Kołowrocki Krzysztof

Soszyńska Joanna
Maritime University, Gdynia, Poland

Testing uniformity of statistical data sets coming from complex systems operation processes

Keywords

complex system, operation process, sojourn times, uniformity testing

Abstract

The method of the statistical data sets uniformity analysis based on Kolmogorov-Smirnov test is presented. The procedure of statistical data sets uniformity testing is proposed to be applied to the empirical sojourn times in operation states coming from the operation processes of the complex technical systems. The proposed procedure is practically applied to the analysis and uniformity testing of the maritime ferry spring and winter sets of realizations of the sojourn times in particular operation states.

1. Introduction

Many real technical systems belong to the class of complex systems. First of all, it is concerned with the large numbers of components and subsystems they are built and with their operating complexity. Modeling of the complicated system operation processes is difficult because of the large number of the operation states, impossibility of their precise defining and because of the impossibility of the exact describing the transitions between these states. The changes of the operation states of the system operations processes cause the changes of these systems reliability structures and also the changes of their components reliability functions. The general semi-markovian model of the complex technical systems operation processes is proposed in [13]-[14]. The reliability models of various multistate complex systems are considered in [10]-[12]. The general joint models linking these system reliability models with the model of their operation processes, allowing us for the reliability and safety analysis of the complex technical systems in variable operations conditions, are constructed in [16], [17], [20]-[23].

To be able to apply these general models practically in the evaluation and prediction the reliability and safety of real complex technical systems it is necessary to elaborate the statistical methods concerned with determining the unknown parameters of the proposed models. Namely, the probabilities of the initials system operation states, the probabilities of transitions between the system operation states and the unknown parameters of distributions of the sojourn times of the system operation process in the particular operation states and also the unknown parameters of the conditional multistate reliability and safety functions of the system components in various operation states should be identified. It is also necessary the elaborating the methods of testing the hypotheses concerned with the conditional sojourn times of the system operation process in particular operation states.

2. Experimental statistical data uniformity analysis

We consider test λ based on Kolmogorov-Smirnov theorem [3] that can be used for testing whether two independent samples of realizations of the conditional sojourn times θ_{bl}, $b,l \in \{1,2,...,\nu\}$, $b \neq l$, in particular operation states of the system operation process are drawn from the population with the same distribution.

We assume that we have two independent samples of non-decreasing ordered realizations

$$\theta_{bl}^{1k}, \;\; k = 1,2,...,n_{bl}^{1}, \;\; b,l \in \{1,2,...,\nu\}, \;\; b \neq l, \qquad (1)$$

and

$$\theta_{bl}^{2k},\ k = 1,2,...,n_{bl}^{2},\ b,l \in \{1,2,...,v\},\ b \neq l, \tag{2}$$

of the sojourn times θ_{bl}^{1} and θ_{bl}^{2} $b,l \in \{1,2,...,v\}$, $b \neq l$, respectively composed of n_{bl}^{1} and n_{bl}^{2} realizations and we mark by

$$H_{bl}^{1}(t) = \frac{1}{n_{bl}^{1}} \#\{k : \theta_{bl}^{1k} < t,\ k \in \{1,2,...,n_{bl}^{1}\}\}, \tag{3}$$

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

and

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

$$t \geq 0,\ b,l \in \{1,2,...,v\},\ b \neq l, \tag{4}$$

their corresponding empirical distribution functions. Then, according to Smirnov theorem, the sequence of distribution functions given by the equation

$$Q_{n_1 n_2}(\lambda) = P(D_{n_1 n_2} < \frac{\lambda}{\sqrt{n}}) \tag{5}$$

defined for $\lambda > 0$, where

$$n_1 = n_{bl}^{1},\ n_2 = n_{bl}^{2},\ n = \frac{n_1 n_2}{n_1 + n_2}, \tag{6}$$

and

$$D_{n_1 n_2} = \max_{-\infty < t < +\infty} \left| H_{bl}^{1}(t) - H_{bl}^{2}(t) \right|, \tag{7}$$

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

$$Q(\lambda) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2\lambda^2},\ \lambda > 0. \tag{8}$$

The distribution function $Q(\lambda)$ given by (8) is called λ distribution and its Tables of values are available. It means that for sufficiently large n_1 and n_2 we may use the following approximate formula

$$Q_{n_1 n_2}(\lambda) \cong Q(\lambda),\ \lambda > 0. \tag{9}$$

Hence it follows that if we define the statistic

$$U_n = D_{n_1 n_2} \sqrt{n}, \tag{10}$$

where $D_{n_1 n_2}$ is defined by (6), then by (4) and (7) we have

$$P(U_n < u) = P(D_{n_1 n_2} \sqrt{n} < u)$$
$$= P(D_{n_1 n_2} < \frac{u}{\sqrt{n}}) = Q_{n_1 n_2}(\lambda) \cong Q(u) \tag{11}$$

for $u > 0$.

