu'$, of the critical infrastructure operation process staying at the operation states at the initial moment $t = 0$, according to the formula

$$p'_b(0) = \frac{n'_b(0)}{n'(0)} \text{ for } b = 1,2,...,\nu', \quad (73)$$

where

$$n'(0) = \sum_{b=1}^{\nu'} n'_b(0), \quad (74)$$

is the number of the realizations of the critical infrastructure operation process starting at the initial moment $t = 0$;

ii) to determine the matrix

$$[p'_{bl}] = \begin{bmatrix} p'_{11} \; p'_{12} \cdots p'_{1\nu'} \\ p'_{21} \; p'_{22} \cdots p'_{2\nu'} \\ \cdots \\ p'_{\nu'1} \; p'_{\nu'2} \cdots p'_{\nu'\nu'} \end{bmatrix}, \quad (75)$$

of the realizations of the probabilities $p'_{bl}$, $b,l = 1,2,...,\nu'$, of the critical infrastructure operation process transitions from the operation state $z_b$ to the operation state $z_l$ according to the formula

$$p'_{bl} = \frac{n'_{bl}}{n'_b} \text{ for } b,l = 1,2,...,\nu', b \neq l, \; p_{bb} = 0 \quad (76)$$

for $b = 1,2,...,\nu'$,

where

$$n'_b = \sum_{b \neq l}^{\nu'} n'_{bl}, \; b = 1,2,...,\nu', \quad (77)$$

is the realization of the total number of the critical infrastructure operation process departures from the operation state $z'_b$ during the experiment time $\Theta'$.

### 3.3. Estimating parameters of distributions of critical infrastructure operation process conditional sojourn times at operation states including operating environment threats

Prior to estimating the parameters of the distributions of the conditional sojourn times of the critical infrastructure operation process at the particular operation states, we have to determine the following empirical characteristics of the realizations of the conditional sojourn time of the critical infrastructure operation process at the particular operation states:

- the realizations of the empirical mean values $\bar{\theta}'_{bl}$ of the conditional sojourn times $\theta'_{bl}$ of the critical infrastructure operation process at the operation state $z'_b$ when the next transition is to the operation state $z'_l$, according to the formula

$$\bar{\theta}'_{bl} = \frac{1}{n'_{bl}} \sum_{k=1}^{n'_{bl}} \theta'^{k}_{bl}, \; b,l = 1,2,...,\nu', b \neq l, \quad (78)$$

- the number $\bar{r}'_{bl}$ of the disjoint intervals $I'_j = \langle a'^{j}_{bl}, b'^{j}_{bl})$, $j = 1,2,...,\bar{r}'_{bl}$, that include the realizations $\theta'^{k}_{bl}$, $k = 1,2,...,n'_{bl}$, of the conditional sojourn times $\theta'_{bl}$ at the operation state $z'_b$ when the next transition is to the operation state $z'_l$, according to the formula

$$\bar{r}'_{bl} \cong \sqrt{n'_{bl}},$$

- the length $d'_{bl}$ of the intervals $I'_j = \langle a'^{j}_{bl}, b'^{j}_{bl})$, $j = 1,2,...,\bar{r}'_{bl}$, according to the formula

$$d'_{bl} = \frac{\bar{R}'_{bl}}{\bar{r}'_{bl} - 1},$$

where

$$\bar{R}'_{bl} = \max_{1 \le k \le n'_{bl}} \theta'^{k}_{bl} - \min_{1 \le k \le n'_{bl}} \theta'^{k}_{bl},$$

- the ends $a'^{j}_{bl}$, $b'^{j}_{bl}$, of the intervals $I'_j = < a'^{j}_{bl}, b'^{j}_{bl})$, $j = 1,2,...,\bar{r}'_{bl}$, according to the formulae

$$a'^{1}_{bl} = \max\{\min_{1 \le k \le n'_{bl}} \theta'^{k}_{bl} - \frac{d'_{bl}}{2}, 0\},$$

$$b'^{j}_{bl} = a'^{1}_{bl} + jd'_{bl},\ j = 1,2,...,\bar{r}'_{bl},$$

$$a'^{j}_{bl} = b'^{j-1}_{bl},\ j = 2,3...,\bar{r}'_{bl},$$

in such a way that

$$I'_1 \cup I'_2 \cup ... \cup I'_{\bar{r}'_{bl}} = < a'^{1}_{bl}, b'^{\bar{r}_{bl}}_{bl})$$

and

$$I'_i \cap I'_j = \emptyset \text{ for all } i \ne j,\ i, j \in \{1,2,...,\bar{r}'_{bl}\},$$

- the numbers $n'^{j}_{bl}$ of the realizations $\theta'^{k}_{bl}$ in the intervals $I'_j = < a'^{j}_{bl}, b'^{j}_{bl})$, $j = 1,2,...,'\bar{r}_{bl}$, according to the formula

$$n'^{j}_{bl} = \#\{k : \theta'^{k}_{bl} \in I'_j, k \in \{1,2,...,n'_{bl}\}\},\ j = 1,2,...,\bar{r}'_{bl},$$

where

$$\sum_{j=1}^{\bar{r}_{bl}} n'^{j}_{bl} = n'_{bl},$$

whereas the symbol # means the number of elements of the set;
To estimate the parameters of the distributions of the conditional sojourn times of the critical infrastructure operation process at the particular operation states distinguished in [EU-CIRCLE Report D2.1-GMU2, 2016], we proceed respectively in the following way:
- for the uniform distribution with the density function given by (2.5) in [EU-CIRCLE Report D2.1-GMU2, 2016], the estimates of the unknown parameters are:

$$x_{bl} = a'^{1}_{bl}\ y_{bl} = x_{bl} + \bar{r}'_{bl} d'_{bl}; \quad (79)$$

- for the triangular distribution with the density function given by (2.6) in [EU-CIRCLE Report D2.1-GMU2, 2016], the estimates of the unknown parameters are:

$$x_{bl} = a'^{1}_{bl},\ y_{bl} = x_{bl} + \bar{r}'_{bl} d'_{bl},\ z_{bl} = \bar{\theta}'_{bl}; \quad (80)$$

- for the double trapezium distribution with the density function given by (2.7) in [EU-CIRCLE Report D2.1-GMU2, 2016], the estimates of the unknown parameters are:

$$x_{bl} = a'^{1}_{bl},\ y_{bl} = x_{bl} + \bar{r}'_{bl} d'_{bl},\ q_{bl} = \frac{n'^{1}_{bl}}{n'_{bl} d'_{bl}},$$
$$w_{bl} = \frac{n'^{\bar{r}'_{bl}}_{bl}}{n'_{bl} d'_{bl}},\ z_{bl} = \bar{\theta}'_{bl}; \quad (81)$$

- for the quasi-trapezium distribution with the density function given by (2.8) in [EU-CIRCLE Report D2.1-GMU2, 2016], the estimates of the unknown parameters are:

$$x_{bl} = a'^{1}_{bl},\ y_{bl} = x_{bl} + \bar{r}'_{bl} d'_{bl},\ q_{bl} = \frac{n'^{1}_{bl}}{n'_{bl} d'_{bl}},$$
$$w_{bl} = \frac{n'^{\bar{r}'_{bl}}_{bl}}{n'_{bl} d'_{bl}},\ z^1_{bl} = \bar{\theta}'^{1}_{bl},\ z^2_{bl} = \bar{\theta}'^{2}_{bl}, \quad (82)$$

