Reliability of a system under two types of failures using a Markovian arrival process

Delia Montoro-Cazorla\textsuperscript{a}, Rafael Pérez-Ocón\textsuperscript{b},\textsuperscript{*}

\textsuperscript{a}Departamento Estadística e I.O., Escuela Universitaria Politécnica de Linares, Universidad de Jaén, 23700 Linares (Jaén), Spain
\textsuperscript{b}Departamento Estadística e I.O., Universidad de Granada, 18071 Granada, Spain

Received 18 May 2005; accepted 6 September 2005
Available online 28 October 2005

Abstract

We consider a system subject to external and internal failures. The operational time has a phase-type distribution (PH-distribution). Failures arrive following a Markovian arrival process (MAP). Some failures require the replacement of the system, and others a minimal repair. This model extends previous papers with arrivals governed by PH-renewal processes. © 2005 Elsevier B.V. All rights reserved.

MSC: Primary 90B25; secondary 62N05

Keywords: Maintenance; Markovian arrival process; Rate of occurrence of failures; Phase-type distribution

1. Introduction

In this paper, we study the arrival of failures for systems operating in a random environment. We allow failures due to different causes. Specifically, a failure can be produced by different causes that can be classified either as internal, due to the inner structure of the system (number of components, aging, quality of materials) or external, due to the environmental conditions in which the system operates (vibrations, humidity, pollution). Both causes of failure can be different, and should be modelled in different ways.

On the other hand, it is frequent that, when a damage is produced, an action must be taken: repair or replacement, depending on the type of damage. Sometimes these are combined in order to obtain better performance of the system. In other cases, when the system undergoes many failures, it is convenient to replace it for safety or economic reasons. The model we present incorporates all these practical considerations and it is analyzed using matrix-analytic methods.

In our model, internal failures cause the fatal failure of the system, therefore it must be replaced. External failures affect the system in two ways: some of them cause damage that can be repaired and others cause fatal failure, and the system must be replaced. We assume that the repair is minimal. When a failure occurs the replacement or minimal repair are instantaneous.

\textsuperscript{*}Corresponding author. Tel.: +34 958 24 31 55; fax: +34 958 24 32 67.
E-mail addresses: dmontoro@ujaen.es (D. Montoro-Cazorla), rperezo@ugr.es (R. Pérez-Ocón).
Due to economic considerations, an imperfect repair that allows the system to be operational even not in a perfect state is sometimes preferred to replacing or repairing completely. For example, electric appliances and cars are not replaced when failures arrive. They are expected to be operational for a certain number of years. When a failure occurs, it is minimally repaired in order to continue operating for a longer period of time. When they have been working for a considerable amount of time, they are replaced or perfectly repaired, depending on the state of the system and costs.

Biswas and Sarkar [1] originally studied this problem under the assumption that operational time had an exponential distribution. He presented formulae for the availability. We have extended it using matrix-analytic methods and also calculating additional reliability measures. There is not an extensive literature on repair and maintenance using these methods. Different types of failure arriving to the system involving PH-distributions have been studied by Neuts et al. [5]. Other papers where PH-distributions play a central role are Pérez-Ocón et al. [6–8], and references contained therein.

We assume that the external failures arrive following a MAP. In the above-mentioned references, the interarrival times between failures are assumed to be independent. We permit a certain dependence relationship between interarrival times: when a failure arrives the MAP enters the absorbing state and it is reinitiated from the transient part but remembers the state prior to absorption. The process restarts randomly, depending on the transient state which produced the absorption. In this way, the interarrival times between failures are not identically or independently distributed.

The MAP, as introduced by Neuts [2], is a generalization of the phase-type distribution, and it is associated with a finite absorbing Markov chain. A methodological paper about these processes is due to Neuts [3]. In Neuts [4] the MAPs are presented as useful models for point processes (pp. 393–395). We use the MAPs in reliability theory, and we introduce a more general arrival model for failures.

