OR Spektrum 2, 227–234 (1981)

OR Spektrum

On the Use of Phase Type Distributions in Reliability Modelling of Systems with Two Components*

Marcel F. Neuts and Kathleen S. Meier

Department of Mathematical Sciences, University of Delaware, Newark, Delaware 19711, USA

Received July 18, 1980 / Accepted December 18, 1980

Summary. Assuming that the time-to-failure and repair time distributions are *of phase type*, a variety of reliability models with a small number of components may be studied in terms of finite-state Markov processes. Although the state spaces of these processes are typically large, their infinitesimal generators are highly structured. By utilizing the formalism of PH-distributions, it is possible to construct efficient algorithms to evaluate a large number of quantities of interest. Some new properties of PH-distributions are also established.

Zusammenfassung. Für Verteilungen der Lebensdauer und der Reparaturdauer vom Phasentyp lassen sich bei kleiner Komponentenanzahl eine Reihe von Zuverlässigkeitsmodellen mit Hilfe von Markoff-Prozessen mit endlichem Zustandsraum untersuchen. Zwar sind die Zustandsräume dieser Prozesse groß, doch sind die Matrizen der Übergangsraten stark strukturiert. Durch Anwendung des Formalismus der Verteilungen vom Phasentyp lassen sich effiziente Algorithmen zur Berechnung einer großen Anzahl interessierender Größen entwickeln. Daneben werden einige neue Eigenschaften der Verteilungen vom Phasentyp dargestellt.

1. Introduction

There is an extensive literature on the reliability of systems with two and three components under varying assumptions on the failures and repairs [2, 3, 7]. With general distributions for the times-to-failure and the repair times, one may at best obtain highly formal expressions for the probability distributions and other quantities of interest. These expressions are rarely suitable for numerical computation. In most cases, analytically explicit expressions are obtainable only under exponential assumptions.

In this paper, we shall demonstrate the utility of the probability distributions *of phase type* (PH-distributions) in the construction of algorithmic solutions to such reliability models. The PH-distributions were introduced in [5]; a detailed discussion of their properties and of their uses in the theory of queues may be found in [6]. Only their most basic properties will be reviewed here.

The distribution $F(\cdot)$ on $[0, \infty)$ is of phase type with representation (α, T) if it is that of the time till absorption in a finite-state Markov process with generator

$$\begin{vmatrix} T & T^{\circ} \\ \boldsymbol{0} & 0 \end{vmatrix},$$

and initial probability vector (α, α_{m+1}). The nonnegative m-vector α satisfies $0 < \alpha e \leqslant 1$. Throughout this paper, e will denote the column vector with all components equal to one, whose length is determined by the context in which it appears. The matrix T of order m is nonsingular, has negative diagonal elements and nonnegative off-diagonal elements and satisfies $-Te = T^{\circ} \geqslant \boldsymbol{0}$. The distribution $F(\cdot)$ is given by

$$F(x) = 1 - \alpha \exp(Tx)\, e, \quad \text{for } x \geqslant 0.$$

In order to avoid uninteresting considerations, we shall assume that, unless otherwise noted, the PH-distributions in this paper do not have a jump at zero. The vector α is then a probability vector. We may assume, without loss of generality, that the generator $T + T^{\circ} \cdot \alpha$ is *irreducible*. The representation (α, T) is then said to be irreducible.

*) This research was supported by the National Science Foundation under Grant No. ENG-7908351 and by the Air Force Office of Scientific Research under Grant No. AFOSR-77-3236

0171-6468/81/0002/0227/$01.60

For use in the sequel, we also recall the following elementary property. If T is an irreducible, stable matrix with eigenvalue of maximum real part $-\xi < 0$, and if α is chosen to be the corresponding left eigenvector, normalized by $\alpha e = 1$, then α is a positive vector and $F(\cdot)$ is the exponential distribution with parameter ξ. This property is useful in the construction of accuracy checks on the algorithm.

Throughout this paper, a specific model will be used to illustrate the general procedure. In Sect. 4, several additional models are briefly described. They may be analyzed in exactly the same manner. Although we discuss only models with two components, the results obtained here may easily be generalized to three or more components. Unfortunately, the high order of the matrices which are obtained limits the practical uses of these results, since a large amount of computer processing time is consumed in attempting to implement algorithmic solutions.

For the specific model, we consider two units 1 and 2 and one repair facility which is utilized on a first-come, first-served basis. We may also motivate this discussion by considering two machines, 1 and 2, which use a single common tool. The times-to-failure of the second interpretation then correspond to the times during which a machine does not require the tool. A repair time corresponds to an interval during which either of the machines has the use of the tool. The second machine may be idle and waiting for the tool.