Where

$$\bar{\theta}'^{1}_{bl} = \frac{1}{n'_{(me)}} \sum_{k=1}^{n'_{(me)}} \theta'^{k}_{bl},\ \bar{\theta}'^{2}_{bl} = \frac{1}{n'_{bl} - n'_{(me)}} \sum_{k=n'_{(me)}+1}^{n'_{bl}} \theta'^{k}_{bl},$$
$$n'_{(me)} = [\frac{n'_{bl}+1}{2}], \quad (83)$$

and $[x]$ denotes the entire part of $x$;
- for the exponential distribution with the density function given by (2.9) in [EU-CIRCLE Report D2.1-GMU2, 2016], the estimates of the unknown parameters are:

$$x_{bl} = a'^{1}_{bl},\ \alpha_{bl} = \frac{1}{\bar{\theta}'_{bl} - x_{bl}}; \quad (84)$$

- for the Weibull's distribution with the density function given by (2.10) in [EU-CIRCLE Report D2.1-GMU2, 2016], the estimates of the unknown parameters are (the expressions for estimates of parameters $\alpha_{bl}$ and $\beta_{bl}$ are not explicit):

$$x_{bl} = a'^{1}_{bl},\ \alpha_{bl} = \frac{n'_{bl}}{\sum_{k=1}^{n'_{bl}} (\theta'^{k}_{bl})^{\beta_{bl}}},$$

$$\alpha_{bl} = \frac{\frac{n'_{bl}}{\beta_{bl}} + \sum_{k=1}^{n'_{bl}} \ln(\theta'^{k}_{bl} - x_{bl})}{\sum_{k=1}^{n'_{bl}} (\theta'^{k}_{bl})^{\beta_{bl}} \ln(\theta'^{k}_{bl} - x_{bl})}; \quad (85)$$

- for the chimney distribution with the density function given by (2.11) in [EU-CIRCLE Report D2.1-GMU2, 2016], the estimates of the unknown parameters are:

$$x_{bl} = a'^{1}_{bl}, \; y_{bl} = x_{bl} + \bar{r}'_{bl} d'_{bl}, \quad (86)$$

and moreover, if

$$\hat{n}'_{bl} = \max_{1 \le j \le \bar{r}'_{bl}} \{n'^{j}_{bl}\} \quad (87)$$

and $i$, $i \in \{1,2,\ldots,\bar{r}'_{bl}\}$, is the number of the interval including the largest number of realizations i.e. such as that

$$n'^{i}_{bl} = \hat{n}'_{bl}, \quad (88)$$

then:

- for $i = 1$

either

$$z^{1}_{bl} = x_{bl} + (i-1)d'_{bl}, \; z^{2}_{bl} = x_{bl} + id'_{bl}, \; A_{bl} = 0,$$

$$C_{bl} = \frac{n'^{i}_{bl}}{n'_{bl}}, \; D_{bl} = \frac{n'^{i+1}_{bl} + \ldots + n'^{\bar{r}'_{bl}}_{bl}}{n'_{bl}}, \quad (89)$$

while

$$n'^{i+1}_{bl} = 0 \text{ or } n'^{i+1}_{bl} \ne 0 \quad (90)$$

and $\frac{n'^{i}_{bl}}{n'^{i+1}_{bl}} \ge 3$,

or

$$z^{1}_{bl} = x_{bl} + (i-1)d'_{bl},$$

$$z^{2}_{bl} = x_{bl} + (i+1)d'_{bl}, \; A_{bl} = 0, \quad (91)$$

$$C_{bl} = \frac{n'^{i}_{bl} + n'^{i+1}_{bl}}{n'_{bl}}, \; D_{bl} = \frac{n'^{i+2}_{bl} + \ldots + n'^{\bar{r}'_{bl}}_{bl}}{n_{bl}}, \quad (92)$$

while

$$n'^{i+1}_{bl} \ne 0 \text{ and } \frac{n'^{i}_{bl}}{n'^{i+1}_{bl}} < 3; \quad (93)$$

- for $i = 2,3,\ldots,\bar{r}'_{bl} - 1$

either

$$z^{1}_{bl} = x_{bl} + (i-1)d'_{bl}, \; z^{2}_{bl} = x_{bl} + id'_{bl},$$

$$A_{bl} = \frac{n'^{1}_{bl} + \ldots + n'^{i-1}_{bl}}{n'_{bl}}, \; C_{bl} = \frac{n'^{i}_{bl}}{n'_{bl}},$$

(94)

$$D_{bl} = \frac{n'^{i+1}_{bl} + \ldots + n'^{\bar{r}_{bl}}_{bl}}{n'_{bl}}, \quad (95)$$

while

$$n'^{i-1}_{bl} = 0 \text{ or } n'^{i-1}_{bl} \ne 0 \text{ and } \frac{n'^{i}_{bl}}{n'^{i-1}_{bl}} \ge 3 \quad (96)$$

and while

$$n'^{i+1}_{bl} = 0 \text{ or } n'^{i+1}_{bl} \ne 0 \text{ and } \frac{n'^{i}_{bl}}{n'^{i+1}_{bl}} \ge 3, \quad (97)$$

or

$$z^{1}_{bl} = x_{bl} + (i-1)d'_{bl}, \; z^{2}_{bl} = x_{bl} + (i+1)d'_{bl},$$

$$A_{bl} = \frac{n'^{1}_{bl} + \ldots + n'^{i-1}_{bl}}{n'_{bl}}, \quad (98)$$

$$C_{bl} = \frac{n'^{i}_{bl} + n'^{i+1}_{bl}}{n'_{bl}}, \; D_{bl} = \frac{n'^{i+2}_{bl} + \ldots + n'^{\bar{r}_{bl}}_{bl}}{n'_{bl}}, \quad (99)$$

while

$$n'^{i-1}_{bl} = 0 \text{ or } n'^{i-1}_{bl} \ne 0 \text{ and } \frac{n'^{i}_{bl}}{n'^{i-1}_{bl}} \ge 3 \quad (100)$$

and while

$$n'^{i+1}_{bl} \ne 0 \text{ and } \frac{n'^{i}_{bl}}{n'^{i+1}_{bl}} < 3, \quad (101)$$

o r

$$z^{1}_{bl} = x_{bl} + (i-2)d'_{bl}, \; z^{2}_{bl} = x_{bl} + id'_{bl},$$

$$A_{bl} = \frac{n'^{1}_{bl} + \ldots + n'^{i-2}_{bl}}{n'_{bl}}, \; C_{bl} = \frac{n'^{i-1}_{bl} + n'^{i}_{bl}}{n'_{bl}}, \quad (102)$$

$$D_{bl} = \frac{n'^{i+1}_{bl} + \ldots + n'^{\bar{r}_{bl}}_{bl}}{n'_{bl}}, \quad (103)$$

while

$$n'^{i-1}_{bl} \ne 0 \text{ and } \frac{n'^{i}_{bl}}{n'^{i-1}_{bl}} < 3 \quad (104)$$

and while

$$n'^{i+1}_{bl} = 0 \text{ or } n'^{i+1}_{bl} \neq 0 \text{ and } \frac{n'^{i}_{bl}}{n'^{i+1}_{bl}} \geq 3, \quad (105)$$

or

$$z^1_{bl} = x_{bl} + (i-2)d'_{bl},\ z^2_{bl} = x_{bl} + (i+1)d'_{bl},$$

$$A_{bl} = \frac{n'^{1}_{bl} + \ldots + n'^{i-2}_{bl}}{n'_{bl}}, \quad (106)$$

$$C_{bl} = \frac{n'^{i-1}_{bl} + n'^{i}_{bl} + n'^{i+1}_{bl}}{n'_{bl}},\ D_{bl} = \frac{n'^{i+2}_{bl} + \ldots + n'^{\bar{r}bl}_{bl}}{n'_{bl}}, \quad (107)$$

while

$$n''^{i-1}_{bl} \neq 0 \text{ and } \frac{n'^{i}_{bl}}{n'^{i-1}_{bl}} < 3 \quad (108)$$

and while

$$n'^{i+1}_{bl} \neq 0 \text{ and } \frac{n'^{i}_{bl}}{n'^{i+1}_{bl}} < 3; \quad (109)$$