We present a system governed by a generalized Markov process that extends the paper of Biswas and Sarkar [1]. The new aspects are: (1) the arrival Markov process for modelling the failure arrivals introduces the Markovian dependence in the interarrival times; (2) three types of failures are considered; (3) the operational time and the number of minimal repairs follow PH-distributions, continuous and discrete, respectively; (4) a policy of repair/replacement depending on the type of failure is considered; (5) two models are studied: a general one and another limiting the number of repairs; and (6) the mean number of failures and replacements per unit time are calculated and the results are presented in well-structured expressions. Throughout the paper, all the expressions are computationally implemented as illustrated in the final example; showing the power of the matrix-analytic methods in the domain of reliability.

The paper is structured as follows. In Section 2 we introduce the assumptions of the general model, and calculate the stationary probability vector and performance measures. In Section 3 we study a model with a maximum number of minimal repairs. Section 4 is dedicated to a numerical application, illustrating the calculations, incorporating costs, and discussing the results for different values of the parameters.

2. The model

The notation to be used and the concepts that appear throughout the present paper are given in the following definitions.

Definition 1. The continuous distribution \( H(\cdot) \) on \([0, \infty)\) is a phase-type distribution (PH-distribution) with representation \((\alpha, T)\), denoted by \(\text{PH}(\alpha, T)\), if it is the distribution of the time until absorption in a finite-state Markov process on the states \(\{1, \ldots, m, m + 1\}\) with generator

\[
\begin{pmatrix}
T & T^0 \\
0 & 0
\end{pmatrix},
\]

and initial probability vector \((\alpha, \alpha_{m+1})\), where \(\alpha\) is a row \(m\)-vector. We assume that the states \(\{1, \ldots, m\}\) are all transient. The matrix \(T\) of order \(m\) is non-singular with negative diagonal entries and non-negative off-diagonal entries and represents the transition rates among transient states. The column vector \(T^0\) of order \(m\) represents the absorption rates from the transient states. Throughout this paper e denotes a column vector with all components equal to one for which the dimension is determined by the context. It is satisfied
\(-\mathbf{T}e = \mathbf{T}^0 \geq 0\). The distribution \(H(\cdot)\) is given by
\[
H(x) = 1 - \alpha \exp(\mathbf{T}x)e, \quad x \geq 0.
\]

**Definition 2.** Let \(\mathbf{D}\) be an irreducible infinitesimal generator of dimension \(m\). We consider a sequence of matrices \(\mathbf{D}_k, k \geq 1\), of dimension \(m\), non-negative and the matrix \(\mathbf{D}_0\) has non-negative off-diagonal entries. The diagonal entries of \(\mathbf{D}_0\) are strictly negative and it is non-singular. The sum of the matrices \(\mathbf{D}_k, k \geq 0\), is the given matrix \(\mathbf{D}\). Consider an \(m\)-state Markov renewal process \(\{(J_n, X_n), n \geq 0\}\) in which each transition epoch has an associated positive number of arrivals \(L_n\). The random variables \(J_n, X_n\) are the states and the sojourn times in states, respectively. The transition probability matrix with \((j, j')\)-entries \(P\{J_n = j', L_n = k, X_n \leq x|J_{n-1} = j\}\) is given by
\[
\int_0^x \exp(\mathbf{D}_0 u) du \mathbf{D}_k \quad \text{for } k \geq 1, \quad x > 0.
\]

The Markovian arrival process (MAP) is defined as the Markov renewal process and a transition probability matrix of the stated particular form.

**Definition 3.** If \(A\) and \(B\) are rectangular matrices of dimensions \(m_1 \times m_2\) and \(n_1 \times n_2\), respectively, their Kronecker product \(A \otimes B\) is the matrix of dimension \(m_1 n_1 \times m_2 n_2\), written in compact form as \((a_{ij} B)\).

**Definition 4.** If \(A\) and \(B\) are matrices of dimensions \(m \times m\) and \(n \times n\), respectively, their Kronecker sum \(A \oplus B\) is the matrix of dimension \(mn \times mn\) written as \(A \oplus B = A \otimes I_n + I_m \otimes B\), where \(I_n\) and \(I_m\) are the identity matrices of order \(n\) and \(m\), respectively.