The probability distributions of the times-to-failure and of the repair times are all of phase type with the following (irreducible) representations

	Time-to-failure	*Order*	*Repair*	*Order*
Unit 1	$[\alpha(1), T(1)]$	m_1	$[\beta(1), S(1)]$	n_1
Unit 2	$[\alpha(2), T(2)]$	m_2	$[\beta(2), S(2)]$	n_2

All times-to-failure and repair times are independent random variables.

2. The Markov Process

The model with two units and a single repair facility may now be studied as a Markov process with $m_1 m_2 + m_1 n_2 + m_2 n_1 + n_1 + n_2$ states. The state space may be partitioned into five sets of states. These sets are denoted by *0*, *1*, *2*, *3* and *4*.

The set *0* consists of all pairs (i, j) with $1 \leqslant i \leqslant m_1$, $1 \leqslant j \leqslant m_2$, written in lexicographic order. It corresponds to the case where both units are functioning. The set *1* corresponds to the pairs (i, j') with $1 \leqslant i \leqslant m_1$, $1 \leqslant j' \leqslant n_2$, in lexicographic order. The set *2* corresponds to the pairs (i', j) with $1 \leqslant i' \leqslant n_1$, $1 \leqslant j \leqslant m_2$. The sets *1* and *2* correspond to the cases where the first, respectively the second, unit are functioning and the other is in repair. The set *3* consists of the single indices i', $1 \leqslant i' \leqslant n_1$. Similarly, *4* consists of the indices j', $1 \leqslant j' \leqslant n_2$. These sets correspond to the cases where Unit 1, respectively Unit 2, is in repair. The other unit is then idle and waiting for repair.

The symbol $\otimes$ will denote *the Kronecker product* of two matrices. Specifically, $L \otimes M$ stands for the matrix made up of the blocks $L_{ij}M$. The properties of Kronecker products are discussed in [1] or [4]. For notational convenience, we also introduce *the Kronecker sum* of two square matrices. If L and M are square matrices of orders m and n, then $L \oplus M$ is defined by $L \otimes I_n + I_m \otimes M$, where I_n and I_m are identity matrices of orders n and m.

The Markov process has the generator Q, given by

$$
\begin{array}{c|ccccc}
 & \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{2} & \boldsymbol{3} & \boldsymbol{4} \\
\boldsymbol{0} & T(1)\oplus T(2) & I\otimes T^{\circ}(2)\beta(2) & T^{\circ}(1)\beta(1)\otimes I & 0 & 0 \\
\boldsymbol{1} & I\otimes S^{\circ}(2)\alpha(2) & T(1)\oplus S(2) & 0 & 0 & T^{\circ}(1)\otimes I \\
\boldsymbol{2} & S^{\circ}(1)\alpha(1)\otimes I & 0 & S(1)\oplus T(2) & I\otimes T^{\circ}(2) & 0 \\
\boldsymbol{3} & 0 & S^{\circ}(1)\alpha(1)\otimes\beta(2) & 0 & S(1) & 0 \\
\boldsymbol{4} & 0 & 0 & \beta(1)\otimes S^{\circ}(2)\alpha(2) & 0 & S(2)
\end{array}
\qquad (1)
$$

Its initial probability vector depends on the initial conditions one wishes to represent. If both machines are started at the beginning of an "up" period, then the (partitioned) initial probability vector is $[\alpha(1)\otimes\alpha(2), \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}]$. By exhaustive verification, it se seen that the generator Q is irreducible.

It may be of interest to show how the entries of Q are obtained. The element $T(1) \oplus T(2)$ corresponds to the case where only a change in the phase of the time-to-failure for Unit 1 ($T(1) \otimes I$) or Unit 2 ($I \otimes T(2)$) may occur. The element $S^{\circ}(1)\alpha(1) \otimes \beta(2)$ in the second column corresponds to the end of the repair for Unit 1. That unit is then instantaneously restarted in a phase of the time-to-failure distribution $[\alpha(1), T(1)]$, according to the vector $\alpha(1)$. Unit 2, which is waiting for repair, is started in a repair phase according to the vector $\beta(2)$. A typical element of that block is of the form $S_i^{\circ}(1)\alpha_i(1)\beta_{j'}(2)$. The interpretation of the other elements of Q is similar.

Many quantities of interest to the study of the Markov process Q will be discussed in Sect. 3. Before doing so, we first prove the following theorem of general interest.

Let $F(\cdot)$ be a PH-distribution with representation (α, T) and let $X_1, \ldots, X_m$ be the total times spent in each of the states $1, \ldots, m$, prior to absorption.