- for $i = \bar{r}'_{bl}$

either

$$z'^{1}_{bl} = x_{bl} + (i-1)d'_{bl},\ z^2_{bl} = x_{bl} + id'_{bl},$$

$$A_{bl} = \frac{n'^{1}_{bl} + \ldots + n'^{i-1}_{bl}}{n'_{bl}}, \quad (110)$$

$$C_{bl} = \frac{n'^{i}_{bl}}{n'_{bl}},\ D_{bl} = 0, \quad (111)$$

while

$$n'^{i-1}_{bl} = 0 \text{ or } n'^{i-1}_{bl} \neq 0 \text{ and } \frac{n'^{i}_{bl}}{n'^{i-1}_{bl}} \geq 3, \quad (112)$$

or

$$z^1_{bl} = x_{bl} + (i-2)d'_{bl},\ z^2_{bl} = x_{bl} + id'_{bl},$$

$$A_{bl} = \frac{n'^{1}_{bl} + \ldots + n'^{i-2}_{bl}}{n'_{bl}}, \quad (113)$$

$$C_{bl} = \frac{n'^{i-1}_{bl} + n'^{i}_{bl}}{n'_{bl}},\ D_{bl} = 0, \quad (114)$$

while

$$n'^{i-1}_{bl} \neq 0 \text{ and } \frac{n'^{i}_{bl}}{n'^{i-1}_{bl}} < 3. \quad (115)$$

### 3.4. Identification of distribution functions of critical infrastructure operation process conditional sojourn times at operation states including operating environment threats

To formulate and next to verify the non-parametric hypothesis concerning the form of the distribution of the critical infrastructure operation process conditional sojourn time $\theta'_{bl}$ at the operation state $z'_b$ when the next transition is to the operation state $z'_l$, on the basis of at least 30 its realizations $\theta'^{k}_{bl}$, $k = 1,2,\ldots,n'_{bl}$, it is due to proceed according to the following scheme:

- to construct and to plot the realization of the histogram of the critical infrastructure operation process conditional sojourn time $\theta'_{bl}$ at the operation state $z_{b'}$, defined by the following formula

$$\bar{h}'_{n'bl}(t) = \frac{n'^{j}_{bl}}{n'_{bl}} \text{ for } t \in I'_j, \quad (116)$$

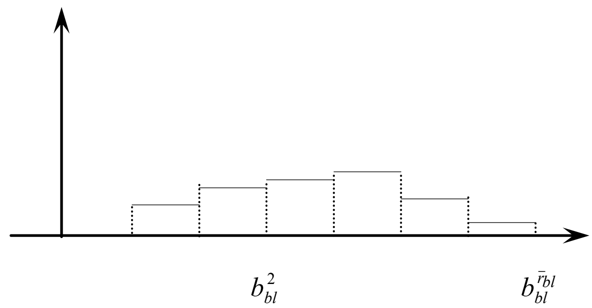


*Figure 4.* The graph of the realization of the histogram of the system operation process conditional sojourn time $\theta'_{bl}$ at the operation state $z'_b$

- to analyse the realization of the histogram $\bar{h}'_{n'bl}(t)$, comparing it with the graphs of the density functions $h_{bl}(t)$ of the previously distinguished in [EU-CIRCLE Report D2.1-GMU2, 2016] distributions, to select one of them and to formulate the null hypothesis $H_0$, concerning the unknown form of the distribution of the conditional sojourn time $\theta'_{bl}$ in the following form:

$H_0$: The critical infrastructure operation process conditional sojourn time $\theta'_{bl}$ at the operation state $z'_b$ when the next transition is to the operation state

$z'_l$, has the distribution with the density function $h_{bl}(t)$;

- to join each of the intervals $I'_j$ that has the number $n'^j_{bl}$ of realizations less than 4 either with the neighbour interval $I'_{j+1}$ or with the neighbour interval $I'_{j-1}$ this way that the numbers of realizations in all intervals are not less than 4;
- to fix a new number of intervals $\bar{\bar{r}}'_{bl}$;
- to determine new intervals

$$\bar{I}'_j = < \bar{a}'^j_{bl}, \bar{b}'^j_{bl}), \; j = 1,2,..,\bar{\bar{r}}'_{bl};$$

- to fix the numbers $\bar{n}'^j_{bl}$ of realizations in new intervals $\bar{I}'_j$, $j = 1,2,..,\bar{\bar{r}}'_{bl}$;
- to calculate the hypothetical probabilities that the variable $\theta'_{bl}$ takes values from the interval $\bar{I}'_j$, under the assumption that the hypothesis $H_0$ is true, i.e. the probabilities

$$p_j = P(\theta'_{bl} \in \bar{I}'_j) = P(\bar{a}'^j_{bl} \le \theta'_{bl} < \bar{b}'^j_{bl})$$
$$= H'_{bl}(\bar{b}'^j_{bl}) - H'_{bl}(\bar{a}'^j_{bl}), \; j = 1,2,..,\bar{\bar{r}}'_{bl}, \qquad (117)$$

where $H'_{bl}(\bar{b}'^j_{bl})$ and $H'_{bl}(\bar{a}'^j_{bl})$ are the values of the distribution function $H'_{bl}(t)$ of the random variable $\theta'_{bl}$ corresponding to the density function $h_{bl}(t)$ assumed in the null hypothesis $H_0$;

- to calculate the realization of the $\chi^2$ (chi-square)-Pearson's statistics $U_{nbl}$, according to the formula

$$u_{nbl} = \sum_{j=1}^{\bar{\bar{r}}'_{bl}} \frac{(\bar{n}'^j_{bl} - n'_{bl}\, p_j)^2}{n'_{bl}\, p_j}; \qquad (118)$$

- to assume the significance level $\alpha$ ($\alpha = 0.01$, $\alpha = 0.02$, $\alpha = 0.05$ or $\alpha = 0.10$) of the test;
- to fix the number $\bar{\bar{r}}'_{bl} - l - 1$ of degrees of freedom, substituting for $l$ dependent on the distinguished in [EU-CIRCLE Report D2.1-GMU2, 2016] distributions respectively the following values: $l = 0$ for the uniform, triangular, double trapezium, quasi-trapezium and chimney distributions, $l = 1$ for the exponential distribution and $l = 2$ for the Weibull's distribution;
- to read from the Tables of the $\chi^2$ – Pearson's distribution the value $u_\alpha$ for the fixed values of the significance level $\alpha$ and the number of degrees of freedom $\bar{\bar{r}}_{bl} - l - 1$ such that the following equality holds

$$P(U_{n\,bl} > u_\alpha) = \alpha, \qquad (119)$$

and next to determine the critical domain in the form of the interval $(u_\alpha, +\infty)$ and the acceptance domain in the form of the interval $< 0, u_\alpha >$,