The arrivals of the external failures are modelled by a MAP\((\mathbf{D}_0, \mathbf{D}_1, \mathbf{D}_2)\), where matrix \(\mathbf{D}_1\) refers to the occurrence of a failure with minimal repair and matrix \(\mathbf{D}_2\) refers to a failure which causes a replacement. Matrix \(\mathbf{D}_0\) records changes that do not produce failures. The MAP assumes an underlying continuous Markov chain: once it has entered the instantaneous absorbing state, the process restarts from the transient part again, depending on the last transient state that reached absorption. We distinguish two types of arrivals, so matrices \(\mathbf{D}_1, \mathbf{D}_2\) must be considered. Matrix \(\mathbf{D}_0\) records the changes among transient states without absorption.

The lifetime of the system due to wear out follows a PH\((\gamma, \mathbf{T})\) distribution. The system can be maintained through several minimal repairs before a replacement. The number of minimal repairs before a replacement follows a discrete PH\(_d\)(\(\gamma, \mathbf{L}\)). We denote by \(\gamma_0\) the probability that the first external failure implies the replacement of the system, \(\gamma_0 = 1 - \gamma e\). This value \(\gamma_0\) is non-null when the vector \(\gamma\) is not a probability vector, then, the system is replaced at the first failure. Repair times are instantaneous.

### 2.1. The generator

In this section we construct the generator of the Markov process that governs the system. We include a counter of the number of imperfect repairs. The system can be either new or operational but not new, after any minimal repair. We must consider the underlying Markov process including the MAP, the continuous PH-distribution, and the counter process. In order to construct the generator, we define two stages: (1) when the system is operational and it has not undergone any minimal repair, we say that it is at stage 0, and (2) when the system has undergone any minimal repair we say that it is at stage 1. With these assignments, the generator will be formed by blocks, that correspond to the transition rates among the stages, so it will be a \(2 \times 2\) block-matrix.

Transition \(0 \rightarrow 0\): If the system is new, then it will continue being new under these circumstances: (1) the wear out time will be ruled by \(\mathbf{T}\) and the MAP by \(\mathbf{D}_0\); (2) an internal failure occurs and the system is instantaneously replaced, the MAP does not change; (3) either the first external failure implies the replacement of the system, or a replacement due to the arrival following \(\mathbf{D}_2\) occurs, and the mechanism of the internal failure initiates. Ordering these events appropriately, we have the block \((\mathbf{T} \oplus \mathbf{D}_0) + \mathbf{T}^0 \alpha \otimes \mathbf{I} + e \alpha \otimes (\gamma_0 \mathbf{D}_1 + \mathbf{D}_2)\).

Transition \(0 \rightarrow 1\): When a minimal repair arrives for the first time, the counter initiates, while the internal failure does not change, then the corresponding block is \(\gamma \otimes \mathbf{I} \otimes \mathbf{D}_1\).

Transition \(1 \rightarrow 0\): If the system has undergone repairs, a replacement occurs under these cases: (1) an internal failure occurs and there is no change in external failures; and (2) an external failure implying replacement occurs, and there is no change in internal failures, in both cases the phase of the counter is...
irrelevant; (3) an external failure occurs and the counter of repairs is absorbed, while the internal failures do not change. Ordering these events appropriately, we have the block \( e \otimes (T^0 \otimes I + e \varepsilon \otimes D_2) + L^0 \otimes e \varepsilon \otimes D_1 \).

Transition 1 \( \rightarrow \) 1. If the system has undergone repairs and it continues being operational, two cases are possible: (1) there is no change in the internal and external failures, (2) a new arrival of imperfect repair without replacement occurs, the counter being operational. The block is \( I \otimes (T \oplus D_0) + L \otimes I \otimes D_1 \).