This result means that in order to formulate and next to verify the hypothesis that the samples of the realizations the system conditional sojourn times θ_{bl}^{1} and θ_{bl}^{2}, $b,l \in \{1,2,...,v\}$, $b \neq l$, 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 (1) and (2) of the conditional sojourn times θ_{bl}^{1} and θ_{bl}^{2} of the system 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 (3) and (4) respectively, in the following forms:

$$H_{bl}^{1}(t) = \begin{cases} \frac{n_{bl}^{11}}{n_{bl}^{1}} = 0, & t \leq \theta_{bl}^{11} \\ \frac{n_{bl}^{12}}{n_{bl}^{1}}, & \theta_{bl}^{11} < t \leq \theta_{bl}^{12} \\ \frac{n_{bl}^{13}}{n_{bl}^{1}}, & \theta_{bl}^{12} < t \leq \theta_{bl}^{13} \\ \cdot\ \cdot\ \cdot & \\ \frac{n_{bl}^{1k}}{n_{bl}^{1}}, & \theta_{bl}^{1k-1} < t \leq \theta_{bl}^{1k} \\ \cdot\ \cdot\ \cdot & \\ \frac{n_{bl}^{1n_{bl}^{1}}}{n_{bl}^{1}}, & \theta_{bl}^{1n_{bl}^{1}-1} < t \leq \theta_{bl}^{1n_{bl}^{1}} \\ \frac{n_{bl}^{1n_{bl}^{1}+1}}{n_{bl}^{1}} = 1, & t \geq \theta_{bl}^{1n_{bl}^{1}} \end{cases}, \tag{12}$$

where

$$n_{bl}^{11} = 0\ \ n_{bl}^{1\,n_{bl}^{1}+1} = n_{bl}^{1}, \tag{13}$$

and

$$n_{bl}^{1k} = \#\{i : \{\theta_{bl}^{1j} < \theta_{bl}^{1k},\ j \in \{1,2,...,n_{bl}^{1}\}\}, \quad (14)$$

$k = 2,3,...,n_{bl}^{1}$,

is the numbers of the sojourn time θ_{bl}^{1} realizations less than its realization θ_{bl}^{1k},

$$H_{bl}^{2}(t) = \begin{cases} \frac{n_{bl}^{21}}{n_{bl}^{2}} = 0, & t \le \theta_{bl}^{21} \\ \frac{n_{bl}^{22}}{n_{bl}^{2}}, & \theta_{bl}^{21} < t \le \theta_{bl}^{22} \\ \frac{n_{bl}^{23}}{n_{bl}^{2}}, & \theta_{bl}^{22} < t \le \theta_{bl}^{23} \\ \cdot\ \cdot\ \cdot & \\ \frac{n_{bl}^{2k}}{n_{bl}^{2}}, & \theta_{bl}^{2k-1} < t \le \theta_{bl}^{2k} \\ \cdot\ \cdot\ \cdot & \\ \frac{n_{bl}^{2n_{bl}^{2}}}{n_{bl}^{2}}, & \theta_{bl}^{2n_{bl}^{2}-1} < t \le \theta_{bl}^{2n_{bl}^{1}} \\ \frac{n_{bl}^{2n_{bl}^{2}+1}}{n_{bl}^{2}} = 1, & t \ge \theta_{bl}^{2n_{bl}^{2}} \end{cases}, \quad (15)$$

where

$$n_{bl}^{21} = 0,\ n_{bl}^{2n_{bl}^{2}+1} = n_{bl}^{2}, \quad (16)$$

and

$$n_{bl}^{2k} = \#\{i : \{\theta_{bl}^{2j} < \theta_{bl}^{2k},\ j \in \{1,2,...,n_{bl}^{2}\}\}, \quad (17)$$

$k = 2,3,...,n_{bl}^{2}$,

is the numbers of the sojourn time θ_{bl}^{2} realizations less than its realization θ_{bl}^{2k},

- to formulate the null hypothesis H_0 and the alternative hypothesis H_A the following form:

H_0: The samples of realizations (1) and (2) are coming from the populations with the same distributions,

H_A: The samples of realizations (1) and (2) are coming from the populations with different distributions;

- to fix the significance level α,
- to read from the Tables of λ distribution the value $u = \lambda_0$ such that the following equality holds

$$P(U_n < u) = Q(u) = Q(\lambda_0) = 1 - \alpha, \quad (18)$$

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

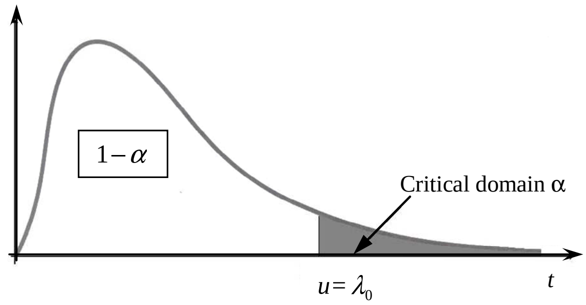

Figure 1. The graphical interpretation of the critical domain and the acceptance domain for the two-sample Smirnov-Kolmogorov test.