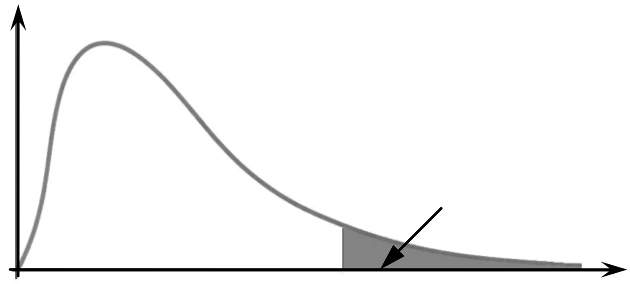

*Figure 5*. The graphical interpretation of the critical interval and the acceptance interval for the chi-square goodness-of-fit test

- to compare the obtained value $u_{n_{bl}}$ of the realization of the statistics $U_{n_{bl}}$ with the read from the Tables critical value $u_\alpha$ of the chi-square random variable and to decide on the previously formulated null hypothesis $H_0$ in the following way: if the value $u_{n_{bl}}$ does not belong these to the critical domain, i.e. when $u_{n_{bl}} \le u_\alpha$, then we do not reject the hypothesis $H_0$, otherwise if the value $u_{n_{bl}}$ belongs to the critical domain, i.e. when $u_{n_{bl}} > u_\alpha$, then we reject the hypothesis $H_0$.

The statistical data that are needed in Section 3.5 for estimating the unknown parameters of the critical infrastructure operation process very often are coming from different experiments of the same operation process and they are collected into separate data sets. Before joining them into one set of data in order to do the unknown parameters evaluation with the methods and procedures described in Section 3.5, we have to make the uniformity testing these statistical data sets.

### *3.5.1. Procedure of critical infrastructure operation process data collection including operating environment threats*

To make the uniformity testing of the statistical data collected in two separate data sets coming from the same system operation process realizations in two different experiments, we should collect necessary statistical data performing the following steps:

i) to fix two independent experiments of the critical infrastructure operation process data collection and their following basic parameters:

- the duration times of the experiments $\Theta'_1$, $\Theta'_2$,
- the critical infrastructure operation processes single realizations,
- the numbers of the investigated (observed) realizations of the critical infrastructure operation process $n'_1(0)$, $n'_2(0)$;

ii) to fix and to collect the following statistical data concerned with the empirical distributions of the conditional sojourn times $\theta'^1_{bl}$ and $\theta'^2_{bl}$, $b,l \in \{1,2,...,\nu'\}$, $b \neq l$, of the critical infrastructure operation process at the particular operation states, respectively in the first experiment and in the second experiment:

- the number of realizations

$$n'^1_{bl},\ b,l \in \{1,2,...,\nu'\},\ b \neq l,$$

of the sojourn time $\theta'^1_{bl}$, $b,l \in \{1,2,...,\nu'\}$, in the first experiment,

- the sample of non-decreasing ordered realizations

$$\theta'^{1k}_{bl},\ k = 1,2,...,n'^1_{bl},\ b \neq l, \tag{120}$$

of the sojourn time $\theta'^1_{bl}$, $b,l \in \{1,2,...,\nu'\}$, in the first experiment,

- the number of realizations

$$n'^2_{bl},\ b,l \in \{1,2,...,\nu'\},\ b \neq l,$$

of the sojourn time $\theta'^2_{bl}$, $b,l \in \{1,2,...,\nu'\}$, in the second experiment,

- the sample of non-decreasing ordered realizations

$$\theta'^{2k}_{bl},\ k = 1,2,...,n'^2_{bl},\ b \neq l, \tag{121}$$

of the sojourn time $\theta'^2_{bl}$, $b,l \in \{1,2,...,\nu'\}$, in the second experiment.

### *3.5.2. Procedure of testing uniformity of distributions of critical infrastructure operation process conditional sojourn times at operation states including operating environment threats*

We consider test $\lambda$ based on Kolmogorov-Smirnov theorem [Kolowrocki, Soszynska, 2010c], [Kołowrocki, Soszyńska-Budny, 2011] that can be used for testing whether two independent samples of realizations of the conditional sojourn time $\theta'_{bl}$, $b,l \in \{1,2,...,\nu'\}$, $b \neq l$, at the particular operation states of the critical infrastructure operation process are drawn from the population with the same distribution.

We assume that we have defined in previous section two independent samples of non-decreasing ordered realizations (120) and (121) of the sojourn times $\theta'^1_{bl}$ and $\theta'^2_{bl}$, $b,l \in \{1,2,...,\nu\}$, $b \neq l$, coming from two different experiments, respectively composed of $n'^1_{bl}$ and $n'^2_{bl}$ realizations and we define their corresponding empirical distribution functions

$$H'^1_{bl}(t) = \frac{1}{n'^1_{bl}} \#\{k : \theta'^{1k}_{bl} < t,\ k \in \{1,2,...,n'^1_{bl}\}\}, \tag{122}$$

$$t \geq 0,\ b,l \in \{1,2,...,\nu'\},\ b \neq l,$$

and

$$H'^2_{bl}(t) = \frac{1}{n'^2_{bl}} \#\{k : \theta'^{2k}_{bl} < t,\ k \in \{1,2,...,n'^2_{bl}\}\}, \tag{123}$$

$$t \geq 0,\ b,l \in \{1,2,...,\nu'\},\ b \neq l.$$

Then, according to Kolmogorov-Smirnov theorem [Kolowrocki, Soszyńska-Budny, 2011], the sequence of distribution functions given by the equation

$$Q_{n_1 n_2}(\lambda) = P(D_{n_1 n_2} < \frac{\lambda}{\sqrt{n}}) \tag{124}$$

defined for $\lambda > 0$, where

$$n_1 = n'^1_{bl},\ n_2 = n'^2_{bl},\ n = \frac{n_1 n_2}{n_1 + n_2}, \tag{125}$$

and

$$D_{n_1 n_2} = \max_{-\infty < t < +\infty} \left| H'^1_{bl}(t) - H'^2_{bl}(t) \right|, \tag{126}$$

is convergent, as $n \to \infty$, to the limit distribution function

$$Q(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2k^2\lambda^2},\ \lambda > 0. \tag{127}$$

The distribution function $Q(\lambda)$ given by (127) is called $\lambda$ distribution and its Tables of values are available.
The convergence of the sequence $Q_{n_1 n_2}(\lambda)$ to the $\lambda$ distribution $Q(\lambda)$ means that for sufficiently large $n_1$ and $n_2$ we may use the following approximate formula

$$Q_{n_1 n_2}(\lambda) \cong Q(\lambda). \tag{128}$$

Hence, it follows that if we define the statistic

$$U_n = D_{n_1 n_2}\sqrt{n}, \tag{129}$$

where $D_{n_1 n_2}$ is defined by (7.126), then by (2.53) and (4.57), we have

$$P(U_n < \lambda) = P(D_{n_1 n_2}\sqrt{n} < \lambda) = P(D_{n_1 n_2} < \frac{\lambda}{\sqrt{n}})$$
$$= Q_{n_1 n_2}(\lambda) \cong Q(\lambda) \quad \text{for} \quad \lambda > 0.$$
(130)

