ELECTRONIC JOURNAL
OF INTERNATIONAL
GROUP ON RELIABILITY

JOURNAL IS REGISTERED
IN THE LIBRARY
OF THE U.S. CONGRESS

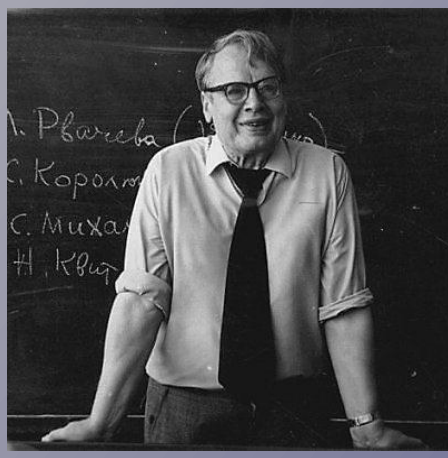
•

ISSN 1932-2321

•

VOL.11 NO.4 (43)
DECEMBER, 2016

Gnedenko Forum Publications



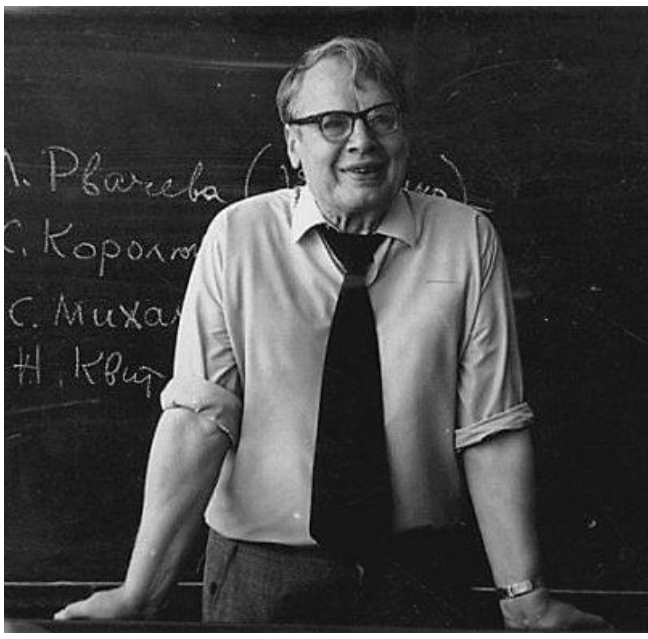
RELIABILITY:
THEORY&APPLICATIONS

San Diego

ISSN 1932-2321

http://www.gnedenko-forum.org/Journal/index.htm





RELIABILITY:

THEORY & APPLICATIONS

Vol.11 No.4 (43),
December, 2016

San Diego
2016

Send your paper

e-Journal ***Reliability: Theory & Applications*** publishes papers, reviews, memoirs, and bibliographical materials on Reliability, Quality Control, Safety, Survivability and Maintenance. Theoretical papers have to contain new problems, finger practical applications and should not be overloaded with clumsy formal solutions.

Priority is given to descriptions of case studies.
General requirements for presented papers

1. Papers have to be presented in English in MS Word format. (Times New Roman, 12 pt, 1 intervals).
2. The total volume of the paper (with illustrations) can be up to 15 pages.
3. A presented paper has to be spell-checked.
4. For those whose language is not English, we kindly recommend to use professional linguistic proofs before sending a paper to the journal.

* * *

The Editor has the right to change the paper title and make editorial corrections.

The authors keep all rights and after the publication can use their materials (re-publish it or present at conferences).

Publication in this e-Journal is equal to publication in other International scientific journals.

Papers directed by Members of the Editorial Boards are accepted without referring.
The Editor has the right to change the paper title and make editorial corrections.

The authors keep all rights and after the publication can use their materials (re-publish it or present at conferences).

Send your papers to
Alexander Bochkov
a.bochkov@gmail.com

Table of Contents

Resilience of Finite Networks Against Simple and Combined Attack on Their Nodes

Ilya B. Gertsbakh

Department of Mathematics, Ben-Gurion University , Beer-Sheva, 84105, Israel
elyager@bezeqint.net; Corresponding author

Radislav Vaisman

School of Mathematics and Physics, The University of Queensland, Australia r.vaisman@uq.edu.au

Abstract

We compare the reliability of two finite networks with the same vertex degree and the same number of nodes; a regular 16x16 grid and a Poisson network. Both networks are subject to random removal of their nodes, and the network failure is defined as the reduction of the maximal component beyond some critical level α. The main tool for comparing the network resilience are the marginal cumulative D-spectra (signatures) of the net- works. It was demonstrated that the regular grid for small α is less reliable than the Poisson network. We study also the situation when multiple hits of the same node are allowed. We demonstrate that finite networks behave similar to infinite random network with regard to the fraction of nodes to be removed to create "similar" giant components containing the same fraction of network nodes. Finally, we consider a combined attack on net- work nodes by two-type of "shells" where the node fails only if it is hit by "shells" of both types. For this case, we derive a formula for determin- ing the minimal number of "shells" which destroy the network with given probability.

Key words:finite network, attack on network node, cumulative D-spectra (signature); combined attack on nodes.

1.Introduction and Preliminaries. D-spectrum

There are many works that study interaction between networks [1,3,4,5,10,11]. The following features are typical for probabilistic models of the network interaction: a) networks are assumed to be very large, formally infinite; b) network N_1 affects another network N_2 by creation of a random connections between their nodes in such a way that a node $a \in N_1$ **hits** a randomly chosen node $b \in N_2$. There are models of interaction in which the choice of node $b \in N_2$ is not random and "hubs" in N_2 have higher probability of being hit than "regular" nodes. Typically, in all these models, the particular structure of network N_2 is not specified, except for the node degree distribution.

In this paper we deal only with finite networks having well-defined structure. In other words, network N subject to an attack is given as a pair $N = (V, E)$, where V is the vertex or node set, $|V| = n$, and E is the edge or link set, $|E| = m$. We assume that if a node is attacked, then all links adjacent to it are erased, while the node remains untouched.

As a measure of network state in the process of node failures we consider the size of its largest component, i.e. the largest set of connected nodes in the network. It is more convenient to

characterize the size of the largest component by the fraction α of nodes n in the network that belong to this component. For example, we will consider a 16x16 grid with $n = 256$. If its largest component has 200 nodes, we will say that $\alpha = 200/256 = 0.78$ We will also define several states of network degradation measured by these fractions $\alpha_9 > \alpha_8 > \ldots > \alpha_1$. We say that the network is in perfect state (State10) if its largest component has size $L \geq \alpha_9$, Generally, we say that the network is in state s if its largest component has size L, $\alpha_s > L \geq \alpha_{s-1}$, $s = 9,8,\ldots 2$. Finally, network state is 1 if $\alpha_1 > L$.

Our main goal is to analyze network probabilistic behavior when network nodes are subject to random node failures. The main tool for this analysis will be so called D-spectra technique.

Denote by $e_1, e_2, \ldots, e_n$ network components subject to failure (the nodes), and let π be a random permutation of components numbers,

$$\pi = (e_{i_1}, e_{i_2}, \ldots, e_{i_n}).$$

Suppose that all these components are *up* and we move along the permutation, from left to right, and turn each component from *up* to *down.* Suppose that network state is controlled after each step. Typically, we will observe exactly 9 occasions when network state has changed: first - from the perfect State10 to state $s = 9$, from $s = 9$ to $s = 8$, and so on, until the transition from $s = 2$ to $s = 1$.

Definition 1. (*The anchors*)

The ordinal number in the permutation π of the component whose turning down causes network state to change from $10 - k$ to $10 - (k + 1)$, $k = 0,1,\ldots,8$ is called the $(k + 1)$-st *anchor* and is denoted by $r_{s+1}(\pi)$. Each permutation has, therefore, k anchors. So, the first anchor signifies the transition $10 \Rightarrow 9$, the second - $9 \Rightarrow 8$,...,the ninth - $2 \Rightarrow 1$

Definition 2. *(Multidimensional D-spectrum)*

Assume that all $n!$ permutations are equally probable. The k-dimensional discrete density

$$f(\delta_1, \delta_2, \ldots, \delta_9) = P(r_i(\pi) = \delta_i, i = 1,2,\ldots,9) =$$

$$\frac{number\ of\ permutations\ with\ \ r_i(\pi)=\delta_i, i=1,2,\ldots,9}{n!} \quad (1)$$

for $1 \leq \delta_1 < \delta_2 < \cdots \leq \delta_9 \leq n$ is called network *multidimensional D-spectrum*.#

A few comments. Letter "D" for the spectrum signifies the process of destruction since we turned from *up* to *down* network components moving along the permutation. In literature, the multidimensional D-spectrum is termed also as a multidimensional signature, see [6,8].

Obviously,

$$\sum_{1\leq\delta_1<\delta_2<\cdots<\delta_9\leq n} f(\delta_1, \delta_2, \ldots, \delta_9) = 1.$$

It is important to stress that the D-spectrum is a *combinatorial parameter* of the network that depends only on network structure and its states definition. It does not depend on probabilistic characterization of the real random mechanism governing network component failures.

Our main interest will be in the probabilistic description of each particular anchor. Formally speaking, our main tool will be the distributions of the positions of each of the 9 anchors.

Definition 3. *The j-th marginal D-spectrum*

The distribution

$$f^{(j)} = (f_1^{(j)}, f_2^{(j)}, \ldots, f_n^{(j)})$$

of the position of the j-th anchor is called the j-th *marginal D-spectrum*.#

Here $f_i^{(j)} = P(\ the\ j-th\ anchor\ \ position\ \ is\ \ \ i)$.

Obviously,

$$f_i^{(j)} = P(r_j(\pi) = i) =$$

$$\sum_{1\leq\delta_1<\delta_2<\ldots\delta_j=i<\ldots<\delta_9\leq n} f(\delta_1, \delta_2, \ldots, \delta_j = i, \ldots, \alpha_9) \quad (2)$$

In what follows, it is more convenient is to operate with a so-called cumulative (marginal) D-

spectra.

Definition 4. *The j-th cumulative D-spectrum*

The cumulative distribution function (cdf) $F^{(j)}(x)$ of the position of the j-th anchor in random permutation π is called the j-th *cumulative D-spectrum*:

$$F^{(j)}(x) = \sum_{i=1}^{x} f_i^{(j)}, x = 1,2,\dots,n.\# \quad (3)$$

Let us clarify the probabilistic meaning of $F^{(j)}(x)$. Divide all network states into two sets U and D. Let all states $J \geq (10 - j)$ belong to U, and all the remaining states - to D, $j = 0,1,2,\dots,9$. Denote by $Y_{(j)}$ the random number of components needed to turn down in the course of the destruction process to cause the transition from U to D. Note that $f_i^{(j)} = P(Y_{(j)} = i)$. Then

$$F^{(j)}(x) = P(Y_{(j)} \leq x).$$

In words: $F^{(j)}(x)$ is the cdf of the number of components to be destroyed to cause the transition from U to D

Remark 1

Suppose that the network has only two states: the UP state, if its largest component has size $L \geq \alpha = 0.7$, and the complementary state D. There will be only one anchor designating the position of the component whose destruction leads to the transition $U \Rightarrow D$. The corresponding D-spectrum is nothing but so-called *signature* introduced by Samaniego [15] and the cumulative D-spectrum is the so-called *cumulative signature*, see [15,16].#

Remark 2

Consider a star network with central node a and three peripheral nodes b, c, d that are connected to a by links $(a, b), (a, c), (a, d)$. If node a fails, the network disintegrates into four isolated components. If network state is defined according to the size of its largest component, we observe a jump from state 4 to state 1.

Formally speaking, it may happen that in the process of component destruction we may observe a transition from state $J - A$ to state $J - A - B$, $B > 1$. Suppose it happens after destructing component standing on the i-th position. Then we put $r_{A+1}(\pi) = \dots = r_{A+B}(\pi) = i$ and therefore formally provide that all permutations have the same number of anchors.#

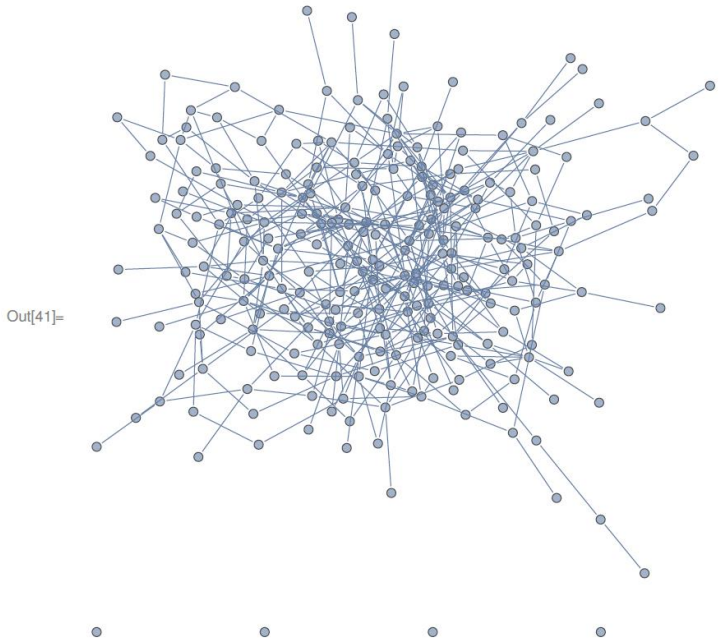


Figure 1: Random graph with $n = 252$ nodes and average node degree $d = 3.75$

2. **Resilience of a Regular Grid vs Random Graph**

In this section we compare the resilience of two finite networks having approximately the same number of nodes and the same average node degree. The first network (call it 'Grid') is a 16x16 regular grid with $n = 256$ nodes and 480 edges, see Figure 3. The average node degree is 960/256=3.75.

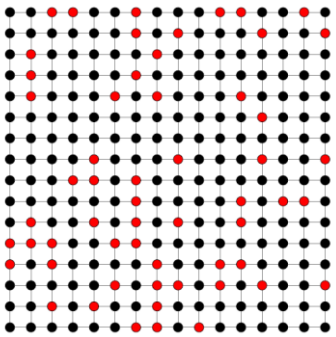


Figure 2: 16x16 Grid. Failed nodes are shown by red

The second network is a random Poisson graph (call it "Map") with 252 nodes and 473 edges thus having the same average node degree 3.75. Each of these networks was subject to random node removal. For Grid and Map we introduced several states according to the *fraction L* of all nodes in the maximal (connected) component. This component is an analogue of the giant component in an infinite network:

$$State10: L \geq 0.9; State9: 0.8 \leq L < 0.9; \ldots; State2: 0.1 \leq L < 0.2; State1: L < 0.1.$$

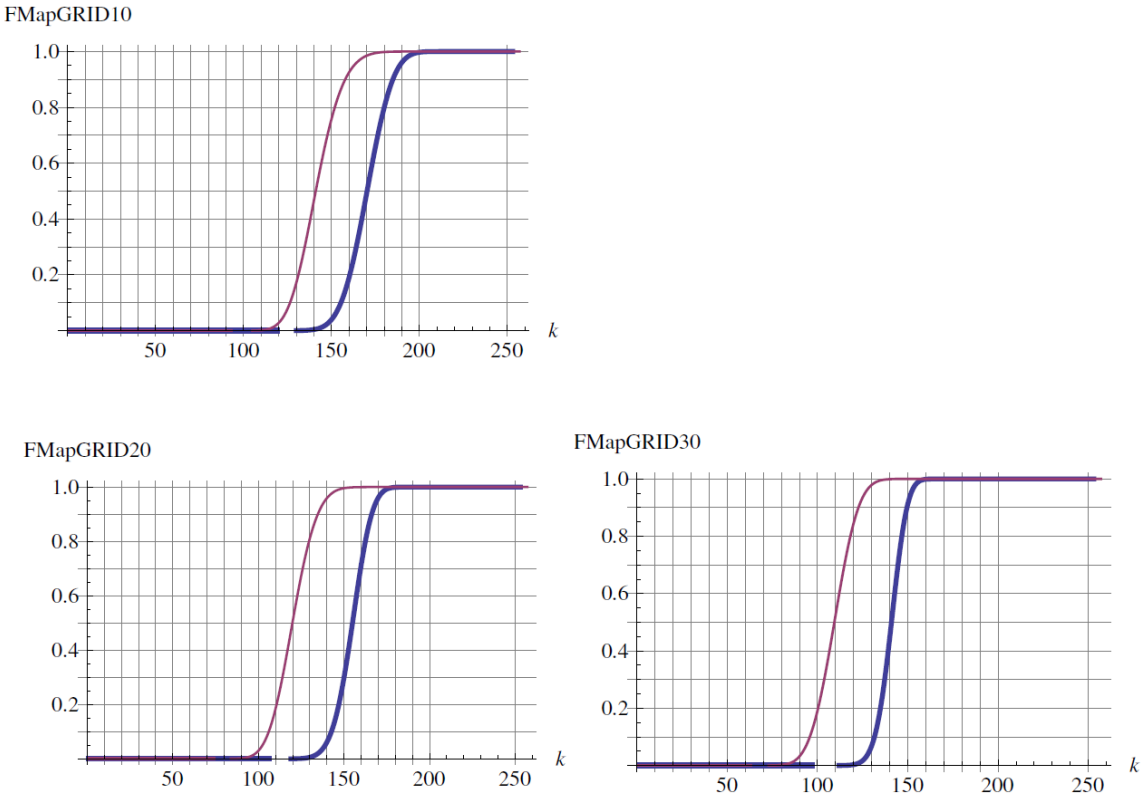


Figure 3: Cumulative D-spectra for Grid vs Map networks. Upper pair is for transition 2 ⇒ 1; left lower - for transition 3 ⇒ 2, right - for transition 4 ⇒ 3. In each pair, the right curve (blue) is for Map, the left - for Grid

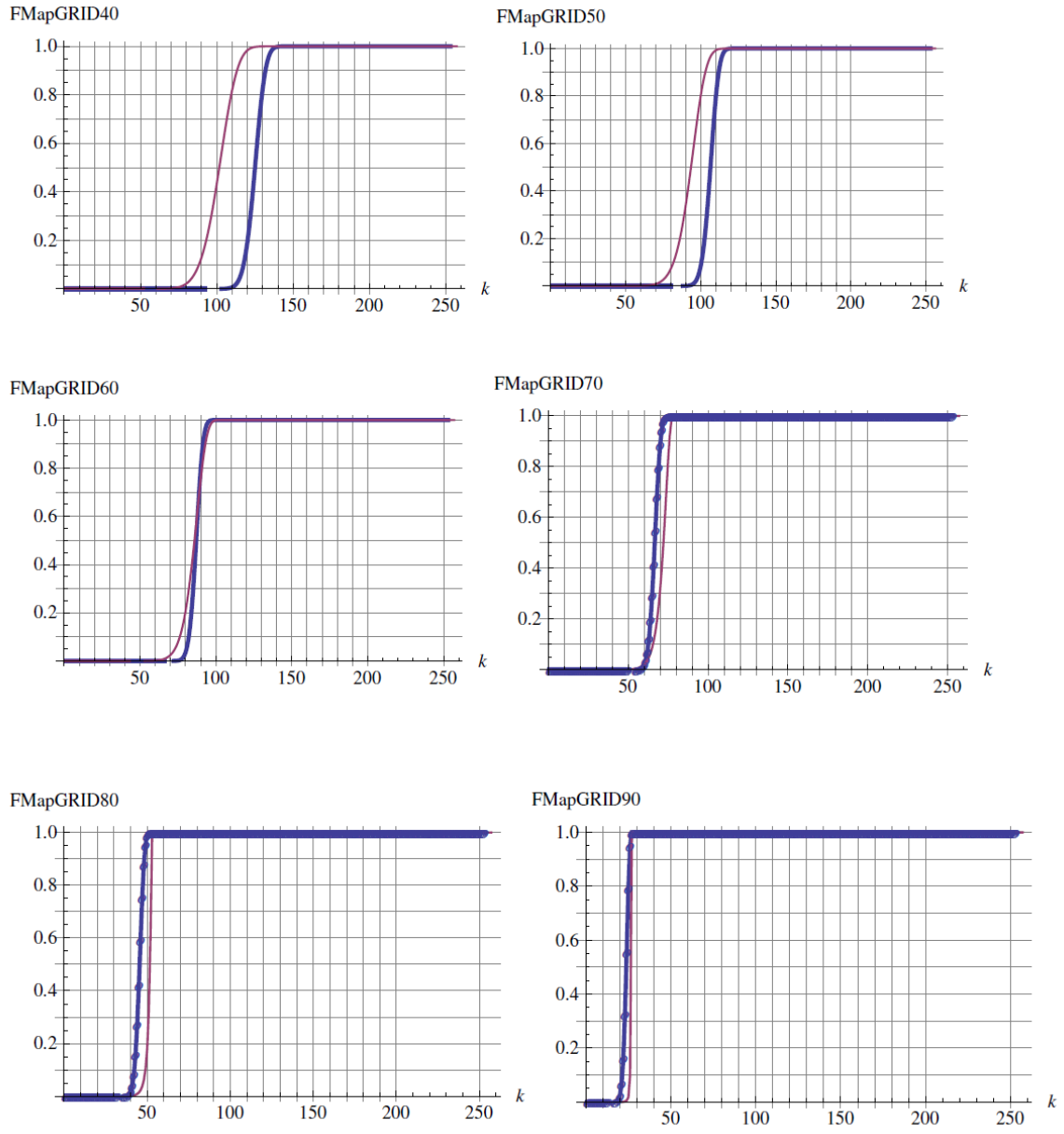


Figure 4: Cumulative D-spectra for Grid vs Map networks. Upper pair is for transition 5 ⇒ 4 (left) and 6 ⇒ 5 (right). The red curve is for Grid, blue - for Map. Here again Map is more resilient. Situation changes for the transitions 7 ⇒ 6,8 ⇒ 7,9 ⇒ 8 and 10 ⇒ 9, see the graphs in the middle and the bottom

.

Random permutation π has therefore nine anchors $r_1(\pi), r_2(\pi), \ldots, r_9(\pi)$ signifying the transition from *State10* to State9, from *State9* to State8, etc. The cumulative marginal D-spectra are shown on Figures 3 and 4. Figure 3 shows Grid vs Map marginal spectra for the transitions 2 ⇒ 1,3 ⇒ 2,4 ⇒ 3.

The spectrum for Grid is in red, for Map -in blue. The most surprising and not expected phenomenon is that the Grid marginal spectra are shifted to the left from the Map spectra. It means that Map is more resilient than the Grid! Let us examine the graph *FMapGrid30*. If about 110 nodes are destroyed, the largest component of Grid with probability about 0.5 has $0.3 \cdot 256$ nodes while Map has not suffered at all. To cause the transition 3 ⇒ 2 with probability 0.5 for Map, one has to

destroy about 140 nodes !

Two upper graphs on Figure 3 present the Map vs Grid spectra for the transitions $5 \Rightarrow 4$ and $6 \Rightarrow 5$. Here again we see that blue curves (Map spectra) are on the right of the Grid spectra.The advantage in resilience of Map vs Grid vanishes when the fraction of nodes in the maximal component becomes ≥ 0.6, see the graphs in the middle row and in the bottom of Figure 4. When the maximal component has 0.9-0.7 faction of all nodes, Grid is slightly more resilient. For $\alpha = 0.6$ both spectra practically coincide, see middle graphs, on the left.

In [7], comparison was made of the resilience for small random networks ($n \leq 40$) vs regular networks, for node degrees $d = 3,4$ and 5. Network failure was defined as the decrease of component becomes below $0.3n$. It was observed that for $d = 5$, the regular network is more resilient, but its advantage over random network became very small when d was 4 or 3.

3. Multiple Hits

When an external source produces a hit on a randomly chosen node of a network that has N nodes, $N \to \infty$, the probability of multiple hits of the same node can be neglected. The situation changes drastically when the network subject to an external attack has a finite number of nodes n. Formally, we are in a situation well-studied in classical probability theory. Suppose that b balls are randomly placed into n boxes. We need to find the probability $p(k|b)$ that there will be exactly k boxes that will contain at least one ball. This problem is known is combinatorics as *occupancy* problem and its solution is given by the famous DeMoivre's formula [2], p 242:

$$p(k|b) = \frac{n!}{k!(n-k)!} \sum_{t=0}^{k} (-1)^t \frac{k!}{t!(k-t)!} \frac{(k-t)^b}{n^b}, k = 1, \ldots, min(n, b). \quad (4)$$

We are interested now in finding network $DOWN$ probability $P(DOWN; b)$ when it is hit by b "balls". (A node that receives more than one hit remains *down*). Suppose that the network entrance into the *DOWN* state is described by j-th marginal cumulative D-spectrum $F^{(j)}(x)$. Using the Total Probability formula, we obtain that

$$P(DOWN; b) = \sum_{k=0}^{\min(n,b)} p(k|b) \cdot F^{(j)}(k), \quad (5)$$

where $p(k|b)$ is given by (4).

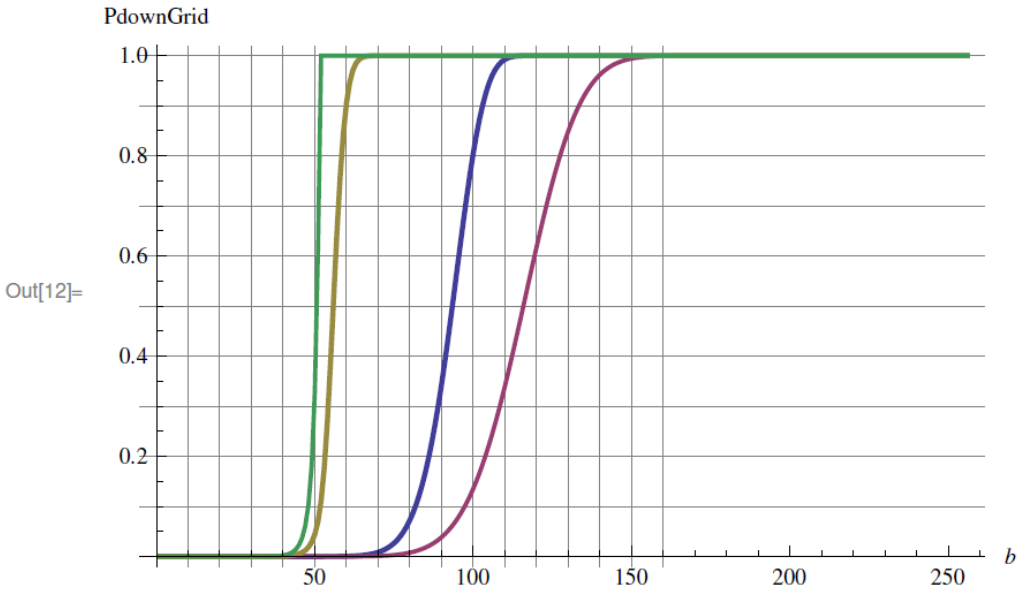


Figure 5: Comparison of $P(DOWN; b)$ (right curve in each pair) with $F^{(j)}(b)$ (left curve in each pair), for GRID. Left pair is for $j = 2$; right pair - for $j = 5$

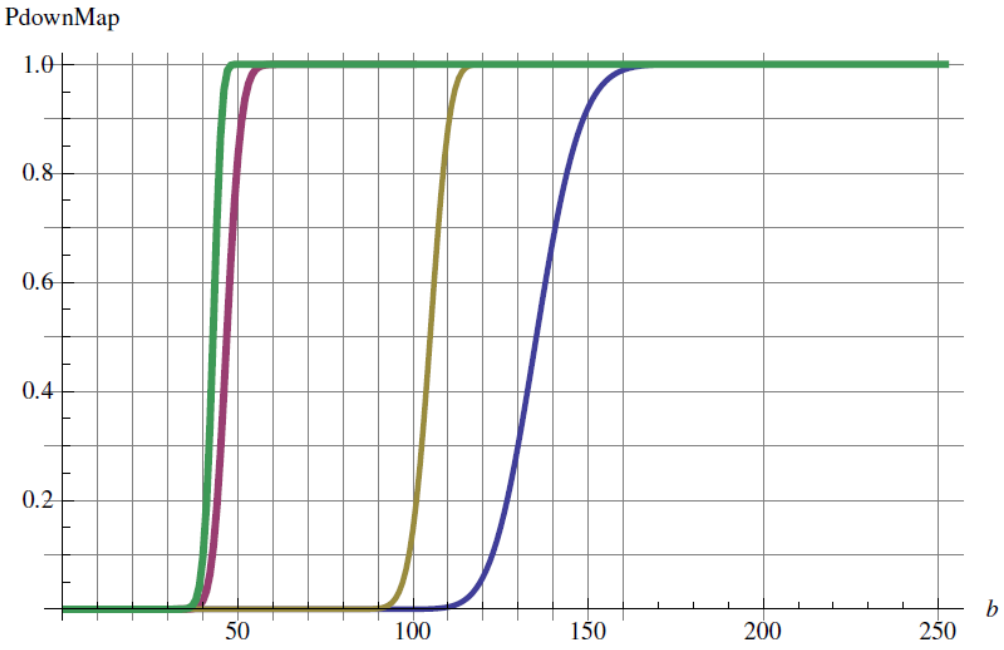


Figure 6: Comparison of $P(DOWN;b)$ (right curve in each pair) with $F^{(j)}(b)$ (left curve in each pair) for Map. Left pair is for $j=2$; right pair - for$j=5$

Examine, for example, the right pair of curves on Figure 5. The left one (in blue) is the cumulative marginal spectrum of GRID, for which the *DOWN* state is defined as the drop of largest component below $0.5n$. The red curve is $P(DOWN;b)$ as a function of the number of "balls" thrown on the nodes of GRID. We see that, e.g., for probability 0.8, the horizontal distance between the curves is about 25, which means that about 25 nodes (out of approximately 125) are hit more than once. The comparison of the green and yellow curve on this figure shows that here the number of nodes with multiple hits is much smaller because the transition $9 \Rightarrow 8$ takes place after considerably smaller number of damaged nodes (50-60). Figure 6 presents a similar picture for Map network.

4. Comparing Giant Component in Infinite Poisson Network With Maximal Component in a Finite Network

Let us consider first a Poisson random infinite network in which a fraction β of its nodes is randomly chosen and removed. Then the size of the giant component G can be found from the following equation, see [12] page 597:

$$G=(1-\beta)(1-e^{-d\cdot G}), \qquad (6)$$

where d is the average node degree. Take $d=3.75$. Let us take $G=0.9(0.1)0.1$ fraction of all nodes and find out from (6) the corresponding values of β. These values are presented in the second row of Table 1.

Table 1: Giant component vs maximal component in the Map

.

G	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
β	0.068	0.203	0.245	0.329	0.409	0.485	0.556	0.621	.680
$q_{0.5}$	0.079	0.0171	0.258	0.341	0.420	0.496	0.559	0.614	0.680

Now let us compare the β-values with the number of nodes hit in the Map in order to provide with probability 0.5 that the maximal component is reduced to $0.9(0.1)0.1$ fraction of 252 nodes. For this purpose we have to look at the values of $q_{0.5}$, the 0.5-quantiles, of the marginal cumulative D-spectra,

$$F^{(j)}(q_{0.5}) \approx 0.5.$$

The values of $q_{0.5}$ are given in the third row of Table 1. We can say that the average fraction of nodes to be removed to create a giant component in an infinite Poisson network is very close to the median fraction of nodes to be removed in finite Map network.

5. Network Attacked by Interacting "Shells"

This title means that the network nodes are hit ("bombarded" or "contaminated") by two different kind of substances ("shells"). The "substances" (it might be infection, computer viruses, explosives, etc.) have the property that they interact between themselves and make damage to a node only if this node is hit *by both substances* .

Using the neutral combinatorial language, M white balls and N black balls are randomly located in n boxes (representing the nodes) The node is hit (fails) if in the corresponding box are balls of different colors. For example, there are 5 boxes numbered 1, 2, 3, 4, 5, boxes 1 and 2 contain one white ball, box 5 has 2 white balls. Also, each box contains one black ball. So, boxes 1,2 and 5 contain balls of different colors and represent the nodes that are hit. Our task is to find out the number B of boxes containing balls of both colors.

Lemma 1.

The mean value of B equals

$$E(B) = n^2(1-(1-\tfrac{1}{n})^M)(1-(1-\tfrac{1}{n})^N).\# \quad (7)$$

Lemma 2.

$$E(B^2) = n(1-(1-1/n)^M(1-(1-1/n)^N) + n(n-1)(1-2(1-1/n)^M + (1-2/n)^M) +$$

$$n(n-1)(1-2(1-1/n)^N + (1-2/n)^N).\# \quad (8)$$

The proof of Lemma 1 and 2 are given in the Appendix [18].

The following theorem follows from Lemmas 1 and 2:

Theorem 1

If $M = \alpha n$, $N = \beta n$, $n \to \infty$ then

$$E(B) = n(1-e^{-\alpha})(1-e^{-\beta}), \quad (9)$$

and

$$Var(B) = n(1-e^{-\alpha})(1-e^{-\beta})(1-(1-e^{-\alpha})(1-e^{-\beta})).\# \quad (10)$$

The following theorem was established by [14], see also [13], Section 3:

Theorem 2

If $n \to \infty$, and $0 < c_1 < \beta < c_2 < \infty$, and $0 < c_1 < \alpha < c_2 < \infty$, the random variable $Y = \frac{B-E(B)}{(Var(B))^{0.5}}$ is asymptotically normal $N(0,1).\#$

Denote by q_ε the ε-quantile of $N(0,1)$. Then we arrive at the following

Corollary

Suppose $M = N$.To guarantee that with probability $1-\varepsilon$ that the number of nodes hit by balls of both colors is at least B_{min}, we have to take $N = \gamma \cdot n_0$ where n_0 is the number of nodes in the network and γ is the root of the following equation:

$$q_\varepsilon n_0^{0.5}((1-e^{-\gamma})(1-(1-e^{-\gamma})^2)^{0.5} + n_0(1-e^{-\gamma})^2 = B_{min}.\# \quad (11)$$

Example

Suppose that we want to guarantee with probability 0.9987 that the maximal component of Grid network will be less or equal 128 nodes, i.e. $\alpha = 0.5$. We see from Figure 4 that this will be provided if the number of failed nodes is at least $120 = B_{min}$.We find that $q_{0.0013} = -3$ and solve the equation (11):

$$-3(256^{0.5})(1-e^{-\gamma})(1-(1-e^{-\gamma})^2)^{0.5} + n(1-e^{-\gamma})^2 = 120,$$

Using *Mathematica* "FindRoot" operator [17], we find that the root equals $\gamma = 1.384$, which means that $M = N = 1.385 \cdot 256 = 354$.

6. Concluding remarks

We demonstrated that the resilience and survivability of finite networks under random attack on their nodes can be efficiently studied using marginal D-spectra techniques. Let us note without going into technical detail that the spectra can be efficiently estimated by well-developed Monte Carlo algorithms, with sufficient accuracy and in short CPU times, see [9].

Comparison between a regular grid and random graph having the same number of nodes and the same node degree reveals that the regular graphs are considerably less resilient for $\alpha \leq 0.5$ and that their inferiority in reliability vanishes when the networks's largest components contain large fraction of the nodes ($\alpha \geq 0.6$)

We demonstrated how to compute network reliability by taking into consideration multiple hits of their nodes.

Our simulation revealed that there are certain similarities between creation of a giant component in infinite random network and the largest component in a finite random network.

Finally, we investigated the case of combined attack on a network nodes with two interacting "substances". In this attack, a node fails only if it is hit by two types of "shells". We showed how one can obtain an estimate of the number of "shells" of both types that guarantee network destruction with given probability.

Appendix [18]

1. Let n be the number of boxes (bins) and M be the number of red balls. Each ball is randomly allocated to one of the boxes. Let

$$R_i = \{\, box \;\; i \;\; contains \;\; at \;\; least \;\; one \;\; red \;\; ball \,\}$$

Denote by 1_{R_i} the 1/0 indicator variable of the event R_i. Obviously,

$$P(R_i) = E[1_{R_i}] = 1 - (((n-1)/n))^M. \quad (12)$$

Let $X = \sum_1^n 1_{R_i}$. Obviously,

$$E[X] = \sum_1^n E[1_{R_i}] = nE[1_{R_i}] = n(((n-1)/n))^M. \quad (13)$$

If $n, M \to \infty$ and $M = \gamma n$, then

$$E[X] = n(1-e^{-\gamma})$$

2. Suppose we have N white balls which, independently of the red balls,are located randomly into the same n boxes (bins). Denote by B the random number of boxes containing balls of both colors. Obviously,

$$B = \sum_{i=1}^n 1_{R_i} 1_{W_i}. \quad (14)$$

From linearity of expectation and independence of events R_i and W_i,

$$E[B] = \sum_{i-1}^n E[1_{R_i}]E[1_{W_i}] = n(((n-1)/n))^M \cdot (((n-1)/n))^N. \quad (15)$$

3. For deriving the expression for $Var[B]$ we need the following formula:

$$E[1_{R_1} \cdot 1_{R_2}] = 1 - 2(\tfrac{n-1}{n})^M + (\tfrac{n-2}{n})^M. \quad (16)$$

Note that

$$E[1_{R_1} \cdot 1_{R_2}] = P(R_1 \wedge R_2) = 1 - P(\overline{R_1} \vee \overline{R_2}) =$$

$$P(\ box\ 1\ or\ box\ 2\ is\ empty\ ) = P(\overline{R_1}) + P(\overline{R_2}) - P(\overline{R_1} \wedge \overline{R_2}) =$$

$$2(\tfrac{n-1}{n})^M + (\tfrac{n-2}{n})^M,$$

and (16) follows.

4.Now we are ready to obtain the expression for $E[B^2]$.

$$E[B^2] = E(\sum_1^n 1_{R_i} \cdot 1_{W_i})^2 = E(\sum_1^n 1_{R_i}^2 \cdot 1_{W_i}^2 + \sum_{i \neq j} 1_{R_i} 1_{R_j} 1_{W_i} 1_{W_j}) =$$

$$nE[1_{R_1} 1_{W_1}] + n(n-1)E[1_{R_1} 1_{W_1} 1_{R_2} 1_{W_2} = n(\tfrac{n-1}{n})^M (\tfrac{n-2}{n})^N +,$$

$$n(n-1)(1 - 2(\tfrac{n-1}{n})^M + 2(\tfrac{n-2}{n}))^M \cdot n(n-1)(1 - 2(\tfrac{n-1}{n})^N + 2(\tfrac{n-2}{n}))^N.$$

Now $Var[B] = E[B^2] - (E[B])^2$. Substituting the expressions for $E[B^2]$ and $E[B]$ we obtain that

$$Var[B] = n(1-x^M)(1-x^N) + n(n-1)(1 - 2x^M + y^M)(1 - 2x^N + y^N) - (n(1-x^M)(1-x^N)))^2, \quad (17)$$

where $x = (n-1)/n$ and $y = (n-2)/n$.

Now assume that $M = N \to \infty$. Then $x \to e^{-\gamma}$ and $y \to e^{-2\gamma}$. After simple algebra we obtain that

$$E[B] = n(1 - e^{-\gamma})^2. \quad (18)$$

and

$$Var[B] = n(1 - e^{-\gamma})^2 \cdot (1 - (1 - e^{-\gamma})^2). \quad (19)$$

It is remarkable that the variance of B is asymptotically of order n, i.e. the st.deviation of B is of order $\sqrt{n}$.#

References:

1. Buldyrev, S.V., Parshani, R, Havlin,S. et al.2010. Catastrophic cascade of failures in interdependent networks. *Nature* 464: 1025- 1028.
2. David, F.N. and D.E.Barton. 1962. *Combinatorial Chance*. Charles Griffin and Co Ltd, London.
3. Dickinson, Mark,Havlin, S. and Stanley,H.2012. Epidemics in interconnected networks.*Physical Review*, E85 066109,
4. Gao, J. et al.2012. Robustness of network formed by n interdependent networks with one-to-one correspondence between nodes. *Phys. Rev*, E85:066134.
5. Gertsbakh, I., Shpungin,Y. and R. Vaisman. 2014.*Ternary Networks: Reliability and Monte Carlo*.Springer
6. Gertsbakh, I. and Y. Shpungin.2011. *Network Reliability and Resilience*. Springer
7. Gertsbakh, I. and Y. Shpungin.2013.Comparing random and regular network resilience against random attack on their nodes. *Journal of Reliability and Stat. Studies.* Vol 6(1), 01-09.
8. Gertsbakh,I., Shpungin, Y. and F. Spizzichino. 2012. Thee-State coherent systems and two-dimensional signatures. *J. Appl. Prob.*,
9. Vaisman, R., Kroese, D.P. and I. Gertsbakh. Improved sampling plans for combinatorial invariants of coherent systems. *IEEE Transaction on Reliability*, Accepted for publication.
10. Juanxi Gao, Daqing Li and Havlin,S.2014.From a single network to a network of networks. *National Science Review Advance Access*,July 16.1-11.
11. Li, W., Bashan, A., et al.2012. Cascading failures in interconnected lattice networks.*Phys. Rev.Lett*. 108(22): 228702.
12. Newman, Mark. 2010. *Networks: an Introduction*. Oxford University Press. New York.
13. Kazuo Nisimura and Masaaki Sibuja, Occupancy with two types of balls, *Ann. Inst. Statist., Math*, Vol40, 1,77-91, 1988 .
14. Popova, T.Y., Limit Theorems in a model of distribution of particles of two types, *Theory of Probability and Its Applications*, 13, 511-516, 1968.
15. Samaniego, F.J. 1985. On closure under ifr formation of coherent systems. *IEEE Trans. Reliab*. 34: 69-72.

16. Samaniego, F.J. 2007. *System signatures and their application in engineering reliability*. Springer, New York.
17. Wolfram, S.1991. *Mathematica: a system for doing mathematics by computer*. 2nd Edition. Addison-Wesley Publishing Company, New York.
18. B. G. Pittel. 2014. Private Communication

Asymptotics of mean-field closed networks

Alexander Vladimirov †

Institute for Information Transmission Problems
The author gratefully acknowledges the support of grants 16-29-09497, 14-01-00379, 14-01-00319 by Russian Foundation for Sciences.

Abstract

We establish the convergence of equilibria of finite symmetric closed networks with the FIFO service discipline and a general service time with bounded second moment to the unique equilibrium of the non-linear Markov process.

1 Introduction

Let us consider a single-class closed network $\mathcal{S}_{N,M}$ with M particles (customers) that live on a complete graph with N identical nodes (servers). Each node is a server with the FIFO service discipline and with a general distribution F of service time s. The evolution of the network goes on as follows.

It is assumed that each particle j waits in the queue at the server i until all the particles that stay ahead of j in the same queue complete their service. Then, immediately, the server i begins to serve the particle j, and the service time s is a random variable distributed as F and independent of anything else. At the end of the service time the particle j jumps to one of the N servers (may be, to the same one) with equal probability $1/N$ and waits for its turn to be served there. This cycle is repeated infinitely many times for each particle.