- to calculate the realization of the statistic U_n defined by (10) according to the formula

$$u_n = d_{n_{bl}^{1} n_{bl}^{2}} \sqrt{n_{bl}}, \quad (19)$$

where

$$d_{n_{bl}^{1} n_{bl}^{2}} = \max\{d^{1}_{n_{bl}^{1} n_{bl}^{2}}, d^{2}_{n_{bl}^{1} n_{bl}^{2}}\}, \quad (20)$$

$$d^{1}_{n_{bl}^{1} n_{bl}^{2}} = \max\{\left|H_{bl}^{1}(\theta_{bl}^{1k}) - H_{bl}^{2}(\theta_{bl}^{1k})\right|, \quad (21)$$

$k \in \{1,2,...,n_{bl}^{1}\}\}$,

$$d^{2}_{n_{bl}^{1} n_{bl}^{2}} = \max\{\left|H_{bl}^{1}(\theta_{bl}^{2k}) - H_{bl}^{2}(\theta_{bl}^{2k})\right|, \quad (22)$$

$k \in \{1,2,...,n_{bl}^{2}\}\}$,

$$n_{bl} = \frac{n_{bl}^{1} n_{bl}^{2}}{n_{bl}^{1} + n_{bl}^{2}}, \quad (23)$$

- to compare the obtained value u_n of the realization of the statistics U_n with the read from the Tables critical value $u = \lambda_0$ and we verify previously formulated the null hypothesis H_0 in the following way:

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

3. The ferry operation process uniformity analysis

We consider a passenger ro-ro ship operating in Baltic Sea between the Gdynia port in Poland and the Karlskrona port in Sweden according to a regular everyday timetable. Taking into account the operation process of the considered ferry we distinguish the following as its eighteen operation states:

- an operation state z_1 – loading at Gdynia Port,
- an operation state z_2 – unmooring operations at Gdynia Port,
- an operation state z_3 – leaving Gdynia Port and navigation to “GD” buoy,
- an operation state z_4 – navigation at restricted waters from “GD” buoy to the end of Traffic Separation Scheme,
- an operation state z_5 – navigation at open waters from the end of Traffic Separation Scheme to “Angoring” buoy,
- an operation state z_6 – navigation at restricted waters from “Angoring” buoy to “Verko” Berth at Karlskrona,
- an operation state z_7 – mooring operations at Karlskrona Port,
- an operation state z_8 – unloading at Karlskrona Port,
- an operation state z_9 – loading at Karlskrona Port,
- an operation state z_{10} – unmooring operations at Karlskrona Port,
- an operation state z_{11} – ferry turning at Karlskrona Port,
- an operation state z_{12} – leaving Karlskrona Port and navigation at restricted waters to “Angoring” buoy,
- an operation state z_{13} – navigation at open waters from “Angoring” buoy to the entering Traffic Separation Scheme,
- an operation state z_{14} – navigation at restricted waters from the entering Traffic Separation Scheme to “GD” buoy,
- an operation state z_{15} – navigation from “GD” buoy to turning area,
- an operation state z_{16} – ferry turning at Gdynia Port,
- an operation state z_{17} – mooring operations at Gdynia Port,
- an operation state z_{18} – unloading at Gdynia Port.

The ferry operation process is very regular in the sense that the operation state changes are from the particular state z_b, $b = 1,2,...,17$, to the neighboring state z_{b+1}, $b = 1,2,...,17$, and from z_{18} to z_1 only.

We will apply two-sample Smirnov-Kolmogorov test described in a previous section to verify the hypotheses that spring and winter data sets consisted of the ferry conditional sojourn times $\theta_{b\,b+1}$, $b = 1,2,...,17$, in particular operation states z_b, $b = 1,2,...,17$, to the neighboring operation state z_{b+1}, $b = 1,2,...,17$, and the ferry conditional sojourn time $\theta_{18\,1}$ from the operation state z_{18} to the operation state z_1 are from the populations with the same distribution.

The procedure of testing the uniformity for data given in the Appendix 5A in *Tables A1-A4* [3] for spring and in *Tables 5-8* [3] for winter suggested in a previous section in particular operation states is exemplary illustrated for the realizations of the ferry conditional sojourn time θ_{12}.