Theorem 1. *The joint Laplace-Stieltjes transform* $\Phi(s) = E\left\{\exp\left[-\sum_{r=1}^{m} s_r X_r\right]\right\}$, *is given by*

$$\Phi(s) = \alpha[\Delta(s) - T]^{-1} T^{\circ}, \tag{2}$$

where $\Delta(s) = \text{diag}\,(s_1, \ldots, s_m)$.

Proof. Formula (2) is easily proved by the method of collective marks [8]. The catastrophe process is a Poisson process with rate s_j, which depends on the state j of the absorbing Markov process. Let $h_{ij}(s, t)$ be the conditional probability that at time t, the Markov process is in the (transient) state j and no catastrophe has occurred in $[0, t]$, given that it has started in the state i at time 0. The matrix $h(s, t)$ has the elements $h_{ij}(s, t)$, $1 \leqslant i, j \leqslant m$.

By conditioning on the time of the first transition in $(0, t)$ (if there is one), we obtain that

$$h_{ij}(s, t) = \delta_{ij} \exp[(T_{ii} - s_i)t] + \sum_{\nu \neq i} \int_0^t \exp[(T_{ii} - s_i)(t - u)] T_{i\nu} h_{\nu j}(s, u)\, du, \tag{3}$$

for $t \geqslant 0$, $s_r \geqslant 0$, $1 \leqslant r \leqslant m$, $1 \leqslant i, j \leqslant m$.

Multiplying both sides of (3) by $\exp[(s_i - T_{ii})t]$ and differentiating with respect to t, readily yields

$$h'_{ij}(s, t) = \sum_{\nu=1}^{m} [T - \Delta(s)]_{i\nu} h_{\nu j}(s, t), \tag{4}$$

which implies that

$$h(s, t) = \exp\{[T - \Delta(s)]t\}, \quad \text{for } t \geqslant 0. \tag{5}$$

Finally $\Phi(s)$, which is also probability that no catastrophe occurs before the Markov process is absorbed, is given by

$$\Phi(s) = \alpha \int_0^{\infty} h(s, t) T^{\circ} dt = \alpha[\Delta(s) - T]^{-1} T^{\circ}. \tag{6}$$

Corollary 1. *Let* $F(\cdot)$ *be a* PH-*distribution. The m phases (transient states) are partitioned into two non-empty sets and the vectors* α, T° *and the matrix* T *are accordingly partitioned as*

$$\alpha = [\tilde{\alpha}(1), \tilde{\alpha}(2)],$$

and

$$T = \begin{vmatrix} T(1,1) & T(1,2) \\ T(2,1) & T(2,2) \end{vmatrix}, \quad T^{\circ} = \begin{vmatrix} T^{\circ}(1) \\ T^{\circ}(2) \end{vmatrix}.$$

Each of the total lengths of time spent in the sets of the partition then has a PH-distribution. For the first set, the representation (γ, L) of that PH-distribution is given by

$$\gamma = \tilde{\alpha}(1) - \tilde{\alpha}(2) T^{-1}(2,2) T(2,1),$$
$$L = T(1,1) - T(1,2) T^{-1}(2,2) T(2,1). \tag{7}$$

For that distribution, the jump at zero is given by $\gamma^{\circ} = -\alpha(2) T^{-1}(2,2) T^{\circ}(2)$. The vector L°, corresponding to L is equal to $L^{\circ} = T^{\circ}(1) - T(1,2) T^{-1}(2,2) T^{\circ}(2)$.

Proof. By Theorem 1, the Laplace-Stieltjes transform of the total time spent in the first subset is given by

$$[\tilde{\alpha}(1), \tilde{\alpha}(2)] \begin{vmatrix} sI - T(1,1) & -T(1,2) \\ -T(2,1) & -T(2,2) \end{vmatrix}^{-1} \begin{vmatrix} T^{\circ}(1) \\ T^{\circ}(2) \end{vmatrix}.$$

By routine matrix manipulations, this may be put into the form $\gamma^{\circ} + \gamma(sI - L)^{-1} L^{\circ}$, where the quantities γ°, γ, L and L° are as defined above.

The nonsingularity of the matrix L is shown as follows. The matrices $T(1, 1)$ and $T(2, 2)$ are nonsingular, since they are principal submatrices of the stable matrix T. The matrix L is readily seen to be semi-stable. The matrix $-T(1, 2)T^{-1}(2, 2)T(2, 1)$ is nonnegative. The eigenvalue of maximum real part of L is therefore at most equal to the corresponding eigenvalue of $T(1, 1)$. Since the latter is negative, the matrix L is nonsingular.