This result means that in order to formulate and next to verify the hypothesis that the two independent samples of the realizations of the critical infrastructure operation process conditional sojourn times $\theta'^{1}_{bl}$ and $\theta'^{2}_{bl}$, $b,l \in \{1,2,...,\nu'\}$, $b \neq l$, at the operation state $z'_b$ when the next transition is to the operation state $z'_l$ are coming from the population with the same distribution, it is necessary to proceed according to the following scheme:
- to fix the numbers of realizations $n^1_{bl}$ and $n^2_{bl}$ in the samples,
- to collect the realizations (120) and (121) of the conditional sojourn times $\theta'^{1}_{bl}$ and $\theta'^{2}_{bl}$ of the critical infrastructure operation process in the samples,
- to find the realization of the empirical distribution functions $H'^{1}_{bl}(t)$ and $H'^{2}_{bl}(t)$ defined by (122) and (123) respectively, in the following forms:

$$H'^{1}_{bl}(t) = \begin{cases} \frac{n'^{11}_{bl}}{n'^{1}_{bl}} = 0, & t \le \theta'^{11}_{bl} \\ \frac{n'^{12}_{bl}}{n'^{1}_{bl}}, & \theta'^{11}_{bl} < t \le \theta'^{12}_{bl} \\ \frac{n'^{13}_{bl}}{n'^{1}_{bl}}, & \theta'^{12}_{bl} < t \le \theta'^{13}_{bl} \\ \cdots \\ \frac{n'^{1k}_{bl}}{n'^{1}_{bl}}, & \theta'^{1k-1}_{bl} < t \le \theta'^{1k}_{bl} \\ \cdots \\ \frac{n'^{1n'^{1}_{bl}}_{bl}}{n'^{1}_{bl}}, & \theta'^{1n'^{1}_{bl}-1}_{bl} < t \le \theta'^{1n'^{1}_{bl}}_{bl} \\ \frac{n'^{1n'^{1}_{bl}+1}_{bl}}{n'^{1}_{bl}} = 1, & t \ge \theta'^{1n'^{1}_{bl}}_{bl} \end{cases}, \tag{131}$$

$$H^{2}_{bl}(t) = \begin{cases} \frac{n^{21}_{bl}}{n^{2}_{bl}} = 0, & t \le \theta^{21}_{bl} \\ \frac{n^{22}_{bl}}{n^{2}_{bl}}, & \theta^{21}_{bl} < t \le \theta^{22}_{bl} \\ \frac{n^{23}_{bl}}{n^{2}_{bl}}, & \theta^{22}_{bl} < t \le \theta^{23}_{bl} \\ \cdots \\ \frac{n^{2k}_{bl}}{n^{2}_{bl}}, & \theta^{2k-1}_{bl} < t \le \theta^{2k}_{bl} \\ \cdots \\ \frac{n^{2n^{2}_{bl}}_{bl}}{n^{2}_{bl}}, & \theta^{2n^{2}_{bl}-1}_{bl} < t \le \theta^{2n^{1}_{bl}}_{bl} \\ \frac{n^{2n^{2}_{bl}+1}_{bl}}{n^{2}_{bl}} = 1, & t \ge \theta^{2n^{2}_{bl}}_{bl} \end{cases}, \tag{132}$$

where

$$n'^{11}_{bl} = 0, \; n'^{1n'^{1}_{bl}+1}_{bl} = n'^{1}_{bl}, \tag{133}$$

and

$$n'^{1k}_{bl} = \#\{j : \theta'^{1j}_{bl} < \theta'^{1k}_{bl}, \; j \in \{1,2,...,n'^{1}_{bl}\}\}, \tag{134}$$
$$k = 2,3,...,n'^{1}_{bl},$$

is the number of the sojourn time $\theta'^{1}_{bl}$ realizations less than its realization $\theta'^{1k}_{bl}$, $k = 2,3,...,n'^{1}_{bl}$, and respectively

$$n'^{21}_{bl} = 0,\; n'^{2\,n'^{2}_{bl}+1}_{bl} = n'^{2}_{bl}, \tag{135}$$

And

$$n'^{2k}_{bl} = \#\{j : \theta'^{2j}_{bl} < \theta'^{2k}_{bl},\; j \in \{1,2,...,n'^{2}_{bl}\}\}, \quad k = 2,3,...,n'^{2}_{bl}, \tag{136}$$

is the number of the sojourn time $\theta'^{2}_{bl}$ realizations less than its realization $\theta'^{2k}_{bl}$, $k = 2,3,...,n'^{2}_{bl}$,

- to calculate the realization of the statistic $u_n$ defined by (129) according to the formula

$$u_n = d_{n^1_{bl} n^2_{bl}} \sqrt{n}, \tag{137}$$

where

$$d_{n^1_{bl} n^2_{bl}} = \max\{ d^1_{n'^1_{bl} n'^2_{bl}},\; d^2_{n'^1_{bl} n'^2_{bl}} \}, \tag{138}$$

$$d^1_{n^1_{bl} n^2_{bl}} = \max\{ |H'^1_{bl}(\theta'^{1k}_{bl}) - H'^2_{bl}(\theta'^{1k}_{bl})|,\; k \in \{1,2,...,n'^1_{bl}\}\}, \tag{139}$$

$$d^2_{n^1_{bl} n^2_{bl}} = \max\{ |H'^1_{bl}(\theta'^{2k}_{bl}) - H'^2_{bl}(\theta'^{2k}_{bl})|,\; k \in \{1,2,...,n'^2_{bl}\}\}, \tag{140}$$

$$n = \frac{n'^1_{bl}\, n'^2_{bl}}{n'^1_{bl} + n'^2_{bl}}, \tag{141}$$

- to formulate the null hypothesis $H_0$ in the following form:

$H_0$ : The samples of realizations (120) and (121) are coming from the populations with the same distributions,

- to fix the significance level $\alpha$ ($\alpha = 0.01$, $\alpha = 0.02$, $\alpha = 0.05$ or $\alpha = 0.10$) of the test,

- to read from the Tables of $\lambda$ distribution, corresponding to $1-\alpha$, the value $\lambda_0$ such that the following equality holds

$$P(U_n < \lambda_0) = Q(\lambda_0) = 1 - \alpha, \tag{142}$$

- to determine the critical domain in the form of the interval $(\lambda_0, +\infty)$ and the acceptance domain in the form of the interval $(0, \lambda_0 >$,

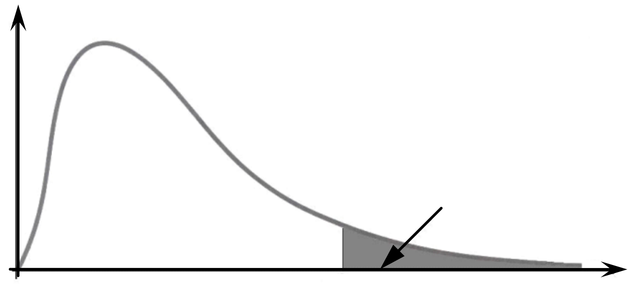

*Figure 6.* The graphical interpretation of the critical domain and the acceptance domain for the two-sample Smirnov-Kolmogorov test

- to compare the obtained value $u_n$ of the realization of the statistics $U_n$ with the read from the Tables critical value $\lambda_0$,

- to decide on the previously formulated null hypothesis $H_0$ in the following way:

if the value $u_n$ does not belong to the critical domain, i.e. when $u_n \le \lambda_0$ then we do not reject the hypothesis $H_0$, otherwise if the value $u_n$ belongs to the critical domain, i.e. when $u_n > \lambda_0$, then we reject the hypothesis $H_0$.