For spring data, given in the Appendix 5A in *Tables 1-4*, [3] the conditional sojourn time θ_{12}^1 has the empirical distribution function

$$H^{1}{}_{12}(t) = \begin{cases} 0, & t \le 15, \\ 1/42, & 15 < t \le 20, \\ 2/42, & 20 < t \le 25, \\ 3/42, & 25 < t \le 33, \\ 4/42, & 33 < t \le 35, \\ 5/42, & 35 < t \le 37, \\ 6/42, & 37 < t \le 40, \\ 8/42, & 40 < t \le 43, \\ 9/42, & 43 < t \le 44, \\ 13/42, & 44 < t \le 45, \\ 14/42, & 45 < t \le 46, \\ 15/42, & 46 < t \le 47, \\ 17/42, & 47 < t \le 50, \\ 18/42, & 50 < t \le 52, \\ 19/42, & 52 < t \le 53, \\ 21/42, & 53 < t \le 55, \\ 22/42, & 55 < t \le 57, \\ 23/42, & 57 < t \le 58, \\ 24/42, & 58 < t \le 59, \\ 26/42, & 59 < t \le 60, \\ 27/42, & 60 < t \le 61, \\ 29/42, & 61 < t \le 62, \\ 30/42, & 62 < t \le 63, \\ 32/42, & 63 < t \le 65, \\ 33/42, & 65 < t \le 67, \\ 34/42, & 67 < t \le 68, \\ 35/42, & 68 < t \le 71, \\ 36/42, & 71 < t \le 72, \\ 38/42, & 72 < t \le 75, \\ 39/42, & 75 < t \le 78, \\ 40/42, & 78 < t \le 84, \\ 41/42, & 84 < t \le 97, \\ 1, & t > 97; \end{cases}$$

whereas for winter data given in the Appendix 5A in *Tables 5-8* [3] , the conditional sojourn time θ_{12}^{2} has the empirical distribution function

$$H^{2}{}_{12}(t) = \begin{cases} 0, & t \le 12, \\ 1/40, & 12 < t \le 15, \\ 3/40, & 15 < t \le 18, \\ 4/40, & 18 < t \le 19, \\ 5/40, & 19 < t \le 20, \\ 6/40, & 20 < t \le 25, \\ 7/40, & 25 < t \le 33, \\ 9/40, & 33 < t \le 34, \\ 10/40, & 34 < t \le 36, \\ 11/40, & 36 < t \le 37, \\ 14/40, & 37 < t \le 40, \\ 15/40, & 40 < t \le 41, \\ 16/40, & 41 < t \le 44, \\ 17/40, & 44 < t \le 46, \\ 18/40, & 46 < t \le 48, \\ 20/40, & 48 < t \le 50, \\ 21/40, & 50 < t \le 53, \\ 22/40, & 53 < t \le 55, \\ 23/40, & 55 < t \le 57, \\ 24/40, & 57 < t \le 59, \\ 25/40, & 59 < t \le 60, \\ 27/40, & 60 < t \le 61, \\ 28/40, & 61 < t \le 62, \\ 29/40, & 62 < t \le 63, \\ 30/40, & 63 < t \le 65, \\ 34/40, & 65 < t \le 67, \\ 35/40, & 67 < t \le 69, \\ 36/40, & 69 < t \le 75, \\ 38/40, & 75 < t \le 80, \\ 39/40, & 80 < t \le 90, \\ 1, & t > 90. \end{cases}$$

The null hypothesis is

H_0: The winter and spring realizations of the ferry conditional sojourn times θ_{12}^{1} and θ_{12}^{2} are from the population with the same distribution.

To verify this hypothesis we will use the two-sample Smirnov-Kolmogorov test λ at the significance level $\alpha = 0.05$. From the table of the λ distribution for the significance level $\alpha = 0.05$ we get the critical value $\lambda_0 = u \cong 1.36$. Using the above empirical distributions we form a common Table composed of all their values. In the *Table 1*, t_k are taken together all realizations θ_{12}^{1k}, $k = 1,2,...,42$, and θ_{12}^{2k}, $k = 1,2,...,40$, of the conditional sojourn times θ_{12}^{1}

and θ_{12}^2 i.e. they represent all discontinuity points of the empirical distribution function $H_{12}^1(t)$ and $H_{12}^2(t)$ were they have jump in their values $H_{12}^1(t_k)$ and $H_{12}^2(t_k)$ respectively.

Table 1.