From (2) a number of useful moment formulas may be derived. Setting $[\Delta(s) - T]^{-1} = M(s)$, we readily obtain that

$$\frac{\partial}{\partial s_i} M(s) = -M(s)\, \Delta(e_i) M(s), \quad \text{for } 1 \leqslant i \leqslant m,$$

and

$$\frac{\partial^2}{\partial s_i \partial s_j} M(s) = M(s)\Delta(e_i)M(s)\Delta(e_j)M(s) + M(s)\Delta(e_j)M(s)\Delta(e_i)M(s),$$

for $1 \leqslant i, j \leqslant m$. The diagonal matrix $\Delta(e_i)$ has the components of the unit e_i on the diagonal.

Corollary 2.

$$E(X_i) = (-\alpha T^{-1})_i, \quad \text{for } 1 \leqslant i \leqslant m,$$

$$E(X_i X_j) = (\alpha T^{-1})_i (T^{-1})_{ij} + (\alpha T^{-1})_j (T^{-1})_{ji}, \tag{8}$$

for $1 \leqslant i, j \leqslant m$.

Proof. By differentiation in Formula (2).

Remark. The matrix Z with $Z_{ij} = (\alpha T^{-1})_i (T^{-1})_{ij}$, may be efficiently computed. We note that $Z = \Delta(\alpha T^{-1}) T^{-1}$, so that Z is evaluated by solving the linear system

$$ZT = \Delta(\alpha T^{-1}), \tag{9}$$

which is suitable for e.g. Gauss-Seidel iteration.

3. Algorithmic Procedures

3.1. The Stationary Distribution

The stationary probability vector $\boldsymbol{x}$ of the generator Q is the unique (positive) solution to the system of equations $\boldsymbol{x}Q = \boldsymbol{0}, \boldsymbol{x}\boldsymbol{e} = 1$. In view of the high order of the matrix Q, it is essential to use its special structure to evaluate the components of $\boldsymbol{x}$. We first partition the vector $\boldsymbol{x}$ as $[\boldsymbol{x}(0), \boldsymbol{x}(1), \boldsymbol{x}(2), \boldsymbol{x}(3), \boldsymbol{x}(4)]$, according to the sets of states $0, 1, 2, 3, 4$, and so obtain

$$\begin{aligned}
&\boldsymbol{x}(0)[T(1) \oplus T(2)] + \boldsymbol{x}(1)[I \otimes \boldsymbol{S}^{\circ}(2)\boldsymbol{\alpha}(2)] \\
&\quad + \boldsymbol{x}(2)[\boldsymbol{S}^{\circ}(1)\boldsymbol{\alpha}(1) \otimes I] = \boldsymbol{0}, \\
&\boldsymbol{x}(0)[I \otimes \boldsymbol{T}^{\circ}(2)\boldsymbol{\beta}(2)] + \boldsymbol{x}(1)[T(1) \oplus S(2)] \\
&\quad + \boldsymbol{x}(3)[\boldsymbol{S}^{\circ}(1)\boldsymbol{\alpha}(1) \otimes \boldsymbol{\beta}(2)] = \boldsymbol{0}, \\
&\boldsymbol{x}(0)[\boldsymbol{T}^{\circ}(1)\boldsymbol{\beta}(1) \otimes I] + \boldsymbol{x}(2)[S(1) \oplus T(2)] \\
&\quad + \boldsymbol{x}(4)[\boldsymbol{\beta}(1) \otimes \boldsymbol{S}^{\circ}(2)\boldsymbol{\alpha}(2)] = \boldsymbol{0}, \\
&\boldsymbol{x}(2)[I \otimes \boldsymbol{T}^{\circ}(2)] + \boldsymbol{x}(3)S(1) = \boldsymbol{0}, \\
&\boldsymbol{x}(1)[\boldsymbol{T}^{\circ}(1) \otimes I] + \boldsymbol{x}(4)S(2) = \boldsymbol{0}.
\end{aligned} \tag{10}$$

Since $S(1)$ and $S(2)$ are nonsingular, the last two equations may be used to express $\boldsymbol{x}(3)$ and $\boldsymbol{x}(4)$ in terms of $\boldsymbol{x}(2)$ and $\boldsymbol{x}(1)$ respectively. Upon substitution into the second and third equations, some worthwhile simplifications occur. For example, we see that

$$\begin{aligned}
&\boldsymbol{x}(3)[\boldsymbol{S}^{\circ}(1)\boldsymbol{\alpha}(1) \otimes \boldsymbol{\beta}(2)] \\
&\quad = -\boldsymbol{x}(2)[I \otimes \boldsymbol{T}^{\circ}(2)]S^{-1}(1)[\boldsymbol{S}^{\circ}(1)\boldsymbol{\alpha}(1) \otimes \boldsymbol{\beta}(2)].
\end{aligned}$$