In the case when the null hypothesis $H_0$ is not rejected we may join the statistical data from the considered two separate sets into one new set of data and if there are no other sets of statistical data including the realizations of the sojourn time $\theta'_{bl}$, we proceed with the data of this new set in the way described in Sections 3.1-3.4. Otherwise, if there are other sets of statistical data including the realizations of the sojourn time $\theta'_{bl}$, we select the next one of them and perform the procedure of this section for data from this set and data from the previously formed new set. We continue this procedure up to the moment when the store of the statistical data sets including the realizations of the sojourn time $\theta'_{bl}$, is exhausted.

## 4. Identification of critical infrastructure operation process including operating environment threats based on expert opinion

We assume, as in [EU-CIRCLE Report D2.1-GMU2, 2016] and [EU-CIRCLE Report D3.3-GMU3, 2016], that the critical infrastructure during its operation including environment threats, at the fixed moment $t$, may be at one of $\nu'$, $\nu' \in N$, different operations states $z'_b$, $b = 1,2,...,\nu'$. Next, we mark by $Z'(t)$,

$t \in <0,+\infty>$, the critical infrastructure operation process related to its operating environment threats, that is a function of a continuous variable $t$, taking discrete values in the set $\{z'_1, z'_2, ..., z'_v\}$ of the critical infrastructure operation states defined in [EU-CIRCLE Report D2.1-GMU2, 2016]. We assume a semi-Markov model [Kolowrocki, 2014], [Kolowrocki, Soszynska, 2009d], [Kołowrocki, Soszyńska-Budny, 2011], [Limnios, Oprisan, 2005], [Limnios et all, 2005], [Macci, 2008], [Mercier, 2008], of the critical infrastructure operation process Z'(t) and we mark by $\theta'_{bl}$ its random conditional sojourn times at the operation states $z'_b$ when its next operation state is $z'_l$, $b, l = 1,2,...,v'$, $b \neq l$.

Under these assumption, the operation process may be described by the vector $[p'_b(0)]_{1 \mathrm{x} v'}$ of probabilities of the critical infrastructure operation process staying at the particular operations states at the initial moment $t = 0$, the matrix $[p'_{bl}(t)]_{v' \mathrm{x} v'}$ of the probabilities of the critical infrastructure operation process transitions between the operation states and the matrix $[H'_{bl}(t)]_{v' \mathrm{x} v'}$ of the distribution functions of the conditional sojourn times $\theta'_{bl}$ of the critical infrastructure operation process at the operation states or equivalently by the matrix $[h'_{bl}(t)]_{v' \mathrm{x} v'}$ of the density functions of the conditional sojourn times $\theta'_{bl}$, $b, l = 1,2,...,v'$, $b \neq l$, of the critical infrastructure operation process at the operation states. These all parameters of the critical infrastructure operation process are unknown and before their use to the prognosis of this process characteristics have to be estimated on the basis of statistical data coming from this operation process realizations.

### 4.1. defining unknown parameters of critical infrastructure operation process including operating environment threats identified by expert opinion

First, before identification of the critical infrastructure operation process, we should perform the following preliminary steps:
i) to analyze the critical infrastructure operation process;
ii) to fix or to define the critical infrastructure operation process following general parameters:
- the number of the operation states of the critical infrastructure operation process $v'$,
- the operation states of the critical infrastructure operation process $z'_1$, $z'_2$, …, $z'_{v'}$;

iii) to fix the possible transitions between the critical infrastructure operation states;
iv) to fix the set of the unknown parameters of the critical infrastructure operation process semi-Markov model;
v) to analyse and recognize the kind of data coming from the critical infrastructure operation process in disposal and to distinguish the following two cases:
- statistical data coming from realizations of the critical infrastructure operation process including the operating environment threats without the possibility of separation of data concerned with those threats and data from expert opinions concerned with this separation;
- data coming from expert opinions only, without of statistical data coming from realizations of the critical infrastructure operation process.

### 4.2. Estimating parameters of critical infrastructure operation process including operating environment threats identified by expert opinion – statistical and expert data

To estimate the unknown parameters of the critical infrastructure operation process in the case of statistical data coming from realizations of the critical infrastructure operation process including the operating environment threats without the possibility of separation of data concerned with those threats and data from expert opinions concerned with this separation, we should get necessary data from expert performing the following steps:
i) to fix (to have in disposal), similarly as in Section 7.2, the following parameters of critical infrastructure operation process $Z(t)$ including the operating environment threats without of separation the operation states including the operating environment threats:
- statistical evaluations of the initial probabilities of the vector $[p_b(0)]_{1 \mathrm{x} v}$,
- statistical evaluations of the probabilities of transitions of the matrix $[p_{bl}]_{v \mathrm{x} v}$,
- statistical evaluations of the conditional sojourn times mean values of the matrix $[M_{bl}]_{v \mathrm{x} v}$:
ii) to get the evaluations of the unknown parameters of the critical infrastructure operation process $Z'(t)$ with included and separated operating threats:
- the vector of initial probabilities $[p'_b(0)]_{1 \mathrm{x} v'}$,
- the matrix of probabilities of transitions $[p'_{bl}]_{v' \mathrm{x} v'}$,
- the matrix of the mean values of the conditional sojourn times $[M'_{bl}]_{v' \mathrm{x} v'}$;

Since according to Section 3.1, the critical infrastructure operation process can be affected by a number $w$, $w \in N$, of unnatural threats $ut_i$, $i = 1,2,...,w$, coming from the critical infrastructure operating environment, we assume that they are random and we mark the probability of the operating environment threat $ut_i$, $i = 1,2,...,w$, appearance at the operation state $z_b$, $b = 1,2,...,v$, by

$$P_b(ut_i),\ i = 1,2,...,w,\ b = 1,2,...,v.$$

Moreover, in this approach, we consider 2 variants:
variant 1 - the probabilities of the operating environment threats $ut_i$, $i = 1,2,...,w$, appearance $P_b(ut_i)$, $i = 1,2,...,w$, $b = 1,2,...,v$, are conditional and concerned with each of the critical infrastructure particular states (they can be different for various operation states);
variant 2 - the probabilities of the operating environment threats $ut_i$, $i = 1,2,...,w$, $b = 1,2,...,v$, are unconditional and concerned with the critical infrastructure operation proces independently of its particular states.
Further, to get the initial probabilities of the vector $[p'_b(0)]$ of the operation process $Z'(t)$ with separated operation states including the operating environment threats, under the assumption that the threats are disjoint (they do not appear simultaneously), we distribute the initial probabilities of the vector $[p_b(0)]$ in the following way [EU-CIRCLE Report D2.1-GMU2, 2016]:

i) variant 1

- if $p_b(0) \neq 0$, $b = 1,2,...,v$,

we replace it by

$$p'_{(w+1)(b-1)+1}(0) = p_b(0) - [P_b(ut_1) + P_b(ut_2) + \dots + P_b(ut_w)], \quad (143)$$

$$p'_{(w+1)(b-1)+1+i}(0) = P_b(ut_i),\ i = 1,2,...,w, \quad (144)$$

for $b = 1,2,...,v$;

- if $p_b(0) = 0$, $b = 1,2,...,v$,

we replace it by

$$p'_{(w+1)(b-1)+1}(0) = 0, \quad (145)$$

$$p'_{(w+1)(b-1)+1+i}(0) = 0,\ i = 1,2,...,w, \quad (146)$$

for $b = 1,2,...,v$.
ii) variant 2

- if $p_b(0) \neq 0$, $b = 1,2,...,v$,

we replace it by

$$p'_{(w+1)(b-1)+1}(0) = p_b(0) - p_b(0[P_b(ut_1) + P_b(ut_2) + \dots + P_b(ut_w)], \quad (147)$$

$$p'_{(w+1)(b-1)+1+i}(0) = p_b(0)P_b(ut_i),\ i = 1,2,...,w, \quad (148)$$

for $b = 1,2,...,v$;

- if $p_b(0) = 0$, $b = 1,2,...,v$,

we replace it by

$$p'_{(w+1)(b-1)+1}(0) = 0, \quad (149)$$

$$p'_{(w+1)(b-1)+1+i}(0) = 0,\ i = 1,2,...,w, \quad (150)$$

for $b = 1,2,...,v$.