Under minimal assumptions each network $\mathcal{S}_{N,M}$ has a unique equilibrium process $\mathcal{E}_{N,M}$ that is a universal attractor. Our goal is to prove the convergence of equilibria $\mathcal{E}_{N,M}$ of the Markov processes on $\mathcal{S}_{N,M}$ to a unique equilibrium $\mathcal{E}$ of a so-called nonlinear Markov process (NLMP) $\mathcal{S}$ as $N \to \infty$ and $M/N \to H$. Note that $\mathcal{E}$ is a single point in the joint state space X of all our processes whence $\mathcal{E}_{N,M}$ is a probability measure on X for each finite N, M.

The NLMP is, in words, the process on the limit network with $N = \infty$ and $M = HN$. Its behavior is in some aspects simpler than that of finite networks. On bounded time intervals, the behavior of $\mathcal{S}_{N,M}$ converges to that of $\mathcal{S}$ as $N \to \infty$ and $M/N \to H$. The convergence of equilibria, nevertheless, is a much harder issue which is the subject of this paper.

We assume $\mathbb{E}s = 1$ and $\mathbb{E}s^2 < \infty$, where the second assumption is necessary for the existence of a nontrivial limit process as $N \to \infty$ and $M/N \to H > 0$. Otherwise it is not hard to see that we get a small number of very long queues for N large and no process at all in the limit.

The main feature of the networks in consideration is symmetry both in N and in M, that is, the system is invariant to all permutations of nodes and of particles. This makes the system a very particular case of a *Jackson-type network*, see [?]. Namely, this is a single-class queueing network with the FIFO service discipline and a general distribution of service time. The specifics of our model, namely, its mean-field nature lies in an especially simple structure of the routing matrix: all the entries of the $N \times N$-matrix P_N are equal to $1/N$.

The stochastic dominance technique had been introduced originally by A. Stolyar [?] for the deterministic service time. In [?] it was extended to a restricted class of general service times. Here we further extended the results to the case of a general service time distribution with the only restriction of finite second moment.

This goal is achieved by means of a new state space that comprises the lengths of queues at

N nodes and the remaining sojourn times of M particles. Similar methods can hopefully be used for the analysis of other mean-field models.

2 Models and parameters

There are different ways to formalize the evolution of $\mathcal{S}_{N,M}$ as a Markov process, that is, to define a state space and a process generator on this space. We begin with a more natural formalization where the current state of the network coincides with the collection of states of its N queues. Then we will give another formalization in terms of N queue heights and M remaining sojourn times. The latter one is more convenient for the proof of our main result though both formalizations describe the same dynamics of $\mathcal{S}_{N,M}$.

Let us begin with a definition of a state space Q of a single FIFO queue. Suppose we know the exogenous inflow of customers to the queue in the future ($t \geq 0$) as a point process of arrival times. Then, in order to know the stochastic queueing process at this server in the future, it suffices to know additionally the current elapsed time of service of the oldest customer in the queue plus the current number of the remaining customers that are currently waiting their turn to be served. Then we may calculate the distribution of future evolution scenarios of the queue.

However, it would be more convenient for several reasons to represent a queue in a more regular way as a finite sequence of *remaining service times* $h_{i,j}$ of all the customers in the queue, in order of their arrivals or in order of their prospective services, which is the same due to the FIFO service discipline. The values of $h_{i,j}$ are not observable, of course, but the dynamics of the process in terms of $h_{i,j}$ has more transparent description than in terms of elapsed service time of the first customer in the queue and the total number of customers in the queue.

Now, the current state of the queue i can be written as

$$q_i(t) = (h_{i,1}, \dots, h_{i,k}),$$

where second indices from 1 up to k mark the order of service of k customers in the queue. Namely, the customer 1 is currently served and the customer k is the last one to be served among the current customers. If new customers arrive, they are, of course, served after the customer k, in the order of their arrival.

If $k = 0$ then the queue is empty. Otherwise the value of $h_{i,1}$ decreases at rate 1 (the first customer is currently served) and all the other values $h_{i,j}$ do not change for a while.

Two kinds of event at the queue i are possible that make the behavior of the state of queue discontinuous. First (exit), as the value of $h_{i,1}$ hits zero at time t', the first customer is released and the length of the queue drops from k to $k-1$. We write then

$$h_{i,1}(t') = h_{i,2}(t'-1), \dots, h_{i,k-1}(t') = h_{i,k}(t'-1).$$

The released customer arrives immediately to one of the N nodes with equal probability $1/N$ and occupies the last position in the queue at this node. Note that with probability $1/N$ this customer returns to the queue i.

Second (arrival), if a new customer arrives to the queue i at time t'', the length of the queue rises from k to $k+1$ and we write

$$h_{i,k+1}(t'') = s,$$

where s is a random service time distributed as F and independent of anything else. As was mentioned, the events of these two kinds happen simultaneously if the released customer returns to the same queue.

The state space of a single queue is, therefore, the union of a countable number of finite-dimensional orthants:

$$Q = \mathbb{R}_+^* = \emptyset \cup \mathbb{R}_+ \cup \mathbb{R}_+^2 \cup \dots.$$

The state space of the whole network is the product of N copies of Q, that is, the vector space Q^N. However, due to the symmetry of the network, the order of vector components in Q^N is immaterial, that is, the dynamics of the system is invariant to permutations of servers. Therefore, we may consider a current configuration of the network as an atomic measure on the space Q,

where each atom has weight $1/N$ and corresponds to one of N single queues. Clearly, Q is a Polish space in the induced topology.

Denote

$$h_i = \sum_{j=1}^{k} h_{i,k},$$

that is, $h_i(t)$ is the total remaining service time of all the customers in the queue. We say that $h_i(t)$ is the current *queue height*.

The dynamics of a single queue height within the network consists of the deterministic (decreasing) part given by the differential equation with discontinuous right-hand side

$$\dot{h}(t) = \begin{cases} -1 & \text{if} \quad h(t) > 0, \\ 0 & \text{if} \quad h(t) = 0. \end{cases}$$

and the stochastic (increasing) part that consist of instant bursts $h(t) = h(t-0) + s$ with random i.i.d. increments s that happen at random times of arrival of new particles.

3 Alternative description

In what follows we will also use another description of the state space in terms of, again, the queue heights h_i at the N nodes and, additionally, the *remaining sojourn times* g_j of the M particles. The remaining sojourn time of the particle j is defined as the remaining time till the exit of customer j from its current queue i, that is,

$$g_j = \sum_{m \le k} h_{i,m},$$

where k is the current position of particle j in the queue i. Note that, because of the FIFO discipline, the value of g_j decreases at rate 1 until the service of particle j at node i is completed and does not change as new particles arrive to this node.

Thus the current state of the process $\mathcal{S}_{N,M}$ is an $(N+M)$-vector $f = (h, g)$ with non-negative components h_i and g_j. Because of the symmetry, we can reduce the state of the process to a couple of atomic measures μ and ν, both on $\mathbb{R}_+$. Namely, μ has N atoms of equal weight $1/N$ and ν has M atoms of equal weight $1/M$. Note that the value of g_j does not associate uniquely the particle j with some server i apart from special cases where, for instance, $g_j = h_i$ and there is no other $h_{i'} = h_i$ and no other $g_{j'} = g_j$.

Of course, the pair of measures μ and ν cannot be arbitrary pair of atomic measures, that is, μ and ν should be consistent. For instance, the upper customer in each queue has the remaining sojourn time equal to the height of the queue. Note, moreover, that the information contained in measures μ and ν is not sufficient to reconstruct the distribution of remaining service times among the customers nor the distribution of population among the queues. Let us regard the following example.

Let $N = 2$, $M = 4$. Let us arrange the components of measures μ and ν in ascending order. Let $h_1 = h_2 = 3$ and let $g_1 = 1$, $g_2 = 2$, and $g_3 = g_4 = 3$. Then either both queues hold two customers and their remaining service times are 1 and 2 but in different order, or one queue holds a single customer with remaining service time 3 and the other queue holds three customers with remaining service time 1 for each customer.

However, the continuous-time Markov process on measures μ and ν is well defined as we will see immediately. Moreover, if we watch the process for some finite time, we get all the information on the distribution of lengths of queues (number of customers in the queue) and on the distribution of remaining service times.

Let us see how the (continuous-time) Markov process on pairs (μ, ν) evolves. Recall that we consider the system $\mathcal{S}_{N,M}$. All the values $h_i(t)$ and $g_j(t)$ are decreasing at rate 1 as long as they are positive. The service events in the network happen exactly as some $g_j(t)$ vanishes. Suppose this is kth service of particle j. Then we denote this time by t_k^j.

Instantly, the particle j is routed to a random queue. Denote the height of its new queue by $h \ge 0$. The particle is allotted an F-distributed service time s there. The values of h and s are

random and independent. The distribution of h is given by the current value of N-atomic measure μ. The values of $h_i = h$ and of g_j both are updated to $h + s$

Then the next jump time for particle j is

$$t_{k+1}^j = t_k^j + h + s$$

and the evolution of $g_j(t)$ and of $h_i(t)$ goes on deterministically on the time interval $(t_k^j, t_{k+1}^j]$ as usual:

$$\dot{g}_j(t) = -1, \quad \dot{h}_i(t) = -1$$

The linear decrease of $g_j(t)$ is completely deterministic while the evolution of $h_i(t)$ may have positive bursts if new particles arrive.

We denote the space of pairs of atomic measures $\mu = \mu_N$ and $\nu = \nu_M$ by $X_{N,M}$. We embed all $X_{N,M}$ into the space X of all pairs of probability measures on $\mathbb{R}_+$. As we will see soon, X is the configuration space of the NLMP. For the NLMP (that is, for $N = \infty$), we have a limit dynamics of general probability measures $\mu(t)$ and $\nu(t)$ on $\mathbb{R}_+$. Again, these measures should be consistent, see below.

Formally, a series of continuous-time Markov processes $\mathcal{S}_{N,M}$ is defined on X and it can be proved that their generators converge to that of the NLMP $\mathcal{S}$ ensuring the convergence of processes on bounded time intervals. We will, however, use other tools for justification of this fact, that is, stochastic dominance methods. We will also see that the NLMP conserves single-point measures (it is a deterministic dynamical system on X) while finite processes $\mathcal{S}_{N,M}$, obviously, do not.

4 The NLMP

As we have mentioned above, the NLMP $\mathcal{S}$ is the limit process for $\mathcal{S}_{N,M}$ as $N \to \infty$ and $M/N \to H$. Formally we will define two limit dynamical systems for each form of Markov process (with two different state spaces) that were used for the description of the evolution of finite systems $\mathcal{S}_{N,M}$ and demonstrate that they are equivalent, that is, they describe the evolution of the same limit process $\mathcal{S}$ (the NLMP).

To begin with, let us use the more intuitive state space based on Q for the primary description of the NLMP. The current state of the process is now a probability measure η on Q. Clearly, the weak limit points of atomic measures on Q cover the space $\mathcal{M}(Q)$ of all probability measures on Q, hence, η is an arbitrary point of $\mathcal{M}(Q)$.

In order to get an intuitive notion of the NLMP, one may imagine the situation where there are infinitely many queues in the system, that is, $N = \infty$ and the distribution of states of these queues is η. Note, however, that we cannot formally define a mean-field routing process on a countable number of servers since there is no uniformly distributed probability measure on such a set. One may think of the queues of the NLMP as of elements of a continuous measurable space, say, of the interval [0,1] with Borel measure but, in fact, we do not need such a specialization.

The non-linear Markov process goes on as follows. Let $\eta(0)$ be given. Let us distinguish a single queue $q = q_\omega$ which is in the state $q(0)$ at $t = 0$. Its evolution within the NLMP is a stochastic process which can be described completely if we know the evolution of the measure $\eta(t)$ for $t \geq 0$. The evolution of different queues are independent.

It would be easier to begin with a definition of the NLMP for a given time-dependent Poisson inflow $\lambda(t)$, that is, without feedback. Let us assume that the inflows to all the nodes are independent Poisson flows of rate $\lambda(t)$ (all the inflows in the closed NLMP are Poisson ones because they have infinitely many additive sources). Then we will have a deterministic dynamics of $\eta(t)$ since the evolution of each particular queue is a Markov process and these processes are independent for different queues.

In turn, if we know $\eta(t)$ for $t \geq 0$, we can find the resulting mean outflow rate

$$b(t) = \lim_{\Delta \to 0} \frac{P(\Delta)}{\Delta},$$

where $P(\Delta)$ is the probability of service event at a random queue during the time interval $[0,\Delta)$. The value of $b(t)$ must be equal to $\lambda(t)$ for all t since the network is closed.

Definition 4.1 *The pair* $(\lambda(t),\eta(t))$ *is a solution of the NLMP if* $\lambda(t)$ *is the outflow generated by* $\eta(t)$ *and* $\eta(t)$ *is the evolution of the measure on queue states generated by the inflow of rate* $\lambda(t)$.

Theorem 4.2 *For each* η_0 *there exists a unique solution* $(\lambda(t),\eta(t))$ *of the NLMP such that* $\eta(0)=\eta_0$.

Proof. We are going to construct a series of processes $\mathcal{P}_n$, $n=0,1,...$, whose solutions converge to the required solution of $\mathcal{S}$. In the process $\mathcal{P}_0$, there is no feedback, that is, the customers that are served do not return to the system. The evolution of the corresponding measure $\eta_0(t)$ is simple: each queue drains out and it is empty forever since time $t=h(0)$.

The resulting outflow normalized by the number of queues has rate $\lambda_0(t)$, $t\geq 0$. Let us now construct the next process $\mathcal{P}_1$ as follows: we only allow particles to return to the system once. Formally, we consider a random queue that is distributed as $\eta(0)$ initially and then receives the inflow of rate $\lambda_0(t)$ (these inflows to different queues are mutually independent). Clearly, it becomes empty eventually with probability 1. We denote the corresponding outflow rate by $\lambda_1(t)$.

Now, we make an important remark: under an appropriate coupling, all the exit events in the process $\mathcal{P}_0$ happen at the same queues at the same time in the process $\mathcal{P}_1$ (because of the FIFO service discipline). Additionally, there are secondary exit events as the secondary particles that have returned to the system after the first service are served the second time in their lives and leave the system forever. Hence, $\lambda_1(\cdot)\geq\lambda_0(\cdot)$ in the following sense:

$$\int_0^t \lambda_1(s)ds \geq \int_0^t \lambda_0(s)ds \quad \text{forall} \quad t\geq 0.$$

Next we recall a simple monotonicity property of a FIFO server.

Lemma 4.3 *Suppose we have two identical FIFO servers* 1 *and* 2 *with the same initial states at* $t=0$. *Let* u_1 *and* u_2 *be point processes on* $\mathbb{R}_+$ *and let* u_1 *dominates* u_2 *stochastically (denoted* $u_1 \pm u_2$*). Denote by* w_1 *and* w_2 *the departure point processes of servers* 1 *and* 2*, respectively, where the server* i *receives the inflow* u_i, $i=1,2$. *Then* $w_1 \pm w_2$.

Recall that u_1 dominates u_2 stochastically if there is a coupling between the two processes such that $t_1^k \geq t_2^k$ for all coupled pairs of configurations $(t_i^1, t_i^2, \dots)$ and all $k=1,2,....$

Then we iterate the construction of λ_n for $n=2,3,....$ From Lemma 4.3, we get inequalities

$$\lambda_n(\cdot)\geq\lambda_{n-1}(\cdot) \quad \text{forall} \quad n=1,2,....$$

On the other hand, there is a finite upper bound $\bar{\lambda}(t)$ for all $\lambda_n(t)$ in the integral sense. Indeed, the maximum of service rate at each queue is 1. Hence there is a convergence and the limit is a solution $(\lambda(t),\eta(t))$ of $\mathcal{S}$.

Suppose there exists another solution $(\lambda'(t),\eta'(t))$. Then, by construction, $\lambda'(t)\geq\lambda(t)$ for all t and the inequality is strict for some $t<\infty$. We come to a contradiction easily since the mean mass of queues at this time t must be different for the inflows $\lambda(\cdot)$ and $\lambda'(\cdot)$.

Moreover, we can construct Markov processes on finite networks $\mathcal{S}_{N,M}$ by the same monotone iteration procedure as in $\mathcal{S}$. As a result, we also conclude that a finite-time convergence of $\mathcal{S}_{N,M}$ to $\mathcal{S}$ takes place.

Namely, it is not hard to prove that the dynamics of $\mu_N(t)$ and $\nu_M(t)$ in the process $\mathcal{S}_{N,M}$ is close to that of $\mu(t)$ and $\nu(t)$ in the NLMP on finite time intervals if N and M are large and if respective initial values of measures are close to each other. Analogous results can be found, for instance, in [?] under stronger assumptions on the service time distribution F.

Theorem 4.4 *Suppose that the sequence of probability measures* $\varphi_{N,M}(0)$ *on* $X_{N,M}\subseteq X$ *converges weakly to a probability measure* $\varphi(0)$ *on* X *as* $N\to\infty$ *and* $M/N\to H$. *Then the solutions of* $\mathcal{S}_{N,M}$ *from the initial states* $\varphi_{N,M}(0)$ *converge to the solution of the NLMP* $\mathcal{S}$ *from the initial state* $\varphi(0)$ *on any bounded time interval* $[0,T]$.

Here we do not prove the theorem and do not specify the notion of convergence of processes since these are rather technical issues.

In the same framework as for the finite systems, we may give a description of the NLMP as evolution of two probability measures on $\mathbb{R}_+$. They are the measure $\mu(t)$ on queue heights and the measure $\nu(t)$ on remaining sojourn times. Clearly, the resulting flow rate $\lambda(t)$ in the (h,g)-representation is the same as in (λ,η)-representation.

Again, there exist some constraints on possible pairs (μ,ν) caused by the fact that the remaining sojourn time of a particle at the top of the queue coincides with the height of this queue. Let us write down this constraint explicitly.

Denote by $\alpha=\alpha(\mu)$ the fraction of empty queues, that is, $\alpha=\mu(\{0\})$. Denote by $\grave{u}=\grave{u}(\mu)$ the measure $(1-\alpha)\frac{N}{M}\mu$ which is not a probability measure, of course. The constraint on the measure ν can be then written as

$$\nu\geq \grave{u},\quad \text{thatis},\quad \nu([a,b])\geq \grave{u}([a,b])\quad \text{whenever}\quad 0\leq a\leq b<\infty.$$

The dynamics of the pair $(\mu(t),\nu(t))$ for $t\geq 0$ is completely defined by the value $(\mu(0),\nu(0))$ as follows. Recall that the mean number of particles per node in the NLMP is exactly H. By the current value of $\nu(t)$ we can find the current service rate $\gamma(t)$ per particle:

$$\gamma(t)=\lim_{Delta\to+0}\frac{\nu[0,\Delta)}{\Delta}.$$

The current service rate per node is then equal to $\lambda(t)=\gamma(t)/H$. Note that the parameter H is already included in the constitutive relations for the process. As soon as we know $\lambda(t)$ and the current distribution of queue heights, we can write evolution equations for $\mu(t)$ and $\nu(t)$.

A rigorous way to prove the existence and uniqueness of a solution from any pair $(\mu(0),\nu(0))$ is, again, to realize a recursive construction where we allow 1, 2,, k lumps to each particle. The same method demonstrates the equivalence of two forms of the NLMP.

Note that an essential difference exists with the dynamics of a pair (μ_N,ν_M) in the finite network $\mathcal{S}_{N,M}$. Namely, the evolution of (μ_N,ν_M) is stochastic for each pair N,M and the evolution of $(\mu(t),\nu(t))$ is deterministic.

Now let us lift the consistency restrictions for the pair (μ,ν). We will define a solution of the NLMP that starts from an arbitrary pair $\mu,\nu\in\mathbb{R}_+$ in the same manner as before. Actually, such a solution has the following physical sense.

We do not suppose any longer that the system contains H particles per node straightaway. Instead we assume that these particles are elsewhere at $t=0$ and that, initially, each particle has its delay g distributed as ν and each queue has its height h distributed as μ. Then the process is started. The queue heights decrease at rate 1 while positive and the particles enter the system at times g.

Then the process goes on exactly by the rules of closed system, that is, the particles, as they jump, are given the new value of g equal to $h+s$ and the value of h is replaced by $h+s$ as well. Clear enough, the dynamics of such a system approaches that of the closed system as the number of particles per node in the system approaches H and the remaining initial height of queues vanishes.

5 Stationary solutions and ergodicity

Note that continuous-time Markov processes $\mathcal{S}_{N,M}$ are ergodic (if F is a non-lattice distribution) since there is a renewal event where all the particles get in the same queue and the oldest one begins its service, see [?, ?]. This event obviously happens with a positive frequency.

In the case of a lattice distribution F, we may consider a discrete-time process and repeat the argument. The discrete-time version of the process is, again, ergodic. For definiteness, in what follows we assume that F is a non-lattice distribution.

Our goal is to study the asymptotic behavior of unique equilibrium solutions $\mathcal{E}_{N,M}$ as the size N tends to infinity and as $M/N\to H$ for some $H>0$ (mean population of a node or, in other

words, mean length of a queue). Namely, we ask if the sequence $\mathcal{E}_{N,M}$ converges to the unique equilibrium $\mathcal{E}$ of the NLMP with the same parameter H. The two parameters of this problem is the service time distribution F and the mean length H of a queue which is time-independent since the networks are closed.

The process $\mathcal{E}_{N,M}$ can be written in (h,g)-notation. Then the unique equilibrium distribution $\varphi_{N,M}$ is defined on the configuration space X_N. Since all X_N are embedded into

$$X = \mathcal{M}(\mathbb{R}_+) \times \mathcal{M}(\mathbb{R}_+),$$

we write $\varphi_{N,M} \in \mathcal{M}(X)$.

The convergence of equilibria $\mathcal{E}_{N,M}$ to $\mathcal{E}$ can be understood in different senses. In particular, one may expect that the stationary measures $\varphi_{N,M}$ converge weakly to the δ-measure on a unique stationary point (μ^*, ν^*) of $\mathcal{S}$ as $N \to \infty$ (in (h,g)-notation). The rest of the paper is concerned with the proof of this fact.

Note that, in general, there is no explicit information on the invariant value of H in a solution $(\mu(\cdot), \nu(\cdot))$ of the NLMP. In equilibrium, however, it is not hard to calculate H from the stationary measure (μ^*, ν^*) as follows. Denote the jump rate per node by λ. Then the jump rate per customer is λ/H. Next, we compare the rates of continuous decrease of h and g. The first one is equal to $1-\alpha$ and the second one is 1.

Since we consider an equilibrium, these rates must be equal to the the rates of increase of h and g at jump events. They are, respectively, λ (since the mean service time is 1) and $\lambda(\mathbb{E}h+1)/H$ (since the mean value of new g after the jump equals the mean height of a queue plus the mean service time). Therefore,

$$H = (1-\alpha)(\mathbb{E}h + 1).$$

Lemma 5.1 *The NLMP has a unique equilibrium and all other solutions approach this equilibrium.*

Proof. For the proof we will use a fundamental result on "smoothing effect" of the FIFO server. It was proved under some additional restrictions on F in [?] and will be proved in the general case in a subsequent paper by the author.

The main idea is the monotonicity argument. Suppose there is no convergence and come to a contradiction. Indeed, in this case the mean population of a queue cannot be constant.

In more details, suppose there is a non-converging solution of the NLMP. Then $\lambda(t)$ does not converge as $t \to \infty$. By limit transition we can construct a non-constant solution $\lambda'(t)$ on the whole time axis $\mathbb{R}$ such that

$$\int_0^1 \lambda'(s)ds \geq \int_t^{t+1} \lambda(s)ds, \quad t \in \mathbb{R}.$$

By monotonicity, there is a coupling between the inflows $\lambda'(t)$ and $\lambda''(t) = \lambda'(t-1)$ on $(-\infty, 0]$ such that all the particles arrive not earlier in the first case and some of them arrive strictly later. Then the mean mass of a queue at $t=0$ is strictly larger than that at $t=-1$, which is a contradiction.

Now, since there is a unique equilibrium measure on solutions of the NLMP (actually, this is a δ-measure on the unique equilibrium solution), it suffices to prove tightness of the family (μ_N^*, ν_M^*), $N = 1,2,\ldots$ in order to derive convergence.

There exists a well-known criterion of tightness for random measures on $Y = \mathcal{M}(X)$, that is, on the space of probability measures on a Polish space X, where the topology of weak convergence of measures generates the measurable structure on Y. Recall that X itself is the space of pairs of probability measures on $\mathbb{R}_+$.

Proposition 5.2 *The family (μ_N^*, ν_M^*) is tight if and only if two following conditions hold. The probability of $h_i^N > K$ tends to zero as $K \to \infty$ uniformly on N. The probability of $g_j^M > K$ tends to zero as $K \to \infty$ uniformly on M.*

We will see that it suffices to get a "stable" upper bound on the inflow that does not

depend on the parameters in order to prove the required tightness.

6 Dominance

We use Baccelli–Foss theorems [?] for Jackson-type networks with implications for closed networks. The following theorem is an extension of Lemma 4.3 from a single FIFO node to a finite open Jackson-type network.

Theorem 6.1 *In a finite open FIFO network, if all the arrivals happen earlier and all the services happen faster, then all the events happen earlier.*

Here the notions "earlier" and "faster" should be understood in the sense of stochastic dominance. For closed networks, we may assume that the particles arrive once from the outside and then circulate within the network.

Theorem 6.2 *Suppose there is an infinite program at each node of the network G, that is, an infinite sequence of pairs (x_k, i_k), where x_k is the service time of kth particle and i_k is its address after the service. Suppose there is another instance G' of the same network with a program (x_k', i_k') that dominates the first one in the following sense: $x_k' \geq x_k$ and $i_k' = i_k$ for each k and each node of the network. Suppose the initial positions of particles coincide in both cases. Then all the events in G' happen not earlier than their counterparts in G.*

In particular, if the programs are identical but the exogenous particles arrive later to the same nodes in the second case, then, again, all the events in the primed case happen later. This assertion can be reduced to Theorem 6.2 by introduction of additional virtual nodes where the exogenous particles reside initially.

Theorem 6.2 can be proved by the following argument. Suppose the contrary. We make a coupling that preserves the order of initial events. Then there exists the first pair of events with the reverse order. And this, clearly, cannot happen.

Our next goal is to find a universal stochastic upper bound on the number of arrivals to a single node of a finite network $\mathcal{S}_{N,M}$ during the time interval $[0,T]$, that is, a bound that does not depend on N (if N is lrge enough) and on the initial state of the network. Suppose this upper bound B satisfies the *stability condition* $\mathbb{E}B < T$. Then the required tightness would follow.

Let us study processes $\mathcal{S}_{N,M}$ and $\mathcal{S}$ on a bounded time interval $[0,T]$ and count the number of services per node that happen within this interval. By coupling and monotonicity, there is the "worst" initial configuration that entails more services than any other initial configuration (either in the sense of stochastic dominance or for any fixed sequence of service times at each node). This is the zero configuration where $h_i = 0$ and $g_j = 0$ for all i and j. In words, all the M particles are jumping immediately (at $t = 0$). Then the process goes on by standard rules.

For given N and M and for the "worst" initial distribution, there is a distribution $B_{N,M,T}$ of the number of arrivals to a given node i during $[0,T]$. This distribution does not depend on i but if we assume the sequence of service times at node i to be known, then the number of arrivals to i has a different distribution, that is, there is a correlation between the service times at node i and the number of arrivals to i.

In order to cope with this inconvenience we consider the "fastest" program at a given node i, that is we assume that all the particles at this node are served immediately. Then the new distribution $B'_{N,M,T}$ of the number of arrivals certainly dominates the distribution $B_{N,M,T}$ since it dominates the distribution $B''_{N,M,T}$ for any other service time sequence at the node i.

Now, in order to find a uniform "stable" upper bound on the inflow to a given node, we need to find such T that distributions $B'_{N,M,T}$ are dominated by some B_T for all relevant pairs N, M (if N is large enough) and that B_T is "T-stable" (its mean total service time is strictly less than T).

Note that the queueing process with a special node i whose service time is always zero coincides in certain sense with the process $\mathcal{S}_{N-1,M}$. Namely, on the nodes different from i, the

process does not differ from $\mathcal{S}_{N-1,M}$. As $N \to \infty$ and $M/N \to H$, the difference between $\mathcal{S}_{N-1,M}$ and $\mathcal{S}_{N,M}$ vanishes. Hence, we will use upper bounds on $B_{N,M,T}$ instead of $B'_{N,M,T}$ and prove that they are all uniformly T-stable for some $T > 0$ and all N large enough.

Then we break the time half-axis $\mathbb{R}_+$ in intervals of length T and study the discrete-time process that dominates all equilibria $\mathcal{E}_{N,M}$ in certain sense. First of all, we look at the NLMP and the corresponding arrival process.

Theorem 6.3 *For the NLMP $\mathcal{S}$, there is a $T > 0$ and $\varepsilon > 0$ such that the mean number of arrivals to a node from any initial state is less than $T - \varepsilon$.*

Proof. We omit detailed proof and give just a bare idea. The mean number of arrivals to a node on the time interval $[0, T]$ is

$$\Lambda(t) = \int_0^T \lambda(t)dt.$$

Let us fix a $\lambda^* < 1$ such that the stationary Poisson arrival of rate λ^* leads to the stationary mean length $H' > H$ of a single queue with service time distribution F. Then, eventually, we have

$$\Lambda(t) < t\lambda^*.$$

Otherwise, by monotonicity argument, we can find $t > 0$ such that the mean length of the queue under the Poisson inflow with rate $\lambda(t)$ exceeds H. This proves the lemma.

Denote by B_T^* the corresponding distribution of arrival events at a given node. By approximation argument, any finite part of this distribution is close to the corresponding part of $B_{N,M,T}$. Then we handle the remaining part by dominance argument again and use some combinatorics to handle the tails.

7 Uniform upper bound for the inflow

The uniform dominance for all $N \geq \bar{N}$ follows from the finite-time convergence Theorem 4.4 and the NLMP stability Theorem 6.3. For the proof, let us prove the dominance separately for medium flows and for large flows.

If the size of flows (number of particles in the flow) in consideration is bounded from above by the same constant D, then we deal with a compact part of the state space. The probability of a single server to receive an inflow of k particles within the time interval $[0, T]$ satisfies the relation

$$\lim_{N\to\infty} P_k^N = P_k,$$

which implies immediately the required dominance.

Next we assume that all the initial queues are infinite and prove the second-moment bounds on the inflows in $\mathcal{S}_{N,M}$, $N \geq N_0$. These bounds dominate the inflows from any initial distribution of M particles among N nodes.

We have N independent identically distributed integer-valued variables m_i^N (numbers of services at individual queues) whose mean values are $T + 1$ (to be sure) and variances are bounded. Their sum

$$M^N = \sum_{i=1}^N m_i^N$$

is then distributed uniformly among N queues, producing the number of arrival n_i^N.

The mean value of n_i^N for any i is equal to $T + 1$, and we are going to find upper bounds for the probabilities of large values of n_i^N. To this end let us first choose a special service time distribution F' that is stochastically dominated by F and then find upper bounds on the probabilities of k arrivals to a single node during the time interval $[0, T]$.

The distribution F' has two atoms, at 0 and at 1. The probability p of 1 is strictly positive. Now note that the probability of k arrivals from N infinite queues with service time distribution F' equals the probability of drawing k ones in a series of $[T + 1]N$ independent draws, where the chance to draw 1 equals p/N.

Next, we find the asymptotic bounds on the probability of massive arrival. We study a discrete-time random walk on the positive orthant $\mathbb{N}_+^2$. Namely, we start at the origin and make one of the three steps: (1,0) with probability u, (0,1) with probability v, and (1,1) with probability $w = 1 - u - v$.

The values of u, v, w depend on the parameter $N = 1,2, \ldots$. We also have a an integer constant $T > 0$ (the length of interval) and a real constant $\alpha > 0$ (probability of service time 1). The process is defined as follows.

Let us make a vertical step with probability α, otherwise we make a horizontal step. Each step is then marked with probability $1/N$, independently. We count the number of marked steps (axis x) and the number of vertical steps (axis y). As a result we draw a path on the lattice $\mathbb{Z}^2$ from (0,0) to infinity.

We are interested in the measure on paths that is induced by the uvw-rules of random walk. Our goal is to assess the slope of the path and the probability that this slope is below certain value as the path reaches some vertical bound. This is a model of the inflow at a given node of the mean field closed FIFO network.

The next step is to draw required upper bounds. Since we need an upper bound, we can use another scheme of path construction: let us replace marked vertical edges by marked horizontal ones. This will be performed for $N > 1$. Then we have the following probability of a horizontal step:

$$\alpha'_N = \frac{(N-1)\alpha+1}{N}, \quad 1 - \alpha'_N = \frac{(N-1)(1-\alpha)}{N}. \tag{7.1}$$

We look for an upper bound on $p(k,n)$, that is, on the probability that the path from the origin passes through the point (k,n).

The event of passing through (k,n) is equivalent to the following one. The first $k+n$ steps contain exactly n vertical and k horizontal steps. The number of paths from (0,0) to (k,n), hence, equals to

$$S(k,n) = \frac{(k+n)!}{k!n!}.$$

All these paths have the same probability $(\alpha_N')^n(1-\alpha_N')^k$, therefore

$$p(k,n) = \frac{(k+n)!}{k!n!}(\alpha_N')^n(1-\alpha_N')^k.$$

Here

$$\alpha_N' = \frac{\alpha(N-1)}{\alpha(N-1)+1}, \quad 1 - \alpha_N' = \frac{1}{\alpha(N-1)+1}$$

For simplicity we may assume $\alpha_N' = (N-1)/N$ and $1 - \alpha_N' = 1/N$ (this will not change the asymptotics). Next, we substitute $n = TN$ and get

$$p(k,TN) = \frac{(k+TN)!}{k!(TN)!}\left(\frac{N-1}{N}\right)^{TN}\left(\frac{1}{N}\right)^k.$$

Actually, we are interested in an upper bound on the sum

$$P(k,n) = \sum_{m=k}^{\infty} p(m,n)$$

or another sum

$$Q(k,n) = \sum_{m=0}^{k} p(m, k+n-m).$$

The second sum is more efficient since no two points on the diagonal

$$D_{k+n} = \{(a,b) \in \mathbb{Z}_+^2 : a + b = k + n\}$$

can lie on the same path from the origin, that is, each path hits D_{k+n} exactly once.

Our goal is to find an appropriate upper bound that is uniform for all $N \geq N_0$ for some finite N_0. Again, for simplicity, let $T = 1$ (for a while). Then we have

$$p(k,N) = \frac{(k+N)!}{k!N!}\left(\frac{N-1}{N}\right)^{N}\left(\frac{1}{N}\right)^k.$$

By means of the Stirling formula, we reduce the bound to

$$p(k,N) \simeq \frac{\sqrt{k+N}(k+N)^{k+N}}{\sqrt{kN}k^kN^{k+N}} \tag{7.2}$$

(up to a multiplicative constant). We need the maximum of (7.2) over $N \geq N_0$. For the main part of (7.2), let us write

$$\frac{(k+N)^{k+N}}{k^kN^{k+N}} = \frac{\left(1+\frac{k}{N}\right)^{k+N}}{k^k} =$$

$$= \left(1+\frac{k}{N}\right)^N \left(\frac{1}{k}+\frac{1}{N}\right)^k \le e^k \left(\frac{1}{k}+\frac{1}{N_0}\right)^k \le \frac{e^k e^{N_0}}{N_0^k}.$$

For $N_0 > e$ this bound vanishes exponentially as $k \to \infty$. The required dominance follows.

References

1. F. Baccelli and S. Foss. Ergodicity of Jackson-type queueing networks. *Queueing Systems*, 17(1-2):5{72, 1994.
2. R. Durrett. *Probability: Theory and Examples*. Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, 2010.
3. Sean Meyn and Richard L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009. With a prologue by Peter W. Glynn.
4. S. Pirogov, A. Rybko, S. Shlosman, and A. Vladimirov. Propagation of Chaos and Poisson Hypothesis. *ArXiv e-prints*, October 2016.
5. A.N. Rybko and S.B. Shlosman. Poisson hypothesis for information networks. *Moscow Math. J.*, 5:679{704, 2005.
6. A. L. Stolyar. The asymptotics of stationary distribution for a closed queueing system. (Russian). Problems Inform. Transmission, 25(4):321{333, 1990.
7. A. A. Vladimirov, A. N. Rybko, and S. B. Shlosman. The self-averaging propert

Probabilistic Models for Reliability Analysis of a System with Three Consecutive Stages of Deterioration

Ibrahim Yusuf
Department of Mathematical Sciences, Bayero University, Kano, Nigeria
iyusuf.mth@buk.edu.ng
Ramatu Idris Gatawa
Department of Mathematics, North West University, Kano, Nigeria
rigatawa@yahoo.com

Abstract

In this paper we present availability and mean time to failure estimation of a system where the deterioration rates follow the Weibull distribution. The paper presents modeling and evaluation of availability and mean time to system failure (MTSF) of a consecutive three stage deteriorating system. The system has three possible modes: working with full capacity, deterioration and failure mode. The three stages of deterioration are minor, medium and major deteriorations. Minor and major maintenance are allowed at minor and medium deterioration states and replacement at system failure. Explicit expressions for the availability and mean time to failure of the system are obtained analytically. Graphs have been plotted to determine the behavior of availability and mean time to system failure with respect to time for different values of deterioration, maintenance and replacement rates. Also, high values of the shape parameter decreases mean time to system failure and availability. The system is analyzed using differential difference equations.
Keywords: list, keywords, enter, here

I. Introduction

In practical engineering applications, most repairable systems are deteriorative that system failure often cannot be as good as new, it is more reasonable for these deteriorating repairable systems to assume that the successive working times of the system after repair will become shorter and shorter while the consecutive repair times of the system after failure will become longer and longer. Most of these systems are subjected to random deterioration which can result in unexpected failures and disastrous effect on the system availability and the prospect of the economy. Therefore it is important to find a way to slow down the deterioration rate, and to prolong the equipment's life span. Maintenance policies are vital in the analysis of deterioration and deteriorating systems as they help in improving reliability and availability of the systems. Maintenance models can assume minor maintenance, major maintenance before system failure, perfect repair (as good as new), minimal repair (as bad as old), imperfect repair and replacement at system failure.

Several models on deteriorating systems under different conditions have been studied by several researchers such as Bérenguer (2008), Frangopol *et al* (2004), Lam and Zhang (2003), Nicolai *et al* (2007), Rani and Sukumari (2014), Vinayak and Dharmaraja (2012), Yuan *et al* (2012), Yuan and Xu (2011). Analysis of reliability and availability model for deteriorating system have been studied under different conditions such as Liu *et al*. (2011) who investigated reliability analysis of a deteriorating system with delayed vacation of repairman, Tuan *et al* (2013).A Reliability-based Opportunistic Predictive Maintenance Model for k-out-of-n Deteriorating Systems, Xiao *et al* (2013) proposed the Bayesian reliability estimation for deteriorating systems with limited samples using the maximum entropy approach, Yusuf *et al* (2012). Presents modelling the reliability and availability characteristics of a system with three stages of deterioration, Zhang and Wang (2007)

deal with the study of deteriorating cold standby repairable system with priority in use.
This paper considers a system with three consecutive stages of deterioration before failure and derived its corresponding mathematical models. Furthermore, we study mean time to system failure and availability using differential difference equation method. The focus of our analysis is primarily to capture the effect of minor maintenance, minor deterioration and shape parameter on mean time to system failure and availability.
The organization of the paper is as follows. Section 2 contains a description of the system under study. Section 3 presents formulations of the models. The results of our numerical simulations are presented and discussed in section 4. Finally, we make some concluding remarks in Section 5.

II. Description and States of the System

In this paper, one unit system is considered. It is assumed that the system most pass through three consecutive stages of deterioration which are minor, medium and major deterioration before failure. The unit is considered to be non repairable. At early state of the system life, the operating unit is exposed to minor deterioration with rate λ_1 **and this deterioration is rectified through minor maintenance** α_1 **which revert the unit to its earliest position before deterioration.** If not maintained, the unit is allowed to continue operating under the condition of minor deterioration which later changes to medium deterioration with rate λ_2. At this stage, the strength of the unit is still strong that it can rectify to early state with rate α_2. However, the system can move to major deterioration stage with rate λ_3 where the strength of the unit has decreases to the extent that it cannot be reverted to its early state, neither that of minor nor medium deterioration stages. Here the unit is allowed to continue operation until it fails with parameter λ_4 and the system is immediately replaced by with a new one with rate α_3. Deteriorating rates follow Weibull distribution $f(t) = \lambda_k \delta t^{\delta-1}$, $k = 1,2,3$, δ **is the shape parameter.**

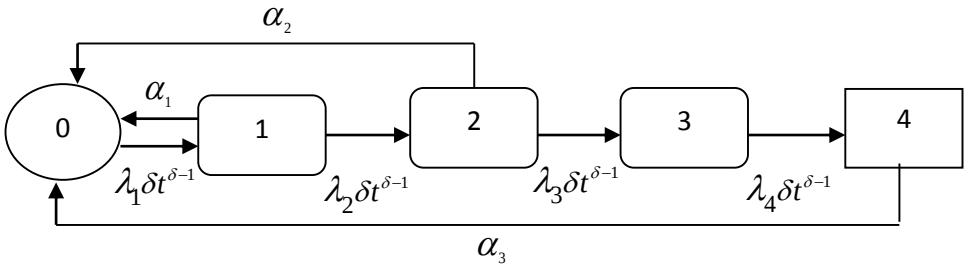


Figure 1: Transition diagram of the system

Table 1: States of the system

State	Description
S_0	Initial state, the system is operative.
S_1	The system is in minor deterioration mode and is under online minor maintenance, and is operative.
S_2	The system is in medium deterioration mode and is under online major maintenance and is operative.
S_3	The system is in major deterioration mode and is operative.
S_4	The system is inoperative.