However,

$$\begin{aligned}
&S^{-1}(1)[\boldsymbol{S}^{\circ}(1)\boldsymbol{\alpha}(1) \otimes \boldsymbol{\beta}(2)] \\
&\quad = S^{-1}(1)\boldsymbol{S}^{\circ}(1)\boldsymbol{\alpha}(1) \otimes \boldsymbol{\beta}(2) = -\boldsymbol{e}\boldsymbol{\alpha}(1) \otimes \boldsymbol{\beta}(2),
\end{aligned}$$

since $S(1)\boldsymbol{e} + \boldsymbol{S}^{\circ}(1) = \boldsymbol{0}$. Therefore, we obtain

$$\boldsymbol{x}(3)[\boldsymbol{S}^{\circ}(1)\boldsymbol{\alpha}(1) \otimes \boldsymbol{\beta}(2)] = \boldsymbol{x}(2)[\boldsymbol{e}\boldsymbol{\alpha}(1) \otimes \boldsymbol{T}^{\circ}(2)\boldsymbol{\beta}(2)].$$

After a similar simplification in the third equation, the second and third equations yield

$$\begin{aligned}
&\boldsymbol{x}(0)[I \otimes \boldsymbol{T}^{\circ}(2)\boldsymbol{\beta}(2)] + \boldsymbol{x}(1)[T(1) \oplus S(2)] \\
&\quad + \boldsymbol{x}(2)[\boldsymbol{e}\boldsymbol{\alpha}(1) \otimes \boldsymbol{T}^{\circ}(2)\boldsymbol{\beta}(2)] = \boldsymbol{0}, \\
&\boldsymbol{x}(0)[\boldsymbol{T}^{\circ}(1)\boldsymbol{\beta}(1) \otimes I] + \boldsymbol{x}(1)[\boldsymbol{T}^{\circ}(1)\boldsymbol{\beta}(1) \otimes \boldsymbol{e}\boldsymbol{\alpha}(2)] \\
&\quad + \boldsymbol{x}(2)[S(1) \oplus T(2)] = \boldsymbol{0}.
\end{aligned}$$

In order to solve the system consisting of the first three equations and the normalizing equation

$$\boldsymbol{x}(0)\boldsymbol{e} + \boldsymbol{x}(1)\boldsymbol{e} + \boldsymbol{x}(2)\boldsymbol{e} + \boldsymbol{x}(3)\boldsymbol{e} + \boldsymbol{x}(4)\boldsymbol{e} = 1, \tag{11}$$

we further partition the vector $\boldsymbol{x}(0)$, $\boldsymbol{x}(1)$ and $\boldsymbol{x}(2)$ as follows

$$\begin{aligned}
&\boldsymbol{x}(0) = [\boldsymbol{x}_1(0), \ldots, \boldsymbol{x}_{m_1}(0)], \quad \boldsymbol{x}_i(0) \text{ is an } m_2\text{-vector}, \\
&\boldsymbol{x}(1) = [\boldsymbol{x}_1(1), \ldots, \boldsymbol{x}_{m_1}(1)], \quad \boldsymbol{x}_i(1) \text{ is an } n_2\text{-vector}, \\
&\boldsymbol{x}(2) = [\boldsymbol{x}_1(2), \ldots, \boldsymbol{x}_{n_1}(2)], \quad \boldsymbol{x}_i(2) \text{ is an } m_2\text{-vector}.
\end{aligned}$$

By using the particular structure of the Kronecker sums and products, which arise as coefficient matrices in these equations, we are led to

$$\begin{aligned}
&\boldsymbol{x}_i(0)[T_{ii}(1)I + T(2)] + \sum_{\substack{j=1 \\ j \neq i}}^{m_1} \boldsymbol{x}_j(0)T_{ji}(1) \\
&+ \boldsymbol{x}_i(1)\boldsymbol{S}^{\circ}(2)\boldsymbol{\alpha}(2) + \sum_{k=1}^{n_1} \boldsymbol{x}_k(2)S_k^{\circ}(1)\alpha_i(1) = \boldsymbol{0}, \\
&\text{for } 1 \leqslant i \leqslant m_1, \\
&\boldsymbol{x}_i(0)\boldsymbol{T}^{\circ}(2)\boldsymbol{\beta}(2) + \boldsymbol{x}_i(1)[T_{ii}(1)I + S(2)] \\
&+ \sum_{\substack{j=1 \\ j \neq i}}^{m_1} \boldsymbol{x}_j(1)T_{ji}(1) + \sum_{k=1}^{n_1} \boldsymbol{x}_k(2)\boldsymbol{T}^{\circ}(2)\boldsymbol{\beta}(2)\alpha_i(1) = \boldsymbol{0}, \\
&\text{for } 1 \leqslant i \leqslant m_1,
\end{aligned} \tag{12}$$