The generator \( Q \) is given by

\[
\begin{pmatrix}
(T \oplus D_0) + T^0 \otimes I + e \varepsilon \otimes (\gamma_0 D_1 + D_2) & \gamma \otimes I \otimes D_1 \\
(e \otimes (T^0 \otimes I + e \varepsilon \otimes D_2) + L^0 \otimes e \varepsilon \otimes D_1 & I \otimes (T \oplus D_0) + L \otimes I \otimes D_1
\end{pmatrix}.
\]

2.2. The stationary probability vector and performance measures

We calculate the stationary probability vector, defined by blocks and denoted by \( \pi = [\pi_0, \pi_1] \), where \( \pi_0 \) denotes the probability that the system is new, and \( \pi_1 \) the probability of undergoing a repair. The orders of this subvectors are those of the corresponding blocks of matrix \( Q \). Throughout the paper they are not relevant. Vector \( \pi e \) satisfies the equation \( \pi Q = 0 \) subject to \( \pi e = 1 \). Operating in the matricial equation we have

\[
\pi_0 \left[ (T \oplus D_0) + T^0 \otimes I + e \varepsilon \otimes (\gamma_0 D_1 + D_2) \right] + \pi_1 \left[ e \otimes (T^0 \otimes I + e \varepsilon \otimes D_2) + L^0 \otimes e \varepsilon \otimes D_1 \right] = 0,
\]

\[
\pi_0 [\gamma \otimes I \otimes D_1] + \pi_1 [I \otimes (T \oplus D_0) + L \otimes I \otimes D_1] = 0.
\]

There is no procedure to solve this linear system of equations in a closed form. In numerical examples it can be calculated in a straightforward way using computational methods. In the next section this will be calculated for a fixed \( k \), and numerically illustrated in Section 4. Given that the system is always operational, the availability is one.

The mean number of repairable failures per unit time is

\[
v = \pi_0 (\gamma \otimes I \otimes D_1) e + \pi_1 (L \otimes I \otimes D_1) e. \tag{2}
\]

The mean number of replacements per unit time is

\[
r = \pi_0 \left[ T^0 \otimes I + e \varepsilon \otimes (\gamma_0 D_1 + D_2) \right] e + \pi_1 \left[ e \otimes (T^0 \otimes I + e \varepsilon \otimes D_2) + L^0 \otimes e \varepsilon \otimes D_1 \right] e. \tag{3}
\]

3. Model with a maximum number of repairs

An interesting case derived from the general model of Section 2 is the one in which the system can undergo a limited number of repairs \( k \), where \( k \geq 1 \). In practice, it is frequent that a system can bear only a determined number of failures, in such a way that when next failure occurs it is replaced.

Let \( k \) be the maximum number of imperfect repairs, \( k \geq 1 \), that the system can undergo. In this case, the representation \((\gamma, L)\) of the PH\(d\)-distribution of the counter process is: \( \gamma = (1, 0, \ldots, 0) \), with order \( k \), and the matrix of order \( k \):

\[
L = \begin{pmatrix}
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
0 & 0 & 0 & \cdots & 0
\end{pmatrix}
\]

The generator for this model is constructed considering the number of repairs that the system has undergone, that is: 0, 1, \ldots, \( k \), named macro-states. The blocks correspond to the transition rates among the states included in these macro-states, and we can see that the generator is \( Q(k) \) given by

\[
\begin{pmatrix}
0 & T \oplus D_0 + T^0 \otimes I + e \varepsilon \otimes D_2 & I \otimes D_1 & \cdots & 0 & k - 1 & k \\
T^0 \otimes I + e \varepsilon \otimes D_2 & I \otimes D_1 & 0 & \cdots & 0 & \cdots & \cdots \\
\vdots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \\
k - 1 & T^0 \otimes I + e \varepsilon \otimes (D_1 + D_2) & 0 & \cdots & T \oplus D_0 & I \otimes D_1 & \cdots \\
k & T^0 \otimes I + e \varepsilon \otimes (D_1 + D_2) & 0 & \cdots & 0 & T \oplus D_0 & I \otimes D_1
\end{pmatrix}
\]
For this matrix the corresponding stationary probability vector by blocks is \( \pi = [\pi_0, \pi_1, \ldots, \pi_{k-1}, \pi_k] \).
This vector satisfies the matricial equation \( \pi Q(k) = 0 \).
The resultant equations are