$t_k = \theta_{12}^1 \vee \theta_{12}^{2k}$	$H_{12}^1(t_k)$	$H_{12}^2(t_k)$	$\lvert H_{12}^1(t_k) - H_{12}^2(t_k) \rvert$
12	0	0	0
15	0	1/40	0.025
18	1/42	3/40	0.051
19	1/42	4/40	0.076
20	1/42	5/40	0.101
25	2/42	6/40	0.102
33	3/42	7/40	0.104
34	4/42	9/40	0.129
35	4/42	10/40	0.156
36	5/42	10/40	0.131
37	5/42	11/40	0.156
40	6/42	14/40	0.207
41	8/42	15/40	0.185
43	8/42	16/40	0.209
44	9/42	16/40	0.186
45	13/42	17/40	0.115
46	14/42	17/40	0.092
47	15/42	18/40	0.093
48	17/42	18/40	0.045
50	17/42	20/40	0.095
52	18/42	21/40	0.096
53	19/42	21/40	0.073
55	21/42	22/40	0.05
57	22/42	23/40	0.051
58	23/42	24/40	0.052
59	24/42	24/40	0.029
60	26/42	25/40	0.006
61	27/42	24/40	0.032
62	29/42	28/40	0.009
63	30/42	29/40	0.011
65	32/42	30/40	0.012
67	33/42	34/40	0.064
68	34/42	35/40	0.065
69	35/42	35/40	0.042
71	35/42	36/40	0.067
72	36/42	36/40	0.043
75	38/42	36/40	0.005
78	39/42	38/40	0.021
80	40/42	38/40	0.002
84	40/42	39/40	0.023
90	41/42	39/40	0.001
97	41/42	1	0.024
>97	1	1	0

Next, according to (20)-(22), from *Table 1*, we get

$$d_{42\,40} = \max_{t_k} \lvert H_{12}^1(t_k) - H_{12}^2(t_k) \rvert \cong 0.209,$$

and according to (23)

$$n_{12} = \frac{42 \cdot 40}{42 + 40} = 20.48.$$

Thus, the realization u_n of the statistics (19) is

$$u_n = d_{42\,40}\sqrt{n_{12}} = 0.209\sqrt{20.48} \cong 0.946.$$

Since

$$u_n \cong 0.946 < u = 1.36,$$

then we do not have arguments to reject the null hypothesis H_0.

After proceeding in an analogous way with data in the remaining operation states we obtained the same results, i.e., the conclusions that the sprig data sets and the winter data sets are coming from the populations with the identical distributions. Consequently, we may join spring and winter statistical data into one common more extensive set of data and analyze them without differing the seasons they are coming.

4. Statistical identification of the ferry operation process on the basis of spring and winter data

From the joined statistical data of the ferry operation process that has been collected during spring and winter given in the Appendix 5A in *Tables 1-8*[3], on the basis of methods and procedures given in [13], the following basic operation process statistical parameters are fixed:

- the number of the ferry operation process states

$$\nu = 18,$$

- the ferry operation process observation/experiment time

$$\Theta = 82 \text{ days},$$

- the number of the ferry operation process realizations

$$n(0) = 82,$$

- the numbers $n_b(0)$ of the ferry operation process transients in the particular operation states z_b at the initial moment $t = 0$

$$n_1(0)=82,\ n_2(0)=0,\ \ldots,\ n_{18}(0)=0,$$

where

$$n_1(0)+\ n_2(0)+\ldots.+n_\nu(0)=82,$$

- the vector of realizations of the numbers of the ferry operation process transients in the particular operation states z_b at the initial moment $t = 0$

$$[n_b(0)]=[n_1(0),n_2(0),...,n_\nu(0)]=[82,0,...,0],$$

- the realization n_{bl} of the numbers of the ferry operation process transitions from the state z_b into the state z_l during the experiment time $\Theta=82$ days

$$n_{11}=\ 0,\ n_{12}=82,\ n_{13}=\ 0,\ ...,\ \ n_{117}=0,\ n_{118}=0,$$

$$n_{21}=0,\ n_{22}=\ 0,\ n_{23}=82,\ \ldots,\ n_{217}=0,\ n_{218}=0,$$

$$\ldots,$$

$$n_{171}=0,n_{172}=0,n_{173}=0,\ \ldots,n_{1717}=0,n_{1718}=82,$$

$$n_{181}=82,n_{182}=0,n_{183}=0,\ \ldots,n_{1817}=0,n_{1818}=0,$$

- the matrix of realizations n_{bl} of the numbers of the ferry operation process transitions from the state z_b into the state z_l during the experiment time $\Theta=82$ days

$$[n_{bl}]=$$

$$\begin{bmatrix} n_{11} & n_{12} & \ldots & n_{118} \\ n_{21} & n_{22} & \ldots & n_{218} \\ \ldots & & & \\ n_{171} & n_{172} & \ldots & n_{1718} \\ n_{181} & n_{182} & \ldots & n_{1818} \end{bmatrix} = \begin{bmatrix} 0\,82 & 0\ldots 0 & 0 \\ 0 \;\; 0 & 82\ldots 0 & 0 \\ \ldots & & \\ 0 \;\; 0 & 0\ldots 0 & 82 \\ 82 \;\; 0 & 0\ldots\ 0 & 0 \end{bmatrix},$$