$$\begin{aligned}
&\sum_{j=1}^{m_1} \boldsymbol{x}_j(0)T_j^{\circ}(1)\beta_i(1) + \sum_{j=1}^{m_1} \boldsymbol{x}_j(1)T_j^{\circ}(1)\beta_i(1)\boldsymbol{e}\boldsymbol{\alpha}(2) \\
&+ \boldsymbol{x}_i(2)[S_{ii}(1)I + T(2)] + \sum_{\substack{k=1 \\ k \neq i}}^{n_1} \boldsymbol{x}_k(2)S_{ki}(1) = \boldsymbol{0}, \\
&\text{for } 1 \leqslant i \leqslant n_1.
\end{aligned}$$

The equations (12) will now be recast into a form which makes them appropriate for solution by block

Gauss-Seidel iteration. We do not wish, however, to store the inverses of all the diagonal blocks $T_{ii}(1)I + T(2)$, $T_{ii}(1)I + S(2)$, for $1 \leqslant i \leqslant m_1$, and $S_{ii}(1)I + T(2)$, for $1 \leqslant i \leqslant n_1$. At the price of slightly slower convergence, we shall evaluate only the three inverses

$$M(1) = -[\tau(1)I + T(2)]^{-1},$$

$$M(2) = -[\tau(1)I + S(2)]^{-1}, \quad \text{where } \tau(1) = \min_i T_{ii}(1), \tag{13}$$

$$M(3) = -[\tau(2)I + T(2)]^{-1}, \quad \text{where } \tau(2) = \min_i S_{ii}(1),$$

and we rewrite the equations (12) as

$$x_i(0) = \left\{ -\tau(1)x_i(0) + \sum_{j=1}^{m_1} x_j(0)T_{ji}(1) \right.$$

$$\left. + [x_i(1)S^{\circ}(2)]\alpha(2) = \alpha_i(1)\left[\sum_{k=1}^{n_1} x_k(2)S_k^{\circ}(1)\right] \right\} M(1),$$

for $1 \leqslant i \leqslant m_1$,

$$x_i(1) = \left\{ [x_i(0)T^{\circ}(2)]\beta(2) - \tau(1)x_i(1) \right.$$

$$\left. + \sum_{j=1}^{m_1} x_j(1)T_{ji}(1) + \left[\sum_{k=1}^{n_1} x_k(2)T^{\circ}(2)\beta(2)\right]\alpha_i(1) \right\} M(2),$$

for $1 \leqslant i \leqslant m_1$, (14)

$$x_i(2) = \left\{ \beta_i(1)\left[\sum_{j=1}^{m_1} x_j(0)T_j^{\circ}(1)\right] \right.$$

$$+ \beta_i(1)\left[\sum_{j=1}^{m_1} x_j(1)eT_j^{\circ}(1)\right]\alpha(2)$$

$$\left. -\tau(2)x_i(2) + \sum_{k=1}^{n_1} x_k(2)S_{ki}(1) \right\} M(3),$$

for $1 \leqslant i \leqslant n_1$.

We solved the equations (14) by block Gauss-Seidel iteration. The successive solution vectors are kept within a compact polytope by forcing them to satisfy the normalizing equation (11). The successive vectors $x(3)$ and $x(4)$ are easily computed from the last two equations in (10). We also note that in (14), several quantities are common to a number of the equations. We have indicated such quantities by enclosing them within square brackets. It is highly advantageous to plan the computer code, so that these quantities are evaluated only once at each iteration.

It is clear that the five quantities $x(0)e, x(1)e, x(2)e$, $x(3)e$ and $x(4)e$ are of particular interest. They are the steady-state probabilities that the system will be found in one of the physically meaningful states, described by the sets *0*, *1*, *2*, *3* and *4*. Some other inner products are also of interest. To give only one example, the conditional steady-state probability that Unit 2 is working at time t, given that Unit 1 fails at that time, is equal to

$$\frac{x(0)[T^{\circ}(1) \otimes e]}{x(0)[T^{\circ}(1) \otimes e] + x(1)[T^{\circ}(1) \otimes e]}.$$

We note that $T^{\circ}(1) \otimes e$ is *not* a common factor in the denominator, since the vector e in the first term is not of the same dimension as in the second term.