To get the probabilities of transitions between the operation states of the matrix $[p'_{bl}]$ of the operation process $Z'(t)$ with separated operation states including the operating environment threats, we distribute the probabilities of transitions between the operation states of the matrix $[p_{bl}]$ in the following way:

i) variant 1

- if $p_{bl} \neq 0$, $b,l = 1,2,...,v$,

we replace it by

$$p'_{(w+1)(b-1)+1\ (w+1)(l-1)+1} = p_{bl} - [P_b(ut_1) + P_b(ut_2) + \dots + P_b(ut_w)], \quad (151)$$

$$p'_{(w+1)(b-1)+1\ (w+1)(l-1)+1+i} = P_b(ut_i),\ i = 1,2,...,w, \quad (152)$$

for $b,l = 1,2,...,v$,

and we additionally assume that

$$p'_{(w+1)(b-1)+1+i\ (w+1)(b-1)+1} = 1,\ i = 1,2,...,w, \quad (153)$$

$$p'_{(w+1)(b-1)+1+i\,j} = 0, \quad (154)$$

$i = 1,2,...,w,\ j = 1,2,...,v2^w,$
and $j \neq (w+1)(b-1)+1$;

- if $p_{bl} = 0,\ b,l = 1,2,...,v,$
we replace it by

$$p'_{(w+1)(b-1)+1\ (w+1)(l-1)+1} = 0, \qquad (155)$$

$$p'_{(w+1)(b-1)+1\ (w+1)(l-1)+1+i} = 0,\ i = 1,2,...,w, \qquad (156)$$

for $b,l = 1,2,...,v$.

wariant 2:

- if $p_{bl} \neq 0,\ b,l = 1,2,...,v,$

we replace it by

$$p'_{(w+1)(b-1)+1\ (w+1)(l-1)+1} = p_{bl} - p_{bl}[P_b(ut_1) + P_b(ut_2) + \ldots + P_b(ut_w)], \qquad (157)$$

$$p'_{(w+1)(b-1)+1\ (w+1)(l-1)+1+i} = p_{bl}P_b(ut_i), \qquad (158)$$
$i = 1,2,...,w,$

and we additionally assume that

$$p'_{(w+1)(b-1)+1+i\ (w+1)(b-1)+1} = 1,\ i = 1,2,...,w, \qquad (159)$$

$$p'_{(w+1)(b-1)+1+i\,j} = 0, \qquad (160)$$
$i = 1,2,...,w,\ j = 1,2,...,v2^w,$

and $j \neq (w+1)(b-1)+1$;
- if $p_{bl} = 0,\ b,l = 1,2,...,v,$

we replace it by

$$p'_{(w+1)(b-1)+1\ (w+1)(l-1)+1} = 0, \qquad (161)$$

$$p'_{(w+1)(b-1)+1\ (w+1)(l-1)+1+i} = 0,\ i = 1,2,...,w, \qquad (162)$$

for $b,l = 1,2,...,v$.

The conditions (153)-(154) and (159)-(160) mean that the transitions from the operation states including the operating environment threats is possible only to the corresponding operation states without the operating envitonment threats.
Finally, as the transformation of the matrix $[H_{bl}(t)]_{v\times v}$ of the critical infrastructure operation process $Z(t)$ conditional sojourn times $\theta_{bl}$, $b,l = 1,2,...,v,$ at the operation states without of separation the operation states including the operating environment threats into the matrix $[H'_{bl}(t)]_{v'\times v'}$ of the distributions of the conditional sojourn times $\theta'_{bl}$, $b,l = 1,2,...,v'$, at the operation states of the critical infrastructure operation process $Z'(t)$ with included and separated operating threats on the basis of expert opinions is practically not possible, we transform the corresponding matrix $[M_{bl}]_{v\times v}$ of the mean values of the conditional sojourn times $\theta_{bl}$, $b,l = 1,2,...,v,$ at the operation states into the matrix $[M'_{bl}]_{v'\times v'}$ of the mean values of the conditional sojourn times $\theta'_{bl}$, $b,l = 1,2,...,v'$. We proceed, for both variants (variant 1 and variant 2), in the following way:

- if $M_{bl} \neq 0,\ b,l = 1,2,...,v,$

we fix the mean values

$$M'_{(w+1)(b-1)+1+i\ (w+1)(b-1)+1}\ i = 1,2,...,w, \qquad (163)$$
$b = 1,2,...,v,$

on the basis of expert opinions and assume

$$M'_{(w+1)(b-1)+1+i\,j} = 0,\ i = 1,2,...,w, \qquad (164)$$
$j = 1,2,...,v2^w$, and $j \neq (w+1)(b-1)+1$,
and

$$M'_{(w+1)(b-1)+1\ (w+1)(l-1)+1} = M_{bl} - \sum_{i=1}^{w} M'_{(w+1)(b-1)+1+i\ (w+1)(b-1)+1}, \qquad (165)$$

for $b,l = 1,2,...,v$;

- if $M_{bl} = 0,\ b,l = 1,2,...,v,$

we replace it by

$$M'_{(w+1)(b-1)+1\ (w+1)(l-1)+1} = 0, \qquad (166)$$

$$M'_{(w+1)(b-1)+1\ (w+1)(l-1)+1+i} = 0,\ i = 1,2,...,w, \qquad (167)$$

for $b,l = 1,2,...,v$.

The distribution of the initial probabilities of the vector $[p_b(0)]_{1\times v}$, the probabilities of transitions between the operation states of the matrix $[p_{bl}]_{v\times v}$ and the mean values of the conditional sojourn times $\theta_{bl}$ at the operation states of the matrix $[M_{bl}(t)]_{v\times v}$ of the operation process $Z(t)$, respectivety into the initial probabilities of the vector $[p'_b(0)]_{1\times v'}$, the probabilities of transitions between the operation states of the matrix $[p'_{bl}]$ and mean values of the conditional sojourn times $\theta'_{bl}$ at the operation states of the matrix $[M'_{bl}(t)]_{v' x v'}$ of the operation process $Z'(t)$ with separated operation states including the operating environment threats, using the pprocedures defined by (143)-(167) , was done under the assumption that the operating environment threats are disjoint (they do not appear simultaneously). It means that the new operation states of the operation process $Z'(t)$ with separated operation states either do not include the operating environment threats or include one of the operating environment threats only. The procedure of this distribution in the case the operating environment threats are not disjoint have to be constructed individually for each specific case.