III. Formulation of the Models

In order to analyze the system availability of the system, we define $P_i(t)$ to be the probability that the system at $t \geq 0$ is in state S_i. Also let $P(t)$ be the row vector of these probabilities at time t.
The initial condition for this problem is:

$$P(0) = [p_0(0), p_1(0), p_2(0), p_3(0), p_4(0)] = [1,0,0,0,0]$$

We obtain the following differential difference equations from Figure 1:

$$\frac{d}{dt}p_0(t) = -\lambda_1\delta t^{\delta-1}p_0(t) + \alpha_1 p_1(t) + \alpha_2 p_2(t) + \alpha_3 p_4(t)$$

$$\frac{d}{dt}p_1(t) = -\left(\alpha_1 + \lambda_2\delta t^{\delta-1}\right)p_1(t) + \lambda_1\delta t^{\delta-1}p_0(t)$$

$$\frac{d}{dt}p_2(t) = -\left(\alpha_2 + \lambda_3\delta t^{\delta-1}\right)p_2(t) + \lambda_2\delta t^{\delta-1}p_1(t)$$

$$\frac{d}{dt}p_3(t) = -\lambda_4\delta t^{\delta-1}p_3(t) + \lambda_3\delta t^{\delta-1}p_2(t)$$

$$\frac{d}{dt}p_4(t) = \alpha_3 p_4(t) + \lambda_4\delta t^{\delta-1}p_3(t) \quad \textbf{(1)}$$

This can be written in the matrix form as

$$\dot{P} = TP, \quad \textbf{(2)}$$

where

$$T = \begin{pmatrix} -\lambda_1\delta t^{\delta-1} & \alpha_1 & \alpha_2 & 0 & \alpha_3 \\ \lambda_1\delta t^{\delta-1} & -\left(\alpha_1 + \lambda_2\delta t^{\delta-1}\right) & 0 & 0 & 0 \\ 0 & \lambda_2\delta t^{\delta-1} & -\left(\alpha_2 + \lambda_3\delta t^{\delta-1}\right) & 0 & 0 \\ 0 & 0 & \lambda_3\delta t^{\delta-1} & -\lambda_4\delta t^{\delta-1} & 0 \\ 0 & 0 & 0 & \lambda_4\delta t^{\delta-1} & -\alpha_3 \end{pmatrix}$$

The steady-state availability (the proportion of time the system is in a functioning condition or equivalently, the sum of the probabilities of operational states) is given by

$$A_V(\infty) = p_0(\infty) + p_1(\infty) + p_2(\infty) + p_3(\infty) = \frac{N}{D} \quad (3)$$

where

$$N = \alpha_3\lambda_4\left(\alpha_1 + \delta\lambda_2 t^{(\delta-1)}\right)\left(\alpha_2 + \delta\lambda_3 t^{(\delta-1)}\right) + \alpha_3\delta\lambda_1\lambda_4 t^{(\delta-1)}\left(\alpha_2 + \delta\lambda_3 t^{(\delta-1)}\right) + \alpha_3\delta^2\lambda_1\lambda_2\lambda_3 t^{(\delta-1)} + \alpha_3\delta^2\lambda_1\lambda_2\lambda_3\left(t^{(\delta-1)}\right)^2$$

$$D = \delta^3\lambda_1\lambda_2\lambda_3\lambda_4 t^{(\delta-1)} + \alpha_3\delta^2\lambda_1\lambda_2\lambda_3\left(t^{(\delta-1)}\right)^2 + \alpha_3\delta^2\lambda_1\lambda_2\lambda_4\left(t^{(\delta-1)}\right)^2 + \alpha_3\alpha_2\delta\lambda_1\lambda_4 t^{(\delta-1)} + \alpha_3\delta^2\lambda_1\lambda_3\lambda_4 t\left(t^{(\delta-1)}\right)^2 +$$

$$\alpha_1\alpha_2\alpha_3\lambda_4 + \alpha_1\alpha_3\delta\lambda_3\lambda_4 t^{(\delta-1)} + \alpha_2\alpha_3\delta\lambda_2\lambda_4 t^{(\delta-1)} + \alpha_3\delta^2\lambda_2\lambda_3\lambda_4\left(t^{(\delta-1)}\right)^2$$

Following Trivedi (2002), Wang and Kuo (2000), Wang et al. (2006) to develop the explicit for MTSF. The procedures require deleting rows and columns of absorbing states of matrix T and take the transpose to produce a new matrix, say M. The expected time to reach an absorbing state is obtained from

$$E\left[T_{P(0)\to P(absorbing)}\right] = P(0)\left(-M^{-1}\right)\begin{pmatrix}1\\1\\1\\1\end{pmatrix} \quad \textbf{(4)}$$

where the initial conditions are given by

$P(0) = [p_0(0), p_1(0), p_2(0), p_3(0)] = [1,0,0,0]$ **and**

$$M = \begin{pmatrix} -\lambda_1\delta t^{\delta-1} & \lambda_1\delta t^{\delta-1} & 0 & 0 \\ \alpha_1 & -\left(\alpha_1+\lambda_2\delta t^{\delta-1}\right) & \lambda_2\delta t^{\delta-1} & 0 \\ \alpha_2 & 0 & -\left(\alpha_2+\lambda_3\delta t^{\delta-1}\right) & \lambda_3\delta t^{\delta-1} \\ 0 & 0 & 0 & -\lambda_4\delta t^{\delta-1} \end{pmatrix}$$

The explicit expression for is given by MTSF

$$MTSF = \frac{\lambda_4\left(\alpha_1+\delta\lambda_2 t^{(\delta-1)}\right)\left(\alpha_2+\delta\lambda_3 t^{(\delta-1)}\right)+\delta\lambda_1\lambda_4\left(\alpha_2+\delta\lambda_3 t^{(\delta-1)}\right)+\delta^2\lambda_1\lambda_2\lambda_4\left(t^{(\delta-1)}\right)^3+\delta^2\lambda_1\lambda_2\lambda_3\left(t^{(\delta-1)}\right)^3}{\delta^3\lambda_1\lambda_2\lambda_3\lambda_4\left(t^{(\delta-1)}\right)^3} \quad (5)$$

IV. Discussion

Numerical examples are presented to demonstrate the impact of repair and failure rates on steady-state availability and net profit of the system based on given values of the parameters. For the purpose of numerical example, the following sets of parameter values are used: $\alpha_1 = 0.1$, $\alpha_2 = 0.3$, $\alpha_3 = 0.5$, $\lambda_1 = 0.4$, $\lambda_2 = 0.5$, $\lambda_3 = 0.6$, $\lambda_4 = 0.2$, $\delta = 0.9$, $0 \le t \le 10$

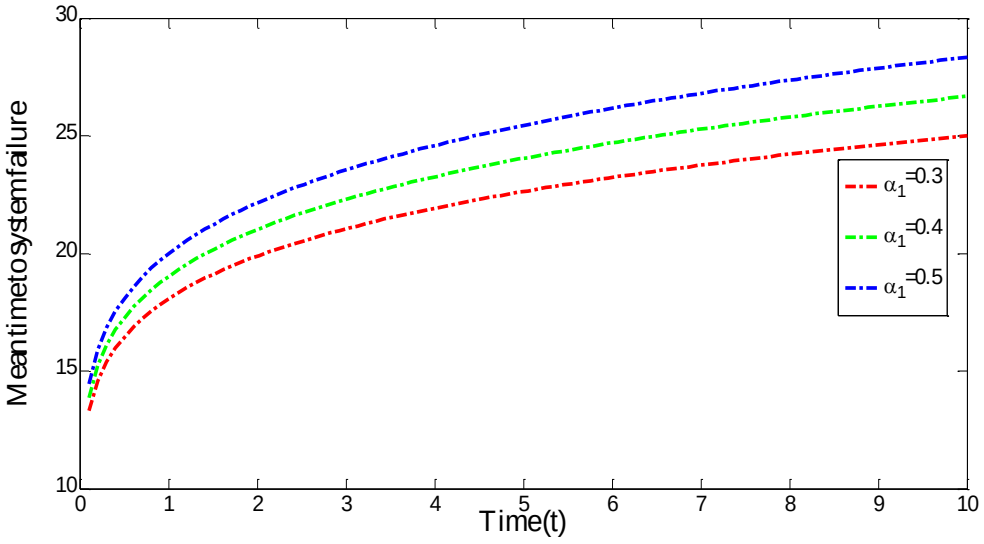


Figure 2: Mean time to system failure against **for** different values of $\alpha_1(0.3, 0.4, 0.5)$

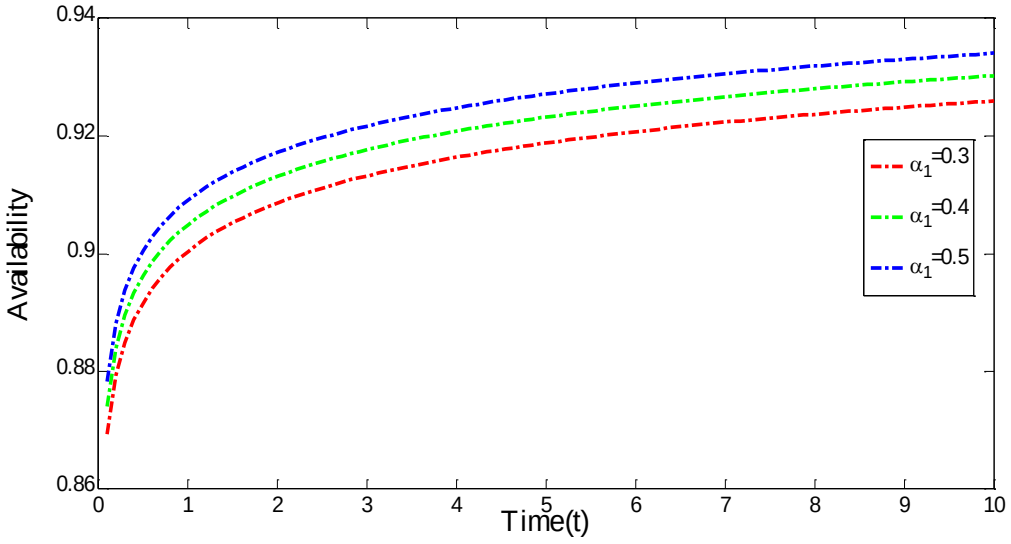


Figure 3: Availability against **for** different values of $\alpha_1(0.3, 0.4, 0.5)$

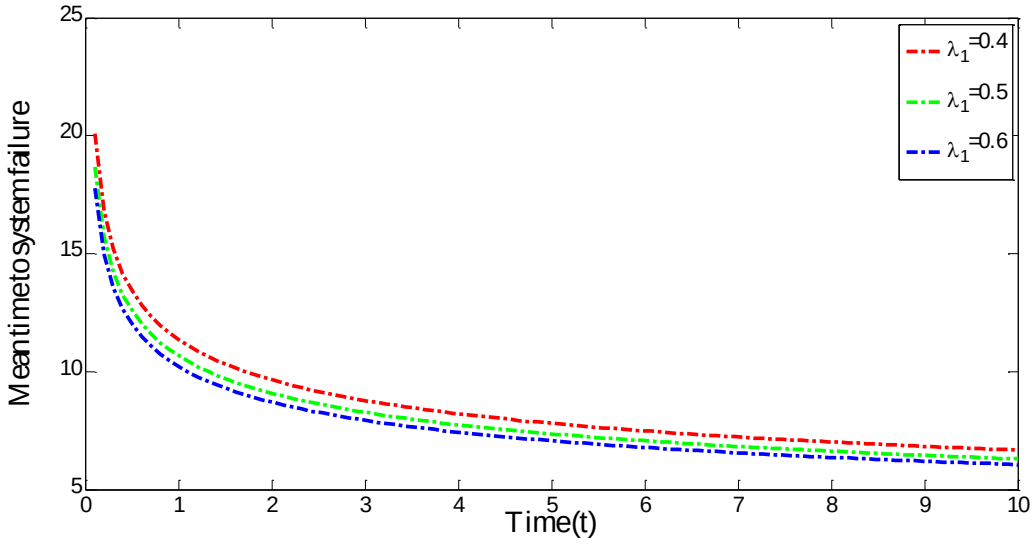


Figure 4: Mean time to system failure against **for** different values of $\lambda_1(0.4, 0.5, 0.6)$

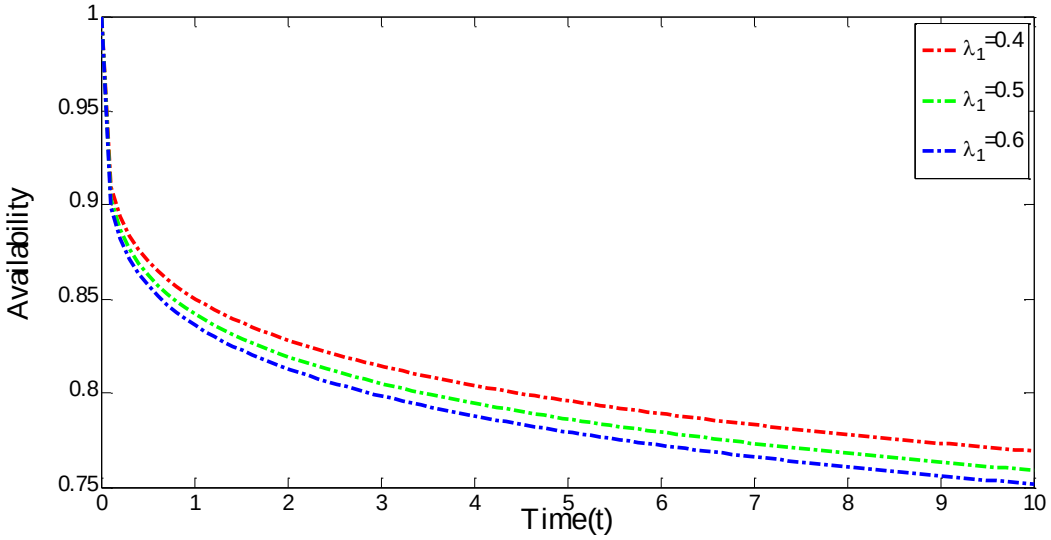


Figure 5: Availability against time for different values of $\lambda_1(0.4, 0.5, 0.6)$

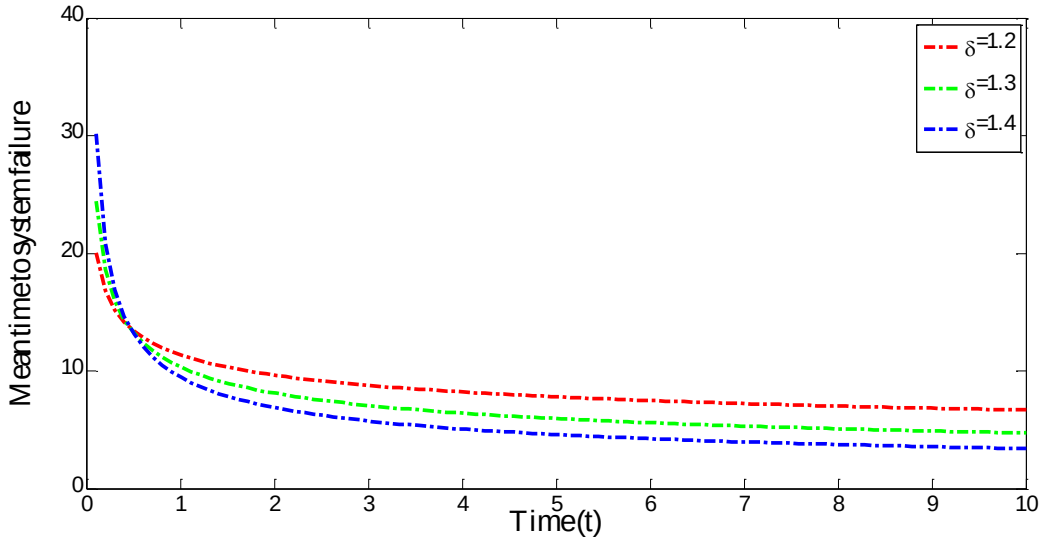


Figure 6: Mean time to system failure against **for** different values of $\lambda_1(0.4, 0.5, 0.6)$

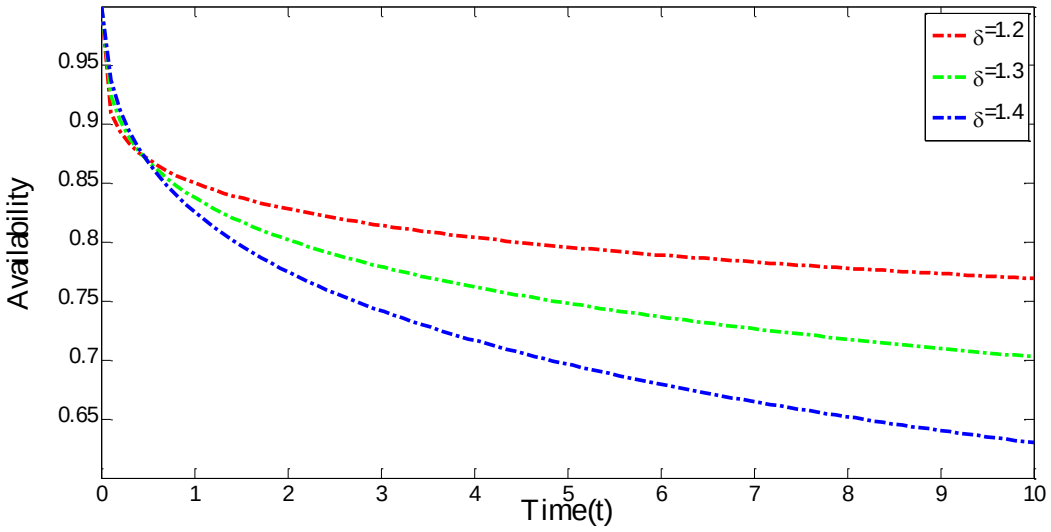


Figure 7: Availability against time for different values of $\delta(1.2,1.3,1.4)$

Numerical results of mean time to system failure and availability with respect to time are depicted in Figures 2 and 3 for different values of minor maintenance rates α_1. In these Figure the mean time to system failure and availability increases as time increases when α_1 increases from 0.3 to 0.5. This sensitivity analysis implies that minor maintenance to the system should be invoked to increase the life span of the system. On the other hand, simulations in Figures 4 and 5, depicts the impact of time on mean time to system failure and availability for different values of minor deterioration λ_1. From these Figures, the mean time to system failure and availability decreases as time increases for different values of λ_1. The above sensitivity analysis depicted the effect of minor maintenance and deterioration rates on mean time to system failure and availability. It can be observe that minor maintenance played a significant role in increasing the mean time to system failure and availability whereas minor deterioration slow down the mean time to system failure. From simulations depicted in Figures 6 and 7, it is evident that the choice of the shape parameter δ influences the time taken for the system to reach the failure state. The higher the value of this shape parameter δ, the less the values of mean time to system failure and availability.

V. Conclusion

This paper studied a one unit system with three consecutive stages of deterioration before failure. Explicit expressions for the mean time to system failure and availability are derived. The numerical simulations presented in Figures 2 – 7 provide a description of the effect of maintenance and deterioration rates on mean time to system failure. On the basis of the numerical results obtained for particular cases, it is suggested that the system availability can be improved significantly by:

- Adding more cold standby units.
- Increasing the maintenance rate.
- Exchange the system at major deterioration with new one before failure.

References

[1] Bérenguer C.(2008). On the mathematical condition-based maintenance modelling for continuously deteriorating systems, International Journal of Materials and Structural Reliability, 6, 133-151.

[2] Frangopol, D.M.; Kallen, M.J. and Van Noortwijk, J.M. (2004). Probabilistic models for life-cycle performance of deteriorating structures: Review and future directions. *Prog. Struct. Eng. Mater.*, *6*, 197–212.

[3] Lam,Y and Zhang, Y. L. (2003). A geometric-process maintenance model for a deteriorating system under a random environment, IEEE Trans. Reliability. 52(1), 83-89.

[4] Liu, D., Xu, G and Mastorakis, N. E. (2011).Reliability analysis of a deteriorating system with delayed vacation of repairman, WSEAS Transactions on Systems, 10(12).

[5] Nicolai, R.P. Dekker, R. and Van Noortwijk, J.M. (2007). A comparison of models for measurable deterioration: An application to coatings on steel structures. *Reliab. Eng. Syst. Saf.* , *92*, 1635–1650.

[6] Pandey, M.D.; Yuan, X.X.; van Noortwijk, J.M. The influence of temporal uncertainty of deterioration on life-cycle management of structures. *Struct. Infrastruct. Eng*. 2009, *5*, 145–156.

[7] Rani, T.C and Sukumari, C. (2014). Optimum replacement time for a deteriorating system, International Journal of Scientific Engineering and Research, 2(1), 32-33.

[8] Tuan K. Huynh, Anne Barros, Christophe Bérenguer. (2013).A Reliability-based Opportunistic Predictive Maintenance Model for k-out-of-n Deteriorating Systems, Chemical Engineering Transactions, 33, 493-498.

[9] Vinayak, R and Dharmaraja. S (2012). Semi-Markov Modeling Approach for Deteriorating Systems with Preventive Maintenance, International Journal of Performability Engineering Vol. 8, No. 5, pp. 515- 526.

[10] Wang, K,-H and Kuo, C,-C. (2000). Cost and probabilistic analysis of series systems with mixed standby components, Applied Mathematical Modelling, 24: 957-967.

[11] Wang, K., Hsieh, C and Liou, C. (2006). Cost benefit analysis of series systems with cold standby components and a repairable service station. Journal of quality technology and quantitative management, 3(1): 77-92.

[12] Xiao, T.C., Li, Y, -F., Wang, Z., Peng, W and Huang, H, -Z. (2013). Bayesian reliability estimation for deteriorating systems with limited samples Using the Maximum Entropy Approach, Entropy, *15*, 5492-5509; doi:10.3390/e15125492.

[13] Yuan, W., Z. and Xu, G. Q. (2012). Modelling of a deteriorating system with repair satisfying general distribution, Applied Mathematics and Computation 218, 6340–6350

[14] Yuan, L and Xu, J. (2011).A deteriorating system with its repairman having multiple vacations, Applied Mathematics and Computation. 217(10), 4980-4989.

[15] Yusuf, I., Suleiman, K., Bala, S.I. and Ali, U.A. (2012). Modelling the reliability and availability characteristics of a system with three stages of deterioration, International Journal of Science and Technology, 1(7) , 329-337.

[16] Zhang, Y.L. and Wang, G. J. (2007). A deteriorating cold standby repairable system with priority in use, European Journal of Operational Research, Vol.183, 1, pp.278–295.

RT&A, No4 (43)
Volume 11, December 2016

Study Of Starting Duty Of Wind Power Plant With Asynchronous Generators

Rauf Mustafayev, Laman Hasanova

•

Azerbaijan Scientific-Research and Disigned-Prospecting Institute of Energetics,
Baku, Azerbaijan AZ1012, Aven. H.Zardabi-94
mustafyevri@mail.ru, gasanovalg@mail.ru

Abstract

Presently, the park of wind power plants (WPP) consists mostly of frequency controlled asynchronous generators. As the generators the squirrel-cage asynchronous machines and generators made on the basis of double fed asynchronous machines (DFAM) are used. When WPPs locate far from the powerful sources of energy generation of power system and they are connected with the power system by "weak" power grids, i.e. by grids, which are not equipped with reactive power sources, then the unwanted voltage dips may occur when connecting the WPPs to the power system in the places of their connection to the power system.

The comparative analysis on the developed three-coordinated mathematical models of asynchronous machines of start by underfrequency relay and connection of WPPs with the above asynchronous generators to the power system has been carried out.

It has been found, that in terms of impact of starting duties on electric power networks the most preferable are the systems of WPPs with squirrel-cage asynchronous generators. The values of starting currents when start by underfrequency relay of WPPs with squirrel-cage asynchronous generators are almost 48% lower than in the system of WPPs with DFAM eactive power compensation of asynchronous generators wind power and small hydroelectric power stations increases the reliability of connecting them to the so-called "weak" power grids of power systems. The methods of reactive power compensation for asynchronous generators of various designs.

Keywords: wind power plants, asynchronous machines, double fed asynchronous machines

I. Introduction

Today small hydropower industry and wind energetics take the leading positions according to the quantity of installed capacities and electric power output, generated by renewable power sources.

Park of modern industrial wind power plants (WPP) consists of highly economical and reliable complexes, based on the latest blade wind motors, high-tech gearboxes and controlled electromechanical converters [1]. For all this the unit capacity of WPPs increases from year to year and now reaches the value of 6–7,5 MW.

II. Controlled induction generators used in wind turbines

The vast majority of electromechanical converters of WPPs consist of the controlled asynchronous machines both with squirrel-cage rotor with a frequency converter, feeding the stator winding of machine, and phase-wound rotor equipped with a frequency converter, feeding a rotor winding of machine (so-called double fed machine (DFAM)).

The advantages and disadvantages of each of above-stated options are well-known. Without going into details of these known ontions let's only note, that in the option of squirrel-cage rotor the simple, reliable generator combines with a frequency converter, made for a full power of generator, and in the option of phase-wound rotor the relatively complicated, more expensive and less reliable generator combines with a frequency converter, feeding the rotor's winding of generator and made for only 20–30% of generator's power.

In both options, these asynchronous generators in steady- state operation modes correspond in full measure to the technological requirements of optimal operation of WPPs, i.e. control of a rotational frequency of wind aggregate (wind motor, gearbox and generator) allows improving its productivity by 15–20% [1].

The purpose of this paper is the study issues of starting duties of WPPs, equipped both with squirrel-cage asynchronous generators and the generators, made on the basis of double fed asynchronous machines. These issues are particularly urgent in those cases, when WPPs and wind parks, containing several dozens of WPPs, are connected to the so-called "weak" power networks, i.e., the networks distanced sufficiently from power centers and insufficiently compensated by reactive powers. In these cases, the connection modes of WPPs with asynchronous generators to them can lead to significant voltage dip of network in startup period [2].

Studies are carried out on the developed three-coordinate models of controlled three-phase asynchronous machines, whose equations are given in [3, 4].

At the first stage let's study the startup issues of WPP, containing a squirrel-cage asynchronous generator, stator winding of which is supplied from a frequency converter, performed on fully controlled power transistors (IGBT-transistors) and controlled on the principles of sinusoidal PDM. The system's diagram of connection to power network is shown in Fig. 1.

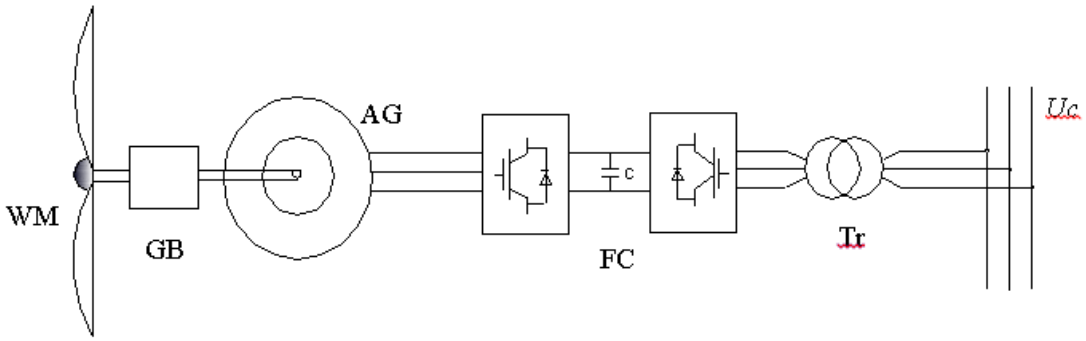


Figure 1: *Grid connection diagram of wind turbines with squirrel-cage induction generator*

Here WM – wind motor, GB – gearbox, AG – squirrel-cage asynchronous generator, FC – frequency converter, operating on power IGBT transistors, Tr – coupling transformer of connection with the system.

As it has been noted in [3], the equations for frequency controlled squirrel-cage asynchronous generator are reasonable to represent in the axes α_s, β_s, γ_s fixed in space, in this case the equations will be presented in the form of:

$$\left.\begin{aligned}
&p\Psi_{s\alpha} = U_{s\alpha} - r_{s\alpha}\cdot i_{s\alpha}\\
&p\Psi_{s\beta} = U_{s\beta} - r_{s\beta}\cdot i_{s\beta}\\
&p\psi_{s\gamma} = U_{s\gamma} - r_{s\gamma}\cdot i_{s\gamma}\\
&p\psi_{r\alpha} = U_{r\alpha} - \frac{1}{\sqrt{3}}\cdot\omega_r\left(\psi_{r\beta} - \psi_{r\gamma}\right) - r_{r\alpha}\cdot i_{r\alpha}\\
&p\psi_{r\beta} = U_{r\beta} - \frac{1}{\sqrt{3}}\cdot\omega_r\left(\psi_{r\gamma} - \psi_{r\alpha}\right) - r_{r\beta}\cdot i_{r\beta}\\
&p\psi_{r\gamma} = U_{r\gamma} - \frac{1}{\sqrt{3}}\cdot\omega_r\left(\psi_{r\alpha} - \psi_{r\beta}\right) - r_{r\gamma}\cdot i_{r\gamma}\\
&m_{эм} = \frac{\sqrt{3}}{2}p_m\cdot x_m\left[\left(i_{s\alpha}\cdot i_{r\gamma} + i_{s\beta}\cdot i_{r\alpha} + i_{s\gamma}\cdot i_{r\beta}\right) - \left(i_{s\alpha}\cdot i_{r\beta} + i_{s\beta}\cdot i_{r\gamma} + i_{s\gamma}\cdot i_{r\alpha}\right)\right]\\
&p\omega_r = \frac{p_m}{J^*}\left(m_{em} - m_{WM}\right)\\
&i_{s\alpha} = k_{s\alpha1}\cdot\psi_{s\alpha} + k_{s\alpha2}\cdot\psi_{s\beta} + k_{s\alpha3}\cdot\Psi_{s\gamma} + k_{s\alpha4}\cdot\psi_{r\alpha} + k_{s\alpha5}\cdot\psi_{r\beta} + k_{s\alpha6}\cdot\psi_{r\gamma}\\
&i_{s\beta} = k_{s\beta1}\cdot\psi_{s\alpha} + k_{s\beta2}\cdot\psi_{s\beta} + k_{s\beta3}\cdot\Psi_{s\gamma} + k_{s\beta4}\cdot\psi_{r\alpha} + k_{s\beta5}\cdot\psi_{r\beta} + k_{s\beta6}\cdot\psi_{r\gamma}\\
&i_{s\gamma} = k_{s\gamma1}\cdot\psi_{s\alpha} + k_{s\gamma2}\cdot\psi_{s\beta} + k_{s\gamma3}\cdot\Psi_{s\gamma} + k_{s\gamma4}\cdot\psi_{r\alpha} + k_{s\gamma5}\cdot\psi_{r\beta} + k_{s\gamma6}\cdot\psi_{r\gamma}\\
&i_{r\alpha} = k_{r\alpha1}\cdot\psi_{s\alpha} + k_{r\alpha2}\cdot\psi_{s\beta} + k_{r\alpha3}\cdot\Psi_{s\gamma} + k_{r\alpha4}\cdot\psi_{r\alpha} + k_{r\alpha5}\cdot\psi_{r\beta} + k_{r\alpha6}\cdot\psi_{r\gamma}\\
&i_{r\beta} = k_{r\beta1}\cdot\psi_{s\alpha} + k_{r\beta2}\cdot\psi_{s\beta} + k_{r\beta3}\cdot\Psi_{s\gamma} + k_{r\beta4}\cdot\psi_{r\alpha} + k_{r\beta5}\cdot\psi_{r\beta} + k_{r\beta6}\cdot\psi_{r\gamma}\\
&i_{r\gamma} = k_{r\gamma1}\cdot\psi_{s\alpha} + k_{r\gamma2}\cdot\psi_{s\beta} + k_{r\gamma3}\cdot\Psi_{s\gamma} + k_{r\gamma4}\cdot\psi_{r\alpha} + k_{r\gamma5}\cdot\psi_{r\beta} + k_{r\gamma6}\cdot\psi_{r\gamma}
\end{aligned}\right\} \qquad (1)$$

where:

$$\begin{bmatrix}
k_{s\alpha1} & k_{s\alpha2} & k_{s\alpha3} & k_{s\alpha4} & k_{s\alpha5} & k_{s\alpha6}\\
k_{s\beta1} & k_{s\beta2} & k_{s\beta3} & k_{s\beta4} & k_{s\beta5} & k_{s\beta6}\\
k_{s\gamma1} & k_{s\gamma2} & k_{s\gamma3} & k_{s\gamma4} & k_{s\gamma5} & k_{s\gamma6}\\
k_{r\alpha1} & k_{r\alpha2} & k_{r\alpha3} & k_{r\alpha4} & k_{r\alpha5} & k_{r\alpha6}\\
k_{r\beta1} & k_{r\beta2} & k_{r\beta3} & k_{r\beta4} & k_{r\beta5} & k_{r\beta6}\\
k_{r\gamma1} & k_{r\gamma2} & k_{r\gamma3} & k_{r\gamma4} & k_{r\gamma5} & k_{r\gamma6}
\end{bmatrix} = \begin{bmatrix}
x_{s\alpha} & -0{,}5\cdot x_m & -0{,}5\cdot x_m & x_m & -0{,}5\cdot x_m & -0{,}5\cdot x_m\\
-0{,}5\cdot x_m & x_{s\beta} & -0{,}5\cdot x_m & -0{,}5\cdot x_m & x_m & -0{,}5\cdot x_m\\
-0{,}5\cdot x_m & -0{,}5\cdot x_m & x_{s\gamma} & -0{,}5\cdot x_m & -0{,}5\cdot x_m & x_m\\
x_m & -0{,}5\cdot x_m & -0{,}5\cdot x_m & x_{r\alpha} & -0{,}5\cdot x_m & -0{,}5\cdot x_m\\
-0{,}5\cdot x_m & x_m & -0{,}5\cdot x_m & -0{,}5\cdot x_m & x_{r\beta} & -0{,}5\cdot x_m\\
-0{,}5\cdot x_m & -0{,}5\cdot x_m & x_m & -0{,}5\cdot x_m & -0{,}5\cdot x_m & x_{r\gamma}
\end{bmatrix}^{-1}$$

$U_{s\alpha}$, $U_{s\beta}$, $U_{s\gamma}$, $U_{r\alpha}$, $U_{r\beta}$, $U_{r\gamma}$ – phase voltages of stator and rotor; $i_{s\alpha}$, $i_{s\beta}$, $i_{s\gamma}$, $i_{r\alpha}$, $i_{r\beta}$, $i_{r\gamma}$ - phase currents of stator and rotor; $\psi_{s\alpha}$, $\psi_{s\beta}$, $\psi_{s\gamma}$, $\psi_{r\alpha}$, $\psi_{r\beta}$, $\psi_{r\gamma}$ – flux linkages of stator and rotor circuits $r_{s\alpha}$, $r_{s\beta}$, $r_{s\gamma}$, $r_{r\alpha}$, $r_{r\beta}$, $r_{r\gamma}$ – resistances of stator and rotor circuits; $x_{s\alpha}$, $x_{s\beta}$, $x_{s\gamma}$, $x_{r\alpha}$, $x_{r\beta}$, $x_{r\gamma}$ – full inductances of stator and rotor windings; x_m – mutual induction reactance; ω_r – rotational frequency of WPP's rotor, p_m – number of pairs of generator's poles; p – differentiation symbol with respect to synchronous time $\tau = 314\cdot t$ sec.

It should be noted that two kinds of frequency control are used - scalar and vector ones [5, 6]. The simplest control of them according to degree of realization is a scalar one. When study the issues of start by underfrequency relay, it is reasonable from the point of view of comparing the results of study of starting duties impact on electric power network to turn to the scalar control.

The asynchronous generator of WPP with the capacity of $P_{\text{WPP}} = 1500$ kW and $U_{\text{WPP}} = 690$ V voltage is taken as the studied object (generator's parameters are given in the Appendix).

At the first stage let's consider the direct start of generator (in practice it isn't used) for different values of driving torque on generator's shaft, which correspond to the specific wind speeds when connecting the generator to electric power network. The studies are conducted by the system of equations (1), which are recorded in the axes α_s, β_s, γ_s, fixed in space. In this process as the machine is with squirrel cage rotor $U_{r\alpha}$, $U_{r\beta}$, $U_{r\gamma} = 0$, and stator voltages are equal to

$$\left.\begin{aligned}
&U_{s\alpha} = U_s\cdot\sin\tau\\
&U_{s\beta} = U_s\cdot\sin\left(\tau - 2{,}094\right)\\
&U_{s\gamma} = U_s\cdot\sin\left(\tau + 2{,}094\right)
\end{aligned}\right\} \qquad (2)$$

III. Study starting modes of wind turbines equipped with squirrel-cage induction generators

The fluktogrammas of change of generator's operating conditions are presented in Fig. 2: rotational frequency of generator's rotor ω_r (Fig. 1,*a*), electromagnetic torque of generator m_{em} (Fig. 2, *b*), and phase stator currents of generator isα, $i_{s\beta}$, $i_{s\gamma}$ (Fig. 2 *c*, *d*, *e*). It is seen from the fluktogrammas, that at a value of driving torque equal to $m_{em} = -0{,}3$ ("minus" sign indicates a generator mode) the time of connecting of generator to network constitutes about $\tau \approx 400$ rad. ($t = 1{,}27$ s.), the values of starting currents isα, $i_{s\beta}$, $i_{s\gamma}$ reach 5,2 of multiple value, while their duration is not less than 1 second. ($\tau_{start} = 316$ rad.). The value of driving torque equal to $m_{\text{WM}} = -0{,}3$ corresponds roughly to the wind speed equal to $\text{V} = 6$ m/s. When connecting the generator to network in the presence of a driving torque $m_{\text{WM}} = -1$ (wind speed is about $\text{V} = 9{,}5$ m/s), the character of process remains the same as in Fig. 1, but the duration of connection to network and correspondingly the duration of starting currents reduces by 33% and constitutes ($\tau_{start} = 210$ rad.).

It is natural, that the significant amounts of starting currents affect negatively on a network voltage, if it is distanced from the other power supply units and is lightly compensated one (it means the lack of reactive power sources). Therefore, this method of connection of asynchronous generator to network isn't used in practice.

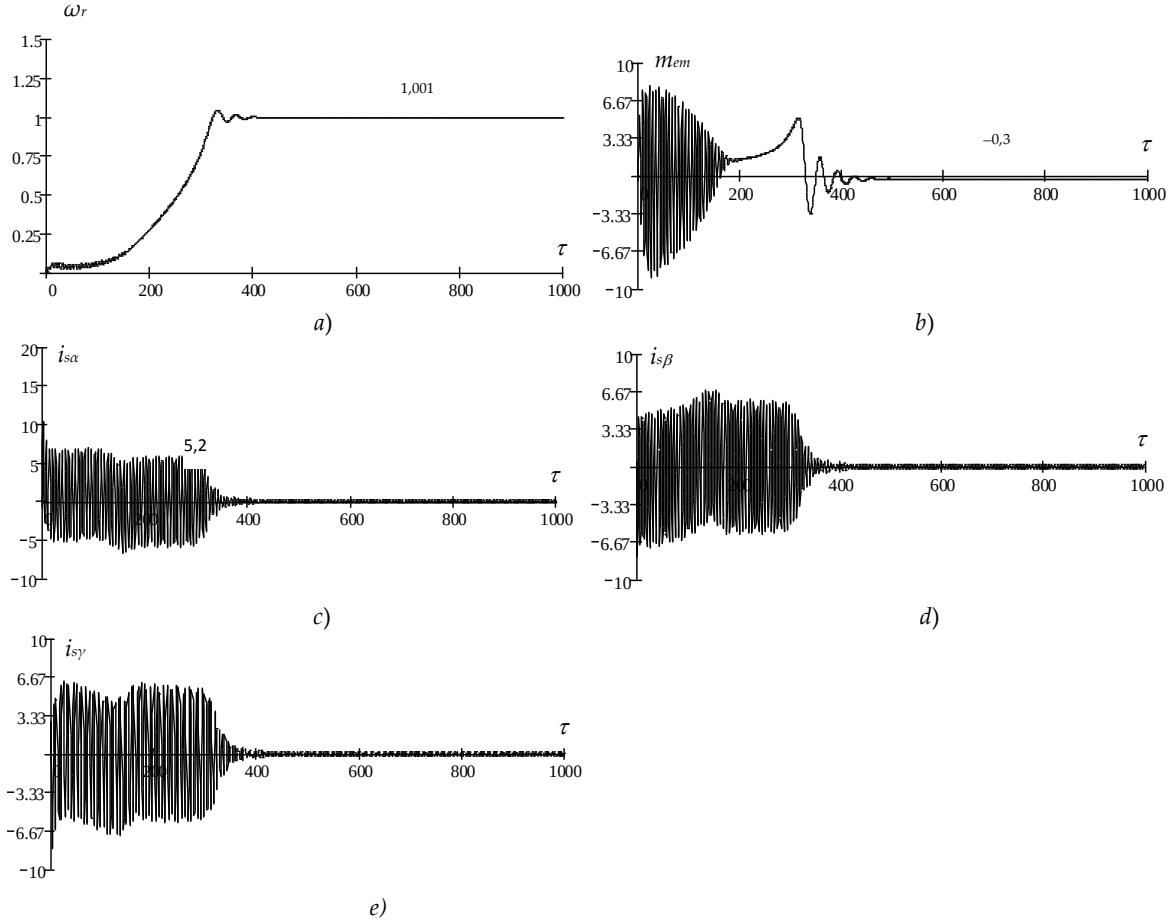


Figure 2. *Fluctogramms of change in operating parameters of induction generator of wind turbines when direct on-line starting*

If there is a frequency converter connected to the stator winding of asynchronous generator, which has a short-circuited rotor, the start by underfrequency relay is carried out. In this process both the amplitude and frequency of supplying the generator voltage should be changed in the equations (2). With linear change of amplitude and frequency, the equations (2) by the same expression will take the form:

$$\left.\begin{aligned} U_{s\alpha} &= k_{us}\cdot\sin k_{fs}\cdot\tau \\ U_{s\beta} &= k_{us}\cdot\sin\left(k_{fs}\cdot\tau-2{,}094\right) \\ U_{s\gamma} &= k_{us}\cdot\sin\left(k_{fs}\cdot\tau+2{,}094\right) \end{aligned}\right\} \qquad (3)$$

where $k_{us}=k_{ou}+k_u\cdot\tau$ and $k_{fs}=k_{of}+k_f\cdot\tau$.

With simultaneous linear change of k_{us} and k_{fs}, $k_{0u}=k_{0f}$ and $k_u=k_f$ their initial value and rates of rise are equal. It is necessary to pay attention to one circumstance: with a linear change of frequency k_{fs} its rate of rise in accordance with [7] should be coordinated with the inertia constant of system and a value of driving torque on the WPP's shaft.