3.2. The Time-dependent Solution

The time-dependent probability vector $\hat{x}(t)$ is found by numerical integration of the Chapman-Kolmogorov equation

$$\hat{x}'(t) = \hat{x}(t)Q, \quad \text{for } t \geqslant 0. \tag{15}$$

The initial conditions need to be chosen so as to reflect the physical condition of the system at time $t = 0$. For example, if both units are "new" at $t = 0$, the vector $\hat{x}(0)$ is chosen as

$$\hat{x}(0) = [\alpha(1) \otimes \alpha(2), 0, 0, 0, 0].$$

The equation (15) may be solved by any one of a number of integration methods. It is clear that the particular structure of the matrix Q could be exploited. We implemented the classical Runge-Kutta procedure of order four and programmed a subroutine for the left multiplication of a row vector and the matrix Q. Computation was halted as soon as the difference between the vector $\hat{x}(t)$ and the stationary probability vector x became small over an interval of sufficient length.

The initial step size h was chosen heuristically to be a small fraction, say 1/100, of an easily computable length L which has a physical significance to the problem at hand. For our computations, we chose L to be the mean time to system failure. How that quantity may be computed is discussed below.

3.3. Distribution of the Time to System Failure

When both machines are inoperational, the system is said to have failed, in order to study the distribution $H(\cdot)$ of the time to failure, we lump together the sets of states *3* and *4* of the Markov process Q into a single absorbing state $*$. We so obtain the absorbing Markov process with generator

$$Q^* = \begin{array}{c|ccc} & \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{2} \\ \boldsymbol{0} & T(1)\oplus T(2) & I\otimes T^\circ(2)\beta(2) & T^\circ(1)\beta(1)\otimes I \\ \boldsymbol{1} & I\otimes S^\circ(2)\alpha(2) & T(1)\oplus S(2) & 0 \\ \boldsymbol{2} & S^\circ(1)\alpha(1)\otimes I & 0 & S(1)\oplus T(2) \\ * & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{array} \quad \begin{array}{c} * \\ \boldsymbol{0} \\ T^\circ(1)\otimes e \\ e\otimes T^\circ(2) \\ \boldsymbol{0} \end{array}. \tag{16}$$

The initial probability vector q^* of that process may again be chosen to reflect initial conditions of interest. With both items "new" q^* is chosen to be $[\alpha(1)\otimes\alpha(2), 0, 0, 0]$. We may also "enter" the stationary version of the process Q and condition on the event that both units are functioning. The vector q^*, which correponds to that choice of initial conditions is given by

$$\left[\frac{x(0)}{x(0)e}, 0, 0, 0\right].$$

It is clear that the probability distribution $H(\cdot)$ is of phase type. Its representation (γ, K) is readily obtained from q^* and Q^*. The distribution $H(\cdot)$ is computed by integrating the highly structured system of differential equations

$$\nu'(x) = \nu(x)K, \quad \nu(0) = \gamma, \tag{17}$$

and setting $H(x) = 1 - \nu(x)e$.

It is advisable to compute the mean and maybe higher moments of $H(\cdot)$ beforehand. Since the mean h_1' of $H(\cdot)$ is given by $h_1' = -\gamma K^{-1}e$, we evaluate the vector $u = -\gamma K^{-1}$, by solving the highly structured system of linear equations $uK = -\gamma$. That system is transformed in exactly the same manner as discussed in 3.1 above and solved by block Gauss-Seidel iteration. The vector u also plays a role in the next discussion.

Knowledge of the mean h_1' also provides us with a physically meaningful choice of the step size for the integration of the differential equations (15) and (17).

3.4. The Total Time that Both Units are Operational before System Failure

Suppose that the time origin is chosen so that at least one of the two units is operational. We may be interested in the total time that *both* machines are operational up to the first time of system failure. This corresponds to specifying an initial probability vector over the sets of states $\boldsymbol{0}$, $\boldsymbol{1}$ and $\boldsymbol{2}$ in Q^* and to studying the total sojourn time X in the set of states $\boldsymbol{0}$, *prior to absorption.*

By Corollary 1, the random variable X has a PH-distribution. Its representation is obtained from the matrix Q^*, but requires a fair amount of computation, which should be planned with care. Before discussing this in greater detail, we note that the *mean* of X is very easily obtained. Let the vector u° be the restriction of the vector u (see 3.3 above) to those components with indices in $\boldsymbol{0}$. Such a component u_i° is the mean time spent in the state $i \in \boldsymbol{0}$, prior to absorption of the Markov process Q^*. The *mean* of X is therefore given by $u^\circ e$.