- the realization n_b of the total numbers of the ferry operation process transitions from the operation state z_b during the experiment time $\Theta=82$ days (the sums of the numbers of the matrix $[n_{bl}]$)

$$n_1=n_{11}+n_{12}+...+n_{118}\ =82,$$

$$n_2=n_{21}+n_{22}+...+n_{218}=82,$$

$$\ldots$$

$$n_{18}=n_{181}+n_{182}+...+n_{1818}\ =82,$$

- the matrix of realizations of the total numbers of the ferry operation process transitions from the operation state z_b during the experiment time $\Theta=82$ days

$$[n_b]=[n_1,n_2,...,n_\nu]=[82,82,...,82].$$

On the basis of the above statistical data it is possible to evaluate the basic parameters of the ferry operation process semi-markovian model:

- the vector of realizations

$$[p(0)]=[1,0,0\ldots,0,0],$$

of the initial probabilities $p_b(0)$, $b=1,2,...,18$, (1) [13] of the ferry operation process transients in the particular operation states z_b at the moment $t = 0$

- the matrix of realizations

$$[p_{bl}]=\begin{bmatrix} 0\,1\,0\ldots 0\,0 \\ 0\,0\,1\ldots 0\,0 \\ \ldots \\ 0\,0\,0\ldots 0\,1 \\ 1\,0\,0\ldots 0\,0 \end{bmatrix},$$

of the transition probabilities p_{bl}, $b,l=1,2,...,18$, (2) [13] of the ferry operation process from the operation state z_b into the operation state z_l during the experiment time $\Theta=82$ days.

In the *Tables 1-8*, [13] there are presented the realizations θ_{bl}^k, $k=1,2,...,82$, for each $b=1,2,...,17$, $l=b+1$ and $b=18$, $l=1$ of the ship operation process conditional sojourn times θ_{bl}, $b=1,2,...,17$, $l=b+1$ and $b=18$, $l=1$ in the state z_b while the next transition is the state z_l during the experiment time $\Theta=82$ days.

These statistical data allow us, applying the methods and procedures given in [13], to formulate and to verify the hypotheses about the conditional distribution functions $H_{bl}(t)$ of the ferry operation process sojourn times θ_{bl}, $b=1,2,...,17$, $l=b+1$

and $b = 18$, $l = 1$ in the state z_b while the next transition is to the state z_l on the base of their realizations θ_{bl}^j, $j = 1,2,...,82$.
Using the methods and procedures given in [3], we may verify the hypotheses on the distributions of the sojourn times and we have the following results:
- the conditional sojourn time θ_{12} has a normal distribution with the density function

$$h_{12}(t) = \frac{1}{18.256\sqrt{2\pi}} \exp[-\frac{(t-51.415)^2}{666.563}],$$

$$t \in (-\infty, \infty),$$

- the conditional sojourn time θ_{23} has a distribution with not identified yet density function (the distribution is none of the distinguished in [13] distributions),

- the conditional sojourn time θ_{34} has a Weibull's distribution with the density function

$$h_{34}(t) = \begin{cases} 0, & t < 25.69, \\ 0.299(t-25.69)^{2.485} \\ \exp[-0.086(t-25.69)^{3.485}], & t \geq 25.69, \end{cases}$$

- the conditional sojourn time θ_{45} has a distribution with not identified yet density function (the distribution is none of the distinguished in [13] distributions),

- the conditional sojourn time θ_{56} has a distribution with not identified yet density function (the distribution is none of the distinguished in [13] distributions),

- the conditional sojourn time θ_{67} has a normal distribution with the density function

$$h_{67}(t) = \frac{1}{2.514\sqrt{2\pi}} \exp[-\frac{(t-37.268)^2}{12.636}],$$

$$t \in (-\infty, \infty),$$

- the conditional sojourn time θ_{78} has a Weibull's distribution with the density function

$$h_{78}(t) = \begin{cases} 0, & t < 2.31, \\ 0.537(t-2.31)^{1.715} \\ \exp[-0.1979(t-2.31)^{2.715}], & t \geq 2.31, \end{cases}$$

- the conditional sojourn time θ_{89} has a normal distribution with the density function

$$h_{89}(t) = \frac{1}{8.73\sqrt{2\pi}} \exp[-\frac{(t-19)^2}{152.44}], \quad t \in (-\infty, \infty),$$

- the conditional sojourn time θ_{910} has a Weibull's distribution with the density function

$$h_{910}(t) = \begin{cases} 0, & t < 3.25, \\ 0.047(t-3.25)^{1.319} \\ \exp[-0.0204(t-3.25)^{2.319}], & t \geq 3.25, \end{cases}$$