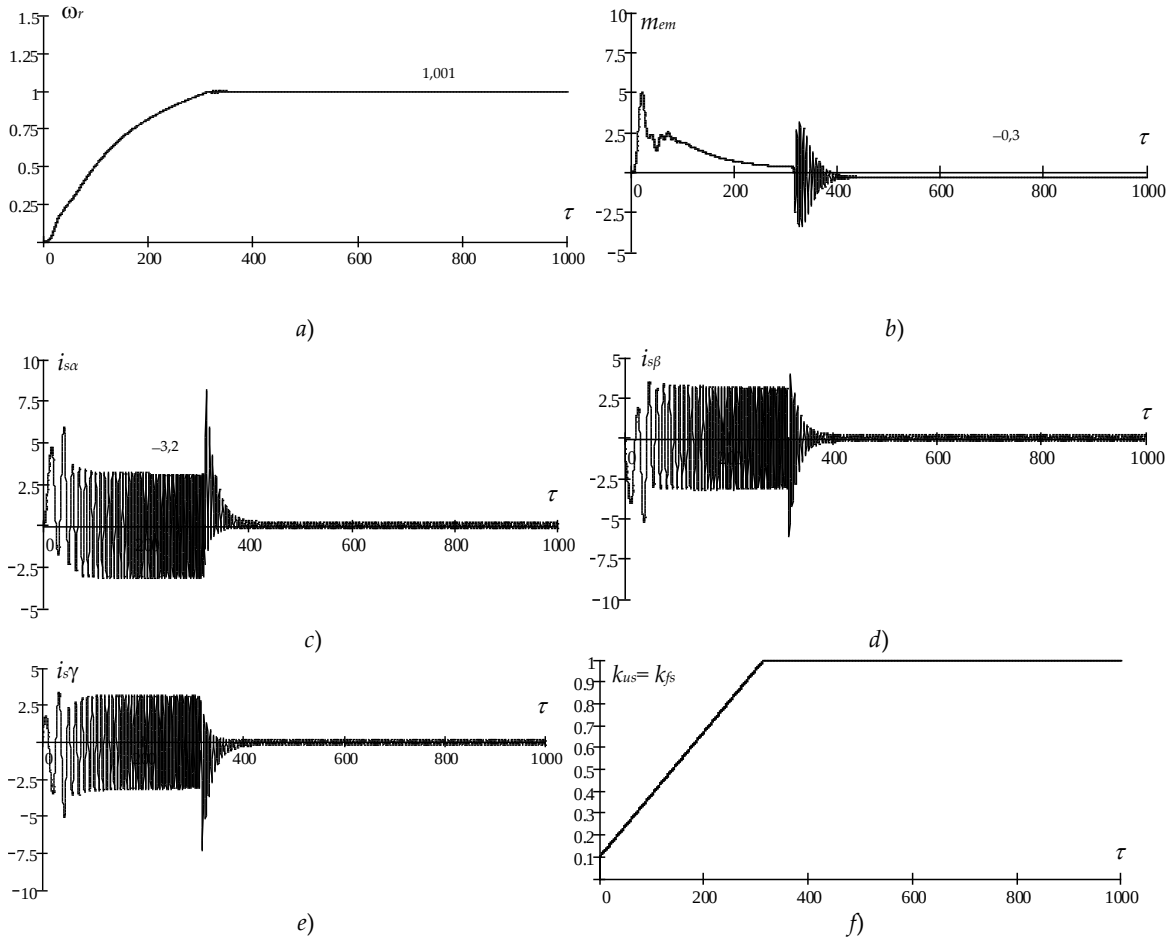


Figure 3. *Fluctogramms of change of operating parameters of the wind turbines with squirrel-cage induction generator in frequency starting*

The fluktogrammas of operating conditions change of the same generator when a start by underfrequency relay according to the expression $k_{us}=k_{fs}=0{,}1+0{,}0028\cdot\tau$ are presented in Fig. 3, herewith a driving torque just as with direct startup was equal to $m_{WM}=-0{,}3$ (wind speed – 5,8-6

m/s). Rotational frequency of WPP ω_r (Fig. 3,*a*) varies practically smoothly from 0 to 1,001, which indicates that a speed of frequency change k_{fs}, and together with it the amplitude of voltage k_{us} (Fig. 3, *f*) has been selected the optimum one [7]. The average value of electromagnetic torque m_{em} (Fig. 3, *b)* at the initial section is less than $m_{em\ av.} = 2{,}5$ (with direct startup $m_{em} = 6{,}7$ (Fig. 2, *b*)). And the most important fact is, that when the start by underfrequency relay the average current values $i_{s\alpha}$, $i_{s\beta}$, $i_{s\gamma}$ in the starting condition practically within the whole startup period do not exceed the values of $i_{s\alpha n} = i_{s\beta n} = i_{s\gamma n} = 3{,}2$ (Fig. 3 *c, d, e*), and their duration constitutes 315–320 rad. Thus, when the start by underfrequency relay with a practically constant duration of action of starting currents their average values reduce almost by 40%. And this allows asserting with a high probability that a voltage drop in the electric power network in the points of WPPs connection will also reduce by~ 40%.

IV. Study starting modes of wind turbines equipped with double-fed induction generator

At the second stage let's consider the issues of startup of WPP's asynchronous generator, carried out on the basis of double fed machine. Diagram of connection to electric power network is shown in Fig. 4 [8].

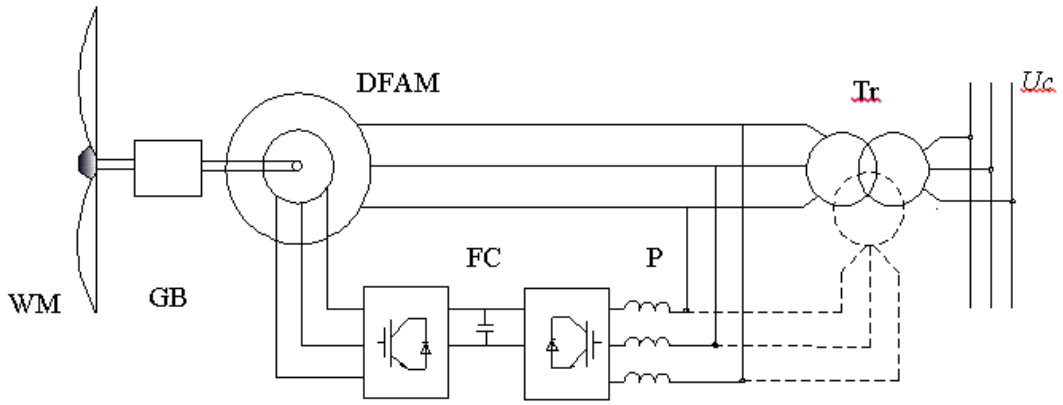


Figure 4: *Grid connection diagram of wind turbines with double fed induction machine*

Here: WD – wind motor, GB – gearbox, DFAM – asynchronous generator made on the basis of double fed asynchronous machine, the FC – frequency converter made on fully controlled power transistors (IGBT – transistors); R – inductive reactor; Tr – coupling transformer of connection with the power system (it can be two-winding one (solid lines) or triple-winding one (dashed lines)). When study the issues of start by underfrequency relay of WPPs with these generators, it is expedient to turn to the form of records of machine's differential equations in three-coordinate system of coordinates α_r, β_r, γ_r, rotating with the rotor speed ω_r [3].

In this case, the first 6 equations of the system (1) are replaced by the following 7 ones, the others remain unchanged:

$$\left.\begin{aligned}
p\Psi_{s\alpha} &= U_{s\alpha}\cdot\sin\theta+\frac{1}{\sqrt{3}}\cdot\omega_r\left(\psi_{s\beta}-\psi_{s\gamma}\right)-r_s\cdot i_{s\alpha}\\
p\Psi_{s\beta} &= U_{s\beta}\cdot\sin\left(\theta-\frac{2\pi}{3}\right)+\frac{1}{\sqrt{3}}\cdot\omega_r\left(\psi_{s\gamma}-\psi_{s\alpha}\right)-r_s\cdot i_{s\beta}\\
p\Psi_{s\gamma} &= U_{s\gamma}\cdot\sin\left(\theta+\frac{2\pi}{3}\right)+\frac{1}{\sqrt{3}}\cdot\omega_r\left(\psi_{s\alpha}-\psi_{s\beta}\right)-r_s\cdot i_{s\gamma}\\
p\Psi_{r\alpha} &= k_{ur}\cdot\sin\left(k_{fr}\cdot\tau\right)-r_r\cdot i_{r\alpha}\\
p\Psi_{r\beta} &= k_{ur}\cdot\sin\left(k_{fr}\cdot\tau-\frac{2\pi}{3}\right)-r_r\cdot i_{r\beta}\\
p\Psi_{r\gamma} &= k_{ur}\cdot\sin\left(k_{fr}\cdot\tau+\frac{2\pi}{3}\right)-r_r\cdot i_{r\gamma}\\
p\theta &= 1-\omega_r
\end{aligned}\right\} \qquad (4)$$

The rest equations (1) of system: electromagnetic torque, movement, connection of the currents with flux linkages, matrix of determination of this connection factors remains unchanged. One variable is added here – θ, the interior angle of machine, i.e. the angle between axis α_{s0} of three-coordinate system of stator's coordinates, moving with the synchronous speed ω_s and axis α_r of three-coordinate system of rotor's coordinates, moving with a speed of machine's rotor ω_r

Technology of start by underfrequency relay of double fed machine of WPP, which can consist of two stages, is the following [2]: at the first stage in the presence of driving torque on the shaft of WPP – m_{WM} the winding of stator is shorted-circuited, and the frequency converter, feeding the winding of rotor of DFAM changes linearly the amplitude and frequency of voltage supplying the voltage winding. But because of the limited capacity of converter and its output parameters, this change does not exceed a value of 15–25% of the total machine's capacity. At the second stage after acceleration of WPP's rotor under the influence of driving torque of WPP – m_{em} and start by underfrequency relay, when the rotor's speed reaches 20–25% of the synchronous one, the rotor's winding of generator is short-circuited, and the full line voltage is supplied to the stator's winding, which is connected directly to electric power network. Upon reaching the synchronous speed the generator rotor's windings come into operation, and WPP operates in a steady-state mode with connecting the automatic control of WPP's rotational frequency with the help of frequency converter as a function of wind speed value in a certain range of its variation.

All of described above has been implemented on the three-coordinate mathematical model of DFAM of WPP [3], the results of which are shown in Fig. 5.

As it has been noted, at the first stage the voltage of stator $U_s = U_{s\alpha}=U_{s\beta}=U_{s\gamma}= 0$ (winding is short-circuited) and the rotor winding is fed from the frequency converter according to the linear expression $k_{us}=k_{fs}=-\left(0{,}01+0{,}00028\cdot\tau\right)$. The fluktogramma of change of rotational frequency of rotor ω_r is given in Fig. 5, *a*. In the process of start by underfrequency relay from the rotor side in the presence of driving torque of wind motor $m_{\text{WM}}=-0{,}3$ (which corresponds to wind speed $\text{V}=6$ m/s) in the range of from 0 to 500 radian the rotational frequency of WPP's rotor rises to $\omega_r=0{,}28$, in this process all operating conditions of generator - the electromagnetic torque m_{em} and currents of stator $i_{s\alpha}$, $i_{s\beta}$, $i_{s\gamma}$ do not exceed 1–1,5 of values in relative units. After 500-th radian the winding of stator is connected to the network voltage U_s, and winding of DFAM's rotor is short-circuited U_r=0. A direct startup of asynchronous generator occurs, but the acceleration begins not from zero, and with the initial rotor's speed equal to $\omega_r=0{,}28$. The time of startup of this stage constitutes about $\tau_{rn.}=120$ rad. (from 500 rad. up to 620 rad.), while the average value of starting electromagnetic torque is equal to $m_{emav.}=3{,}3$, stator currents $i_{s\alpha}$ (Fig. 5, *c*), $i_{s\beta}$ (Fig. 5, *d*) $i_{s\gamma}$ (Fig. 5, e) reach the value of $i_{s\alpha}\approx i_{s\beta}\approx i_{s\gamma}\approx 5{,}8$.

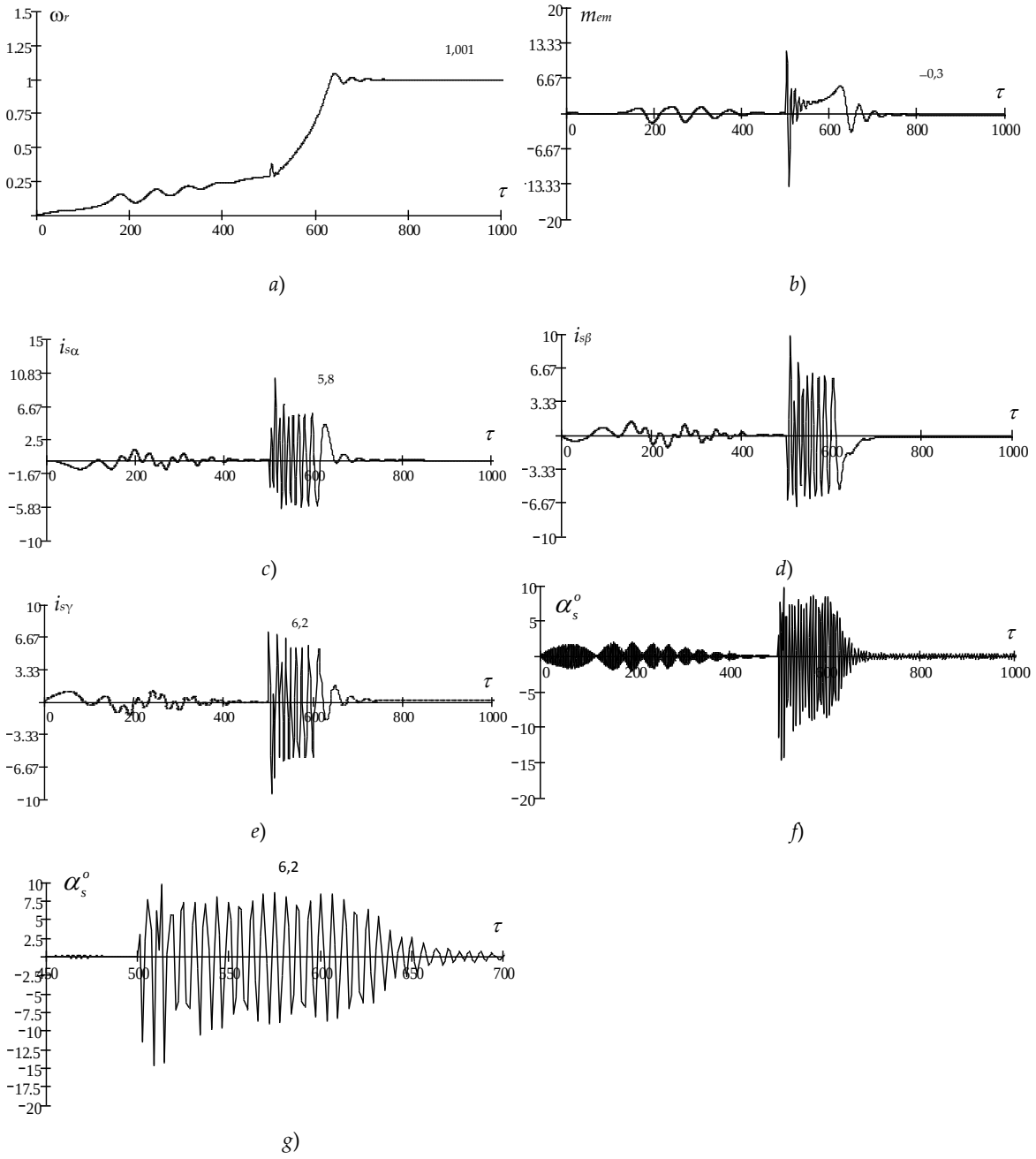


Figure 5. *Fluctogramms of change of operating parameters of double-fed induction generator of wind turbine when frequency-direct on-line starting*

In order to carry out a comparative analysis of start by underfrequency relay processes of asynchronous generators of WPPs, made with the squirrel-cage rotor and winding of rotor (DFAM) it needs to transfer the currents received on mathematical model of DFAM into the system of fixed coordinates. By trivial conversions let's determine, for example, a current in the fixed in space system of coordinate α_s^o, it is determined by the correlation [9]:

$$i_{s\alpha}^o = i_{s\alpha} \cdot \cos(\theta + \omega_s \cdot \tau) + i_{s\beta} \cdot \cos\left(\frac{2\pi}{3} - \theta - \omega_s \cdot \tau\right) - i_{s\gamma} \cdot \cos\left(\frac{\pi}{3} - \theta - \omega_s \cdot \tau\right) \quad (5)$$

Where $i_{s\alpha}^o$ – reduced current of DFAM's stator along the axis α_s, fixed in space, $i_{s\alpha}$, $i_{s\beta}$, $i_{s\gamma}$ – currents of DFAM's stator recorded in the axes rotating with the rotor speed ω_r. θ –angle between the axis rotating with the rotor speed ω_r and the axis rotating with synchronous speed ω_s, and

finally $\omega_s \cdot \tau$ – angle between the fixed in space axis and axis rotating with synchronous speed $\omega_s = 1$.

Current in one phase α_s^o determined by the expression (5) is presented in Fig.5, *f*, (unfolded fluktogramma of this current is shown in Fig. 5, *g*).

As it is seen from the fluktogramma the average starting value of this current reaches the value of $i_{s\alpha cp} = 6{,}2$ and its duration is equal to 130 radian.

Thus, comparing the results of studies of start by underfrequency relay modes of frequency controlled squirrel-cage asynchronous generator of WPP,and frequency controlled from the rotor side DFAM of WPP, it should be noted, that in terms of impact on electric power network, the starting currents of the first generator constitute of the order of 3,2 relative units with the time of their action $\tau_n = 315$ radian, and the starting currents of DFAM constitute of the order of 6,2 relative units with the time of their action $\tau_n = 130$ radian. In the first case the total starting time constitutes 400 radian, and in the second case – 620 radian.

Conclusion

1. The methods of comparative analysis of start by underfrequency relay of WPP, containing the frequency controlled squirrel-cage asynchronous generator, and WPP, containing the asynchronous generator, made on the basis of double fed machine, the rotor windings of which is supplied from a frequency converter, has been developed.

2. The most preferable option out of the investigated ones, in terms of impact on the electric power networks in the place of WPP's connection, is the one with the frequency controlled generators with squirrel-cage rotors, starting current of which when the start by underfrequency relay is almost 2 times lower than the starting current when startup of WPP with DFAM, although its duration is 2.4 times longer than the starting current of DFAM. However, the total time of output to steady-state modes for WPP with squirrel-cage asynchronous generator is 1,5 times less than for the generator with DFAM.

3. And finally, the energy expended for startup process in the first case is 55% lower than the one expended for startup of DFAM.

References

[1] World Wind Resource Assessment Report by the WWEA Technical Committee. December 2014, WWEA Technical Paper Series (TP-01-14), Bonn, Germany.

[2] www.elespec-ltd.com. Elspec Power Quality Solutions. Wind Energy: Reactive Power Compensation Systems.

[3] Mustafayev R., Hasanova L. Mathematical models of controlled three-phase asynchronous machines. Electricity, №5, 2016.

[4] Kopylov I.P. Mathematical modeling of electrical machine. M. High school. 1987.

[5] Dejan Schreiber High power converters for renewable power sources. Translation of A.Kolpakov. Power electronics, №5, 2010.

[6] Guzeev B.V. Khakimyanov M.I. Modern industrial high-voltage frequency converters for control of asynchronous and synchronous motors. "Oil and Gas Business" Electronic scientific journal, № 3, 2011.

[7] Guzeev B.V. Khakimyanov M.I. Modern industrial high-voltage frequency converters for control of asynchronous and synchronous motors. "Oil and Gas Business" Electronic scientific journal, № 3, 2011.

Appendix

Item 1

a) Parameters of WPP's generator; resistances in relative units:

$P_{\text{n}} = 1500$ kW	r_s=0,0086	x_m=3,459
U= 690 V	r_r=0,01	x_s=3,5399≈3,54
p_m=2	$x_{\sigma s}$=0,0809	x_r=3,5461
$J_{total} = 52{,}4$ kgm^2 (WPP's rotor)	$x_{\sigma r}$=0,871	
$n_н$=1500 $^{rev}/_{\min.}$	$M_{ном} = \frac{9550 \cdot P_н}{n_н} = 9550$ Nm	

b) Calculated and basic values of parameters

U_{phase}=398 B $\quad P_{баз} = \frac{3}{2} \cdot U_{баз} \cdot I_{баз} = 1664{,}256$ кW

$U_{basic} = \sqrt{2} \cdot U_{фаз} = 562{,}8$ B $\quad M_{баз} = \frac{P_{баз}}{\omega_{баз}} = 5300{,}17$ Nm

$I_{phase} = 1394$ A

$I_{basic} = \sqrt{2} \cdot I_{фаз} = 1971{,}4$ A $\quad J_{баз} = \frac{M_{баз}}{\omega_{баз}^2} = 0{,}0537$ kgm^2

c) Values of parameters in relative units

$$m_H^* = \frac{M_{ном}}{M_{баз}} = 1{,}8 \qquad J^* = \frac{J_{общ}}{J_{баз}} = 975{,}6 \qquad \frac{p_m}{J^*} = \frac{2}{975{,}6} = 0{,}00205$$

Item 2. Algorithm of study on the three-phase model of start by underfrequency relay of squirrel-cage asynchronous machine with a linear change of amplitude and frequency of the stator's voltage. Amplitude of the phase voltage of stator winding for all three phases changes according to the ratio: $k_{us} = k_{fs} = k_0 + k \cdot \tau = 0{,}1 + 0{,}00286 \cdot \tau$.

$$D(\tau, Y) = \begin{bmatrix} (0{,}1+0{,}00286\cdot\tau)\cdot\sin\left[(0{,}1+0{,}00286\cdot\tau)\cdot\tau\right] - 0{,}0086\cdot I_{s\alpha} \\ (0{,}1+0{,}00286\cdot\tau)\cdot\sin\left[0{,}1+0{,}00286\cdot\tau)\cdot\tau - 2{,}094\right] - 0{,}0086\cdot I_{s\beta} \\ (0{,}1+0{,}00286\cdot\tau)\cdot\sin\left[0{,}1+0{,}00286\cdot\tau)\cdot\tau + 2{,}094\right] - 0{,}0086\cdot I_{s\lambda} \\ -0{,}01\cdot I_{r\alpha} - 0{,}578\cdot \mathrm{Y}_4\cdot \mathrm{Y}_6 + 0{,}578\cdot \mathrm{Y}_5\cdot \mathrm{Y}_6 \\ -0{,}01\cdot I_{r\beta} - 0{,}578\cdot \mathrm{Y}_5\cdot \mathrm{Y}_6 + 0{,}578\cdot \mathrm{Y}_3\cdot \mathrm{Y}_6 \\ -0{,}01\cdot I_{r\gamma} - 0{,}578\cdot \mathrm{Y}_3\cdot \mathrm{Y}_6 + 0{,}578\cdot \mathrm{Y}_4\cdot \mathrm{Y}_6 \\ 0{,}00205\cdot\left[5{,}99\cdot\left[(I_{s\alpha}\cdot I_{r\gamma} + I_{s\beta}\cdot I_{r\alpha} + I_{s\gamma}\cdot I_{r\beta}) - (I_{s\alpha}\cdot I_{r\beta} + I_{s\beta}\cdot I_{r\gamma} + I_{s\gamma}\cdot I_{r\alpha})\right] - 0{,}00205\cdot(-0{,}3)\right] \end{bmatrix}$$

$$Y_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $Y_0 = \psi_{s\alpha}$; $Y_1 = \psi_{s\beta}$; $Y_2 = \psi_{s\gamma}$; $Y_3 = \psi_{r\alpha}$; $Y_4 = \psi_{r\beta}$; $Y_5 = \psi_{r\gamma}$; $Y_6 = \omega_r$.

$$I_{s\alpha} = 8{,}123 \cdot Y_0 + 2{,}119 \cdot Y_1 + 2{,}119 \cdot Y_2 - 3{,}936 \cdot Y_3 + 1{,}968 \cdot Y_4 + 1{,}968 \cdot Y_5$$
$$I_{s\beta} = 2{,}119 \cdot Y_0 + 8{,}123 \cdot Y_1 + 2{,}119 \cdot Y_2 + 1{,}968 \cdot Y_3 - 3{,}936 \cdot Y_4 + 1{,}968 \cdot Y_5$$
$$I_{s\gamma} = 2{,}119 \cdot Y_0 + 2{,}119 \cdot Y_1 + 8{,}123 \cdot Y_2 + 1{,}968 \cdot Y_3 + 1{,}968 \cdot Y_4 - 3{,}936 \cdot Y_5$$
$$I_{r\alpha} = -3{,}936 \cdot Y_0 + 1{,}968 \cdot Y_1 + 1{,}968 \cdot Y_2 + 7{,}825 \cdot Y_3 + 1{,}828 \cdot Y_4 + 1{,}828 \cdot Y_5$$
$$I_{r\beta} = 1{,}968 \cdot Y_0 - 3{,}936 \cdot Y_1 + 1{,}968 \cdot Y_2 + 1{,}828 \cdot Y_3 + 7{,}825 \cdot Y_4 + 1{,}828 \cdot Y_5$$
$$I_{r\gamma} = 1{,}968 \cdot Y_0 + 1{,}968 \cdot Y_1 - 3{,}936 \cdot Y_2 + 1{,}828 \cdot Y_3 + 1{,}828 \cdot Y_4 + 7{,}825 \cdot Y_5$$

The factors linking the currents with flux linkages are determined from the inverse matrix consisting the machine's parameters, i.e. from the matrix of equation (12):

$$\begin{bmatrix} 3{,}54 & -1{,}73 & -1{,}73 & 3{,}46 & -1{,}73 & -1{,}73 \\ -1{,}73 & 3{,}54 & -1{,}73 & -1{,}73 & 3{,}46 & -1{,}73 \\ -1{,}73 & -1{,}73 & 3{,}54 & -1{,}73 & -1{,}73 & 3{,}46 \\ 3{,}46 & -1{,}73 & -1{,}73 & 3{,}546 & -1{,}73 & -1{,}73 \\ -1{,}73 & 3{,}46 & -1{,}73 & -1{,}73 & 3{,}546 & -1{,}73 \\ -1{,}73 & -1{,}73 & 3{,}46 & -1{,}73 & -1{,}73 & 3{,}546 \end{bmatrix}^{-1}$$

Item 3. Algorithm of study on the three-phase mathematical model of double fed machine, the equations of which are recorded in three-phase coordinate system α_r, β_r, γ_r, rotating with a rotor speed of machine is:

$$D(\tau, Y) = \begin{bmatrix} \sin(Y_6) + 0{,}577 \cdot Y_7 \cdot (Y_1 - Y_2) - 0{,}0086 \cdot I_{s\alpha} \\ \sin(Y_6 - 2{,}09439) + 0{,}577 \cdot Y_7 \cdot (Y_2 - Y_0) - 0{,}0086 \cdot I_{s\beta} \\ \sin(Y_6 + 2{,}09439) + 0{,}577 \cdot Y_7 \cdot (Y_0 - Y_1) - 0{,}0086 \cdot I_{s\lambda} \\ (-0{,}01 - 0{,}000286 \cdot \tau) \cdot \sin[(-0{,}01 - 0{,}00286 \cdot \tau) \cdot \tau] - 0{,}01 \cdot I_{r\alpha} \\ (-0{,}01 - 0{,}00286 \cdot \tau) \cdot \sin[-0{,}01 - 0{,}00286 \cdot \tau) \cdot \tau - 2{,}094] - 0{,}01 \cdot I_{r\beta} \\ (-0{,}01 - 0{,}00286 \cdot \tau) \cdot \sin[-0{,}01 - 0{,}00286 \cdot \tau) \cdot \tau + 2{,}094] - 0{,}01 \cdot I_{r\gamma} \\ 1 - Y_7 \\ 0{,}00205 \cdot [5{,}99 \cdot [(I_{s\alpha} \cdot I_{r\gamma} + I_{s\beta} \cdot I_{r\alpha} + I_{s\gamma} \cdot I_{r\beta}) - (I_{s\alpha} \cdot I_{r\beta} + I_{s\beta} \cdot I_{r\gamma} + I_{s\gamma} \cdot I_{r\alpha})] - 0{,}00205 \cdot (-0{,}3)] \end{bmatrix}$$

Simulation Modelling Of A Sporadic Demand Applying A Bootstraping

Alexej Chovanec, Alena Breznická
•
Faculty of special technology
alexej.chovanec@tnuni.sk
alena.breznicka@tnuni.sk

Abstract

This technique bootraping has been successfully used in various applied statistical problems, although not many applications have been reported in the area of time series. In this paper we present a new application of Bootstrap to time series. A fundamental aspect of supply chain management is accurate demand forecasting. We address the problem of forecasting intermittent (or irregular) demand, i.e. random demand with a large proportion of zero values. Items of spare parts with sporadic consumption can make a significant, up to 60% portion of the value of supplies in service and workshop inventory areas of many industrial segments. An understanding of key features of demand data is important when developing computer systems for forecasting and inventory control.

Keywords: Simulation modelling, sporadic demand, bootstraping, forecasting

I. Introduction

So called driving systems are the most common from advanced approaches in management and optimization of inventory management. They are included into stochastic and dynamic inventory models defined by a random demand. As input random variables are generated data on consumed amount of material gained from a statistical probability distribution. The arithmetic, moving average or a weighted moving means. An exponential averaging of the first and the second degree are used in a trend development of a demand or a linear regression. Auto correlation and identification models are used as well. However arrays of empirical data on a sporadic demand include a random variety of null values with no nulls. It might provide for variable results in defining needed amount in forecast of a mean, a standard deviation or dispersion in a very simple parametric way in mathematical operations. Due to deviousness of input data the distribution of random variable (demand) obviously does not meet standard probability distribution. An applicable option, being used, is a non-parametric method using past data on a sporadic demand called bootstrapping. We classify it among MC simulation statistical methods, based on a stochastic forecast of a future demand from data on a recent demand. Numerous methods of bootstrapping work with random data on demand, from which an experimental pdf, cdf functions of a distribution of random variable (demand) are generated through a computer experiment applicable for assignment of parameters for modelling par a level inventory management.

Bootstraping is a method aiming to increase an accuracy value of statistic estimations. The results are dependent only on bootstrapping samples. We do not need to know the basic distribution of a random variable. Bootstraping creates a large amount of random choices from input data of a bootstraping sample and it calculates improved statistics on each of such choice. In addition to numerical characteristics it provides data for statistical characteristics in form of frequency histograms and choices probability histograms. From a data set being reviewed we generate bootstrapping y random choices several thousands times so that we choose with

repetition (by a substitution of chosen data) from a data set being reviewed x = (x1, x2,.....,xn), a needed amount of m data y = (y1, y2,.....,ym). The chosen numerical values yi are inter independent and they are chosen for a bootstraping sample with the same probability /uniform distribution/. The samples usually differ from each other and they differ from a base data set being reviewed. As we sample with repetitions, it is possible, that some xi data appear several times in a sample or that we do not choose them ever. In case of a specification of a future demand within a delivery term /a delivery leadtime - LT/ we choose from a bootstrapping sample a number of data corresponding with LT.

II. Simulation model of a sporadic demand

The presented simulation model applies stochastic and dynamic principles of inventories modeling. The simulation model was created based on simple algorithms and MATLAB language commands. It consists of two parts: The first part of a model applies the principle of a bootstrapping aiming to define an optimal stock level for an item with a following sequence:

- Downloading the array and a bootstrapping sample of demand data. Fig. 1.
- Computation of numerical characteristics for a bootstrapping sample of a demand for any term, number of data, min, max, mean, std.
- Specification of simulation input data – number of bootstrapping choices, number of chosen periods for a delivery time period, specification of a quantile of a demanded logistic support of delivery.
- Generating a matrix of indices for bootstrapping samples of uniform distribution.
- Transference of indices matrix into a matrix of demand of bootstrapping choices.
- Sum of values in a row of demand matrix of bootstrapping choices.
- Graphic and statistical processing of output data for a definition of a size of an optimal stock.

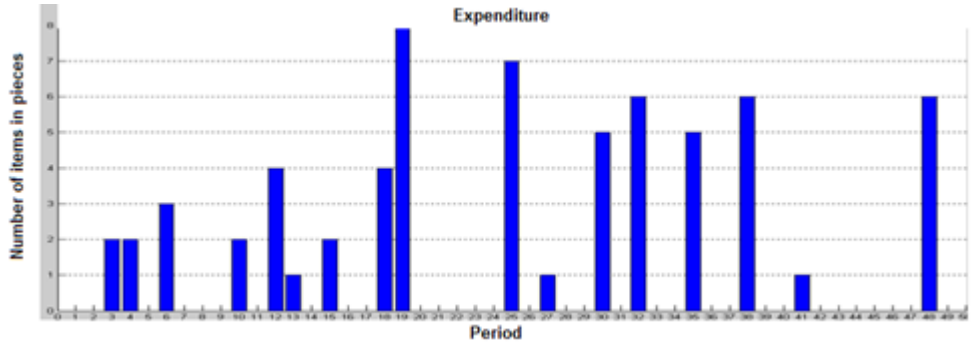


Figure 1: *Sample of demand data for 50 periods*

We simulate choices from a demand sample for 50 time periods. For each choice we randomly choose a number of values corresponding to a lead time. For simplicity LT=2 time periods. Each choice from a sample is represented by a row of a matrix, each column represents a delivery time period. Table1 ... displays a generating of values of bootstrapping indices matrix on 10 choices with a uniform distribution, transformation of the indices matrix into a matrix of bootstrapping choices demand and a sum of values of the rows from a bootstrapping choices demand matrix.

Table 1: *1 Generating 10 choices of indices of an item being reviewed and a demand for LT=2*

Choice number	Index 1	Index 2	expenditure1	expenditure2	SUM expenditure
1	26	12	0	4	4
2	31	46	0	0	0
3	45	34	0	0	0
4	7	37	0	0	0
5	15	35	2	5	7
6	4	40	2	0	2
7	8	27	0	1	1
8	4	38	2	6	8
9	10	26	2	0	2
10	39	46	0	0	0

Number of choices – simulations has an impact on a provision of a same probability that an item index will be chosen by which we assign a demand in pieces Fig. 2. A requirement of an uniform distribution does not become evident at a small number of simulations, amount of choices of indices lines up with an increasing number of simulations and it confirms an algorithm rightness of a procedure for a generator of a uniform distribution from a choice.

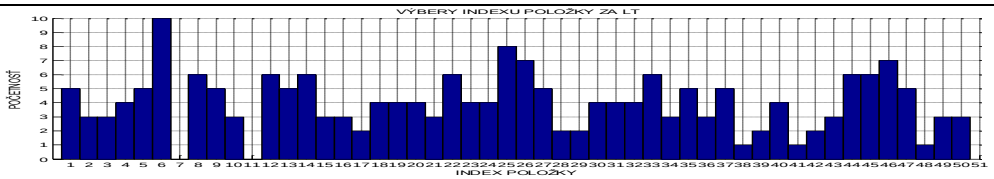
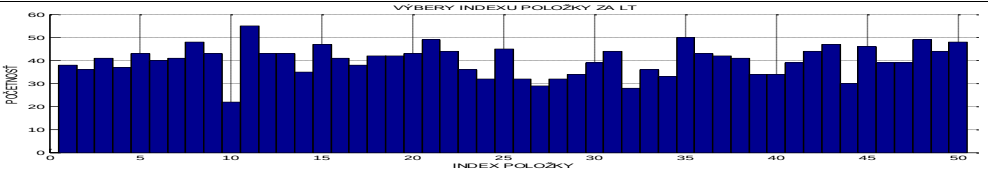
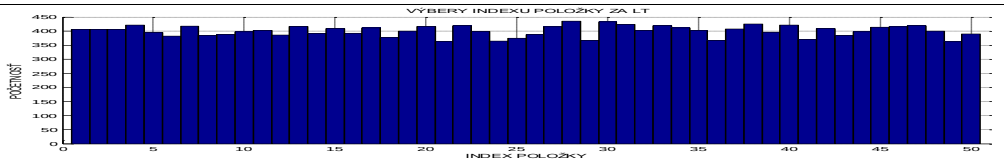
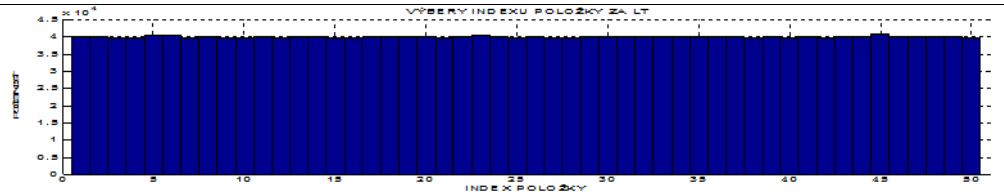
Number of choices	For a delivery time period /LT 2 periods
100	
1000	
10000	
1000000	

Figure 2: *Choice of indices of time periods being modelled and a verification of a uniform distribution*

We develop a histogram from a sum of values from the rows of a demand matrix made based on bootstrapping choices. For 100 choices of the delivery time period 2 periods with a probability 0.95, in the Fig. 3, we see that the intervals with a null demand are represented with the greatest

frequency. Statistics of the set: min: 0, max: 13, mean: 2.63, median: 1, std: 3.25

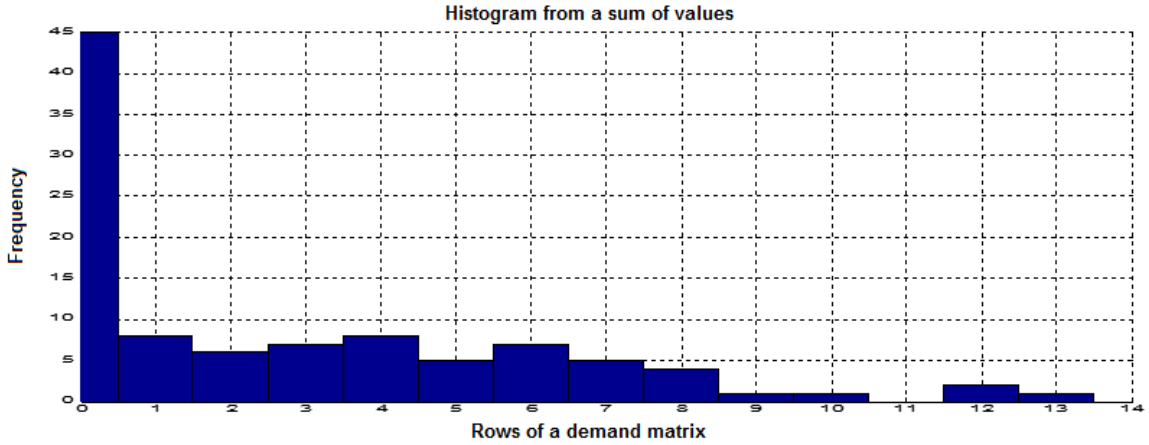


Figure 3: Histogram of a sum of choices frequencies

The second part of a model implements a simulation algorithm of a supplying process with a variable time step.
Characteristics of model parameters:

- Lead Time, a number of time units from sending an order until delivery of an item. A delivery time of an order is defined by contract terms. It may be adjusted by a discharge time of the delivery / a logistic delay.
- Provision probability – Service Level, that a demand will not exceed an offer during implementation with a specified probability. Requested provision probability / Service level is specified ranging from 0.95 - 0.99 by an item criticality.
- Level or stock ordering - Reorder Level is specified as an optimal level with respect to a lead time – is specified as an optimal level with respect to a reorder level and service level. It should ensure that a level of stock during a service level will not drop below zero. Optimal reorder level is specified by a bootstrapping in accordance with a demand forecasting during a lead time of a supplier rounded to the nearest higher ordered amount. Fig. 4. In a moment when a reorder level is intersected, the information system generates an order to a supplier marked with a red asterisk. The above mentioned approach allows a setting of a reorder level and a moment for drawing an offer to refill the stocks in accordance with a specified level of logistic provision.
- Safety stock is created due to an unstable demand / or a lead time as a protection against an item shortage. A safety stock is not created in case of a bootstrapping definition of an optimal stock. A safety factor should be taken into consideration by a Service Level.

Sequence of a simulation algorithm:

- It takes the random data over from bootstrapping choices in order to define a demand of a time period from the first part of the model.
- It monitors a decrease of a stock level / a blue color
- It matches when the stock ordering level reaches a reorder level / green level.

At a moment when a reorder level is reached, or the stock is below the ordering level, it orders an optimal amount of stock, that have been defined in the first part of the model by bootstrapping. Time to draw an order is a random variable.

- It monitors a lead time.
- It carries out a model delivery of an item and it increases a stock level / a red vertical line.
- It collects needed data for computations.

- It computes the costs when a simulation time is shifted.
- It creates graphs of stock and costs courses.
- It repeats a procedure in line with a defined number of time periods in an experiment.

The model allows changing of input values level / delivery time period - LT, number of bootstrapping choices, number of time periods of a demand simulation, a needed level of probability for logistic support, and an initial stock level.
Graphic outputs of a simulation of a short time period are shown in Fig. 4 and Fig. 5.

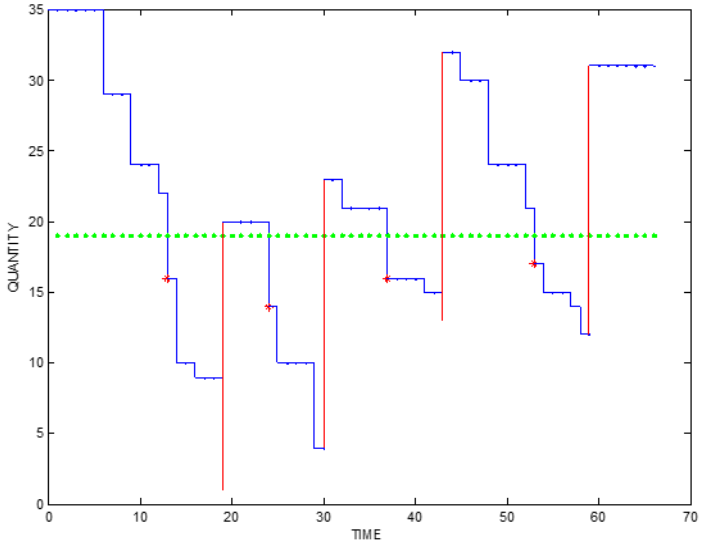


Figure 4: *Course of simulation of stock item movement for 67 time periods at LT=7*

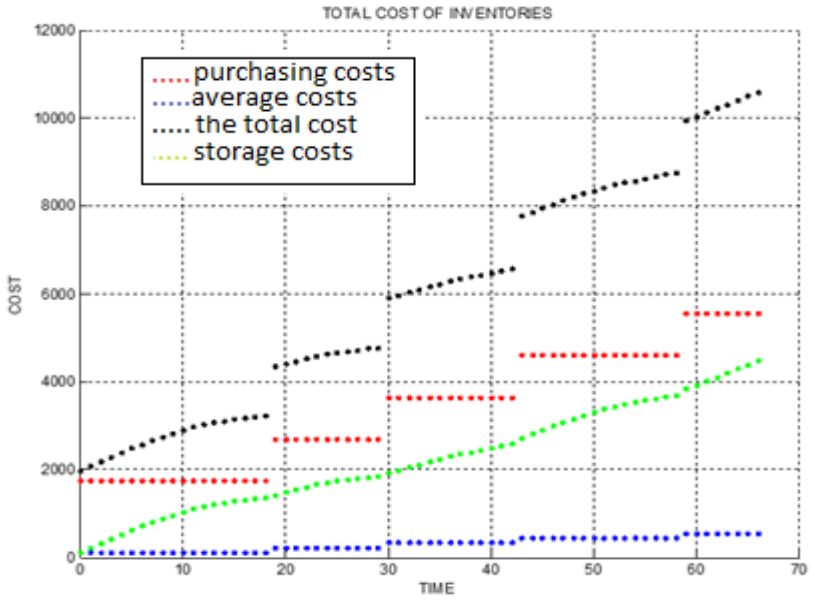


Figure 5: *Course of simulation of stock costs for an item for 67 time periods*

III. Discussion

For evaluation and prediction of the consumption is used arithmetic, moving or weighted moving average. For the trend development of consumption is used exponential equalizing of first and second degree, or linear regression. Auto correlation and identification models are also used. Empirical data arrays of sporadic consumption, however, contain randomly substituting zero values with non-zero. This may, by use of calculations provide variable results for determining the

required amount. The bootstrap distribution of a point estimator of a parameter has been used to produce a bootstrapped confidence interval for the parameter's true value, if the parameter can be written as a function of the distribution. Parameters are estimated with many point estimators. A Bayesian point estimator and a maximum-likelihood estimator have good performance when the sample size is infinite, according to asymptotic theory. For practical problems with finite samples, other estimators may be preferable. Asymptotic theory suggests techniques that often improve the performance of bootstrapped estimators; the bootstrapping of a maximum-likelihood estimator may often be improved using transformations related to pivotal quantities [6]. It is obvious from the above mentioned results, that the increased demand for logistic support of an optimum stock / delivery causes an increased level of an optimum stock /Service level and naturally the costs as well. It is interesting, that costs of acquisition are about on the same level, the transportation costs decrease and storage costs increase.