\[
\begin{align*}
\pi_0 \left[ T \otimes D_0 + T^0 \alpha \otimes I + e \alpha \otimes D_2 \right] \\
+ \cdots + \pi_k \left[ T^0 \alpha \otimes I + e \alpha \otimes D_1 + D_2 \right] &= 0,
\end{align*}
\]

\[
\pi_j [I \otimes D_1] + \pi_{j+1} [T \otimes D_0] = 0,
\]

\[
j = 0, 1, \ldots, k - 2,
\]

\[
\pi_{k-1} [I \otimes D_1] + \pi_k [T \otimes D_0] = 0.
\]

By recurrence, it is possible to express every \( \pi_j \), \( j = 1, \ldots, k \), in terms of \( \pi_0 \) by using the above equations (except the first one). Replacing these values in the first equation, we have the following geometric matricial solution:

\[
\pi_j = \pi_0 R^j, \quad j = 0, \ldots, k,
\] (4)

with

\[
R = -[I \otimes D_1] [T \otimes D_0]^{-1},
\]

and \( \pi_0 \) being the solution of the equation:

\[
\pi_0 \left[ T \otimes D_0 + (I - R^{k+1})(I - R)^{-1} \right. \\
\left. \times (T^0 \alpha \otimes I + e \alpha \otimes D_2) + R^k (e \alpha \otimes D_1) \right] = 0.
\]

The normalization condition is \( \sum_{j=0}^{k} \pi_j e = 1 \). Given that \( \sum_{j=0}^{k} R^j = (I - R^{k+1})(I - R)^{-1} \), it can be written

\[
\pi_0 (I - R^{k+1})(I - R)^{-1} e = 1.
\]

The performance measures (2), (3), are, preserving the notation, the following: the mean number of repairable failures per unit time is

\[
v = \sum_{j=0}^{k-1} \pi_j (I \otimes D_1) e
\]

\[
= \pi_0 (I - R^k)(I - R)^{-1} (I \otimes D_1) e,
\] (5)

and the mean number of replacements per unit time is

\[
r = \sum_{j=0}^{k} \pi_j (T^0 \alpha \otimes I + e \alpha \otimes D_2) e + \pi_k (e \alpha \otimes D_1) e
\]

\[
= \pi_0 \left[ (I - R^{k+1})(I - R)^{-1} (T^0 \alpha \otimes I + e \alpha \otimes D_2) e \\
+ R^k (e \alpha \otimes D_1) e \right].
\] (6)

4. A numerical illustration

We illustrate the model with a maximum number of repairs numerically. We write a simple example to show that the elemental situations can be performed by an MAP, in spite of its complex appearance. Since the PH-distributions can be considered a particular MAP, we include the arrival of external failures by using PH-distributions. We assume that the maximum number of repairs is \( k = 4 \), and consider three phases in the random times involved. The operational random time follows a PH-distribution with representation \( (\alpha, T) \), given by

\[
\alpha = (1, 0, 0), \quad T = \begin{pmatrix} -3 & 3 & 0 \\ 0 & -2 & 2 \\ 0 & 0 & -1 \end{pmatrix}
\]

with \( T^0 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \), and the occurrence of an external failure follows a PH-distribution with representation \( (\beta, S) \), being

\[
\beta = (1, 0, 0), \quad S = \begin{pmatrix} -27 & 27 & 0 \\ 3 & -18 & 15 \\ 0 & 1 & -9 \end{pmatrix}
\]

with \( S^0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \).

The form of matrix \( T \) indicates that the operational phases are successively occupied as in a pure birth process, and the one of matrix \( S \) indicates that only instantaneous transitions among adjacent phases as in a birth-and-death process are allowed. Vectors \( T^0, S^0 \) represent the absorption rates from the transient ones (Definition 1), and their form indicates that the wear out and the external failures occur from the last transient phase.