## 4.3. Estimating parameters of critical infrastructure operation process including operating environment threats identified by expert opinion - expert data only

In the case of lack of statistical data collection, together with experienced experts operating the critical infrastructure, it is possible to estimate approximately the unknown parameters of the critical infrastructure operation process including operating environment threats performing the following steps:
i) to determine the vector

$$[p'(0)] = [p'_1(0), p'_2(0), \ldots, p'_{v'}(0)], \qquad (168)$$

of expert evaluations of the probabilities $p'_b(0)$, $b = 1,2,\ldots,v'$, of the critical infrastructure operation process staying at the operation states at the initial moment $t = 0$, after explanation to the expert practical meaning of the formula

$$p'_b(0) = \frac{n'_b(0)}{n'(0)} \text{ for } b = 1,2,\ldots,v'; \qquad (169)$$

ii) to determine the matrix

$$[p'_{bl}] = \begin{bmatrix} p'_{11} \ p'_{12} \cdots p'_{1v'} \\ p'_{21} \ p'_{22} \cdots p'_{2v'} \\ \cdots \\ p'_{v'1} \ p'_{v'2} \cdots p'_{v'v'} \end{bmatrix}, \qquad (170)$$

of expert evaluations of the probabilities $p'_{bl}$, $b,l = 1,2,\ldots,v'$, of the critical infrastructure operation process transitions from the operation state $z'_b$ to the operation state $z'_l$, after explanation to the expert practical meaning of the formula

$$p'_{bl} = \frac{n'_{bl}}{n'_b} \text{ for } b,l = 1,2,\ldots,v', b \neq l, \ p'_{bb} = 0 \qquad (171)$$

for $b = 1,2,\ldots,v'$;

iii) to determine the matrix

$$[M'_{bl}] = \begin{bmatrix} M'_{11} \ M'_{12} \ldots M'_{1v'} \\ M'_{21} \ M'_{22} \ldots M'_{2v'} \\ \ldots \\ M'_{v'1} \ M'_{v'2} \ldots M'_{v'v'} \end{bmatrix}, \qquad (172)$$

of expert evaluations of the mean values $M'_{bl}$, $b,l = 1,2,\ldots,v'$, of the critical infrastructure operation process conditional sojourn times $\theta'_{bl}$, $b, l = 1,2,\ldots,v'$, at the operation state $z'_b$ when the next operation state is $z'_l$, after explanation to the expert practical meaning of these parameters.

## 5. Conclusions

The proposed statistical methods of identification of the unknown parameters of the critical infrastructure operation process allow us for the identification of the unknown parameters of models presented in [EU-CIRCLE Report D2.1-GMU2, 2016]. After this identification, the models can be used in the critical infrastructure operation process characteristics evaluation and prediction.

## Acknowledgements

The paper presents the results developed in the scope of the EU-CIRCLE project titled “A pan – European framework for strengthening Critical Infrastructure resilience to climate change” that has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 653824. http://www.eu-circle.eu/

## References

Barbu V., Limnios N., Empirical estimation for discrete-time semi-Markov processes with applications in reliability. Journal of Nonparametric Statistics, Vol. 18, No. 7-8, 483-498, 2006

EU-CIRCLE Report D2.1-GMU2, Modelling outside dependences influence on Critical Infrastructure Safety (CIS) – Modelling Critical Infrastructure Operation Process (CIOP) including Operating Environment Threats (OET), 2016

EU-CIRCLE Report D2.1-GMU3, Modelling outside dependences influence on Critical Infrastructure Safety (CIS) – Modelling Climate-Weather Change Process (C-WCP) including Extreme Weather Hazards (EWH), 2016

EU-CIRCLE Report D3.3-GMU3, Modelling inside and outside dependences influence on safety of complex multistate ageing systems (critical infrastructures) – Integrated Model of Critical Infrastructure Safety (IMCIS) related to its operation process including operating environment threats (with other critical infrastructures influence, without climate-weather change influence), 2016

Ferreira F., Pacheco A., Comparison of level-crossing times for Markov and semi-Markov processes. Statistics and Probability Letters, Vol. 7, No 2, 151-157, 2007

Glynn P.W., Haas P.J., Laws of large numbers and functional central limit theorems for generalized semi
Markov processes. Stochastic Models, Vol. 22, No 2, 201-231, 2006

Grabski F., (2002) Semi-Markov Models of Systems Reliability and Operations Analysis. System Research Institute, Polish Academy of Science, 2002 (*in Polish*)

Kołowrocki K., Reliability of Large and Complex Systems, Amsterdam, Boston, Heidelberd, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sidney, Tokyo, Elsevier, 2014

Kołowrocki K., Soszyńska J., A general model of industrial systems operation processes related to their environment and infrastructure. Summer Safety & Reliability Seminars. Journal of Polish Safety and Reliability Association, Issue 2, Vol. 2, 223-226, 2008

Kolowrocki K., Soszynska J., Modeling environment and infrastructure influence on reliability and operation process of port oil transportation system. Electronic Journal Reliability & Risk Analysis: Theory & Applications, Vol. 2, No 3, 131-142, 2009a

Kolowrocki K., Soszynska J., Safety and risk evaluation of Stena Baltica ferry in variable operation conditions. Electronic Journal Reliability & Risk Analysis: Theory & Applications, Vol. 2, No 4, 168-180, 2009b

Kołowrocki K., Soszyńska J., Statistical identification and prediction of the port oil pipeline system's operation process and its reliability and risk evaluation. Summer Safety & Reliability Seminars. Journal of Polish Safety and Reliability Association, Issue 3, Vol. 2, 241-250, 2009c

Kołowrocki K., Soszyńska J., Methods and algorithms for evaluating unknown parameters of operation processes of complex technical systems. Summer Safety & Reliability Seminars. Journal of Polish Safety and Reliability Association, Issue 3, Vol. 1, 2, 211-222, 2009d

Kołowrocki K., Soszyńska J., Statistical identification and prediction of the port oil pipeline system's operation process and its reliability and risk evaluation. Summer Safety & Reliability Seminars. Journal of Polish Safety and Reliability Association, Issue 4, Vol. 2, 241-250, 2009e

Kolowrocki K., Soszynska J., Testing uniformity of statistical data sets coming from complex systems operation processes. Summer Safety & Reliability Seminars. Journal of Polish Safety and Reliability Association, Issue 4, Vol. 1,123-132, 2010b

Kołowrocki K., Soszyńska-Budny J., Reliability and Safety of Complex Technical Systems and Processes: Modeling - Identification - Prediction - Optimization, London, Dordrecht, Heildeberg, New York, Springer, 2011

Kołowrocki K., Soszyńska-Budny J., Introduction to safety analysis of critical infrastructures. Proc. International Conference on Quality, Reliability, Risk, Maintenance and Safety Engineering - QR2MSE-2012, Chendgu, China, 1-6, 2012a

Limnios N., Oprisan G., Semi-Markov Processes and Reliability. Birkhauser, Boston, 2005

Limnios N., Ouhbi B., Sadek A., Empirical estimator of stationary distribution for semi-Markov processes. Communications in Statistics-Theory and Methods, Vol. 34, No. 4, 987-995 12, 2005

Macci C., Large deviations for empirical estimators of the stationary distribution of a semi-Markov process with finite state space. Communications in Statistics-Theory and Methods, Vol. 37, No. 19,3077-3089, 2008

Mercier S., Numerical bounds for semi-Markovian quantities and application to reliability. Methodology and Computing in Applied Probability, Vol. 10, No. 2, 179-198, 2008

Rice J. A., Mathematical statistics and data analysis. Duxbury. Thomson Brooks/Cole. University of California. Berkeley, 2007