Acknowledgements
This publication was created in the frame of the project Research of a technological base for draft of application of renewable sources of energy in practice, ITMS code 26220220083, of the Operational Program Research and Development funded from the European Fund of Regional Development.

References

[1] Efron, B.; Tibshirani, R. An Introduction to the Bootstrap. ISBN 0-412-04231-2., 2003, Boca Raton, FL: Chapman & Hall/CRC.

[2] Bradley E. Second Thoughts on the Bootstrap, 2003.

[3] Rever, M., Wunderink, S., Dekker, K., Schorr, B.: Inventory control based on advanced probability theory, an application, 2005, European Journal of Operational Research, ISBN 0377-2217

[4] D iCiccio TJ, Efron B . Bootstrap confidence intervals, 1996 (with Discussion). Statistical Science 11

[5] Weisstein, E. W. Bootstrap Methods,2015, From MathWorld--A Wolfram Web Resource.http://mathworld.wolfram.com/BootstrapMethods.htm

[6] Davison, A. C.; Hinkley, D. V. Bootstrap methods and their application. ISBN 0-521-57391-2, 1997, Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

[7] Raffai, P., Novotný, P., MAršálek, O. Numerical Calculation of Mechanical Losses of the Piston Ring Pack of Internal Combustion Engines. Journal of the Balkan Tribological Association. 2015, no. 4, p. 125–145. ISSN 1310-4772.

[8] Novotný, P., Prokop, A., Zubík, M., Řehák, K. Investigating the influence of computational model complexity on noise and vibration modeling of powertrain. Journal of Vibroengineering. 2016, vol. 22, no. 4, p. 277–392. ISSN 1392-8716.

[9] Raffai, P., Novotný, P., Maršálek, O. Numerical Calculation of Mechanical Losses of the Piston Ring Pack of Internal Combustion Engines. Journal of the Balkan Tribological Association. 2015, no. 4, p. 125–145. ISSN 1310-4772

X-EXPONENTIAL BATHTUB FAILURE RATE MODEL

V.M. Chacko

•

Department of Statistics
St.Thomas College(Autonomous)
Thrissur-680001, Kerala, India
chackovm@gmail.com

Abstract

The properties of x-Exponantial Bathtub shaped failure rate model are discussed. Estimation process and failure rate behavior is explained.

Keywords: Bathtub failure rate

I. Introduction

There are many distributions for modeling lifetime data. Among the known parametric models, the most popular are the Lindley, Gamma, log-Normal, Exponentiated Exponential and the Weibull distributions. These five distributions are suffer from a number of drawbacks. None of them exhibit bathtub shape for their failure rate functions. The distributions exhibit only monotonically increasing, montonically decreasing or constant failure rates. Most real life system exhibit bathtub shapes for their failure rate functions. Generalized Lindley (GL), Generalized Gamma (GG) and Exponentiated Weibull (EW) distributions are proposed for modeling lifetime data having bathtub shaped failue rate model. In this paper we consider a simple model but exhibiting bathtub shaped failure rate, x-Exponential distribution, and discuss the failure rate behavior of these distributions. The x-Exponential distribution has properties similar to Generalized Lindley, but it is more simple and can be used instead of Generalized Lindley, Generalized Gamma and Exponentiated Weibull. The inference procedure also become simple than these distributions.

Section II, discussed x-Exponential distribution and their properties, Generalized Lindley distribution, Generalised Weibull distribution, discussed Generalized Gamma distribution and conclusions are given at the final section.

II. Bathtub shaped failure rate models

I. X-Exponential Distribution

In this section, consider a simplified form of distribution function,

$$F(x) = (1 - (1 + \lambda x^2)e\hat{}(-\lambda x))^{\alpha}, x > 0, \lambda > 0, \alpha > 0. \quad (1)$$

It is an alternative model GL, GG, EW distributions. A life time random variable X has X-Exponential distribution if its cumulative distribution function is (1), [2].

Clearly F(0)=0, F(∞) = 1, F is non-decreasing and right continuous. More over F is absolutely continuous.

The probability density function (pdf) of a x-Exponential random variable X, with scale parameter λ is given by

$$f(x) = \alpha e^{-\lambda x}(\lambda^2 x^2 - 2\lambda x + \lambda)(1 - (1 + \lambda x^2)e\text{^}(-\lambda x))^{\alpha-1}, x > 0, \lambda > 0, \alpha > 0.$$

It is positively skewed distribution. Failure rate function of x-Exponential distribution is

$$h(x) = \frac{\alpha e\text{^}(-\lambda x)\ (\lambda^2 x^2 - 2\lambda x + \lambda)(1 - (1 + \lambda x^2)e\text{^}(-\lambda x))\text{^}(\alpha - 1)}{1 - (1 - (1 + \lambda x^2)e\text{^}(-\lambda x))^{\alpha}}, x > 0, \lambda > 0, \alpha > 0.$$

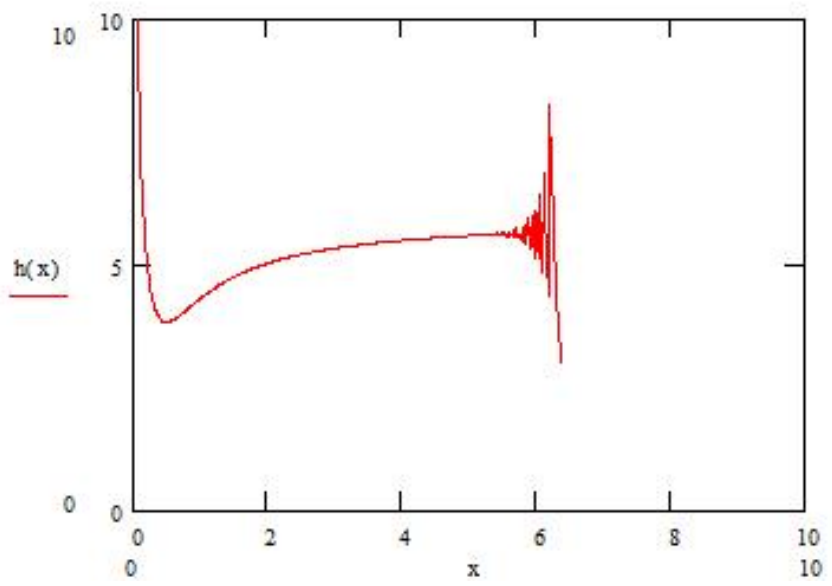


Figure 1. Failure rate function of x-Exponential distribution for **α**=0.01 and λ= 6

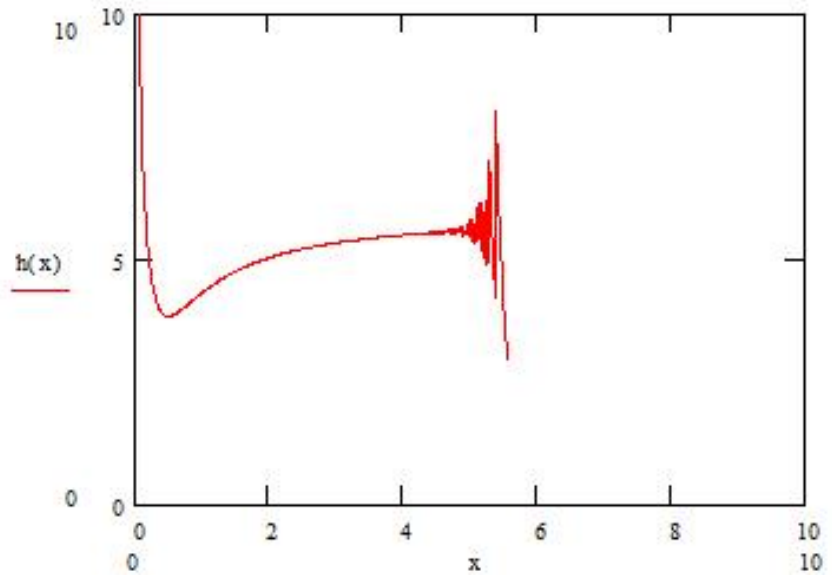


Figure 2. Failure rate function of x-Exponential distribution for **α**=0.0001 and λ= 6

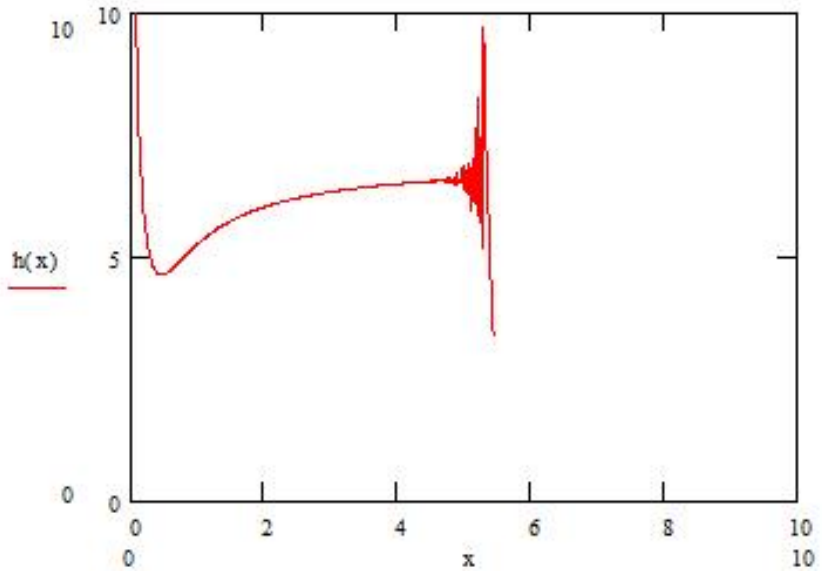


Figure 3. Failure rate function of x-Exponential distribution for **α**=0.001 and λ= 10

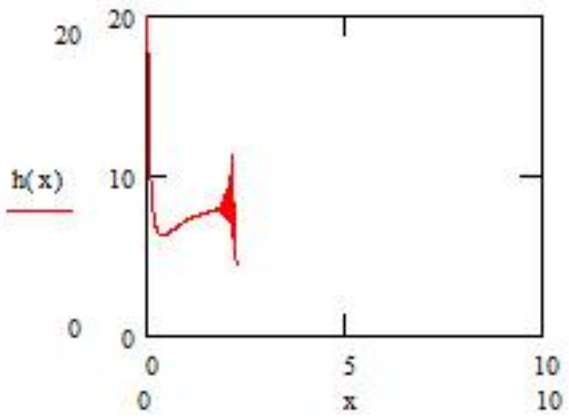


Figure 4. Failure rate function of x-Exponential distribution for **α**=0.000000001 and λ= 9

From Figure 1,2,3 and 4, the shape of the hazard rate function appears monotonically decreasing or to initially decrease and then increase , a bathtub shape if $\alpha < 1$; the shape appears monotonically increasing if $\alpha \geq 1$. So the proposed distribution allows for monotonically decreasing, monotonically increasing and bathtub shapes for its hazard rate function. As **α** decreases from 1 to 0, the graph shift above whereas if λ increases from 1 to ∞ the shape of the graph concentrate near to 0. It is the distribution of the failure of a series system with independent components. The equation (1) has two parameters, α and λ just like the Gamma, log Normal, Weibull and Exponentiated Exponential distributions.

Moments

Calculating moments of X requires the following lemma.

Lemma 2.1: For **α**>0, λ>0, x>0, $K(\boldsymbol{\alpha},\lambda,c)=\int_0^\infty x^c \left[1-(1+\lambda x^2)e^{-\lambda x}\right]^{\alpha-1} e^{-\lambda x}dx$, Then,

$$K(\alpha,\lambda,c)=\sum_{i=0}^{\alpha-1} C_i^{\alpha-1}(-1)^i \sum_{j=0}^{i} C_j^i \lambda^j \left[\int_0^\infty x^{2j+c}\ e^{-i\lambda x}e^{-\lambda x}dx \quad \right]$$

Proof:

We know $(1-z)^{\alpha-1} = \sum_{i=0}^{\alpha-1} C_i^{\alpha-1}(-1)^i z^i$. Therefore

$$K(\boldsymbol{\alpha},\lambda,c)=\int_0^\infty x^c \left[1-(1+\lambda x^2)e^{-\lambda x}\right]^{\alpha-1} e^{-\lambda x}dx$$

$$= \int_0^\infty x^c \sum_{i=0}^{\alpha-1} C_i^{\alpha-1}(-1)^i \left[(1+\lambda x^2)e^{-\lambda x}\right]^i e^{-\lambda x}dx$$

$$= \sum_{i=0}^{\alpha-1} C_i^{\alpha-1}(-1)^i \int_0^\infty x^c \quad \left[(1+\lambda x^2)e^{-\lambda x}\right]^i e^{-\lambda x}dx$$

$$= \sum_{i=0}^{\alpha-1} C_i^{\alpha-1}(-1)^i \int_0^\infty x^c \quad [(1+\lambda x^2)]^i e^{-i\lambda x}e^{-\lambda x}dx$$

$$= \sum_{i=0}^{\alpha-1} C_i^{\alpha-1}(-1)^i \int_0^\infty x^c \quad \sum_{j=0}^{i} C_j^i(\lambda x^2)^j \quad e^{-(i+1)\lambda x} \quad dx$$

$$= \sum_{i=0}^{\alpha-1} C_i^{\alpha-1}(-1)^i \quad \sum_{j=0}^{i} C_j^i \lambda^j \int_0^\infty x^c \quad (x^2)^j \quad e^{-(i+1)\lambda x}dx$$

$$= \sum_{i=0}^{\alpha-1} C_i^{\alpha-1}(-1)^i \quad \sum_{j=0}^{i} C_j^i \lambda^j \int_0^\infty x^c\, x^{2j} \qquad e^{-(i+1)\lambda x}dx$$

$$= \sum_{i=0}^{\alpha-1} C_i^{\alpha-1}(-1)^i \quad \sum_{j=0}^{i} C_j^i \lambda^j \int_0^\infty x^{2j+c} \qquad e^{-(i+1)\lambda x}dx$$

$$= \sum_{i=0}^{\alpha-1} C_i^{\alpha-1}(-1)^i \quad \sum_{j=0}^{i} C_j^i \lambda^j \left[\frac{\Gamma(2j+c+1)}{((i+1)\lambda)^{2j+c+1}} \qquad \right]$$

$$K(\alpha,\lambda,c) = \sum_{i=0}^{\alpha-1} C_i^{\alpha-1}(-1)^i \quad \sum_{j=0}^{i} C_j^i \lambda^j \left[\frac{\Gamma(2j+c+1)}{((i+1)\lambda)^{j+c+1}} \qquad \right]$$

It follows that

$$E(X) = \int_0^\infty x\alpha e^{-\lambda x}(\lambda^2x^2-2\lambda x+\lambda)(1-(1+\lambda x^2)e\hat{}(-\lambda x))^{\alpha-1}dx$$

$$E(X^n) = \int_0^\infty x^n\alpha e^{-\lambda x}(\lambda^2x^2-2\lambda x+\lambda)(1-(1+\lambda x^2)e\hat{}(-\lambda x))^{\alpha-1}dx$$

$$E(X^1) = \alpha\lambda^2 K(\alpha,\lambda,3) - 2\alpha\lambda K(\alpha,\lambda,2) + \boldsymbol{\alpha\lambda} K(\alpha,\lambda,1)$$

The moments are

$$E(X^n) = \alpha\lambda^2 K(\alpha,\lambda,n+2) - 2\alpha\lambda \quad K(\alpha,\lambda,n+1) + \boldsymbol{\alpha\lambda} K(\alpha,\lambda,n),\ n=1,2,3,\ldots$$

Moment Generating Function

$$M_X(t) = \int_0^\infty e^{tx} \quad \alpha(\lambda^2x^2-2\lambda x+\lambda)\left[1-(1+\lambda x^2)e^{-\lambda x}\right]^{\alpha-1} e^{-\lambda x}dx$$

$$M_X(t) = \int_0^\infty \alpha(\lambda^2x^2-2\lambda x+\lambda)\left[1-(1+\lambda x^2)e^{-\lambda x}\right]^{\alpha-1} e^{-(\lambda-t)x}dx$$

$$M_X(t) = \alpha\lambda^2 K(\alpha,\lambda-t,3) - 2\alpha\lambda \quad K(\alpha,\lambda-t,2) + \boldsymbol{\alpha\lambda} K(\alpha,\lambda-t,1)$$

Characteristic Function

$$\Phi_X(t) = \int_0^\infty e^{itx} \quad \alpha(\lambda^2x^2-2\lambda x+\lambda)\left[1-(1+\lambda x^2)e^{-\lambda x}\right]^{\alpha-1} e^{-\lambda x}dx$$

$$\Phi_X(t) = \int_0^\infty \alpha(\lambda^2x^2-2\lambda x+\lambda)\left[1-(1+\lambda x^2)e^{-\lambda x}\right]^{\alpha-1} e^{-(\lambda-it)x}dx$$

$$\Phi_X(t) = \alpha\lambda^2 \mathrm{K}(\alpha, \lambda - \mathrm{it}, 3) - 2\alpha\lambda \ \mathrm{K}(\alpha, \lambda - \mathrm{it}, 2) + \boldsymbol{\alpha\lambda}\mathrm{K}(\alpha, \lambda - \mathrm{it}, 1)$$

Shape of the density function

Consider probability density function,

$$f(x) = \alpha e^{-\lambda x}(\lambda^2 x^2 - 2\lambda x + \lambda)(1 - (1 + \lambda x^2)e\hat{}(-\lambda x))^{\alpha-1}, x > 0, \lambda > 0, \alpha > 0$$

$$logf(x) = \log(\alpha) - \lambda x + log\ (\lambda^2 x^2 - 2\lambda x + \lambda) + \ (\alpha - 1)\ log(1 - (1 + \lambda x^2)e\hat{}(-\lambda x)) \ \ , x, \lambda, \alpha > 0$$

$$\frac{\mathrm{d}}{\mathrm{dx}} log\ f(x) = -\lambda + \frac{(2\lambda^2 x \ \ - 2\lambda)}{(\lambda^2 x^2 - 2\lambda x + \lambda)} + \ \frac{(\alpha - 1)}{(1 - (1 + \lambda x^2)e\hat{}(-\lambda x))}((1 + \lambda x^2)(-\lambda)e^{-\lambda x} \ + 2\lambda x e^{-\lambda x})$$

$$\frac{\mathrm{d}^2}{\mathrm{dx}^2} log\ f(x) = -\frac{1}{x^2} + \ \frac{(\alpha-1)}{(1-(1+\lambda x^2)e^{-\lambda x})^1}((-\lambda)((1 + \lambda x^2)(-\lambda)e^{-\lambda x} \ + 2\lambda x e^{-\lambda x}) \ \ - \frac{(\lambda^2 x^2 - 2\lambda x)^2 e^{-2\lambda x}}{(1-(1+\lambda x^2)e^{-\lambda x})^2}) \ \ , x > 0, \lambda > 0, \alpha > 0.$$

Here f(x) first increases and then decreases, it is unimodel.

Mean Deviation about Mean

The amount of scatter in a population is evidently measured to some extent by the totality of deviations from the mean and median. Mean deviation about the mean defined by

$$\mathrm{MD(Mean)}=2\mu F(\mu) - 2\mu + 2\int_{\mu}^{\infty} xf(x)dx$$

$$MD(Mean) = 2\mu F(\mu) - 2\mu + 2(\alpha\lambda^2 \mathrm{L}(\alpha, \lambda, 3, \mu) - 2\alpha\lambda \ \mathrm{L}(\alpha, \lambda, 2, \mu) + \boldsymbol{\alpha\lambda}\mathrm{L}(\alpha, \lambda, 1, \mu)$$

where $L(\alpha, \lambda, c, \mu) = \int_{\mu}^{\infty} x^c \left[1 - (1 + \lambda x^2)e^{-\lambda x}\right]^{\alpha-1} e^{-\lambda x} dx$

$$= \sum_{i=0}^{\alpha-1} C_i^{\alpha-1}(-1)^i \ \sum_{j=0}^{i} C_j^i \lambda^j \left[\int_{\mu}^{\infty} x^{2j+c+1} e^{-(j+1)\lambda x}\, dx \ \ \right].$$

Mean deviation about the Median defined by

$$\mathrm{MD(Median)}=-M + 2\int_{M}^{\infty} xf(x)dx$$

$$MD(Mean) = -M + 2(\alpha\lambda^2 \mathrm{L}(\alpha, \lambda, 3, \mathrm{M}) - 2\alpha\lambda \ \mathrm{L}(\alpha, \lambda, 2, \mathrm{M}) + \boldsymbol{\alpha\lambda}\mathrm{L}(\alpha, \lambda, 1, \mathrm{M})$$

Estimation

Here, we consider estimation by the methods of moments and maximum likelihood. We also consider estimation issues for censored data. Let $X_1, X_2, \ldots, X_n$ are random sample taken from x-Exponential distribution. Let $m_1 = \frac{1}{n}\sum_{i=1}^{n} x_i \quad m_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$. Equating sample moments to population moments we get moment estimators for parameters.

$$m_1 = \alpha\lambda^2 \mathrm{K}(\alpha, \lambda, 3) - 2\alpha\lambda \mathrm{K}(\alpha, \lambda, 2) + \boldsymbol{\alpha\lambda}\mathrm{K}(\alpha, \lambda, 1)$$

$$m_2 = \alpha\lambda^2 \mathrm{K}(\alpha, \lambda, 4) - 2\alpha\lambda \mathrm{K}(\alpha, \lambda, 3) + \boldsymbol{\alpha\lambda}\mathrm{K}(\alpha, \lambda, 2)$$

The solution of these equations are moment estimators.

To find maximum likelihood estimator, consider likelihood function as,

$$L(\alpha, \lambda) = \prod_{i=1}^{n} f(x_i)$$

$$L(\alpha, \lambda) = \prod_{i=1}^{n} \alpha e^{-\lambda x_i}(\lambda^2 {x_i}^2 - 2\lambda x_i + \lambda)(1 - (1 + \lambda {x_i}^2)e\hat{}(-\lambda x_i))^{\alpha-1}$$

$$L(\alpha, \lambda) = (\alpha)^n \ e^{-\lambda \sum_{i=1}^{n} x_i} \prod_{i=1}^{n}(\lambda^2 {x_i}^2 - 2\lambda x_i + \lambda) \prod_{i=1}^{n} \ (1 - (1 + \lambda {x_i}^2)e\hat{}(-\lambda x_i))^{\alpha-1}$$

$$logL(\alpha, \lambda) = nlog\ (\alpha) \quad - \lambda \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} log(\lambda^2 {x_i}^2 - 2\lambda x_i + \lambda)$$

$$+ \sum_{i=1}^{n} \ (\alpha - 1) log\ (1 - (1 + \lambda {x_i}^2)e\hat{}(-\lambda x_i))$$

$$= nlog\ (\alpha) \quad - \lambda \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} log(\lambda^2 {x_i}^2 - 2\lambda x_i + \lambda) + (\alpha - 1) \sum_{i=1}^{n} log\ (1 - (1 + \lambda {x_i}^2)e\hat{}(-\lambda x_i))$$

$$\frac{\partial}{\partial\alpha}\ log\ L(\alpha,\lambda) = \frac{n}{\alpha} + \sum_{i=1}^{n} log\ (1-(1+\lambda {x_i}^2)e\hat{}(-\lambda x_i))$$

$$\hat{\alpha} = -1/n \sum_{i=1}^{n} log\ (1-(1+\lambda {x_i}^2)e\hat{}(-\lambda x_i))$$

$$\frac{\partial}{\partial\lambda}\ log\ L(\alpha,\lambda) = -\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} \frac{(2\lambda^2 x_i\ \ -2x_i)}{(\lambda^2 {x_i}^2 - 2\lambda x_i + \lambda)} + (\alpha-1)\sum_{i=1}^{n} \frac{1}{(1-(1+\lambda {x_i}^2)e\hat{}(-\lambda x_i))}\ ((1+\lambda {x_i}^2)(-x_i)e^{-\lambda x_i} + {x_i}^2 e^{-\lambda x_i})$$

The MLE of λ will be solution of the following non-linear equation.

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \frac{(2\lambda^2 x_i\ \ -2x_i)}{(\lambda^2 {x_i}^2 - 2\lambda x_i + \lambda)} + (\alpha-1)\sum_{i=1}^{n} \frac{1}{(1-(1+\lambda {x_i}^2)e\hat{}(-\lambda x_i))}\ ((1+\lambda {x_i}^2)(-x_i)e^{-\lambda x_i} + {x_i}^2 e^{-\lambda x_i})$$

II. Generalized Lindley Distribution

Suppose $X_1, X_2, \dots, X_n$ are independent random variables distributed according to Lindley distribution and $T = \min(X_1, X_2, \dots, X_n)$ represent the failure time of the components of a series system, assumed to be independent, [2]. Then the probability that the system will fail before time x is given by

$$F(x) = [1-\ (1+\lambda+\lambda x)/(1+\lambda)\ e\hat{}(-\lambda x)]^n, x>0, \lambda>0.$$

It is the distribution of the failure of a series system with independent components. The cumulative distribution function and pdf of Generalized Lindley distribution are

$$F(x) = [1-\ (1+\lambda+\lambda x)/(1+\lambda)\ e\hat{}(-\lambda x)]^{\alpha}, x>0, \lambda>0, \boldsymbol{\alpha}>0$$

$$f(x) = \frac{\alpha\lambda(1+x)}{1+\lambda}[1-\ (1+\lambda+\lambda x)/(1+\lambda)\ e\hat{}(-\lambda x)]^{\alpha-1} e^{-\lambda x}, x>0, \lambda>0, \boldsymbol{\alpha}>0$$

The equation has two parameters, λ and $\boldsymbol{\alpha}$ just like the Gamma, log Normal, Weibull and Exponentiated Exponential distribution. For n= 1 it reduces to Lindley distribution.

The failure rate function is

$$h(x) = \frac{\frac{\alpha\lambda(1+x)}{1+\lambda}\left[1-\frac{1+\lambda+\lambda x}{1+\lambda}e^{-\lambda x}\right]^{\alpha-1} e^{-\lambda x}}{1-\left[1-\frac{1+\lambda+\lambda x}{1+\lambda}e^{-\lambda x}\right]^{\alpha}}, \quad x>0, \lambda>0, \boldsymbol{\alpha}>0$$

The shape of the failure rate function appears monotonically decreasing or to initially decrease and then increase, a bathtub shape if $\boldsymbol{\alpha}<1$, the shape appears monotonically increasing if $\boldsymbol{\alpha}\geq 1$. So the Generalized Lindley distribution allows for monotonically decreasing, monotonically increasing and bathtub shapes for its failure rate function.

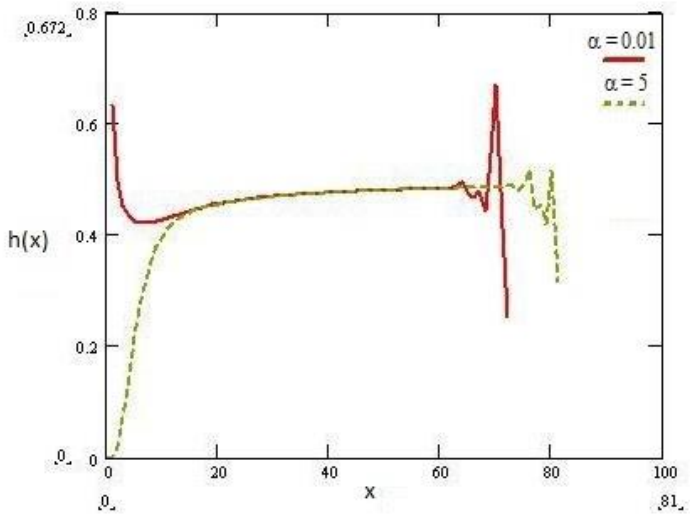


Figure 5. Failure rate function of Generalized Lindley distribution

III. Exponentiated Weibull Distribution

Exponentiated Weibull (EW) distribution has a scale parameter and two shape parameters, [4]. The Weibull family and the Exponentiated Exponential (EE) family are found to be particular cases of this family. The cumulative distribution function of the EW distribution is given by

$$F(x) = \left(1 - e^{-\left(\frac{x}{\beta}\right)^{\alpha}}\right)^{\lambda}, \lambda > 0, \boldsymbol{\alpha} > 0, \beta > 0.$$

Here λ and $\boldsymbol{\alpha}$ denote the shape parameters and β is the scale parameter. For When λ= 1, the distribution reduces to the Weibull Distribution with parameters. When β = 1, $\boldsymbol{\alpha}$=1 it represents the EE family. Thus, EW is a generalization of EE family as well as the Weibull family.
Then the corresponding density function is

$$f(x) = \left(\frac{\alpha\theta}{\sigma}\right)[1 - \exp\{-(x/\sigma)^{\alpha}\}]^{\theta-1}\exp\{-(\frac{x}{\sigma})^{\alpha}\}(\frac{x}{\sigma})^{\alpha-1}, x \geq 0.$$

The failure rate function is

$$h(x) = \frac{\left(\frac{\alpha\theta}{\sigma}\right)\left[1 - exp\left\{-\left(\left[\left(\frac{x}{\sigma}\right)\right]^{\alpha}\right\}\right]^{\theta-1} exp\left\{-\left(\frac{x}{\sigma}\right)^{\alpha}\right\}\left(\frac{x}{\sigma}\right)^{\alpha-1}}{1 - \left[1 - exp\left\{-\left(\left[\left(\frac{x}{\sigma}\right)\right]^{\alpha}\right\}\right]^{\theta}}, x \geq 0, \alpha, \theta, \sigma > 0.$$

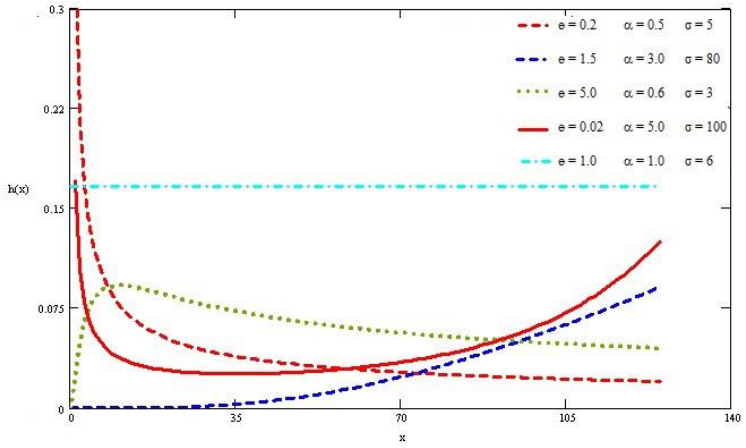


Figure 6: Plot of the failure rate function of EW distribution

The EW distribution is constant for **α**= 1 and = 1. The EW distribution is IFR for **α** > 1 and **θ**≥1. The EW distribution is DFR for **α**< 1 and **θ**≤1. The EW distribution is BT(Bathtub) for **α**> 1 and **θ**< 1. The EW distribution is UBT (Upside down Bathtub) for **α** < 1 and **θ** > 1.

IV. Exponentiated Gamma Distribution

The Gamma distribution is the most popular model for analyzing skewed data and hydrological processes, [3]. This model is flexible enough to accommodate both monotonic as well as non-monotonic failure rates. The Exponentiated Gamma (EG) distribution is one of the important families of distributions in lifetime tests. The EG distribution has been introduced as an alternative to Gamma and Weibull distributions.

The Cumulative Distribution function of the Exponentiated Gamma distribution is given by

$$G(x) = [1 - \exp\{-\lambda x\}(1 + \lambda x)]^{\theta}, x > 0, \lambda, \boldsymbol{\theta} > 0.$$

where λ and **θ** are scale and shape parameters respectively. Then the corresponding probability density function (pdf) is given by

$$g(x) = \theta\lambda^2 x \exp\{-\lambda x\}([1 - \exp\{-\lambda x\}(1 + \lambda x)]^{\theta-1}, x > 0, \lambda, \boldsymbol{\theta} > 0.$$

The failure rate function is

$$h(x) = \frac{\theta\lambda\text{^}2\ x\ exp\{-\lambda x\}([1 - exp\{-\lambda x\}\ (1 + \lambda x)\ ]\text{^}(\theta - 1)}{1 - [1 - exp\{-\lambda x\}\ (1 + \lambda x)\ ]\text{^}\theta}, x > 0, \lambda, \boldsymbol{\theta} > 0.$$

Then the other advantage is that it has various shapes of failure function for different values of . It has increasing failure function when **θ**≥ 1/2 and its failure function takes Bath-tub shape for **θ**<1/2.

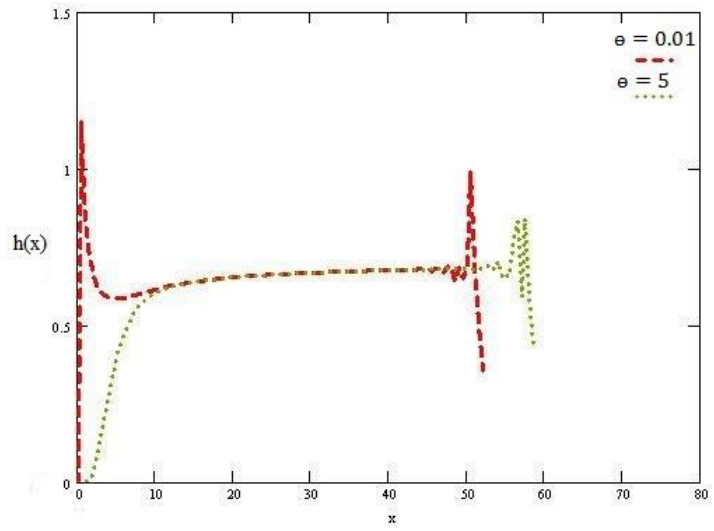


Figure 7: Failure rate function of EG distribution.

III. Generalized X-Exponential Class Distribution

Consider the Distribution function,

$$F(x) = (1 - (\beta + \lambda x^2)e\text{^}(-\lambda x))^{\alpha}, x > 0, \lambda > 0, \alpha > 0, \beta > 0.$$

The failure rate function, provided various Bathtub shaped models as see in Figure 8,9,10. For **α**=0.001, λ=6 and β=5, the failure rate function is

$$h(x) = \frac{\alpha e\text{^}(-\lambda x)\ (\lambda^2 x^2 - 2\lambda x + \lambda)(1 - (\beta + \lambda x^2)e\text{^}(-\lambda x))\text{^}(\alpha - 1)}{1 - (1 - (\beta + \lambda x^2)e\text{^}(-\lambda x))^{\alpha}}, x > 0, \lambda > 0, \alpha > 0, \beta > 0.$$

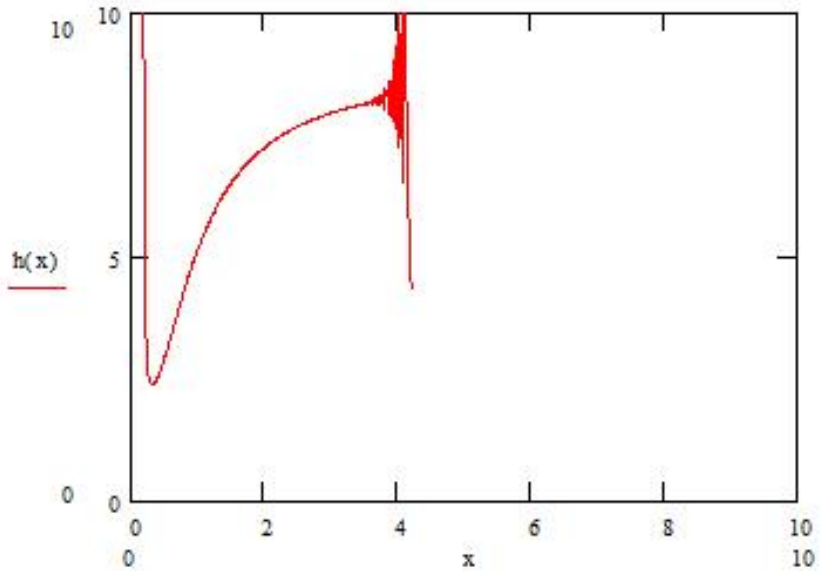


Figure 8. Generalized X-Exponential failure rate function **α**=0.01, λ=9 and β=5

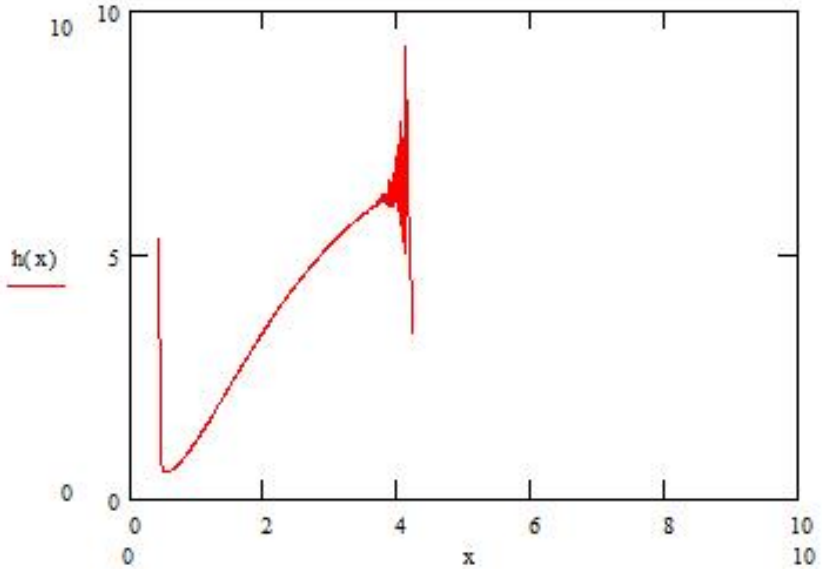


Figure 9. Generalized X-Exponential failure rate function **α**=0.01, λ=9 and β=50

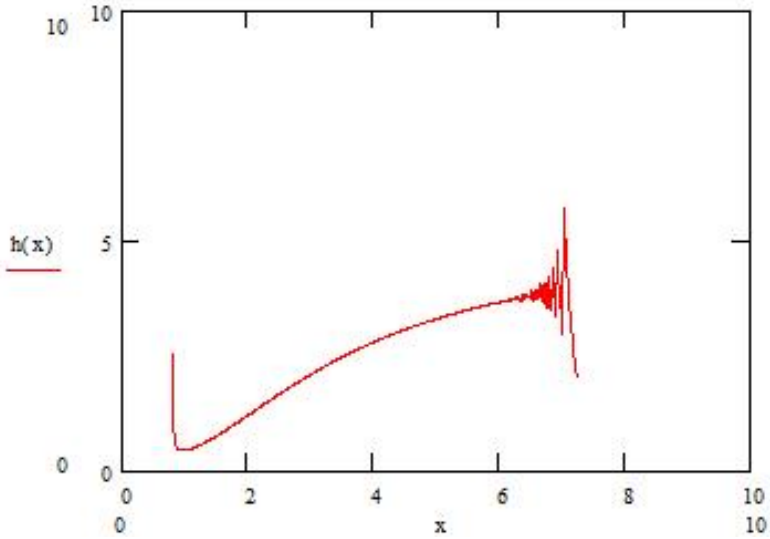


Figure 10. Generalized X-Exponential failure rate function **α**=0.001, λ=5 and β=50

Upside down Bathtub shaped failure rate viewed for $F(x) = (1 - (\beta + \lambda x)e^{\wedge}(-\lambda x))^{\alpha}, x > 0, \lambda > 0, \alpha > 0, \beta > 0$ **α**=0.001, λ=6 and β<1, see figure 11.

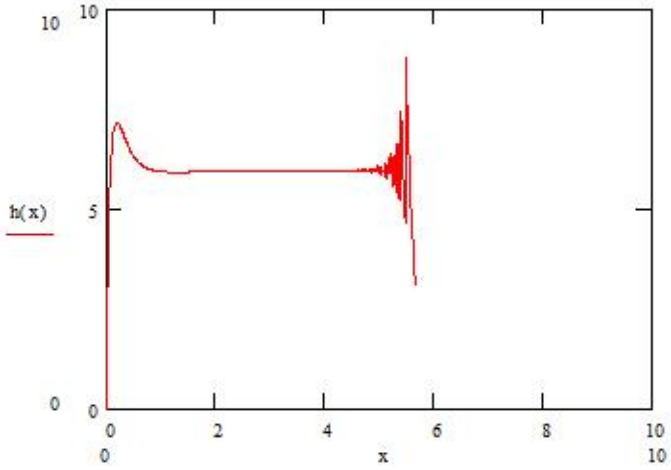


Figure 11. Generalized X-Exponential failure rate function **α**=0.001, λ=6 and β=0.1

All the procedure for finding moments, moment generating function, characteristic function, and estimation are same as that of X-Exponential distribution. If we insert one more parameter **θ** in the model still we get beautiful Bathtub and Upside down bathtub shapes for its failure rate functions as seen below. For $F(x) = (1 - (\beta + \lambda x + \theta x^2)e^{\wedge}(-\lambda x))^{\alpha}, x > 0, \lambda > 0, \alpha > 0, \beta > 0, \theta > 0$.