We will assume that the external failures are repairable with probability \( p = 0.75 \) and non-repairable
Table 1

<table>
<thead>
<tr>
<th>( k )</th>
<th>( v )</th>
<th>( r )</th>
<th>( c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>4.1530</td>
<td>0.3771</td>
</tr>
<tr>
<td>1</td>
<td>1.7873</td>
<td>2.4277</td>
<td>0.4571</td>
</tr>
<tr>
<td>2</td>
<td>2.3788</td>
<td>1.8884</td>
<td>0.4789</td>
</tr>
<tr>
<td>3</td>
<td>2.6630</td>
<td>1.6442</td>
<td>0.4871</td>
</tr>
<tr>
<td>4</td>
<td>2.8201</td>
<td>1.5162</td>
<td>0.4906</td>
</tr>
<tr>
<td>5</td>
<td>2.9126</td>
<td>1.4441</td>
<td>0.4921</td>
</tr>
<tr>
<td>10</td>
<td>3.0445</td>
<td>1.3475</td>
<td>0.4934</td>
</tr>
<tr>
<td>20</td>
<td>3.0565</td>
<td>1.3394</td>
<td>0.4934</td>
</tr>
<tr>
<td>30</td>
<td>3.0566</td>
<td>1.3393</td>
<td>0.4934</td>
</tr>
<tr>
<td>50</td>
<td>3.0566</td>
<td>1.3393</td>
<td>0.4934</td>
</tr>
<tr>
<td>100</td>
<td>3.0566</td>
<td>1.3393</td>
<td>0.4934</td>
</tr>
</tbody>
</table>

with probability \( q = 0.25 \). Then, the matrices \( D_0, D_1, D_2 \) are

\[
D_0 = S, \quad D_1 = S^0 \beta p, \quad D_2 = S^0 \beta q.
\]

The mean time up to the occurrence of a wear out failure is 1.8333 unit times, and the mean time up to the occurrence of an external failure is 0.2454 unit times. For these values, and taking \( k = 4 \), we calculate the performance measures of the system. Moreover, we include costs, defining and assigning: \( c_0 = 1 \): benefit per unit time while operational, \( c_I = -0.10 \): cost per unit time in imperfect repair, \( c_R = -0.5 \): cost per replacement. The mean cost of the system is \( c = c_0 + v c_I + r c_R \).

For \( k = 4 \), using (4) we can calculate the stationary probability vector, we have

\[
\pi_0 = [0.0505, 0.0733, 0.1041, 0.0078, 0.0246, 0.0619, 0.0011, 0.0052, 0.0200],
\]

\[
\pi_1 = [0.0241, 0.0330, 0.0412, 0.0187, 0.0330, 0.0563, 0.074, 0.0158, 0.0349],
\]

\[
\pi_2 = [0.0095, 0.0130, 0.0163, 0.0151, 0.0243, 0.0376, 0.0107, 0.0197, 0.0370],
\]

\[
\pi_3 = [0.0038, 0.0052, 0.0065, 0.0097, 0.0150, 0.0222, 0.0106, 0.0183, 0.0319],
\]

\[
\pi_4 = [0.0015, 0.0020, 0.0026, 0.0056, 0.0085, 0.0123, 0.0088, 0.0147, 0.0245].
\]

The performance measures (5), (6), and the costs, are given by

\[
v = 2.8201, \quad r = 1.5162, \quad c = 0.4906.
\]

We summarize the results in Table 1 for different values of \( k \). Note that when \( k \) increases, so does \( v \), the mean number of replacements \( r \) decreases, and the benefits increase. For \( k \geq 10 \) the numbers reach stability.

Acknowledgement

This work was partially supported by Ministerio de Educación y Ciencia, Spain, under Grant MTM2004-03672. The authors would like to thank the Associate Editor for his detailed comments that have helped us to improve an earlier version of this paper.

References