- the conditional sojourn time θ_{1011} has a exponential distribution with the density function

$$h_{1011}(t) = \begin{cases} 0, & t \leq 1.69, \\ 0.3534 \exp[-0.3534(t-1.69)], & t > 1.69, \end{cases}$$

- the conditional sojourn time θ_{1112} has a chimney distribution with the density function

$$h_{1112}(t) = \begin{cases} 0, & t < 2.81, \\ 0.0434, & 2.81 \leq t < 3.94, \\ 1.6585, & 3.94 \leq t < 4.31, \\ 0.1756, & 4.31 \leq t < 6.19, \\ 0, & t > 6.19, \end{cases}$$

- the conditional sojourn time θ_{1213} has a Weibull's distribution with the density function

$$h_{1213}(t) = \begin{cases} 0, & t < 19, \\ 0.344(t-19)^{1.364} \\ \exp[-0.146(t-19)^{2.364}], & t \geq 19, \end{cases}$$

- the conditional sojourn time θ_{1314} has a chimney distribution with the density function

$$h_{1314}(t) = \begin{cases} 0, & t < 416.5, \\ 0.0014, & 416.5 \le t < 479.5, \\ 0.0151, & 479.5 \le t < 521.5, \\ 0.0033, & 521.5 \le t < 605.5, \\ 0, & t \ge 605.5, \end{cases}$$

- the conditional sojourn time θ_{1415} has a Weibull's distribution with the density function

$$h_{1415}(t) = \begin{cases} 0, & t < 38, \\ 0.1466(t-38)^{1.181} \\ \exp[-0.067(t-38)^{2.181}], & t \ge 38, \end{cases}$$

- the conditional sojourn time θ_{1516} has a chimney distribution with the density function

$$h_{1516}(t) = \begin{cases} 0, & t < 27.94, \\ 0.043, & 27.94 \le t < 32.19, \\ 0.1836, & 32.19 \le t < 36.44, \\ 0.0034, & 36.44 \le t < 47.06, \\ 0, & t \ge 47.06, \end{cases}$$

- the conditional sojourn time θ_{1617} has a distribution with not identified yet density function (the distribution is none of the distinguished in [13] distributions),

- the conditional sojourn time θ_{1718} has a distribution with not identified yet density function (the distribution is none of the distinguished in [13] distributions),

- the conditional sojourn time θ_{181} has a Weibull's distribution with the density function

$$h_{181}(t) = \begin{cases} 0, & t < 0, \\ 0.93t^{0.884}\exp[-0.049t^{1.884}], & t \ge 0. \end{cases}$$

Next, for the above distributions, the mean values $M_{bl} = E[\theta_{bl}],\ b,l = 1,2,...,18,\ b \ne l,$ (11) [13] of the ferry operation process conditional sojourn times in particular operation states may be determined and they are as follows:

$$M_{12} = 51.415,\ M_{34} = 36.176,\ M_{67} = 37.268,$$

$$M_{78} = 6.807,\ M_{89} = 19,\ M_{910} = 46.614,$$

$$M_{1011} = 2.829,\ M_{1112} = 4.459,\ M_{1213} = 25.091,$$

$$M_{1314} = 513.689,\ M_{1415} = 51.182,\ M_{1516} = 33.807,$$

$$M_{181} = 18.039.$$

In the remaining cases the mean values $M_{bl} = E[\theta_{bl}]$ after successful uniformity testing their approximate values are:

$$M_{23} = 2.533,\ M_{45} = 52.393,\ M_{56} = 530.188,$$

$$M_{1617} = 4.448,\ M_{1718} = 5.473.$$

5. Conclusion

The procedure of statistical data sets uniformity analysis based on Kolmogorov-Smirnov test is proposed to be applied to the empirical sojourn times coming from the operation processes of complex technical systems. The proposed procedure is practically applied to the analysis and uniformity testing of the maritime ferry spring and winter sets of realizations of the sojourn times in particular operation states. Next, after successful uniformity testing, the spring and winter data coming from the ferry operation process were joined into common data sets and the identification of the this process parameters was performed.

Acknowledgements

The paper presents partial results of the work in the Poland-Singapore Joint Research Project titled "Safety and Reliability of Complex Industrial Systems and Processes" supported by grants from the Poland's Ministry of Science and Higher Education (MSHE grant No. 63/N-Singapore/2007/0) and the Agency for Science, Technology and Research of Singapore (A*STAR SERC grant No. 072 1340050).

References

[1] Barbu, V. & Limnios, N. (2006). Epirical estimation for discrete-time semi-Markov processes with applications in reliability. *Journal of Nonparametric Statistics*, Vol. 18, No. 7-8, 483-498.