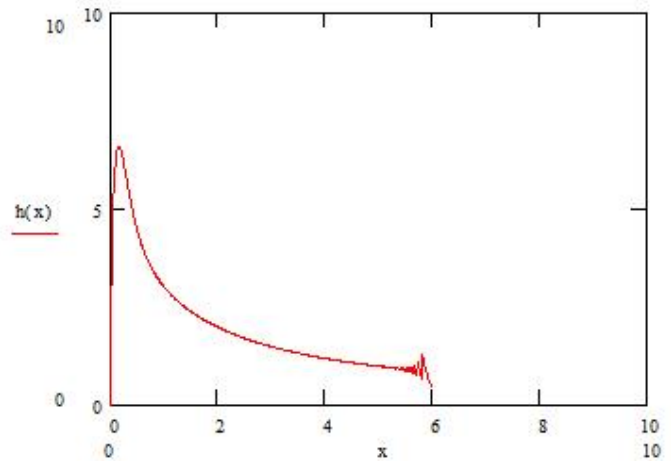


Figure 11. Generalized X-Exponential failure rate function **α**=0.001, λ=6, β=0.1, **θ**=6.

Generalized Lindely distribution is a special case of Generalized X-Exponential distribution.

IV. Conclusions

There are many distributions in reliability which exhibit Bathtub shaped failure rate model, but most of them are complicated in finding the moments, reliability etc. Moreover the increased number of parameters make complication and difficulty in estimation process. The proposed model is similar to Generalized Lindley, so all the computational procedures are like GL distribution. The complication in using GL,GG,GE distributions is reduced in the proposed model. Moreover MLE of **α** is readily available and that of λ can be computed numerically. Generalized X-

Exponential distribution provided various Bathtub shaped and Upside down Bathtub shaped failure rates.

References

[1] Chacko, V.M. (2016). A new bathtub shaped failure rate model, *Reliability: Theory and Applications*, RT&A, No.1(40), Vol.11, March 2016.

[2] Nadarajah, S., Bakouch, H.S., and R. Tahmasbi. 2011. *A generalized Lindley distribution*. Technical Report, School of Mathematics, University of Manchester, UK.

[3] Nadarajah, S., and A.K. Gupta. 2007. The exponentiated gamma distribution with application to drought data. *Calcutta Statistical Association Bulletin* 59:233–234.

[4] Pal, M. M. Ali, J.Woo- Exponentiated Weibull Distribution, Statistica, anno LXVI, n.2,2006.

Metastability Of Large Networks With Mobile Servers

F. Baccelli [1], A. Rybko [2], Senya Shlosman [2,3,4], A. Vladimirov [2]

[1]UT Austin, Department of Mathematics, USA
[2]Institute for Information Transmission Problems, RAS, Moscow
[3]Aix Marseille Université, Université de Toulon
[4]Skolkovo Institute of Science and Technology, Moscow
mailto:@smith.comvladim@iitp.ru

Abstract

We study symmetric queueing networks with moving servers and FIFO service discipline. The mean-field limit dynamics demonstrates unexpected behavior which we attribute to the metastability phenomenon. Large enough finite symmetric networks on regular graphs such as cycles are proved to be transient for arbitrary small inflow rates. However, the limiting non-linear Markov process possesses at least two stationary solutions. The proof of transience is based on martingale technique.[1]

Keywords: ad hoc network, transience, metastability, mean field

I Introduction

In this paper we consider networks with moving servers. The setting is the following: the network is living on a finite or countable graph $G = (V, E)$, at every node $v \in V$ of which one server s is located at any time. For every server, there are two incoming flows of customers: the exogenous customers, who come from the outside, and the transit customers, who come from some other servers. Every customer c coming into the network (through some initial server $s(c)$) is assigned a destination $D(c) \in V$ according to some randomized rule. If a customer c is served by a server located at $v \in V$, then it jumps to a server at the node $v' \in V$, such that $\mathrm{dist}\big(v', D(c)\big) = \mathrm{dist}\big(v, D(c)\big) - 1$, thereby coming closer to its destination. If there are several such v', one is chosen uniformly. There the customer c waits in the FIFO queue until his service starts. If a customer c completes his service by the server located at v, and it so happens that $\mathrm{dist}\big(v, D(c)\big)$ is 1 or 0, the customer is declared to have reached its destination and leaves the network.

The important feature of our model is that the servers of our network are themselves moving over the graph G. Namely, we suppose that any two servers s, s' located at adjacent nodes of G exchange their positions as the alarm clock associated to the edge rings. The time intervals between the rings of each alarm clock are i.i.d. exponential with rate β. When this happens, each of the two servers takes all the customer, waiting in its buffer or being served, to the new location. In particular, it can happen that after such a swap, the distance between the location of the customer c and its destination $D(c)$ increases (at most by one). We assume that the service times of all customers at all servers are i.i.d. exponential with rate 1.

The motivation for this model comes from opportunistic multihop routing in mobile ad hoc wireless networks, see [5, 9, 7, 1, 3, 4]. Within this context, the servers represent mobile wireless devices. Each device moves randomly on the graph G which represents the phase space of device locations. The random swaps represent the random mobile motions on this phase space.

[1] The authors gratefully acknowledge the support of grants 16-29-09497, 14-01-00379, 14-01-00319, 13-01-12410 by Russian Foundation for Sciences.

Each node $v \in G$ of the phase space generates an exogenous traffic (packetized information) with rate λ_v corresponding to the exogenous customers alluded to above. Each such packet has some destination, which is some node of G. In opportunistic routing, each wireless device adopts the following greedy routing policy: any given packet scheduled for wireless transmission is sent to the neighboring node which is the closest to the packet destination. The neighbor condition represents in a simple way the wireless constraints. It implies a multihop route in general. This routing policy is the most natural one to use in view of the lack of knowledge of future random swaps.

In this paper we restrict consideration to cyclic graphs $C_K = \mathbb{Z}^1/K\mathbb{Z}^1$ and their *mean-field* versions, see below. Our main results, however, can be easily extended to much wider classes of networks.

The interest in mean-field versions is both of mathematical and practical nature. The mathematical interest of the mean-field version of a network is well documented. There are also practical motivations for analyzing such networks: their properties are crucial for understanding the long-time behavior of finite size networks.

The results we obtain look somewhat surprising. First of all, we find that for finite graphs the network is transient once the diameter of the graph is large enough. For example, consider the network on the graph C_K with Poisson inflows with rate $\lambda > 0$ at all nodes, exponential service times with rate 1, FIFO discipline and node swap rate $\beta > 0$. Then for all $K \geq K(\lambda, \beta)$ the queues at all servers tend to infinity as time grows. In words this means that the network is unstable for any λ, however small it is – once the network is large enough.

The same picture takes place for mean-field graphs C_K^N with N finite. They consist of N "parallel" copies of C_K such that two nodes in different copies are adjacent if and only if the projections of these nodes to a single copy of C_K are adjacent. However, the limiting picture, for $N = \infty$, is different: the corresponding NLM process on C_K has stationary distributions, provided $0 < \lambda \leq \lambda_{cr}(K, \beta)$, with $\lambda_{cr}(K, \beta) < \infty$ for all $K \leq \infty$. Moreover, for all $\lambda < \lambda_{cr}$ there are at least two different stationary distributions, see Sect. 4 for more details. We demonstrate results of numerical modeling that suggest existence of three equilibria in some cases.

On the other hand, the general convergence result of [2] claims the convergence of the networks on C_K^N to the one on C_K^∞ as $N \to \infty$, which seem to contradict to the statements above. The explanation of this 'contradiction' is that the convergence in [2] holds only on finite time intervals $[0, T]$.

That is, for any T there exists a value $N = N(T)$, such that the network on C_K^N is close to the limiting network on C_K^∞ for all $t \in [0, T]$, provided $N \geq N(T)$. Putting it differently, the C_K^N network behaves like the limiting C_K^∞ network – and might even look as a stationary process – for quite a long time, depending on N, but eventually it departs from such regime and gets into the divergent one. Clearly, the picture we have is an instance of metastable behavior. We believe that more can be said about the metastable phase of our networks, including the formation of critical regions of servers with oversized queues, in the spirit of statistical mechanics, see e.g. [8], but we will not elaborate here on that topic.

II Finite networks

2.1 The C_K network

The only case of a finite network we study here is the cyclic graph $C_K = \mathbb{Z}^1/K\mathbb{Z}^1$. As was mentioned, our main results proved for this graph are easily extendable to much wider classes of networks. We use notation $C_K = (V_K, E)$, where $V_K = \{1, \dots, K\}$ and $E = \{(1,2), \dots, (K-1, K), (K, 1)\}$. For simplicity we take K to be odd.

We study a continuous-time Markov process on a countable state Q, related to the graph C_K. Namely,

$$Q = \{q^v : v \in V_K\} = (V_K^*)^{V_K},$$

where V_K^* is the set of all finite words in the alphabet V_K, including the empty word $\emptyset$.

The queue $q^v \in V_K^*$ at a server located at $v \in V_K$ consists of a finite (≥ 0) number of customers which are ordered by their arrival times (FIFO service discipline) and are marked by their destinations which are vertices of the graph C_K. Since the destination of the customer is its only relevant feature, in our notations we sometime will identify the customers with their destinations.

2.1.1 Dynamics

Let us introduce the continuous-time Markov process $\mathcal{M} \ll \mathcal{M}(t)$ with the state space Q. Let h_v be the length of the queue q^v at node v. We have $q^v = \{q_1^v, \dots, q_{h_v}^v\}$ if $h_v > 0$ and $q^v = \emptyset$ if $h_v = 0$.

The following events may happen in the process $\mathcal{M}$.

An arrival event at node v changes the queue at this node. If the newly arrived customer has for its destination the node w, then the queue changes from q^v to $q^v \oplus w$, that is, to $\{q_1^v, \dots, q_{h_v}^v, w\}$ if $h_v > 0$ or from $\emptyset$ to $\{w\}$ if $h_v = 0$.

In this paper we consider the situation where each exogenous customer acquires its destination at the moment of first arrival to the system, in a translation-invariant manner: the probability to get destination w while arriving to our network at the node v depends only on $w - v$ mod K. The case $w = v$ is not excluded. We thus have the rates $\lambda_{v,w}$, $v, w \in C_K$, and the jump from q^v to $q^v \oplus w$, corresponding to the arrival to v of the exogenous customer with final destination w happens with the rate $\lambda_{v,w}$. We introduce the rate λ of exogenous customers as

$$\lambda = \sum_w \lambda_{v,w} \quad (1)$$

(according to our definitions it does not depend on v).

Each node is equipped with an independent Poisson clock with parameter 1 (the service rate). As it rings, the service of the customer q_1^v is over, provided $h_v > 0$; nothing happens if $h_v = 0$. In the former case the queue at node v changes from q^v to

$$q_-^v = \{q_2^v, \dots, q_{h_v}^v\}$$

(we also define $\emptyset_- = \emptyset$) and immediately one of the two things happen: either the customer q_1^v leaves the network, or it jumps to one of the two neighboring queues, $q^{v\pm1}$. The customer q_1^v leaves the network only if its current position, v, is at distance ≤ 1 from its destination, i.e. iff $q_1^v = v - 1$, v, or $v + 1$. (This is just one of many possible choices we make for simplicity.) Otherwise it jumps to one of the neighboring vertices $w = v \pm 1$, which is the closest to its destination, i.e. to the one which satisfy: $\mathrm{dist}(w, q_1^v) = \mathrm{dist}(v, q_1^v) - 1$ (there is a unique such $w \in V_K$ since we assume K to be odd. The case of even K requires small changes).

The last type of event is the swap of two neighboring servers. Namely, there is an independent Poisson clock at each edge $uv \in E$ of C_K, with rate $\beta > 0$. As it rings, the queues at the vertices u and v swap their positions, that is,

$$q^v(t_+) = q^u(t), \qquad q^u(t_+) = q^v(t).$$

2.1.2 Submartingales

Here we introduce some martingale technique that will be used for the proof of transience of $\mathcal{M}$ for K large enough. To begin with, we label the K servers by the index $k = 1, \dots, K$; this labelling will not change during the evolution. Together with the original continuous-time Markov process $\mathcal{M}(t)$ we will consider the embedded discrete time process $M(n)$, which is the value of $\mathcal{M}(t)$ immediately after the n-th event. The state of the process M consists of the states of all K

servers and all their locations.

The general theorem below will be applied to the quantities X_n^k, which are, roughly speaking, the lengths of the queues at the servers k, $k = 1, \dots, K$, of the process $M(\Lambda n)$. The integer parameter $\Lambda = \Lambda(K, \lambda, \beta)$ will be chosen large enough, so that, in particular, after time Λ, the locations of the servers are well mixed on the graph C_K, and the joint distribution of their location on C_K is close to the uniform one. Moreover, we want the expectations of all the differences $X_{n+1}^k - X_n^k$ to be uniformly positive.

We start with the following theorem.

Theorem 1 *Let $\mathcal{F} = \mathcal{F}_n$, $n = 0,1, \dots$, be a filtration and let X_n^k, $k = 1, \dots, K$, be a finite family of non-negative integer-valued submartingales adapted to $\mathcal{F}$, such that for all $k = 1, \dots, K$, and all $n = 0,1, \dots$, the following assumptions hold:*

(1) For some $\rho > 0$ the inequality

$$\mathbb{E}_{\mathcal{F}_n}(X_{n+1}^k - X_n^k) \geq \rho \tag{2}$$

holds whenever $X_n^k > 0$.

(2) The increments are bounded by a constant R:

$$|X_{n+1}^k - X_n^k| \leq R \quad \text{a. s.} \tag{3}$$

Then there exists an initial state $(X_0^1, \dots, X_0^K)$ such that, with positive probability, $X_n^k \to +\infty$ as $n \to +\infty$ for all $k = 1, \dots, K$.

In order to prove the theorem we begin with an auxiliary lemma.

Lemma 2 *Let $\mathcal{Y}^k = \{Y_n^k : n = 0,1, \dots, k = 1, \dots, K$, be a finite family of submartingales adapted to the same filtration $\mathcal{F}$ and such that $Y_n^k \in 0,1]$ for all k, n. Suppose also that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that*

$$\mathbb{E}(Y_{n+1}^k - Y_n^k) > \delta \quad \text{once} \quad 0 < Y_n^k < 1 - \varepsilon$$

for all k and n. Suppose that the initial vector $Y_0 \in A = [0,1]^K$ is deterministic and satisfies the condition

$$\sum_{k=1}^K Y_0^k > K - 1., \tag{4}$$

Then, with positive probability, $Y_n^k \to 1$ as $n \to \infty$, for all $k = 1, \dots, K$.

Proof. Since all submartingales Y^k are bounded, there is a limit $\lim_{n\to\infty} Y_n^k$ almost surely for all k, see the Martingale Convergence Theorem in [6]. The value of this limit vector with probability 1 is either the 'maximal' vertex $(1, \dots, 1)$ of the cube A or a point a on the 'lower boundary' B of A: $B = \{a : \min_{k=1,\dots,K} a_k = 0\}$. Indeed, for all other vectors $v \in A$, we have

$$\mathbb{E}(Y_{n+1}^k - Y_n^k) > 0 \quad \text{if} \quad Y_n^k = v_k = v, \quad k = 1, \dots, K.$$

Note that

$$\sum_{k=1}^K b_k \leq K - 1 \tag{5}$$

for any vector $b \in B$. By the submartingale property, we conclude that

$$\mathbb{E} \sum_{k=1}^K Y_n^k \geq \sum_{k=1}^K Y_0^k > K - 1 \tag{6}$$

for all $n = 1, \dots$. Inequalities (4)-(6) rule out the option that the limit of Y_n belongs to B with probability 1.

Now, in order to derive Theorem 1 from Lemma 2, we make the following change of variables for submartingales X_n^k. For a positive parameter $\alpha < 1$ we define an 'irregular lattice' $h_i \in \mathbb{R}_+$, by

$$h_0 = 0, \quad h_{i+1} = h_i + \alpha^i, \quad i = 0,1, \dots.$$

We get $\lim_{i\to\infty} h_i = H = (1 - \alpha)^{-1} < \infty$. Now, for each $k = 1, \dots, K$, we define the process Y_n^k on the same filtration $\mathcal{F}$ by the relation

$$Y_n^k(\omega) = h_{X_n^k(\omega)}.$$

The processes Y_n^k take values at the 'lattice' $\{h_i\}$ for $k = 1, \dots, K$. They are still submartingales if $1 - \alpha$ is small enough. Indeed, for such α the local structure of the lattice in an R-neighborhood of a given point is modified only slightly. Since $|X_{n+1}^k - X_n^k| \le R$ and $\mathbb{E}(X_{n+1}^k - X_n^k) \ge \rho > 0$, we conclude that the submartingale property is preserved.

Then the hypothesis of Lemma 2 holds (up to a constant factor H) and Theorem 1 is proved.

2.1.3 Transience

Let us return to the process $\mathcal{M}(t)$. Suppose that the parameters $\lambda > 0$ and $\beta > 0$ are fixed. We remind the reader that our service rate is set to 1.

Theorem 3 *For each $\lambda > 0$ and $\beta > 0$, there exists $K^* \in \mathbb{Z}_+$ such that for any $K \ge K^*$ the process $\mathcal{M}$ is transient.*

Proof. First of all, we construct a discrete time Markov chain $\mathcal{D}$ on the state space Q. To define it, we start with the embedded Markov chain $M(n)$, defined earlier, and then pass to the chain $M(\Lambda n)$, with the integer Λ to be specified later. To get the chain $\mathcal{D} \ll \{\mathcal{D}_n\}$, we modify the chain $M(\Lambda n)$ as follows: if for some n at least one of the K queues is at most Λ, we add to all such queues extra customers, to make these queues to be of length exactly Λ and then stop the process forever. Otherwise we do no changes. The obtained Markov chain is denoted by $\mathcal{D}$.

We start the process $\mathcal{D}$ at some configuration Q_0 with all queues longer than Λ.

We now prove the following statement: if Λ is large enough, the queue length process $\mathcal{D}$ at any given server is a submartingale satisfying the conditions of Theorem 1, with respect to the filtration defined by our discrete-time Markov chain $M(n)$ (the individual queue length processes are clearly adapted to this filtration). This completes the proof because of Theorem 1. We need the following lemmas.

Lemma 4 *(1) Let us consider the following function $\pi(t)$ of the process $\mathcal{M}$. At each $t \ge 0$, $\pi(t)$ is the current permutation of indices of K servers with respect to indices of K nodes. Then the evolution of $\pi(t)$ is a continuous time Markov process, independent of service and arrival processes, and, as $t \to \infty$, the distribution of $\pi(t)$ converges to the uniform one on the set S_K of all permutations.*

(2) Let us fix index i and denote by $v(i,t)$ the position of the server i at time t. Then the distribution of $v(i,t)$ converges to the uniform distribution on $\{1, \dots, K\}$ as $t \to \infty$.

Proof. Let us introduce the graph structure on the permutation group S_K. Namely, we consider all the transpositions $\tau \in S_K$ corresponding to the exchanges of pairs of neigboring servers, and we call two permutations π', π'' to be connected by an edge iff $\pi' = \pi''\tau$ for some τ.

The resulting graph on S_K is connected – because G is connected. The process of migration of servers is, obviously, a random walk on this graph, that is, a reversible process. Hence, as $t \to \infty$, the distribution of permutations converges to the uniform one uniformly on all initial states. The assertion of the lemma clearly follows.

Lemma 5 *For any initial state Q_0, the probability of a customer with position $H > 0$ in the queue to leave the network after being served, tends to $3/K$ as $H \to \infty$, uniformly in Q_0.*

Proof. As the waiting time of the customer tends to infinity with $H \to \infty$, the distribution of its server on V_K tends to the uniform one on C_K (see Lemma 4). In order for the customer c to exit the network, the last server of c has to be located at this moment at one of the three nodes: $D(c) + 1$, $D(c)$, or $D(c) - 1$. The lemma follows.

Now we see that for all the customers in the initial queues whose positions are at least H, the mean chance of exit approaches $3/K$ as $H \to \infty$, and the rate of this approach does not depend on the particularities of the initial state Q_0, but only on H.

The next remark is that if a customer is served and then jumps to a different server, then the index j of that server is distributed almost uniformly over the remaining $K-1$ indices. This fact follows from Lemma 5. Again, the rate of convergence is independent of Q_0 because the servers swap positions independently of anything else. So we have established a lemma, analogous to Lemma 5:

Lemma 6 *The probability of a customer with position H on server i to jump to server j tends to $1/(K-1)$ as $H \to \infty$ uniformly in i, j, and in the initial states $Q_0 \in Q$.*

We need a third combinatorial lemma, and we start with some definitions, and then formulate and prove it. Let $\{u, v\} \subset C_K$ be an ordered pair of elements. We define the map T from the set of all such pairs into the union $V_K \cup \{*\}$, by

$$T\{u,v\} = \begin{cases} w & \text{for} w \text{defined by} |u-w| = 1,\ |v-w| = |u-v| - 1, \\ & \text{provided} |u-v| > 1, \\ * & \text{otherwise.} \end{cases}$$

For K odd the map T is well-defined. In case $T\{u,v\} = w$ we say that a customer transits through w (on his way from u to v).

Let $D: V_K \to V_K$ be an arbitrary map. We want to compute the quantity

$$p_K = \frac{1}{K!} \sum_{\pi \in S_K, i \in V_K} \mathbb{I}_{\{T\{\pi(i), D(i)\} = \pi(j)\}}, \tag{7}$$

where S_K is the symmetric group, π runs over all permutations from S_N, while i and j are taken from some fixed labelling of the elements of V_K. Thus p_K is the probability of transit through the node $\pi(j)$ in the ensemble defined by the uniform distribution on S_K. Of course, it does not depend on j. (Note that we consider the action of S_K on pairs $\{u, v\}$ given by $\pi\{u, v\} = \{\pi u, v\}$.)

Lemma 7

$$p_K = \frac{K-3}{K}.$$

Proof of the Lemma. Let $1,2,\dots,K$ be the labelling fixed; without loss of genelarily we can take $j = 1$. Instead of performing the summation in (7) over whole group S_K, we partition S_K into $(K-2)!$ subsets A_π, and perform the summation over each A_π separately. If the result will not depent on π, we are done. Here $\pi \in S_K$, and, needless to say, for π, π' different we have either $A_\pi = A_{\pi'}$ or $A_\pi \cap A_{\pi'} = \emptyset$.

Let us describe the elements of the partition $\{A_\pi\}$. So let π is given, and the string $i_1, i_2, i_3, \dots, i_l, i_{l+1}, \dots, i_K$ is the result of applying the permutation π to the string $1,2,\dots,K$. Then we include into A_π the permutation π, and also $K-1$ other permutations, which correspond to the cyclic permutations, e.g. we add to A_π the strings $i_K, i_1, i_2, i_3, \dots, i_l, i_{l+1}, \dots$, $i_{K-1}, i_K, i_1, i_2, i_3, \dots, i_l, i_{l+1}, \dots$, and so on. We call these transformations 'cyclic moves'. Now with each of K permutations already listed we include into A_π also $K-2$ other permutations, where the element i_1 does not move, and the rest of the elements is permuted cyclically, i.e., for example from $i_{K-1}, i_K, i_1, i_2, i_3, \dots, i_l, i_{l+1}, \dots$, we get $i_K, i_2, i_1, i_3, \dots, i_l, i_{l+1}, \dots, i_{K-1}$, $i_2, i_3, i_1, \dots, i_l, i_{l+1}, \dots, i_{K-1}, i_K$, and so on. We call these transformations 'restricted cyclic moves'. The main property of thus defined classes of configurations is the following: Let $a \neq b \in \{1,2,\dots,K\}$ be two arbitrary indices, and $l \in \{2,\dots,K\}$ be an arbitrary index, different from 1. Then in every class A_π there exists exactly one permutation π', for which $i_1 = a$ and $i_l = b$.

Given π, take the customer $l \neq 1 (= j)$, and its destination, $D(l)$. If we already know the position i_1 of customer 1 on the circle C_K, then in the class A_π there are exactly $K-1$ elements, each

of them corresponds to a different position of the server l on C_K. If it so happens that $i_1 = D(l)$, then for no position of the server l the transit from l through i_1 happens. The same also holds if $i_1 = \left(D(l) + \frac{K-1}{2}\right) mod K$ or $i_1 = \left(D(l) + \frac{K+1}{2}\right) mod K$. For all other $K-3$ values of i_1 the transit from l through i_1 happens precisely for one position of l (among $K-1$ possibilities). Totally, within A_π we have $(K-1)(K-3)$ transit events. Since $|A_\pi| = K(K-1)$, the lemma follows. +

End of the proof of the theorem. Now we define the submartingales X_n^k and show that they satisfy all the properties of Theorem 1. We define X_n^k to be the length of the queue of the k-th server in the process $\mathcal{D}_n$, from which the constant Λ is subtracted. Clearly, $X_n^k \geq 0$. We now show that if K and Λ are both suitably large, then the properties (1) and (2) of Theorem 1 hold.

Relation (3) is evidently satisfied with $R = \Lambda$. Let us check (2). Let us start the process M at a configuration where all the queue lengths are of the form $X_0^k + \Lambda$ with $X_0^k > 0$, $k = 1, \dots, K$. We want to show that after time Λ, we have $\mathbb{E}(X_1^k - X_0^k) \geq \rho$, for some $\rho > 0$. Let $H = H(K)$ be the time after which the distribution of the K servers is almost uniform on C_K, see Lemma 5. Before this moment, we do not know much about our network, so we bound the lengths of the queues M_H^k roughly, by $M_H^k \geq M_0^k - H$. After the time H the probability that a customer leaving a server leaves the network is almost $1/K$, and the probabilities that it jumps to the left or the right are both close to $\frac{K-1}{2K}$.

More precisely, by Lemma 7, the rate of arrival to every server after time H is almost $\lambda + (K-3)/K$, which is higher than the exit rate, 1, provided K is large enough (namely, $K > K^* = 3/\lambda$). Hence the expected queue lengths in the process M grow linearly in time, at least after time H, which implies the existence of $\Lambda > 0$ such that $\mathbb{E}\left(M_\Lambda^k\right) \geq M_0^k + \rho$. So, Theorem 1 applies.

III Infinite networks

3.1 NLMP on $\mathbb{Z}^1$

In this section we consider the limit of the network $(\mathbb{Z}^1)^N$ as $N \to \infty$, i.e. the NLMP on $\mathbb{Z}^1$. The limit of the network C_K^N can be studied in the same way. This NLMP is described in details in [2], and we use the notations therein. Here we are interested in its stationary distributions.

The NLMP is the evolution of the measure $\otimes \mu_v$ on the states (queues q_v) of the (jumping) servers at the nodes $v \in \mathbb{Z}^1$, given by the equations

$$\frac{d}{dt}\mu_v(q_v, t) = \mathcal{A} + \mathcal{B} + \mathcal{C} + \mathcal{D} + \mathcal{E} \tag{8}$$

with

$$\mathcal{A} = -\frac{d}{dr_{i^*(q_v)}(q_v)}\mu_v(q_v, t) \tag{9}$$

being the derivative along the direction $r(q_v)$ (in our case of the exponential service time with rate 1 we have, of course, that $\frac{d}{dr_{i^*(q_v)}(q_v)}\mu_v(q_v, t) = \mu_v(q_v, t)$)

$$\mathcal{B} = \delta\left(0, \tau\left(e(q_v)\right)\right)\mu_v(q_v!\, e(q_v), t)[\sigma_{tr}(q_v!\, e(q_v), q_v) + \sigma_e(q_v!\, e(q_v), q_v)] \tag{10}$$

where q_v is created from $q_v!\, e(q_v)$ by the arrival of $e(q_v)$ from v', and $\delta\left(0, \tau\left(e(q_v)\right)\right)$ takes into account the fact that if the last customer $e(q_v)$ has already received some amount of service, then he cannot arrive from the outside;

$$\mathcal{C} = -\mu_v(q_v, t)\sum_{q_v'}[\sigma_{tr}(q_v, q_v') + \sigma_e(q_v, q_v')], \tag{11}$$

which corresponds to changes in queue q_v due to customers arriving from other servers and from the outside (in the notations of (1), $\sigma_e(q_v, q^v \oplus w) = \lambda_{v,w}$);

$$\mathcal{D} = \int_{q_v' : q_v'!\, C(q_v') = q_v} d\mu_v(q_v', t)\sigma_f\left(q_v', q_v'!\, C(q_v')\right) - \mu_v(q_v, t)\sigma_f\left(q_v, q_v!\, C(q_v)\right), \tag{12}$$

where the first term describes the situation where the queue q_v arises after a customer was served

in a queue q'_v (longer by one customer), and $q'_v!\,C(q'_v) = q_v$, while the second term describes the completion of service of a customer in q_v;

$$\mathcal{E} = \sum_{v'\,\text{n.n.}\,v} \beta_{vv'}[\mu_{v'}(q_v, t) - \mu_v(q_v, t)], \tag{13}$$

where the β-s are the rates of exchange of the servers.

For the convenience of the reader we repeat the equation $(8-813)$ once more:

$$\begin{aligned}
&\tfrac{d}{dt}\mu_v(q_v, t) = -\tfrac{d}{dr_{i^*(q_v)}(q_v)}\mu_v(q_v, t)\\
&+\delta\left(0, \tau\big(e(q_v)\big)\right)\mu_v\big(q_v!\,e(q_v)\big)[\sigma_{tr}(q_v!\,e(q_v), q_v) + \sigma_e(q_v!\,e(q_v), q_v)]\\
&-\mu_v(q_v, t)\textstyle\sum_{q'_v}[\sigma_{tr}(q_v, q'_v) + \sigma_e(q_v, q'_v)] + \int_{q'_v: q'_v!C(q'_v)=q_v} d\mu_v(q'_v)\sigma_f\big(q'_v, q'_v!\,C(q'_v)\big)\\
&-\mu_v(q_v)\sigma_f\big(q_v, q_v!\,C(q_v)\big) + \textstyle\sum_{v'\,\text{n.n.}\,v}\beta_{vv'}[\mu_{v'}(q_v) - \mu_v(q_v)].
\end{aligned} \tag{14}$$

We are looking for the fixed points μ of the evolution (14). Then the measures δ_μ (on measures) will be stationary measures of our NLMP. Note that the dynamical system (14) might have other stationary measures (on measures) then those corresponding to the fixed points. We will simplify our setting. Namely, we make the following changes:

1. for the graph G we take the lattice $\mathbb{Z}^1$;

2. all the customers have the same class;

3. the service time distribution η is exponential, with the mean value 1;

4. the service discipline considered is FIFO;

5. the exogenous customer c arriving to the node v has for its destination the same node v, i.e. $D(v) = v$; inflow rates at all the nodes are constant, equal to λ;

6. the two servers at v, v', which are neighbors in $\mathbb{Z}^1$ can exchange their positions with the same rate $\beta \equiv \beta_{vv'}$;

The queue q_v can in this setting be identified with the sequence of destinations $D(c_i)$ of its customers. The equation for the fixed point then becomes:

$$\begin{aligned}
&0 = \mu_v\big(q_v!\,e(q_v)\big)[\sigma_{tr}(q_v!\,e(q_v), q_v) + \sigma_e(q_v!\,e(q_v), q_v)]\\
&-\mu_v(q_v)\textstyle\sum_{q'_v}[\sigma_{tr}(q_v, q'_v) + \lambda] + \sum_{q'_v: q'_v!C(q'_v)=q_v}\mu_v(q'_v)\\
&-\mu_v(q_v)\mathbb{I}_{q_v\neq\emptyset} + \textstyle\sum_{v'=v\pm1}\beta[\mu_{v'}(q_v) - \mu_v(q_v)].
\end{aligned}$$

We are interested in translation-invariant solutions. In that case the queue q_v can be identified with the sequence of (signed) distances between the node v and the destinations $D(c_i)$ of its customers, so it becomes a finite integer sequence $\mathcal{N} \equiv \{n_1, \ldots, n_l; n_i \in \mathbb{Z}^1\}$, where $l \geq 0$ is the length of the queue q_v. The rate of the arrival of the transit customer, $\sigma_{tr}(q_v!\,e(q_v), q_v) \equiv \sigma_{tr}([n_1, \ldots, n_{l-1}], [n_1, \ldots, , n_{l-1}, n_l])$ is then a function of one integer, n_l, and so we adopt the notation

$$\lambda_{n_l} \equiv \sigma_{tr}([n_1, \ldots, n_{l-1}], [n_1, \ldots, , n_{l-1}, n_l]).$$

According to our definitions, we thus have

$$\lambda_k \equiv \lambda_k(\mu) = \begin{cases} \sum_{\mathcal{N}}\mu(k+1, \mathcal{N}) & \text{if} k > 0,\\ \sum_{\mathcal{N}}\mu(k-1, \mathcal{N}) & \text{if} k < 0,\\ 0 & \text{if} k = 0. \end{cases} \tag{15}$$

In what follows we look only for states μ which have symmetric rates λ_k :

$$\lambda_k = \lambda_{-k}. \tag{16}$$

The probability $\mu_v(q_v)$ then turns into $\mu(\mathcal{N})$; note, however, that for $v' = v \pm 1$ we need to interpret $\mu_{v'}(q_v)$ as $\mu(n_1 \mp 1, \ldots, n_l \mp 1)$. The equation now becomes:

$$\mu(n_1, \ldots, n_{l-1})\big[\lambda_{n_l} + \lambda\delta(n_l, 0)\big] - \mu(n_1, \ldots, n_l)(\textstyle\sum_k \lambda_k + \lambda)$$

$$+\sum_k \mu(k,n_1,\dots,n_l)-\mu(n_1,\dots,n_l)\mathbb{I}_{l\neq 0} \qquad (17)$$
$$+\beta[\mu(n_1+1,\dots,n_l+1)+\mu(n_1-1,\dots,n_l-1)-2\mu(n_1,\dots,n_l)]=0.$$

As we see later, the equations $(15)-(1517)$ can have several solutions, one solution or no solution, depending on the value of the parameter λ. If μ is a solution of the equations $(15)-(1517)$ for some λ, then we denote by

$$\nu(\mu)=\sum_k \lambda_k(\mu)$$

the rate of the transit customers to every node in the state μ, and by $\eta(\mu)$ the rate of the total flow to every node in the state μ:

$$\eta(\mu)=\nu(\mu)+\lambda.$$

Theorem 8 *For every positive $\eta<1$ there exist a unique value $\lambda(\eta)$ of the exogenous flow rate λ and the state μ_η on the set of the queues $\{\mathcal{N}\}$, satisfying the equations $(15)-(1517)$ with $\lambda=\lambda(\eta)$, such that*

$$\eta(\mu_\eta)=\eta.$$

Proof. Consider the process which is described just by the relation (17), with arbitrary parameters λ_k, $k=0,\pm1,\dots$ and λ. This is an ordinary queuing system with a single server and with infinitely many types of customers. The customer of type k arrives with rate λ_k (and with rate $\lambda_0+\lambda$ for $k=0$). Consider the random variable ξ_η, which is the total time a customer spends in such a server in the stationary state. It has exponential distribution, which depends only on $\eta=\sum_k \lambda_k+\lambda$ (and which does not depend on the type of the customer), namely, $\mathbb{E}(\xi_\eta)=(1-\eta)^{-1}$.

Suppose a customer of type k arrives to such a server. When it leaves the server, its type is changed to $k+\tau_\eta$, where τ_η is a random (integer valued) variable. That change happens due to the β-terms in (17). By symmetry, $\mathbb{E}(\tau_\eta)=0$. The distribution of τ_η is the following. Consider a random walker $W(t)$, living on $\mathbb{Z}^1$, which starts at 0 – i.e. $W(0)=0$, and which makes ±1 jumps with rates β. Then $\tau_\eta=W(\xi_\eta)$.

We now are going to present a choice of the rates λ_k and λ in such a way that the equations (15) is satisfied as well. Our choice of the rates λ_k,λ is related to the stationary distribution of a certain ergodic Markov process on $\mathbb{Z}^1$, which we describe now. Define the matrix of transition probabilities $P_1=\{\pi_{st}\}$ by $\pi_{st}=\Pr(\tau_\eta=s-t)$. Of course, this Markov chain on $\mathbb{Z}^1$ is not positive recurrent since its mean drift is zero. Let P_2 be the second Markov chain, with transition probabilities

$$\rho_{st}=\begin{cases}1 & \text{for} t>0,\ s=t+1,\\ 1 & \text{for} t=0,\ s=1,0,-1,\\ 1 & \text{for} t<0,\ s=t-1,\\ 0 & \text{in other cases.}\end{cases}$$

I.e. P_2 is non-random map of $\mathbb{Z}_1$ into itself. Consider the composition Markov chain, with transition matrix Q being the product, $Q=P_1P_2$. This chain is, in contrast, positive recurrent (it has a positive drift towards the origin), and it has a stationary state $q=\{q_k,k\in\mathbb{Z}^1\}$. We take

$$\lambda_k=\eta q_k,\ k\neq 0;\ \lambda=\eta q_0. \qquad (18)$$

The relations (15) are satisfied since the process Q describes the evolution of the type of the customer in the stationary state of the process $(15)-(1517)$.

We now state some properties of the function $\lambda(\eta)$ as the parameter η varies in (0,1).

Proposition 9 *There is a $\lambda_+>0$ such that, for any positive $\lambda<\lambda_+$, there are at least two different velues $\eta=\eta_-(\lambda)$ and $\eta=\eta_+(\lambda)$ satisfying the relation $\lambda(\eta)=\lambda$ and such that $\eta_-(\lambda)\to 0$ and $\eta_+(\lambda)\to 1$ as $\lambda\to 0$.*

Proof. Clearly, $\lambda(\eta)\to 0$ as $\eta\to 0$. We want to argue that $\lambda(\eta)\to 0$ also when $\eta\to 1$. Indeed, in this regime every customer spends more and more time waiting in the queue, so for every k the

probability $\Pr(\xi_\eta \le k) \to 0$ as $\eta \to 1$. Therefore the distribution of the random variable τ_η becomes more and more spread out: for every k, $\Pr(|\tau|_\eta \le k) \to 0$ as $\eta \to 1$. Therefore the same property holds for the stationary distribution q, and the claim follows from relations (18) and Proposition 9. In particular, it means that the following equation on η:

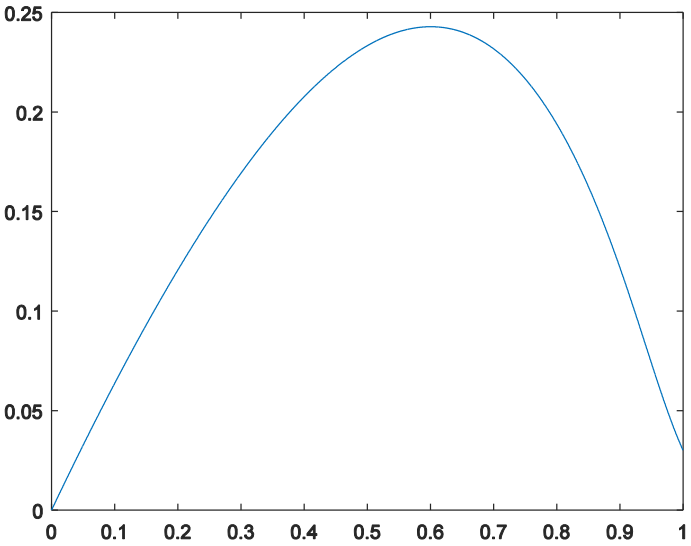
$$\lambda(\eta) = a > 0$$

has at least two solutions for small a: the corresponding η is either small or close to 1. Indeed, this follows from the continuity of $\lambda(\eta)$.

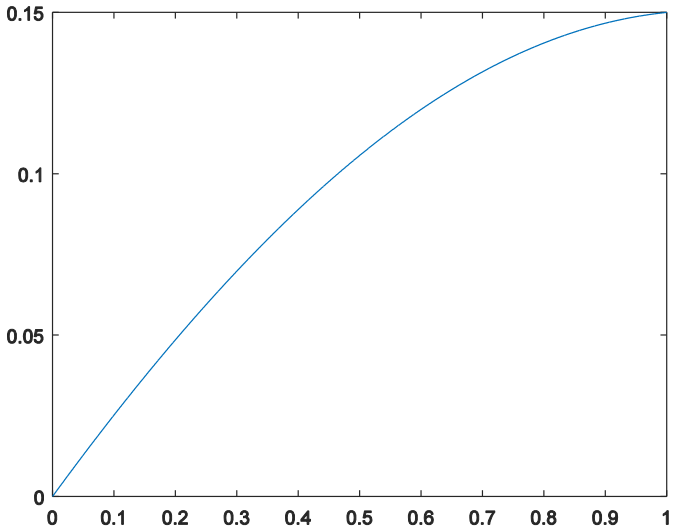
IV Some examples

Let us look at the function $\lambda(\eta)$ for some finite cyclic graphs C_K with different parameters, that is, with different randomized rules for the destination assignment. As we see, there are cases with one, two, and three equilibrium solutions.

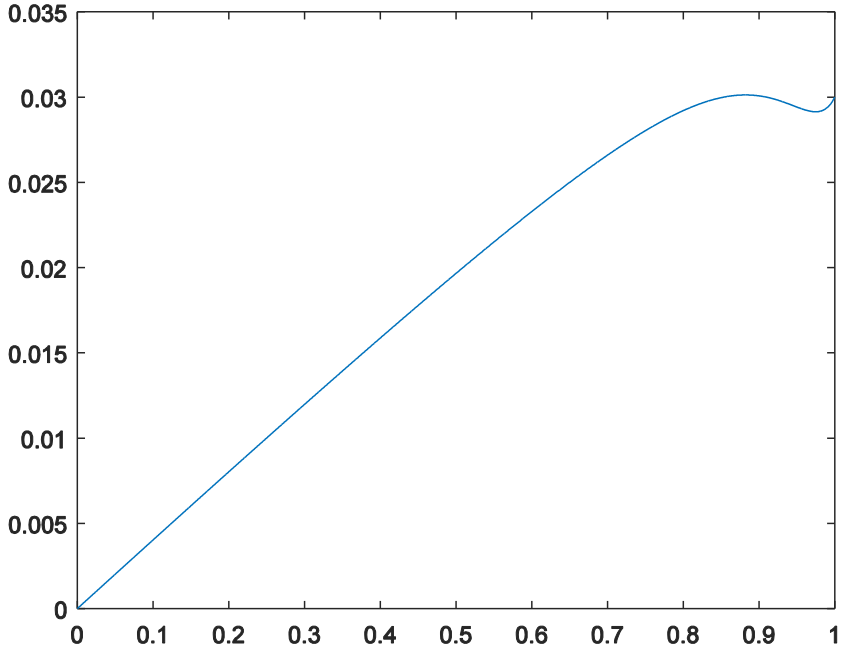
Typical case: two solutions



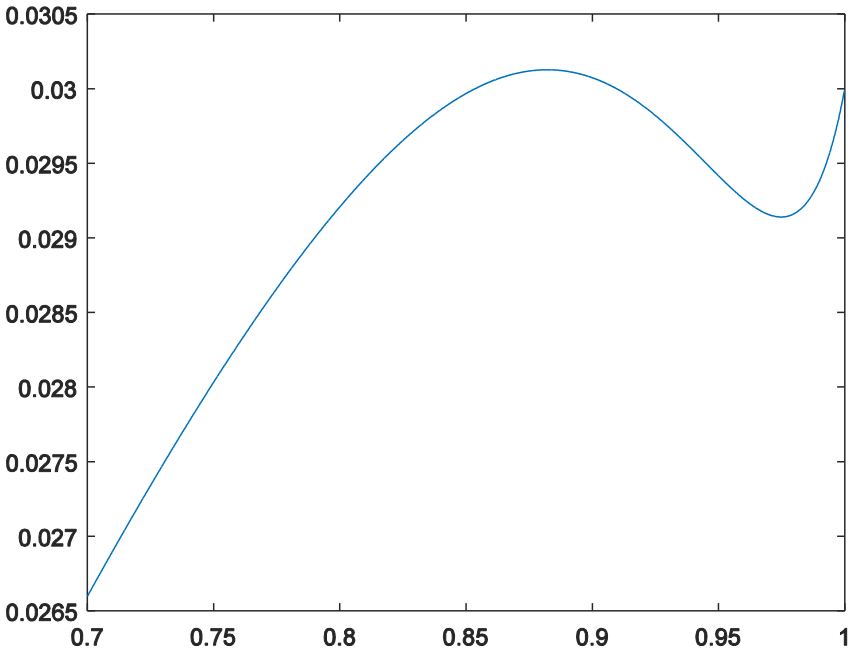
Single solution



Three solutions



Three solutions: closer look



References

[1] Emmanuel Baccelli and Charles E. Perkins. Multi-hop Ad Hoc Wireless Communication. Internet-Draft draft-baccelli-manet-multihop-communication-04, Internet Engineering Task Force, September 2014. Work in Progress.

[2] F. Baccelli, A. Rybko, and S. Shlosman. Queuing Networks with Varying Topology – A Mean-Field Approach. *ArXiv e-prints*, November 2013.

[3] François Baccelli and Bartlomiej Blaszczyszyn. Stochastic geometry and wireless networks, volume 1: Theory. *Foundations and Trends in Networking*, 3(3-4):249–449, 2009.

[4] François Baccelli and Bartlomiej Blaszczyszyn. Stochastic geometry and wireless networks, volume 2: Applications. *Foundations and Trends in Networking*, 4(1-2):1–312, 2009.

[5] Josh Broch, David A. Maltz, David B. Johnson, Yih chun Hu, and Jorjeta Jetcheva. A performance comparison of multi-hop wireless ad hoc network routing protocols. pages 85–97, 1998.

[6] R. Durrett. *Probability: Theory and Examples*. Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, 2010.

[7] C.S.R. Murthy and B.S. Manoj. *Ad Hoc wireless networks: architectures and protocols*. Prentice Hall communications engineering and emerging technologies series. Prentice Hall PTR, 2004.

[8] Roberto H Schonmann and Senya B Shlosman. Wulff droplets and the metastable relaxation of kinetic ising models. *Communications in mathematical physics*, 194(2):389–462, 1998.

[9] C.K. Toh. *Ad Hoc Mobile Wireless Networks: Protocols and Systems*. Prentice Hall PTR, 2002.

Effective Way of Conducting Highly Accelerated Life Testing – Linking the Failure Mode Effects Analysis and Finite Element Analysis

Vikas Pandey[1], Pravin Kadekodi[2]

•

Eaton Corporation
[1]vikaspandey@eaton.com, [2]pravinkadekodi@eaton.com

Abstract

In today's competitive marketplace, the design phase presents a perfect opportunity to test a product to find its maximum limitations and weak links. On the same context HALT (Highly Accelerated Life Test) has been adopted by many industries. HALT is a destructive stress testing methodology for accelerating product reliability during the engineering development process. It is a great process used for precipitating failure mechanisms in an electronics hardware design and product which may occur into the field.

The traditional HALT process which is followed by most of the industries, deals with destructive stress testing and subjective approach to fix the design weaknesses based on experience, followed by iterative HALT to check the robustness against the design fixes done which may not be relevant fixes.

This paper summarizes the effective way of conducting HALT by emphasizes on the "Analysis First" approach, the FMEA (Failure Mode Effect Analysis) and FEA (Finite Element Analysis) which will help identifying the critical functions along with associated components to be monitored during HALT and reduces the iteration of HALT by analyzing the board robustness against the stresses i.e. temperature and vibration prior to HALT respectively. And also presents the specification limits derived based on the product specification and chamber standard deviation, up to which the root cause and design fixes needs to be done, eliminating the subjectivity around it.

Keywords: HALT, FMEA, FEA

I. Introduction

HALT (Highly Accelerated Life Testing) is a qualitative approach to identify the design weaknesses and improve on the design margins of the electronics assemblies. Extensive studies have been done to understand the HALT methodology and its benefits by many industries. Conducting conventional HALT requires a detailed plan with understanding of applicable stress profiles, functional test, cross functional team responsibilities [1]. This paper will explain in brief the effective HALT methodology including the FMEA and FEA activities prior to HALT to reduce design iteration and also derive specification limits [2] for the root cause analysis. The placements of thermocouple and accelerometer have also been discussed to take a decision for design fixes.

HALT tests methods are developed to find design defects and weaknesses in electronic and electromechanical assemblies so that a more reliable product can be released to market in rapid order. The product specification always can be defined as per the Figure 1 below which shows how the product can be used in unknown environment (red in color).

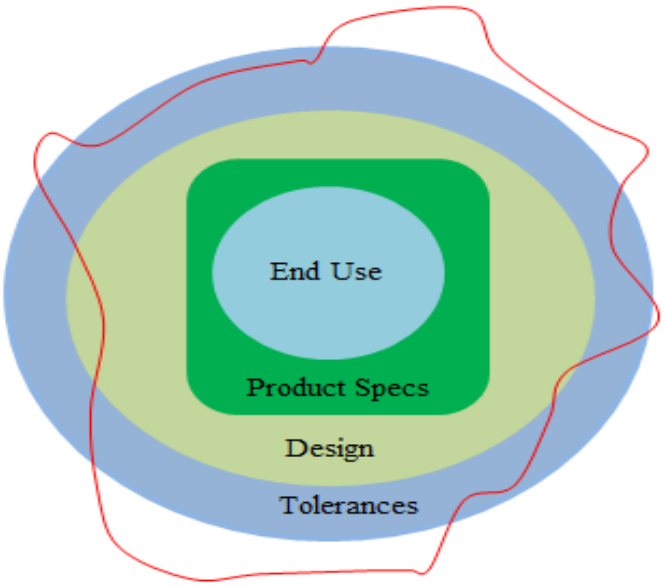


Figure 1: *Product Specifications & Field Environment*

Based on the field profile and applicable stress study, the engineered systems deliver stresses of temperature, rapid thermal cycling, and random, tri-axial vibration which are used to rapidly reveal design weaknesses in electronics and electromechanical product assemblies. Hence the conventional HALT is performed considering the sequence of stresses as follows:

- Thermal Cold Step Stress
- Thermal Hot Step Stress
- Rapid Thermal Transition Cycling
- Vibration Step Stress
- Combined Environment (Thermal Transition Cycling and Vibration Step Stress)

Traditionally in the HALT process, product design/material limitations may be discovered for each stress that is applied. Each of these limitations should have their root-cause understood and corrective action implemented based on product specification. The specification limits for doing the RCA (Root Cause Analysis) for all issues which should be corrected are subjective in nature.

The process of conventional HALT and proposed effective way of conducting HALT is explained in below paragraphs with their limitation and advantages respectively.

II. Conventional HALT Methodology

The conventional HALT methodology deals with the five different stress profiles as mentioned above. The detail test is conducted based on the profiles with limited inputs from the FMEA and FEA. The conventional HALT process is depicted in the below Figure 2.

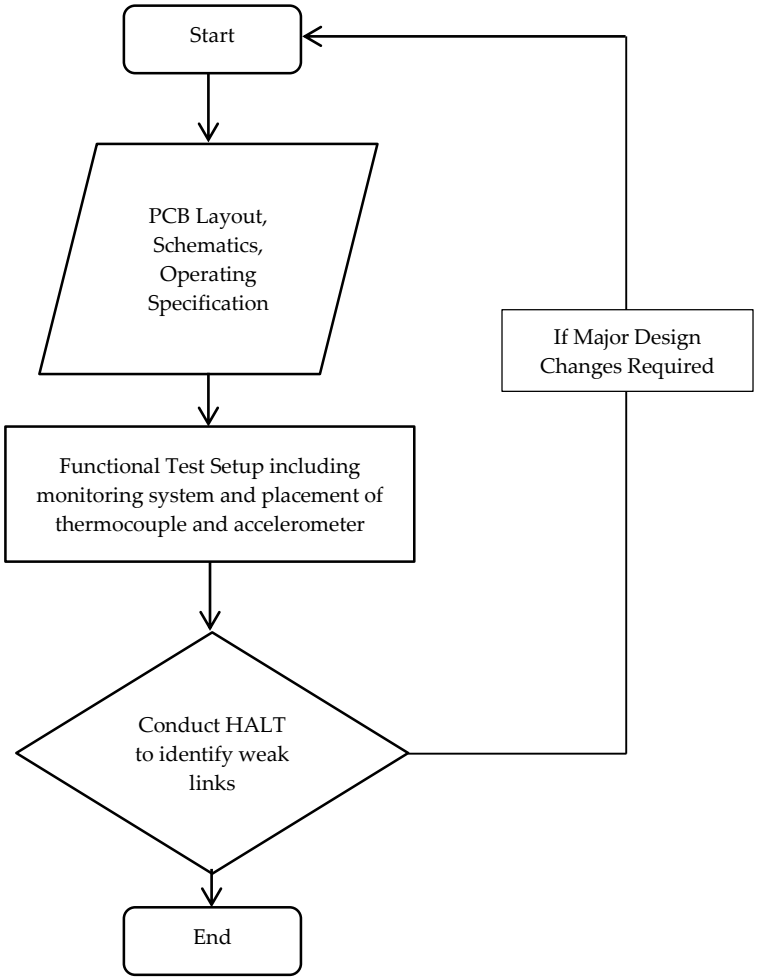


Figure 2: *Conventional HALT Methodology*

III. Proposed Effective HALT Methodology

The effective HALT methodology incorporates inclusion of FMEA outputs for identification of critical components and functions to be monitored along with FEA to validate the PCB design against the temperature cycling and the random vibration stresses prior to testing. This approach helps to identify the weak links in the design through analysis and reduces the iteration of HALT.
The advantages of Effective HALT are as follows:

- The inclusion of FMEA helps in identification of critical function which needs to be monitored during HALT and hence cross verify the performance of all the component as per the functional requirement and its criticality ranking.
- Inclusion of FEA helps in identification of design flaws prior to HALT and reduce the design iteration post HALT.
- Placement of thermocouple and accelerometer can be identified based on the FEA results i.e. most heat dissipating component and most resonant component respectively.
- Identification of specification limit helps in taking decisions for RCA needs to be done and

with its limits.

- The effective HALT helps in identification of weak links and improve the design robustness by involving cross functional team inputs and recommendations.

In the process of doing FMEA and FEA there may be chances of getting multiple number of functions and component which needs to monitored during HALT. The functional test of associated component along with the placement of thermocouple and accelerometer can be finalized by doing tradeoff between the severity ranking (can be rated in scale of 1 to 10) for the functions in FMEA and occurrence ranking (can be rated in scale of 1 to 10) for the identified stressed component in FEA. The proposed approach enables to identify the functions to be monitored and component based on the tradeoff on severity*occurrence rating.

The flow chart shown in Figure 3 summarizes the steps to conduct HALT effectively with brief explanation of key activities.

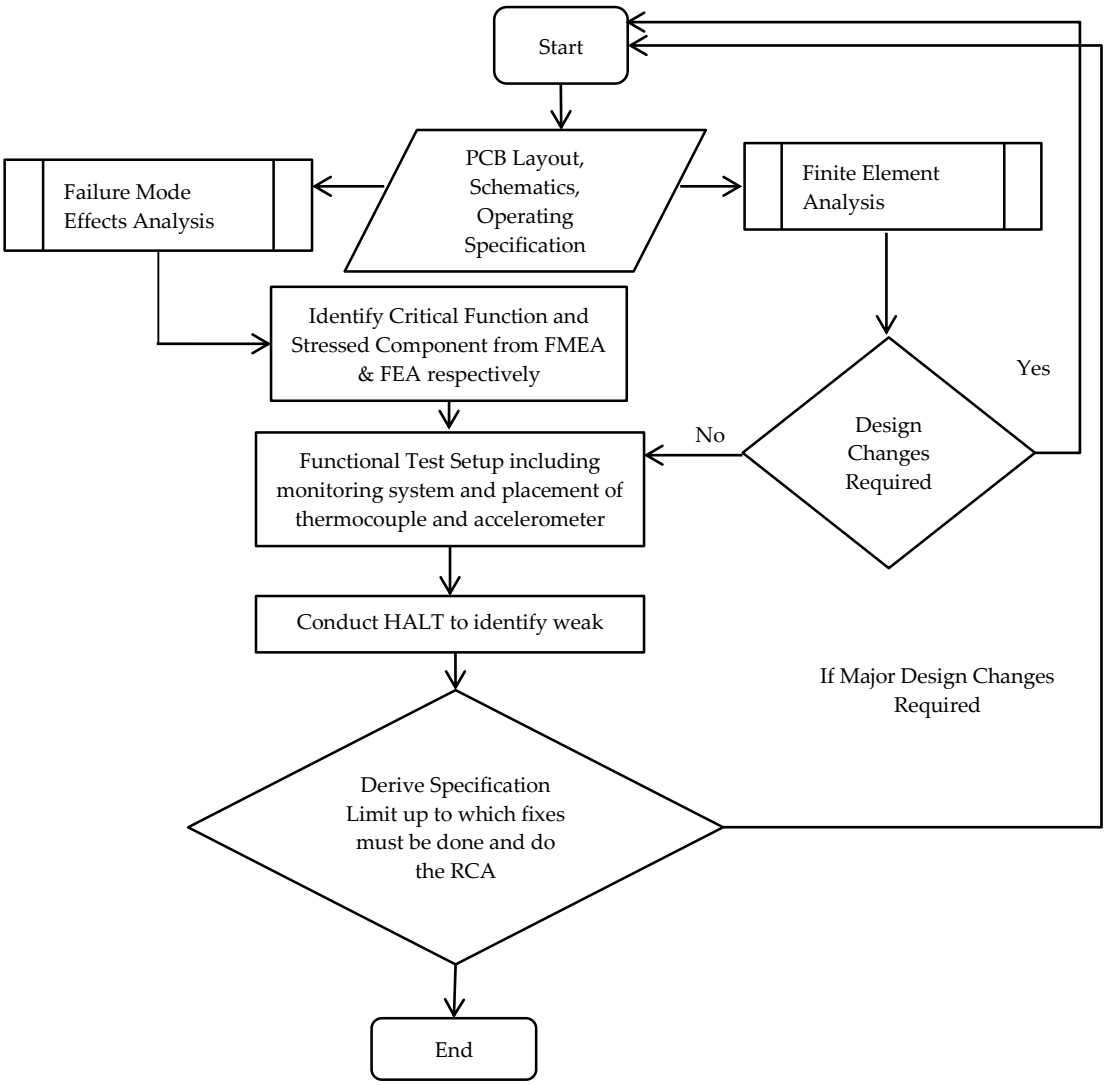


Figure 3: *Effective HALT Methodology*

I. Failure Mode Effect Analysis

The main objective of Failure Mode Effect Analysis (FMEA) is to thoroughly analyzes product design against all failure modes and reduce the associated risks. During each stress profile, functional testing is performed on the product sample to evaluate its operation performance. To verify the functionality of the board the FMEA is best way to identify the safety or mission critical functions and the respective associated failure mode of the components. For conducting effective HALT this exercise helps in identification of critical functions of the product which needs to be monitored during HALT and respective components. The functional test setup during the HALT should be made such that it covers all the critical functionalities of the product.

II. Finite Element Analysis

Finite Element Analysis (FEA) as applied in engineering is a computational tool for performing engineering analysis. This analysis is carried out for electronic and electromechanical assemblies to verify its performance against the temperature and vibration profile. These exercises help on identification of the stressed components and also design flows, if any prior to HALT to reduce the HALT iteration. The placement of thermocouples and accelerometers also becomes very important. The outputs of both needs to be analyzed to take a decision for the fixes. So the placement of thermocouple and accelerometer is also needs to be done based on the stressed component identified during the analysis.

III. Specification Limits

After conducting HALT two concerns always arise. The first is: which issues should be corrected; and the second is: should all issues be corrected? To resolve this issues in HALT the operating limits are considered, and the specification limits are derived for which the root cause analysis must be done, considering the components specification and operational limits with ±3σ range [3]. An example is shown in below Table 1 to derive the specification limits considering 2°C & 2 Grms as standard deviation for the thermocouple and accelerometer reading respectively.

Table 1: *Specification Limits Calculation*

S. No.	Factors	Values
1	Hot Temperature Specification	85°C
2	Hot Temperature Operating Limit	98°C
3	Proposed Hot Temperature Specification Limit	98°C + 6°C (i.e. 3*(σ)) = 104°C
4	Cold Temperature Specification	-40°C
5	Cold Temperature Operating Limit	-54°C
6	Proposed Cold Temperature Specification Limit	-54°C +(-6°C) (i.e. 3*(σ)) = -60°C
7	Vibration Specification	2.24 Grms
8	Vibration Operating Limit	5 Grms
9	Proposed Vibration Specification Limit	5 Grms +(6 Grms) (i.e. 3*(σ)) = 11 Grms

The above derived specification limits are shown in graphical form in the below Figure 4.

Temperature

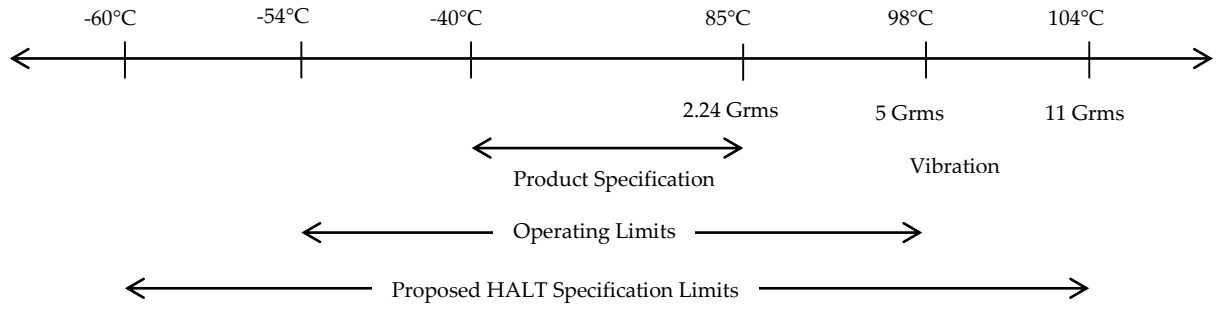


Figure 4: *Proposed HALT Specification Limits*

III. Conclusions

In this paper effective way of conducting HALT is explained in detail. Inclusion of FMEA and FEA with specification limits up to which RCA needs to be conducted are mentioned with the respective advantages. The above mentioned way of conducting HALT helps in identifying the weak links more accurately and reduce the iteration of HALT also the design.

References

[1] Qualmark. HALT Testing Guideline Document: 933-0336 Rev. 04.

[2] Doertenbach, N. Highly Accelerated Life Testing – Testing With a Different Purpose: White Paper at *http://www.qualmark.com/resources/library-documents*.

[3] Mclean, H. HALT, HASS, AND HASA EXPLAINED: American Society of Quality, *Quality Press*, 2009.

Simulation Of Reliability For Electronic Means With Regard To Temperature Fields

Artyukhova M., Polesskiy S., Linetskiy B., Ivanov I.

•

National Research University Higher School of Economics, Moscow, Russia
mayaartyukhova@gmail.com, spolessky@hse.ru, blinetskiy@hse.ru, i.ivanov@hse.ru

Abstract

The paper considers the technique of modeling of electronic reliability based on modeling electrical components environment temperature. As experience of the simulation and exploitation of electronic shows, one of the main factors that significantly affect the reliability characteristics is the thermal effect. This is confirmed by the statistics of a number of companies. In the paper for the simulation were used systems ASONIKA-K and ASONIKA-TM. On the example of a real electronic mean proved the need for a point temperature estimate for each electrical component and the account of these temperatures, instead of the average values in predicting the reliability indices. Such approach will significantly improve (20% - 40%) the accuracy of estimates of the mean time to failure. Developed engineering method to predict reliability, built on the "downward" hierarchical circuit simulation.

The reported study was supported by RFBR, research project No. 14-07-00422 a.

Keywords: reliability, printed circuit boards, thermal analysis

I. Introduction

Reliability is a complex electronic device property, which, depending on the purpose and conditions of its application consists of a combination of properties: dependability, durability, maintainability and conservation.

Today the actual direction of the reliability theory is the prediction of indicators of reliability in the early stages of design. The direction uses different approaches, one of the key is a methodology for the synthesis of highly reliable electronic means on the criteria of reliability.

Practice of design and operation shows that the greatest impact on the reliability by climatic, mechanical and electrical effects [1]. General failure rate model of the printing assembly of the electronic means in the mode of operation is as follows [2]:

$$\Lambda_{PA\,EM} = K_a \cdot \sum_{j=1}^{m} \sum_{i=1}^{n} \lambda_{eij} \qquad (1)$$

where: K_a – quality factor of production equipment, relative units; λ_{eij} – operational failure rate of the *i*-th type of product *j*-th group (see model below), 1/h; *n* – the number of products *j*-th group, items; *m* – number of product groups, items.

The model λ_{eij} in general for standard electronic components (chip resistors, chip capacitors, etc.) is as follows [2]:

$$\lambda_{eij} = \lambda_b(\lambda_{b.g.}) \cdot K_r(K_t) \cdot K_e \cdot \prod_{i=1}^{n} K_i \quad (2)$$

where: $\lambda_b(\lambda_{b.g.})$ – basic failure rate of type (group) of electrical components, calculated according to the results of tests on the electrical component reliability, durability, life, 1/hr; $K_r(K_t)$ – mode coefficient (temperature) takes into account the magnitude of the electrical load and (or) the ambient temperature (the product's enclosure), relative units; K_e – operating factor takes into account the severity of operating conditions, relative units; K_i – coefficient taking into account changes in operational failure rate depending on various other factors, relative units; n - number of factors taken into account, items.

Affecting electronic factors can be divided into four types of effects, as shown in Table 1.

Table 1: *List of external influencing factors*

№	Effects	Factor name
1	Climatic	• high pressure air or gas • reduced atmospheric pressure • changes in atmospheric pressure • Low ambient temperature • Increased ambient temperature • high humidity • atmospheric condensed precipitation • low air humidity • salt mist • solar radiation
2	Mechanical	• Broadband random vibration • acoustic noise • linear acceleration • seismic shock • Mechanical shock of single action • Mechanical shock of repeated action
3	Biological	• mold fungi • insects • rodents
4	Other	• static dust • Dynamic dust • aggressive environment (ozone, ammonia, nitrogen dioxide, sulfur dioxide, hydrogen sulfide)

Objective factors are determined by the time and conditions of use and include the operation time; climatic factors; mechanical factors; biological factors; operating modes. The typical distribution of electrical component failure due to objective reasons shown in Figure 1.

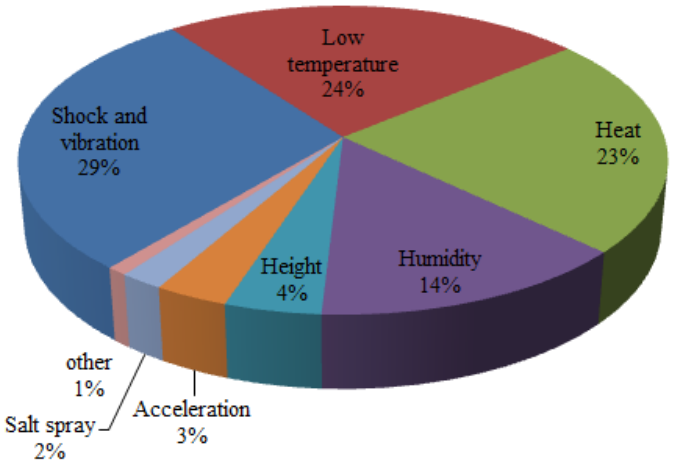


Figure 1: *The electrical component failure rate of various objective factors*

As seen from the model (2) and the real statistical failure (see. Figure 2) for each electrical component makes the largest contribution $K_r(K_t)$, and it, in turn, is determined by the point modeling of ambient temperature (or shell) of the element or of experimental investigations [3]. As shown in Figure 2 for a typical printing assembly change in the ambient temperature of +25 ° C to +80 ° C leads to a change in failure rate of more than 1.66 times.

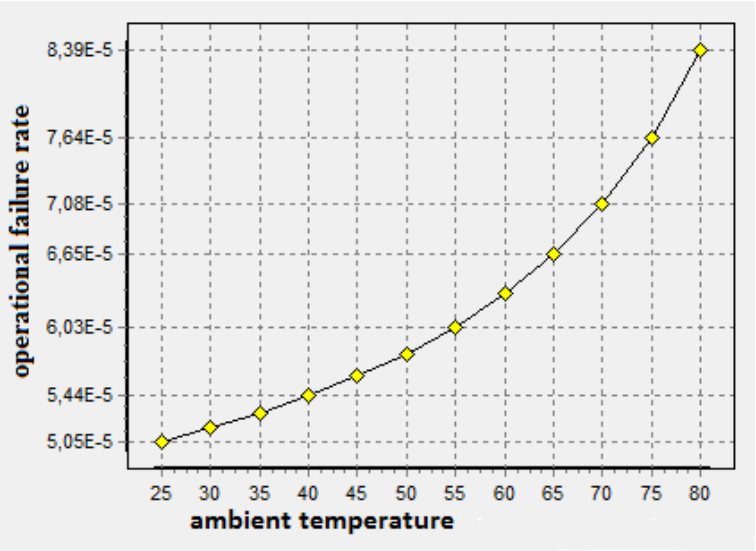


Figure 3: *Graph of operational failure rate of a typical printing assembly from ambient temperature*

II. Thermal analysis of electronic equipment

As shown in [3] thermal modeling reveals the weakness of the development, to correct them and protect from heat. This approach also allows one to get a more accurate value of time to failure under the conditions of use of the object prior to disposal. Thermal analysis will allow at a stage of simulations increase the value of system reliability indicators in the possible reduction of cost and the geometric dimensions. The fact that the higher the temperature, the lower the reliability. In the presence of operational data can be predicted the mean time to failure for a newly developed product.

Thermal modeling relevant because:

1. Analysis of temperature fields of electronic means - is rapidly expanding area of research;
2. Thermal analysis is applicable to many areas of design;

3. Thermal analysis is very important for engineering research.

Model of the failure rate of the temperature λ_d. The failure rate for any reference temperature T_r can be calculated using the following equation, and with known T_b and λ_b:

$$\lambda_d = \lambda_b / exp \frac{-AE\left(\frac{1}{T_b} - \frac{1}{T_d}\right)}{8.61735 \cdot 10^{-5}} \quad (3)$$

The failure rate is doubled by raising to 10 °C ambient temperature (K=293 °C) for AE=0.53. Most electronic solid state components have AE=0.4, and failure rate is doubled when the temperature rises to 13,5 °C.

Cooling systems [3] should be designed to control the temperature of the components. By varying the cooling systems in board electronic means in some cases could increase by 500% the average time to failure.

Implementation of the requirements of the thermal analysis leads to an additional increase in the cost of the design. However, the average cost of a heat-resistant electronic means compensated by saving operating costs.

A thermal analysis of the electronic means should be performed at the system level. Without it, it can happen that parts and components will continue to refuse. Components can be designed to work in normal conditions, but due to the low heat transmission from different heat generators, they can not work at increased temperatures.

There are two main areas in the thermal analysis of electronic means: 1. Knowing electrical component temperature and therefore to quantify the degradation of electrical parameters; 2. Reduce the temperature of the electronic components that improve system reliability. The first may predict "hot" spots in the development through detailed analytical prediction or through direct measurement of heat. The second allows local cooling of these areas that will significantly increase the component life time.

To select the mathematical models for calculating the reliability of foreign and national reference books were analyzed. For the basics reference [6] was taken as the most used and reliable.

Thermal modeling was carried out on the example of a typical printed board assembly of electronic means.

The task is this: to calculate the printing assembly for given thermal actions. Based on the analysis of the thermal characteristics of printed assembly conclude that the technical requirements for electrical components for thermal characteristics performed.

Data for calculation.

The initial data for the calculation of blueprints printed board assembly and output PCAD system files have been received, as well as maps electrical component operating modes. Design printed board assembly subsystem ASONIKA-TM, is shown in Fig. 3 (first side) and Fig. 4 (second side).

The capacity of heat generation electrical component in the PCA: B2 – 0,6 mW; R7 – 30 mW; R8 – 40 mW; R14 – 200 mW; R15 – 200 mW; R17 – 30 mW; R20 – 110 mW; R21 – 10 mW; R23 – 20 mW; R24 – 110 mW; R25 – 10 mW; R26 – 20 mW; R27 – 110 mW; R28 – 80 mW; R29 – 30 mW; D5 – 1500 mW; D6 – 1500 mW; D7 – 1500 mW; D20 – 157 mW; VT1 … VT3 - 40 mW; **Total 5777,6 mW.**

According to the results of thermal calculation unit in the subsystem ASONIKA-T obtained the following air temperature inside the unit:

- for natural convection 100,2 °C;
- with forced convection blowing speed of 1 m/s 53 °C.

Use the data the temperature values as the boundary conditions for the thermal design of printed assembly.

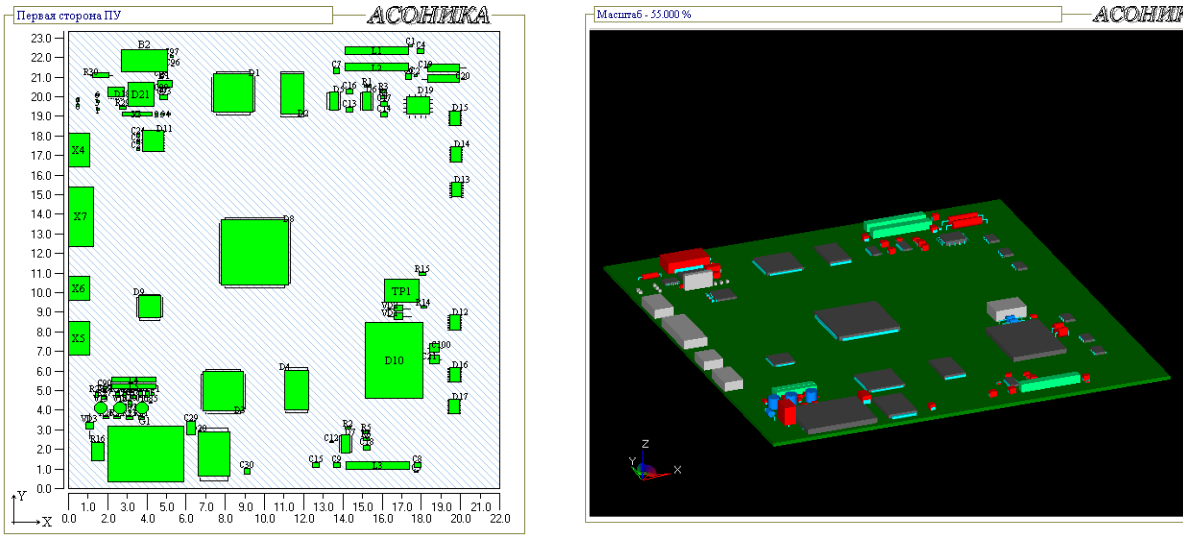


Figure 3: *Design of printed assembly in the subsystem ASONIKA-TM (side 1)*

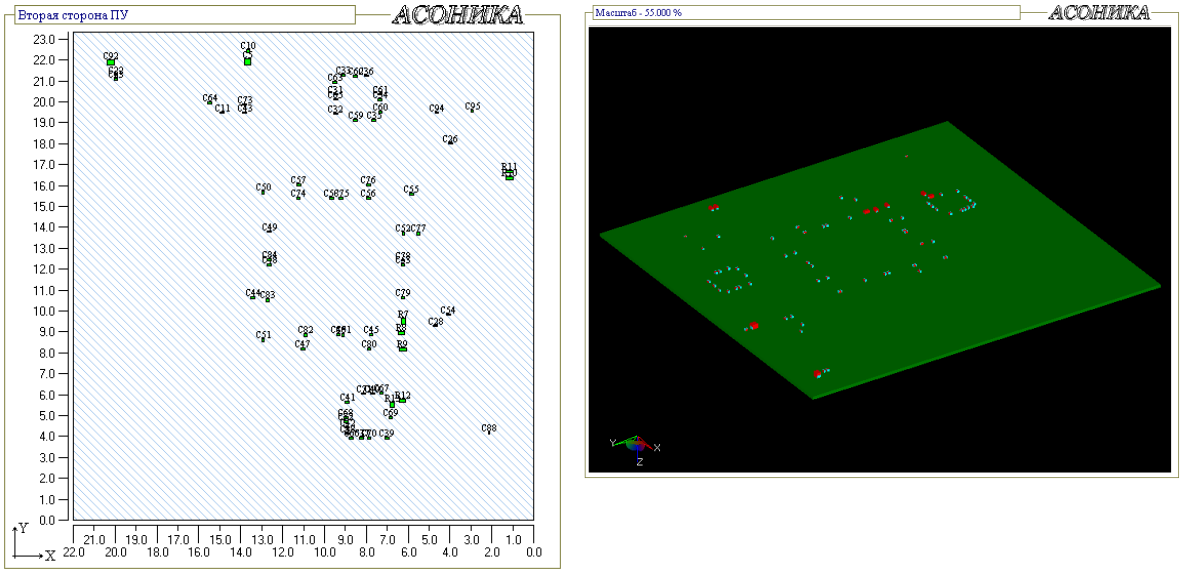


Figure 4: *Design of printed assembly in the subsystem ASONIKA-TM (Side 2)*

Results of thermal analysis.

Calculation of thermal characteristics of printed board assembly was held in an automated subsystem ASONIKA-TM. Fig. 5 and Fig. 6 shows obtained thermal characteristics for printed board assembly mode 1 in operation (the air inside the unit for natural convection 100.2 °C) and mode 2 (air inside the unit in a forced convection blowing speed of 1 m/s 53 °C). Maps of thermal modes of electrical component are presented in tables 2 and 3.

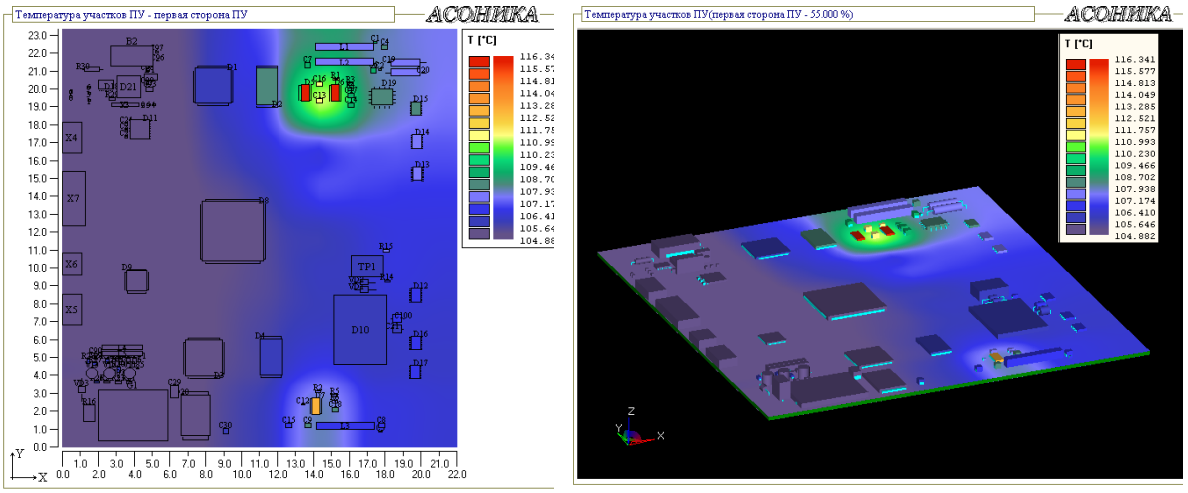


Figure 5: *Temperature Field for printed board assembly in mode 1*

Table 2: *Section of the map of thermal modes of electrical component (when stationary thermal action) for of printed assembly in mode 1*

№	Symbol of electrical components	side	The temperature of electrical components		Coefficient of thermal load, [relative units]	Overheat, [°C]
			Estimated, [°C]	Maximum permissible, [°C]		
1	R1	1	111.222	100.000	1.112	11.222
2	R17	1	105.574	100.000	1.056	5.574
3	R18	1	105.445	100.000	1.054	5.445
4	R19	1	105.418	100.000	1.054	5.418
5	R2	1	107.819	100.000	1.078	7.819
6	R20	1	106.025	100.000	1.060	6.025
7	R21	1	105.487	100.000	1.055	5.487
8	R22	1	105.418	100.000	1.054	5.418
9	R23	1	105.479	100.000	1.055	5.479
10	R24	1	106.004	100.000	1.060	6.004
.....						
20	C1	1	108.730	85.000	1.279	23.730
.....						
179	C95	2	105.153	100.000	1.052	5.153

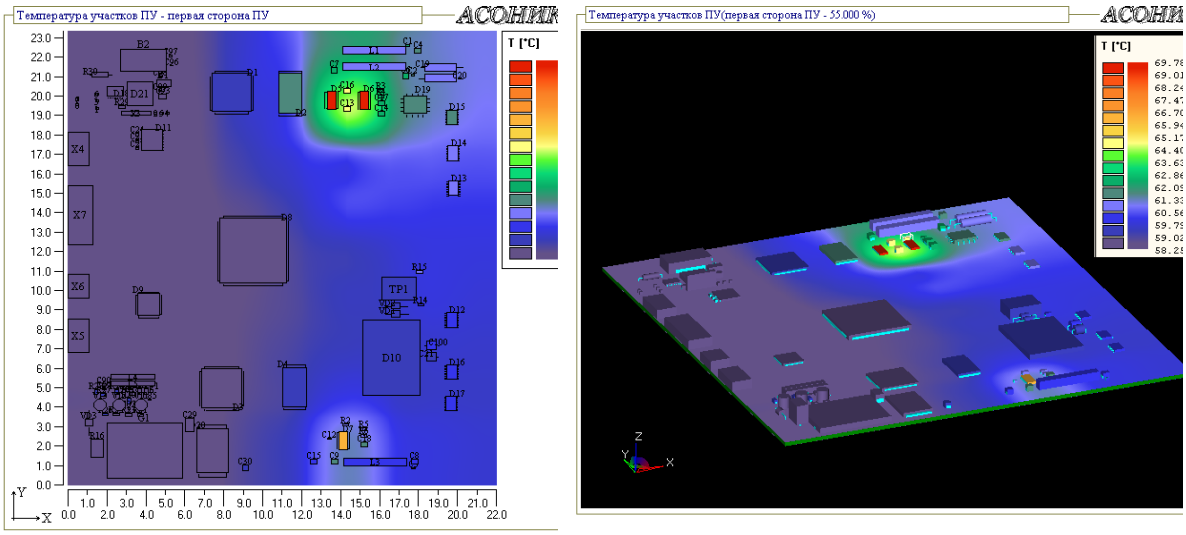


Figure 6: *The temperature field for a printed assembly in mode 2*

Table 3: *Section of the map of thermal modes of electrical component (when stationary thermal action) for the printing unit in mode 2*

№	Symbol of electrical components	side	The temperature of electrical components		Coefficient of thermal load, [relative units]	Overheat, [°C]
			Estimated, [°C]	Maximum permissible, [°C]		
1	R1	1	64.652	100.000	0.647	
2	R17	1	58.946	100.000	0.589	
3	R18	1	58.820	100.000	0.588	
4	R19	1	58.790	100.000	0.588	
5	R2	1	61.206	100.000	0.612	
6	R20	1	59.401	100.000	0.594	
7	R21	1	58.863	100.000	0.589	
8	R22	1	58.790	100.000	0.588	
9	R23	1	58.855	100.000	0.589	
10	R24	1	59.376	100.000	0.594	
			……...			
20	C1	1	62.169	85.000	0.731	
			……...			
179	C95	2	58.530	100.000	0.585	

III. Calculation of reliability printed board assembly

The first and one of the main steps of calculating the reliability of the printed board assembly is to identify the electrical component parameters.

Under the parameter identification should be understand the process of determining the parameters of a mathematical model of reliability calculation, for each specific type of electronic components. The process of identification of electronic components can be represented schematically in the form of an algorithm, illustrated in Figure 7. When performing the

identification of electronic components, according to the algorithm, some points should be noted. Secondly, found in prior specifications, not always this information is sufficient for the calculation of reliability, in such cases, according to the block 9, was searched averaged parameters of technological groups and subgroups in the directory of the reliability of foreign-made product when checking the adequacy of the information. As a result of the identification of all part types from the list, we were assigned to a particular class of electrical component, and in line with the previously selected mathematical models, all the necessary parameters have been found.

Calculation of reliability of the printed board assembly.

An indicator of reliability of printed board assembly is its mean time to failure with no recovery in the process. Reliability of printed assembly is characterized by a set of failure rates of its components (electrical component). The scheme of calculating the reliability of printed board assembly corresponding to a predetermined criterion of failure, is a serial connection of a technologically and functionally combined electrical component groups.

Figure 8 shows the sequence diagram for calculating the reliability of the device included in printed assembly on the level of technology combined electrical component groups.

Operational electrical component failure rate was calculated according to the corresponding reference books [7-9] and a set of maps the correct application of electrical component.

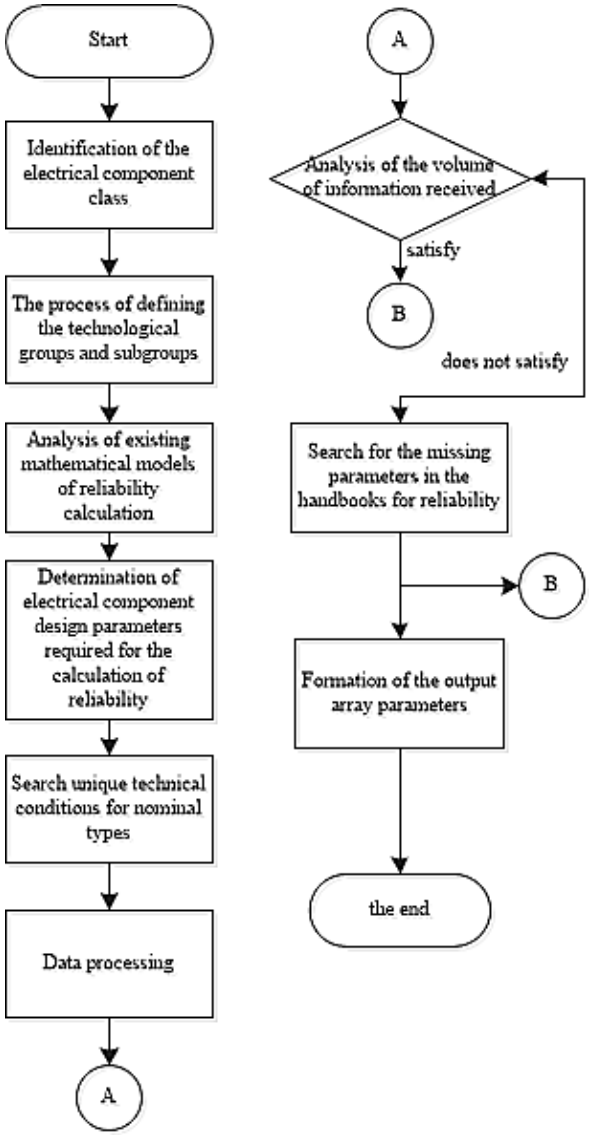


Figure 7: *Algorithm for electrical component identification process*

Figure 9 shows the window ASONIKA-K system with the results of printed assembly calculation (estimation).