[2] Black, M., Brint, A.T. & Brailsford, J.R. (2005). A semi-Markov approach for modelling asset deterioration. *Journal of the Operational Research Society*, Vol. 56, No. 11, 1241-1249.

[3] Blokus-Roszkowska, A., Guze, S., Kołowrocki, K., Jurdziński, M., Kwiatuszewska-Sarnecka, B., Milczek, B., Soszyńska, J., Salahuddin, H., M. & Fu Xiuju. (2009). Data mining for identification

and prediction of operation, safety and reliability characteristics of complex industrial systems and processes. WP6. Methods and algorithms for evaluating unknown parameters of system operation, reliability and safety models and their applications in port, shipyard and maritime transport. WP6 - Task 6.1.3 – English – 30.11.2009. Poland-Singapore Joint Project. MSHE Decision No. 63/N-Singapore/2007/0.

[4] Dziula, P., Jurdziński, M., Kołowrocki, K. & Soszyńska, J. (2007). On multi-state safety analysis in shipping. *International Journal of Gnedenko e-Forum "Reliability: Theory & Application"*, Vol. 2, No 3-4, 40-53.

[5] Ferreira, F. & Pacheco, A. (2007). Comparison of level-crossing times for Markov and semi-Markov processes. *Statistics & Probability Letters*, Vol. 77, No. 2, 151-157.

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

[7] Giudici, P. & Figini, S. (2009). *Applied data mining for business and industry*. John Wiley & Sons Ltd.

[8] Habibullah, M. S., Lumanpauw, E., Kolowrocki, K., Soszynska, J. & Ming, N. G. (2009). A computional tool for general model of industrial systems operation processes. *Electronic Journal Reliability & Risk Analysis: Theory & Applications*, Vol. 2, No 4, 181-191.

[9] Hryniewicz, O. (1995). Lifetime tests for imprecise data and fuzzy reliability requirements. Reliability and Safety Analyses under Fuzziness. Onisawa T. and Kacprzyk J., Eds., Physica Verlag, Heidelberg, 169-182.

[10] Kołowrocki, K. (2003). Asymptotic approach to reliability analysis of large systems with degrading components. *International Journal of Reliability, Quality and Safety Engineering*, Vol. 10, No 3, 249-288.

[11] Kołowrocki, K. (2004). *Reliability of Large Systems*. Amsterdam - Boston - Heidelberg - London - New York - Oxford - Paris - San Diego - San Francisco - Singapore - Sydney - Tokyo, Elsevier.

[12] Kołowrocki, K. (2008). Reliability of large systems, Section in Encyclopedia of Quantitative Risk Analysis and Assessment, John Wiley & Sons, Vol. 4, 1466-1471.

[13] Kołowrocki, K. & Soszyńska, J. (2008). A general model of industrial systems operation processes related to their environment and infrastructure. *Proc. 2nd Summer Safety and Reliability Seminars – SSARS 2008*. Vol. 2, 223-226.

[14] Kolowrocki, K. & Soszynska, J. (2009). Modelling 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.

[15] Kolowrocki, K. & Soszynska, J. (2009). Reliability, risk and availability based optimization of complex technical systems operation processes. Part 1. Theoretical backgrounds. *Electronic Journal Reliability & Risk Analysis: Theory & Applications*, Vol. 2, No 4, 2009, 141-152.

[16] Kolowrocki, K. & Soszynska, J. (2009). 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.

[17] Kołowrocki, K., Soszyńska, J., Jurdziński, M. & Dziula, P. (2007). On multi-state safety analysis in shipping. *International Journal of Reliability, Quality and Safety Engineering. System Reliability and Safety*, Vol. 14, No 6, 547-567.

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

[19] Ross, S. M. (2007). *Introduction to probability models.* Academic Press, Elsevier, San Diego.

[20] Soszyńska, J. (2006). Reliability of large series-parallel system in variable operation conditions. *International Journal of Automation and Computing*, Vol. 3, No 2, 199-206.

[21] Soszyńska, J. (2006). Reliability evaluation of a port oil transportation system in variable operation conditions. *International Journal of Pressure Vessels and Piping*, Vol. 83, Issue 4, 304-310.

[22] Soszyńska, J. (2007). Systems reliability analysis in variable operation conditions. *International Journal of Reliability, Quality and Safety Engineering. System Reliability and Safety*, Vol. 14, No 6, 617-634.

[23] Soszynska, J. (2009). Asymptotic approach to reliability evaluation of large "m out of l"- series system in variable operation conditions. *Journal Reliability & Risk Analysis: Theory & Applications*, Vol. 2, 9-43.

[24] Wilson, A.G., Graves, T.L., Hamada, M.S. et al. (2006). Advances in data combination, analysis and collection for system reliability assessment. *Statistical Science*, Vol. 21, No. 4, 514-531.