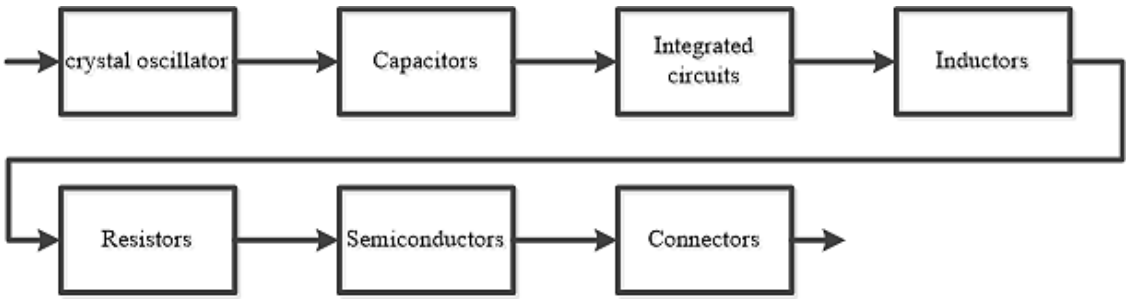


Figure 8: *The scheme of calculating the reliability of printed assembly*

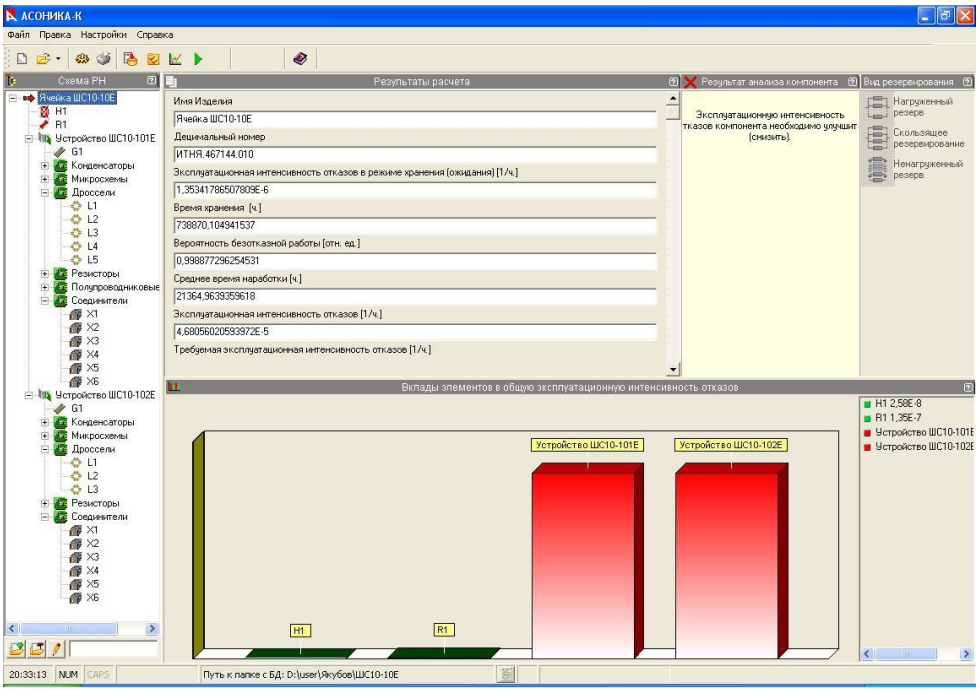


Figure 9: *ASONIKA-K system: the results of printed assembly calculation (estimation)*

As can be seen from Fig. 9, obtained by calculating the value of the average operating time of printed assembly is ≈21,364 thousand [hours] (Electric Load coefficients varying depending on the type of electric components from 0.1 to 0.7 at a temperature of 65 [°C]) that does not satisfy the technical requirements (T_0=150 thousand [hours]).

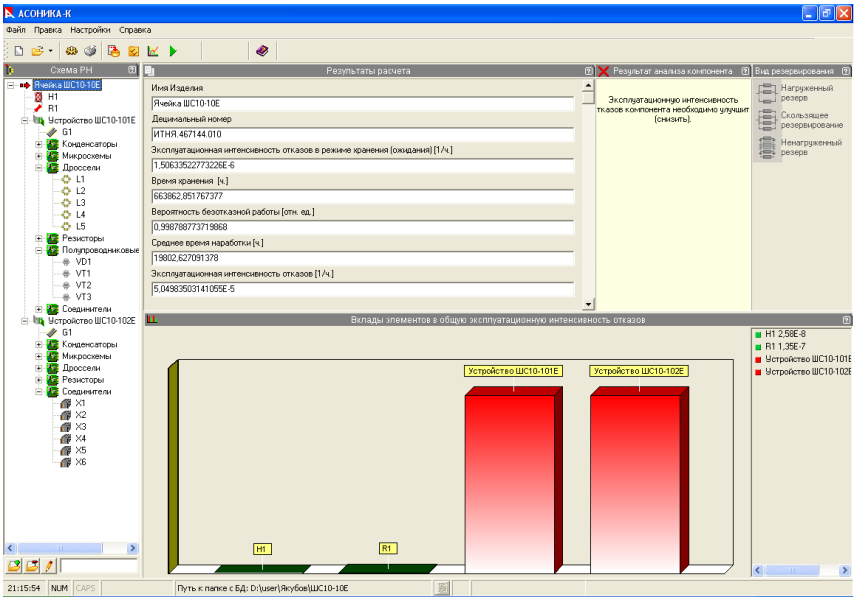


Figure 10: *System ASONIKA-K: The results of printed assembly calculation (adjusted calculation)*

Adjusted calculation printed board assembly reliability.

Adjusted calculation operating electrical component failure rate was based on electrical component temperature, the resulting heat-transfer simulation using subsystem ASONIKA-TM, and other data about the electrical component of printed assembly were taken from the set of maps of the correct application of electrical component.

Fig. 10 shows the window ASONIKA-K system with the results of printed assembly calculation (adjusted calculation).

As can be seen from Fig. 10 obtained by calculating the average value of use of printed assembly is ≈ 19,802 thousand [hours] (Load for electric coefficients varying depending on the type of electrical component from 0.1 to 0.7 at temperatures electro obtained by subsystem ASONIKA-TM), which does not meet the technical requirements (T0=150 thousand. [hours]).

Analysis of the results of calculations.

To assess the influence of ambient temperature environment was constructed operational temperature dependence of printed board assembly failure rate in the temperature range +25 ... + 85 [°C] for given values of electrical load coefficient depending on the type of electrical component from 0.1 to 0.7 (see Fig. 11).

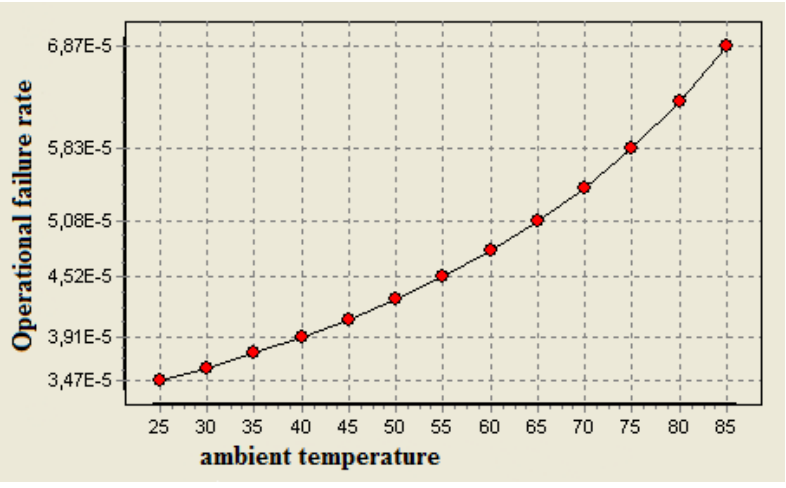


Figure 11: *Dependence of operational failure rate of printed assembly of temperature*

As can be seen from Fig. 11, a simultaneous change of electrical component temperature in the range +25 ... + 85 [°C] causes a change in the intensity of printed board assembly failures in 2 times.

Assessing the impact of the specific characteristics of reliability of electrical component on operational intensity printed board assembly failures carried out directly during the calculation.

Figure 12 shows the contribution classes printed board assembly electrical component to the total failure rate.

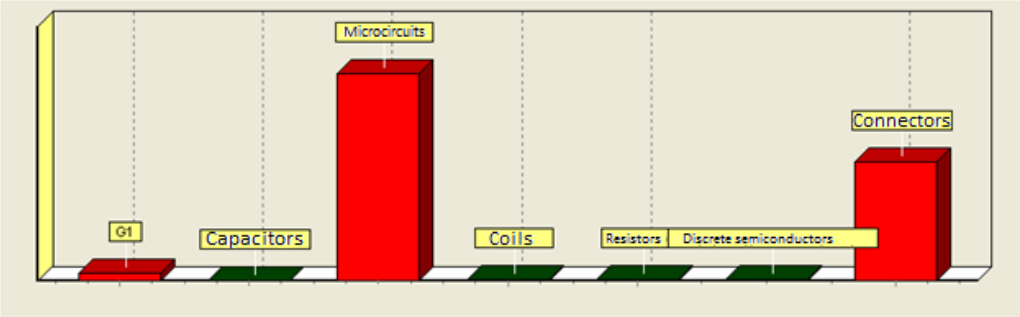


Figure 12: *Contributions of electrical component classes to the total intensity of the printed board assembly failure*

As shown in Figure 12 of the most unreliable class electrical component is a class "Integrated circuits" and "connectors".

Figure 13 shows the contribution of electrical component class "Integrated circuits" to the total failure rate.

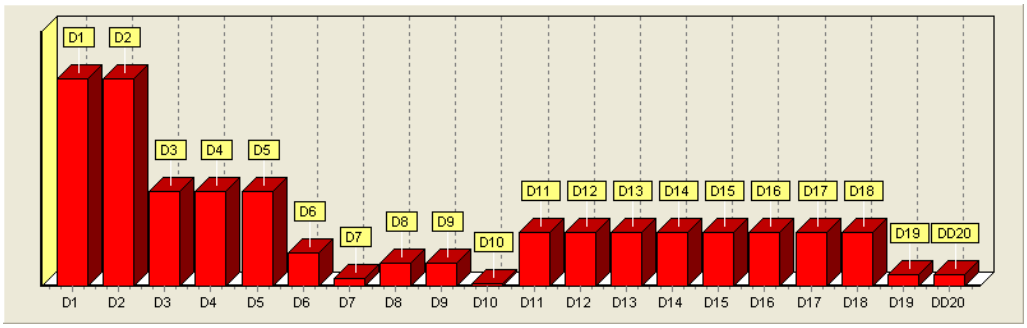


Figure 13: *Contributions of class "Integrated circuits" electrical components to the total failure rate*

As it follows from Fig. 13 unreliable chips are chips D1, D2 type TMS320VC5416PGE160 and D3-D5 type TPS73HD301PWPR, TPS73HD325PWPR.

Figure 14 shows the contribution of electrical component class "connectors" to the total failure rate. As it follows from Fig. 14, the connectors are unreliable connectors X1, X2 type C 6921 03164.

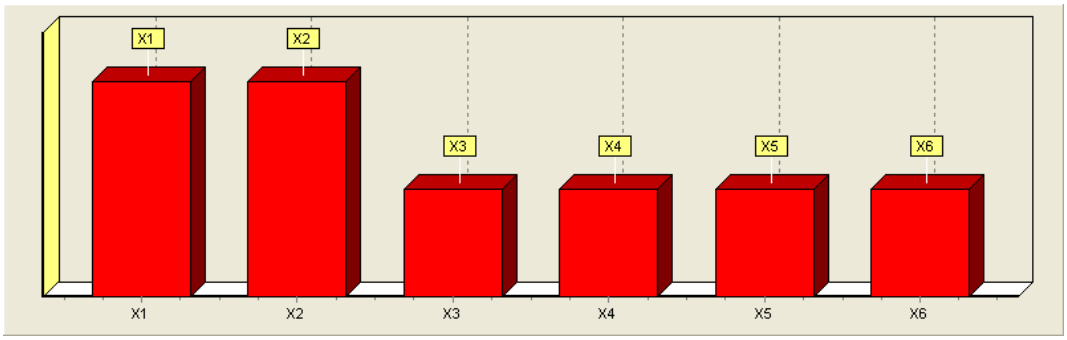


Figure 14: *The electrical component class contributions of "connectors" to the total failure rate*

IV. Conclusion

The calculation of printed assembly reliability has shown that:

- At a temperature of 65 [° C] average time to failure is not less than 21.364 thousand [hours.]. Electric load factor depending on the type of electrical component that varies in the range from 0.1 to 0.7;

- At temperatures of electrical component derived from simulations using subsystem ASONIKA-TM, the average time to failure is not less than 19.802 thousand [hours.], For the electric load factor depending on the type of electrical component, varying in the range of 0.1 to 0.7.

Options considered analysis printed board assembly reliability showed that the reliability of the product does not meet the requirements (mean time to failure is to be not less than 150000 hours). The most unreliable electrical component classes are the class of "Integrated circuits" and "connectors". To improve reliability, we can recommend the following measures:

- change the type of electrical component (use electrical components with less λ_b);
- to facilitate the operation of the electronic components (lower operating thermal and electrical load);
- reduce the number of electrical component (use the chip higher degree of integration);
- use electrical components with a high level of quality;
- reduce the ambient temperature (to increase the efficiency of the cooling system).

Using the concept of mathematical modeling of complex heterogeneous physical processes in the development of printed board assemblies within systems ASONIKA-K and ASONIKA-TM allows one to improve the accuracy of reliability parameters modeling;

In this paper: 1) proved by the example of printed board assembly need for differential evaluation of the temperature of each electronic components and their integration in predicting reliability, rather than the averaged temperature values; 2) developed a technique of mathematical modeling of reliability for thermal printed board assemblies that the example has proved its effectiveness.

References

[1] GOST RV 20.39.302-98. KSOTT. Equipment, instruments, devices and equipment for military purposes. Requirements for the program to ensure reliability.

[2] RDV 319.01.10-98. KSOTT. Equipment, instruments, devices and equipment for military purposes. Methods of reliability based design and manufacturing of electronics.

[3] V.V. Goldin Information Support lifecycle of electronic / V.V. Goldin, V.G. Zhuravsky, A.V. Sarafanov etc. - M.: Radio and Communications, 2002. - 386 p.

[4] GOST RV 20.39.304-98. KSOTT. Equipment, instruments, devices and location-ment for military purposes. Requirements for resistance to external factors.

[5] OST 4G 0.012.242-84. Radio-electronic equipment. The methodology of calculation of reliability indicators.

[6] Reliability of electrical component: Handbook / 22 TSNIII Defense Ministry. - M.: Ministry of Defense of the Russian Federation, 2006. - 641 p.

[7] RDV 319.01.20-98. The position of the directory "Reliability of Electrical products".

[8] B.S. Sotskov. Fundamentals of the theory and calculation of reliability of elements and arrangements of automation and computer technology/ B.S. Sotskov. - M .: Higher School, 1970 - 270 p.

[9] A.S. Shalumov. ASONIKA automated system for designing highly reliable radio electronic means on the principles of CALS-technologies: Volume 1 / A.S. Shalumov, Y.N. Kofanov, V.V. Zhadnov and others// Ed. Y.N. Kofanov, N.V. Malyutin, A.S. Shalumov. - M.: Energoatomisdat, 2007. - 538 p.

[10] Pisarev, A.V. Test System - the basis of ensuring the reliability / B. Pisarev, Kritenko M., V. Postnov// "Electronica: NTB», № 5, 2002 - p. 32-35.

[11] Reliability prediction of electronic equipment: Military Handbook. MIL-HDBK-217F (Notice 1, 2), 1991 - 205 p.

Parameter Estimation of Mukherjee-Islam Model under Step Stress Partially Accelerated Life Tests with Failure Constraint

Ahmadur Rahman*, Showkat Ahmad Lone, Arif-Ul-Islam

Deptt of Statistics & Operations Research,
Aligarh Muslim University, Aligarh-202002
*Email: kahef.ahmad@gmail.com

Abstract:

In this paper, we have studied the estimation of parameters under failure censored data using step stress partially accelerated life testing. The lifetimes of test items are assumed to follow Mukherjee-Islam distribution. The estimation of different parameters and acceleration factor are obtained by Maximum Likelihood Method. Relative absolute bias (RAB), mean squared error (MSE), relative error (RE), standard deviation and confidence intervals are also obtained. Asymptotic variance-covariance matrix and also test method are given. Simulation studies have been introduced to illustrate the performance of all the statistical assumptions.

Keywords: Mukherjee-Islam distribution, Step-stress partially accelerated life test, Maximum likelihood, failure censoring.

Introduction:

The present era is the era of high reliability. The products and items made nowadays are too much reliable. Usually they do not fail early at normal use condition. So it is not easy to get reasonable amount of failure data under use condition for a given period of time. For this reason, Accelerated Life Testing is the modest procedure to be applied. By using it, products would fail early and at the end of test we have sufficient failure data to study the behaviour of products. ALT quickens the procedure that's why it costs less money and consume less time.

In ALT, the relationship between lifetime stress is known either in the form of acceleration factor or there exists a mathematical model. But in many situations neither acceleration factor is known nor there exist any such model. Then partially accelerated life test is the better option to use. In PALT, the acceleration factor and mathematical model which sustain the relationship between the life time and stress are not known and cannot be assumed any type of such model.

Nelson [19] introduced that the stress can be applied on test item in various ways, commonly used are step-stress and constant-stress. Under step-stress PALT, a test unit is first subjected to run at normal use condition and if it does not fail for a specified time, then it is run at accelerated use condition until failure occurs or the observation is censored. But in constant stress PALT each item runs at daily use condition or at accelerated condition.

DeGroot and Goel [12] have introduced the concept of step-stress PALT in which a

test item is first run at use condition and, if it does not fail for a specified time 'τ', then it is run at accelerated condition until failure.

A lot of literature is available on SS-PALT analysis, for example, see Goel [13], Bhattacharyya and Soejoeti [11], Bai and Chung [10], Abdel-Ghani [8] and Abdel-Ghaly et al. [6,7], Abdel-Ghani [9]. Ismail [15] studied the estimation and optimal design problems for the Gompertz distribution in SS-PALT with type I censored data. P.W. Srivastava and N. Mittal [20] considered optimum step stress partially accelerated life tests for the truncated logistic distribution with censoring. This article include type I and type II both censoring. S. Hyun and J. Lee [21] used constant stress partially accelerated life testing for log logistic distribution with censored data. F. K. Wang et al [22] have studied partially accelerated life tests for the Weibull distribution under multiply censored data. Recently, Showkat Ahmad Lone et al [23] studied estimation in step stress partially accelerated life tests for the Mukherjee-Islam distribution using time constraint. For a brief knowledge of step-stress ALT, one should go through [14, 17, 16, and 18].

II. Test Methods and Model

Mukherjee-Islam failure model is introduced by Mukherjee and Islam [1]. It is finite range distribution which is one of the most important property of it in recent time in reliability analysis. Its mathematical form is simple and can be handled easily, that is why, it is preferred to use over more complex distribution such as normal, Weibull, beta etc. The pdf of the distribution is given as

$$f(x,\alpha,\lambda)=\frac{\alpha}{\lambda^{\alpha}}x^{\alpha-1},\quad 0\le x\le\lambda,\ \alpha>0,\ \lambda>0$$

Where λ is the scale parameter and α is the shape parameter.

The cdf is given as

$$F(x)=\left(\frac{x}{\lambda}\right)^{\alpha},\ 0\le x\le\lambda,\ \alpha>0,\ \lambda>0$$

And the Reliability function of finite range model is given as

$$R(x)=1-\left(\frac{x}{\lambda}\right)^{\alpha}$$

In SS-PALT, all of the n units are tested first under normal condition, if the unit does not fail for a pre-specified time τ, then it runs at accelerated condition until failure. This means that if the item has not fail by some pre-specified time τ, the test is switched to the higher level of stress and it is continued until items fail. The effect of this switch is to multiply the remaining lifetime of the item by the inverse of the acceleration factor β. In this case, switching to the higher stress level will shorten the life of the test item. Thus the total lifetime of a test item, denoted by Y, passes through two stages, which are the normal and accelerated conditions.

The lifetime of the test unit in SSPALT is given as follows

$$Y=\begin{cases} T & if\quad T\le\tau \\ \tau+\beta^{-1}(T-\tau) & if\ T>\tau \end{cases}$$

Where T is the lifetime of item at use condition.

Therefore, the probability density function of total lifetime Y of an item is given by

$$f(y)=\begin{cases}0 & y<0\\ f_1(y) & 0<y\le\tau\\ f_2(y) & y>\tau\end{cases}$$

Where

$$f_1(y)=\frac{\alpha}{\lambda^{\alpha}}y^{\alpha-1} \qquad \alpha>0,\lambda>0$$

$$f_2(y)=\frac{\beta\alpha}{\lambda^{\alpha}}[\tau+\beta(y-\tau)]^{\alpha-1} \qquad \beta>1,\quad \alpha>0$$

III. Estimation Procedure

The maximum likelihood estimation method is used here because it is very robust and gives the estimates of the parameters with good statistical properties such as consistency, asymptotic unbiasedness, asymptotic efficiency and asymptotic normality. In this section, point and interval estimation for the parameters and acceleration factor of Mukherjee-Islam distribution based on type II censoring are evaluated using this method.

3.1. Point estimates

In type II censoring scheme, we set the number of units or subject to the experiment and stop the experiment at a predetermined number of failure. The observed values of the total lifetime Y are $y_{(1)}<\ldots<y_{(n_u)}\le\tau<y_{(n_u+1)}<\ldots<y_{(n_u+n_a-1)}\le y_{(r)}$, where n_u and n_a are the number of subjects or items failed at normal conditions and accelerated conditions respectively. Let φ_{1i} and φ_{2i} be indicator functions, such that

$$\varphi_{1i}=\begin{cases}1 & y_{(i)}\le\tau\\ 0 & elsewhere\end{cases} \quad i=1,2,\ldots,n$$

And,

$$\varphi_{2i}=\begin{cases}1 & \tau<y_{(i)}\le y_{(r)}\\ 0 & elsewhere\end{cases} \quad i=1,2,\ldots,n$$

For our convenience, $y_{(i)}$ is written as y_i. The likelihood function of independent and identically distributed random variables $y_1,\ldots,y_n$, the life times of the items is given by

$$L(y;\beta,\lambda,\alpha)=\prod_{i=1}^{n}\left\{\frac{\alpha}{\lambda^{\alpha}}y_i^{\alpha-1}\right\}^{\varphi_{1i}}\left[\frac{\beta\alpha}{\lambda^{\alpha}}\{\tau+\beta(y_i-\tau)\}^{\alpha-1}\right]^{\varphi_{2i}}\left[1-\left\{\frac{\tau+\beta(y_{(r)}-\tau)}{\lambda}\right\}^{\alpha}\right]^{\bar{\varphi}_{1i}\,\bar{\varphi}_{2i}} \qquad (3.1.1)$$

Where $\bar{\varphi}_{1i}=1-\varphi_{1i}$ and $\bar{\varphi}_{2i}=1-\varphi_{2i}$, We take the logarithm of the likelihood function and write it as follows;

$$\ln L=r\log\alpha-r\alpha\log\lambda+(\alpha-1)\left\{\sum_{i=1}^{n}\varphi_{1i}\log y_i+\sum_{i=1}^{n}\varphi_{2i}\log A\right\}+n_a\log\beta+(n-r)\log(1-B^{\alpha}) \qquad (3.1.2)$$

Where, $A=[\tau+\beta(y_i-\tau)]$, $B=\left[\frac{\tau+\beta(y_{(r)}-\tau)}{\lambda}\right]$, $\sum_{i=1}^{n}\varphi_{1i}=n_u$, $\sum_{i=1}^{n}\varphi_{2i}=n_a$,

$$\sum_{i=1}^{n}\varphi_{1i}\varphi_{2i} = n - n_u - n_a \quad \& \quad r = n_u + n_a$$

The maximum likelihood estimates of β, α and λ can be obtained by solving the system of equations which are the first partial derivatives of the above log likelihood equation, and equate to zero with respect to β, α and λ respectively.

Thus the system of solutions are given as:

$$\frac{\partial InL}{\partial\beta} = \frac{n_a}{\beta} + (\alpha-1)\sum_{i=1}^{n}\varphi_{2i}A^{-1}(y_i-\tau) - (n-r)rB^{\alpha-1}(1-B^{\alpha})^{-1}\frac{(y_{(r)}-\tau)}{\lambda} = 0 \qquad (3.1.3)$$

$$\frac{\partial InL}{\partial\alpha} = \frac{r}{\alpha} - rIn\lambda + \sum_{i=1}^{n}\phi_{1i}In(y_i) + \sum_{i=1}^{n}\phi_{2i}InA - \frac{(n-r)}{(1-B^{\alpha})}B^{\alpha}InB = 0 \qquad (3.1.4)$$

$$\frac{\partial InL}{\partial\lambda} = -\frac{r\alpha}{\lambda} + \frac{(n-r)}{\lambda}(1-B^{\alpha})^{-1}\alpha B^{\alpha} = 0 \qquad (3.1.5)$$

The solution of above equations cannot be obtained in closed form because the equations are non linear in three unknown parameters β, α and λ. Therefore, to find numerical solution we use an iterative method. Newton-Raphson Method is used to obtain the numerical estimate of the parameters β, α and λ.

3.2. Interval estimates

If $L_\omega = L_\omega(y_1, y_2, ..., y_n)$ and $U_\omega = U_\omega(y_1, y_2, ..., y_n)$ are functions of the sample data $y_1, ..., y_n$, then the confidence interval for a population parameter ω is given by

$$p[L_\omega \le \omega \le U_\omega] = \pi \qquad (3.2.1)$$

Where, L_ω and U_ω are the lower and upper confidence limits which enclose ω with probability π. The interval $[L_\omega, U_\omega]$ is called a $100\pi\%$ confidence interval for ω. For large sample size, the maximum likelihood estimates, under appropriate regularity conditions, are consistent and asymptotically normally distributed. Therefore, the approximate $100\pi\%$ Confidence limits for the maximum likelihood estimate $\hat{\omega}$ of a population parameter ω can be constructed, such that

$$p\left[-Z \le \frac{\hat{\omega}-\omega}{\sigma(\hat{\omega})} \le Z\right] = \pi \qquad (3.2.2)$$

Where, Z is the $\left[\frac{100(1-\pi)}{2}\right]$ standard normal percentile. Therefore, the approximate $100\pi\%$ confidence limits for a population parameter ω can be obtained, such that

$$p[\omega - Z\sigma(\hat{\omega}) \le \hat{\omega} \le \omega + Z\sigma(\hat{\omega})] = \pi \qquad (3.2.3)$$

Then, the approximate confidence limits for β, α and λ will be constructed using Eq. (3.2.3) with confidence levels 95% and 99%.

IV. Asymptotic variances and covariance of Estimates

The asymptotic variances of maximum likelihood estimates are the elements of the inverse of the Fisher information matrix $I_{IJ}(\varphi) \cong E\{-\partial^2 InL/\partial\varphi_i\partial\varphi_j\}$. The exact mathematical expressions for the above expectation are very difficult to obtain. Therefore, the observed Fisher information matrix is given by $I_{IJ}(\varphi) \cong \{-\partial^2 InL/\partial\varphi_i\partial\varphi_j\}$ which is

obtained by dropping the expectation on operation E . The approximate (observed) asymptotic variance–covariance matrix F for the maximum likelihood estimates can be written as follows

$$F = \left[I_{ij}(\varphi)\right] \quad i, j = 1,2,3 \qquad (\varphi) = (\beta, \alpha, \lambda) \tag{4.1}$$

The second partial derivatives of the maximum likelihood function are given by the following

$$\frac{\partial^2 InL}{\partial \beta^2} = \frac{-n_a}{\beta^2} - (\alpha - 1)\sum_{i=1}^{n} A^{-2}\phi_{2i}(y_i - \tau)^2 - \alpha(n-r)\frac{(y_{(r)} - \tau)^2}{\lambda^2}\left[\alpha B^{2\alpha-2}\left(1 - B^{\alpha}\right)^{-2} + (\alpha - 1)B^{\alpha-2}\left(1 - B^{\alpha}\right)^{-1}\right] \tag{4.2}$$

$$\frac{\partial^2 InL}{\partial \alpha^2} = \frac{-r}{\alpha^2} - (n-r)B^{\alpha}\left(\frac{InB}{1 - B^{\alpha}}\right)^2 \tag{4.3}$$

$$\frac{\partial^2 InL}{\partial \lambda^2} = \frac{r\alpha}{\lambda^2} - \frac{\alpha(n-r)}{\lambda^2}\left[\alpha B^{2\alpha}\left(1 - B^{\alpha}\right)^{-2} + \alpha B^{\alpha}(1 - B^{\alpha})^{-1} + B^{\alpha}(1 - B^{\alpha})^{-1}\right] \tag{4.4}$$

$$\frac{\partial^2 InL}{\partial \beta \partial \alpha} = \sum_{i=1}^{n} \phi_{2i} A^{-1}(y_i - \tau) - (n-r)\left(\frac{y_{(r)} - \tau}{\lambda}\right)\left[\alpha B^{2\alpha-1}\left(1 - B^{\alpha}\right)^{-2} InB + \left(1 - B^{\alpha}\right)^{-1} B^{\alpha-1}(\alpha InB + 1)\right] \tag{4.5}$$

$$\frac{\partial^2 InL}{\partial \beta \partial \lambda} = \alpha^2(n-r)B^{\alpha-1}\left(1 - B^{\alpha}\right)^{-1}\left(\frac{y_{(r)} - \tau}{\lambda^2}\right)\left[1 + B^{\alpha}\left(1 - B^{\alpha}\right)^{-1}\right] \tag{4.6}$$

$$\frac{\partial^2 InL}{\partial \alpha \partial \lambda} = -\frac{r}{\lambda} + \left(\frac{n-r}{\lambda}\right)B^{\alpha}\left(1 - B^{\alpha}\right)^{-1}\left[1 + \alpha InB + \alpha InB\left(1 - B^{\alpha}\right)^{-1} B^{\alpha}\right] \tag{4.7}$$

Consequently, the maximum likelihood estimators of β, α and λ have an asymptotic variance–covariance matrix defined by inverting the Fisher information matrix F and by substituting $\hat{\beta}$ for β, $\hat{\alpha}$ for α and $\hat{\lambda}$ for λ.

V. Simulation studies

Simulation studies are very important part of the study. It has been performed to illustrate the precision and consistency of the theoretical results of estimation parameters. R software is used in simulation studies. Absolute relative bias (RAB), mean square error (MSE) and relative error (RE) are the main measure to check the performance of resulting estimators. The detailed steps of procedures are presented below:

Step 1. 1000 random samples of sizes 50, 75, 100, 125 and 150 were generated from the Mukherjee-Islam distribution. The data generation of the Mukherjee-Islam distribution is very simple, if U has a uniform (0, 1) random number, and then $Y = [\lambda.u\text{^}(1/\alpha)]$ follows a Mukherjee-Islam distribution. The true parameter values are selected as $(\alpha = 1.6,\ \lambda = 2,\ \beta = 1.05)$ and $(\alpha = 1.5,\ \lambda = 2,\ \beta = 1.1)$.

Step 2. Choosing the censoring time τ at the normal condition to be τ=1 and the total number of failure in PALT is to be r=0.75*n

Step 3. For each sample and for the two sets of parameters, the acceleration factor and the parameters of distribution were estimated in SS-PALT under type II censored sample.

Step 4. The Newton–Raphson method was used for solving the nonlinear equations given in (3.1.3), (3.1.4) and (3.1.5).

Step 5. The RABias, MSEs, and REs of the estimators for acceleration factor and other

parameters for all sample sizes were tabulated.

Step 6. The confidence limit with confidence level γ=0.95 and γ =0.99 of the acceleration factor and other parameters were constructed.

The results are summarized in Tables 1 and 2. Table 1 presents the RABias, MSEs, and REs of the estimators. The approximated confidence limits at 95% and 99% for the parameters and acceleration factor are presented in Table 2.

Following are the observations can be made from the tabulated data on the performance of SS-PALT parameter estimation of the above used lifetime distribution:

1. For the second set of parameters $(\alpha = 1.5,\ \lambda = 2,\ \beta = 1.1)$, the maximum likelihood estimators have good statistical properties than the first set of parameters $(\alpha = 1.6,\ \lambda = 2,\ \beta = 1.05)$ for all sample sizes (see Table 1)
2. As the acceleration factor increases the estimates have smaller MSE, and RE. As the sample size increases the RABias and MSEs of the estimated parameters decreases. Hence the estimates provide asymptotically normally distributed and consistent estimators for the acceleration factor and other parameters.
3. The interval of the estimators decreases when the sample size is increasing. Also, the range of the interval estimate at γ=0.95 is smaller than the range of the interval estimate at γ=0.99 (see Table 2).

Table 1 The RABias, MSE and RE of the parameters (α, λ, β) for different sample sizes under type II censoring

N	Parameters (α, λ, β)	(1.6,2,1.05)			(1.5,2,1.1)		
		RABias	MSE	RE	RABias	MSE	RE
50	α	0.0102	0.0616	0.0385	0.0008	0.0634	0.0422
	λ	0.0576	0.0829	0.0414	0.0157	0.0861	0.0430
	β	0.0013	0.0655	0.0013	0.0702	0.0850	0.0773
75	α	0.0094	0.0591	0.0369	0.0019	0.0411	0.0274
	λ	0.0484	0.1006	0.0503	0.1139	0.0448	0.0569
	β	0.0471	0.0747	0.0711	0.0148	0.0761	0.0692
100	α	0.0280	0.0451	0.0282	0.0211	0.0328	0.0219
	λ	0.0371	0.0674	0.0337	0.0448	0.1139	0.0569
	β	0.0184	0.0550	0.0524	0.0289	0.0643	0.0584
125	α	0.0061	0.0392	0.0245	0.0066	0.0458	.0305
	λ	0.0641	0.1070	0.0535	0.0331	0.0951	0.0475
	β	0.0807	0.1287	0.1225	0.0325	0.0755	0.0687
150	α	0.0125	0.0185	0.0115	0.0425	0.0568	0.0379
	λ	0.0304	0.0284	0.0142	0.0065	0.0748	0.0374
	β	0.0012	0.0264	0.0251	0.0255	0.1080	0.0982

Table 2 Confidence bounds of the estimates at confidence levels 0.95 and 0.99

N	Parameters (α, λ, β)	(1.6,2,1.05)			(1.5,2,1.1)		
		Standard Deviation	Lower Bound	Upper Bound	Standard Deviation	Lower Bound	Upper Bound
50	α	0.2547	1.0842 0.9263	2.0829 2.2409	0.2584	0.9948 0.8346	2.0078 2.1680
	λ	0.2955	1.5359	2.6944	0.3011	1.4413	2.6218
	β	0.2627	1.3527 0.5364	2.8776 1.5664	0.2991	1.2546 0.4363	2.8085 1.6091
			0.3735	1.7293		0.2508	1.7946
75	α	0.2495	1.0956 0.9409	2.0739 2.2287	0.2080	1.0950 0.9660	1.9107 2.0397
	λ	0.3255	1.4588	2.7349	0.3463	1.4110	2.7685
	β	0.2804	1.2570 0.5498	2.9367 1.6491	0.2830	1.1962 0.5615	2.9832 1.6712
			0.3759	1.8230		0.3860	1.8467
100	α	0.2179	1.1277 0.9926	1.9823 2.1174	0.1860	1.1035 0.9881	1.8329 1.9482
	λ	0.2664	1.5521	2.5966	0.3463	1.4110	2.7685
	β	0.2408	1.3869 0.5973	2.7618 1.5414	0.2602	1.1962 0.6218	2.9832 1.6419
			0.4480	1.6907		0.4604	1.8033
125	α	0.2032	1.2115 1.0855	2.0081 2.1341	0.2199	1.0789 0.9425	1.9411 2.0774
	λ	0.3356	1.4705	2.7862	0.3169	1.4449	2.6875
	β	0.3680	1.2624 0.4132	2.9943 1.8562	0.2824	1.2484 0.5105	2.8840 1.6179
			0.1850	2.0844		0.3354	1.7930
150	α	0.1397	1.3060 1.2194	1.8537 1.9404	0.2446	1.0842 0.9325	2.0433 2.1950
	λ	0.1731	1.7214	2.4003	0.2806	1.4629	2.5631
	β	0.1667	1.6140 0.7218	2.5077 1.3754	0.3372	1.2889 0.4109	2.7372 1.7328
			0.6184	1.4788		0.2018	1.9419

Conclusion

Here, we have used type II censored data to obtain the likelihood estimation for performance of Mukherjee-Islam distribution and the acceleration factor under SSPALT. The censoring scheme along with SSPALT will be feasible to the experimenter and he/she will have a fixed number of failure prior to the experiment. That will save cost as well as money. The maximum likelihood estimation technique is used to estimate the parameters and the estimators of it are consistent and asymptotically normally distributed. The current research shows that the second set of parameters has good statistical properties than the first set of parameters. As the sample size increases, the confidence interval become narrower. It can also be noted that interval of the estimators at γ=0.99 is greater than the corresponding at γ=0.95. Therefore, it can be inferred that the model and test method used in this study works satisfactorily for step stress partially accelerated life testing under the certain assumptions.

References

[1] S. P. Mukherjee, & A. Islam, A Finite-Range Distribution of Finite Times, *Naval Research Logistics Quarterly* 30 (1983): 487-491.
[2] C. D. Lai and S. P. Mukherjee, A Note on "A finite range distribution of failure times", *Microelectronics Reliability (Elsevier)*, 26 (1986): 183-189.
[3] S. A. Siddiqui, S. Jain, S. N. Andrabi, I. Siddiqui, M. Alam, and P. A. Chalkoo, Moments of concomitants of order statistics for a new finite range bi-variate distribution (FRBD), *Journal of Reliability and Statistical Studies*, 5 (2012): 77-84
[4] S. A. Siddiqui, M. Subharwal, The Mukherjee-Islam Failure Model, *Microelectronic Reliability*, 32 (1992): 923-924.
[6] A.A. Abdel-Ghaly, A.F. Attia and M.M. Abdel-Ghani, The maximum likelihood estimates in step partially accelerated life tests for the Weibull parameters in censored data, *Communication in Statistics-Theory and Methods*, 31 (4) (2002): 551–573.
[7] A.A. Abdel-Ghaly, E.H. El-Khodary, A.A. Ismail, Maximum likelihood estimation and optimal design in step stress partially accelerated life tests for the Pareto distribution with type I censoring, in: Proceeding of the 14th Annual Conference on Statistics and Computer Modelling in Human and Social Sciences, Faculty of Economics and Political Science, Cairo University, 2002, pp. 16–29.
[8] M.M. Abdel-Ghani, Investigations of some lifetime models under partially accelerated life tests, Ph.D. Thesis, Department of Statistics, Faculty of Economics and Political Science, Cairo University, Egypt, 1998.
[9] M.M. Abdel-Ghani, The estimation problem of the log logistic parameters in step partially accelerated life tests using type I censored data, *The National Review of Social Sciences*, 41 (2) (2004): 1–19.
[10] D.S. Bai and S.W. Chung, Optimal design of partially accelerated life tests for the exponential distribution under type I censoring, *IEEE Transactions on Reliability* 41 (1992): 400–406.
[11] G.K. Bhattacharyya, Z. Soejoeti, A tampered failure rate model for step-stress accelerated life test, *Communication in Statistics-Theory and Methods* 8 (1989) 1627–1644.
[12] M.H. DeGroot, P.K. Goel, Bayesian estimation and optimal design in partially

accelerated life testing, *Naval Research Logistics Quarterly* 16 (2) (1979): 223–235.

[13] P.K. Goel, Some estimation problems in the study of tampered random variables, Ph.D. Thesis, Department of Statistics, Cranegie-Mellon University, Pittsburgh, Pennsylvania, 1971.

[14] E. Gouno, A. Sen and N. Balakrishnan, Optimal step-stress test under progressive type-I censoring, *IEEE Transactions on Reliability* 53 (3) (2004): 388–393.

[15] A.A. Ismail, On the optimal design of step-stress partially accelerated life tests for the Gompertz distribution with type I censoring, *InterStat, Electronic Journal* (2006): 1-15.

[16] C. Li and N. Fard, Optimum bi-variate step-stress accelerated life test for censored data, *IEEE Transactions on Reliability*, 56 (1) (2007): 77–84.

[17] W. Nelson, A bibliography of accelerated test plans, *IEEE transactions on Reliability*, 54 (3) (2005): 370–373.

[18] G. Yang, Life Cycle Reliability Engineering, *John Wiley and Sons, New Jersey*, 2007.

[19] W. Nelson, Accelerated Life Testing: Statistical Models, Test Plans and Data Analysis, *John, Wiley and Sons, New York*, 1990.

[20] P. W. Srivastava and N. Mittal. Optimum step-stress partially accelerated life tests for the truncated logistic distribution with censoring, *Applied Mathematical Modelling*,34(2010) 3166-3178

[21] S. Hyun and J. Lee, Constant-Stress Partially Accelerated Life Testing for Log-Logistic Distribution with Censored Data, *Journal of Statistics Applications & Probability* 4 No.2, 193-201(2015)

[22] F. K. Wang, Y.F. Cheng and W.L. Lu, Partially Accelerated life Tests for the Weibull Distribution Under Multiply Censored Data, *Communication in Statistics-Simulation and Computation,* vol 41(2012) 1667-1678

[23] Showkat Ahmad Lone, Ahmadur Rahman and Arif-Ul-Islam, Estimation in Step-Stress Partially Accelerated Life Tests for the Mukherjee-Islam Distribution Using Time Constrain , *International Journal of Modern Mathematical Sciences,* 2016, 14(3):**227-238**

ISSN 1932-2321