By Corollary 1, the representation $(\hat{\gamma}, L)$ of the distribution $G(\cdot)$ of X is given by

$$L = T(1,1) - T(1,2)T^{-1}(2,2)T(2,1), \tag{18}$$

where

$$T(1,1) = T(1)\oplus T(2),$$

$$T(1,2) = |I\otimes T^\circ(2)\beta(2) \quad T^\circ(1)\beta(1)\otimes I|,$$

$$T(2,2) = \begin{vmatrix} T(1)\oplus S(2) & 0 \\ 0 & S(1)\oplus T(2) \end{vmatrix},$$

$$T(2,1) = \begin{vmatrix} I\otimes S^\circ(2)\alpha(2) \\ S^\circ(1)\alpha(1)\otimes I \end{vmatrix}.$$

The initial probability vector $\hat{\gamma}$ may again be chosen to reflect physically meaningful initial conditions. In particular, if both items are started "new", then $\hat{\gamma} = \alpha(1)\otimes\alpha(2)$.

The computation and the application of the representation $(\hat{\gamma}, L)$ present some interesting features. The matrix L will, in general, not have any particular structure left. Since the order of L is $m_1 m_2$, the computation of the distribution $G(\cdot)$ may require substantial processing time.

In evaluating L, there is merit in computing the nonnegative matrix

$$\begin{vmatrix} U \\ V \end{vmatrix} = -T^{-1}(2,2)T(2,1),$$

by solving the linear system

$$[T(1)\oplus S(2)]U = -I\otimes S^\circ(2)\alpha(2),$$

$$[S(1)\oplus T(2)]V = -S^\circ(1)\alpha(1)\otimes I,$$

through judicious use of its particular structure. The di-

mensions of the matrices U and V are $m_1 n_2 \times m_1 m_2$ and $m_2 n_1 \times m_1 m_2$ respectively. The matrix U is partitioned into an $m_1 \times m_1$ array of blocks of dimensions $n_2 \times m_2$. The matrix V is partitioned into an $n_1 \times m_1$ array of square blocks of order m_2.

As in Sect. 3.1 above, the linear equations (20) can be rewritten into a form which is well-suited for block Gauss-Seidel iteration. The details are entirely similar and will be omitted.

The computation of $G(\cdot)$ itself again requires the numerical integration of a system of linear differential equations with constant coefficients. As we may wish to consider *several* initial probability vectors $\hat{\gamma}$, there is clearly an advantage in setting

$$\nu(x) = \exp(Lx)e, \quad \text{for } x \geqslant 0$$

and in solving the system of differential equations

$$\nu'(x) = L\nu(x), \quad x \geqslant 0, \quad \nu(0) = e.$$

The distribution $G(\cdot)$ is then given by

$$G(x) = 1 - \hat{\gamma}\nu(x), \quad \text{for } x \geqslant 0.$$

Remark. Corollary 1 may similarly be used to characterize the distributions of other times of interest, such as the total time one unit (or both) have spent in repair prior to system failure.

3.5. Accuracy Checks

Algorithms for general PH-distributions have a powerful accuracy check, which is based on the property mentioned in Sect. 1. The models, in which PH-distributions are used, may usually be solved explicitly under exponential assumptions. We may do this in two ways. With the exponential distributions in their usual simple form, we first obtain the analytic solution explicitly. Next, we implement the general algorithm, but choose the representations of the various PH-distributions *so that they are in fact exponential.* The general algorithm does not utilize this fact in any manner, but the two sets of numerical results should of course agree.

For the sake of illustration, let $T(1)$, $T(2)$, $S(1)$ and $S(2)$ be the scalars $-\lambda_1$, $-\lambda_2$, $-\mu_1$, and $-\mu_2$ respectively and $\alpha(1) = \alpha(2) = \beta(1) = \beta(2) = 1$. The matrix Q of (1) then becomes

$$Q = \begin{vmatrix} -\lambda_1-\lambda_2 & \lambda_2 & \lambda_1 & 0 & 0 \\ \mu_2 & -\lambda_1-\mu_2 & 0 & 0 & \lambda_1 \\ \mu_1 & 0 & -\mu_1-\lambda_2 & \lambda_2 & 0 \\ 0 & \mu_1 & 0 & -\mu_1 & 0 \\ 0 & 0 & \mu_2 & 0 & -\mu_2 \end{vmatrix},$$

and its stationary probability vector x is given by

$$x_0 = \frac{(\mu_1+\lambda_2)(\mu_1\mu_2+\lambda_1\mu_2+\lambda_1\mu_1)}{\lambda_1^2\mu_1} x_2 - \frac{\mu_2}{\lambda_1},$$

$$x_1 = \frac{\mu_2}{\lambda_1} - \frac{\mu_2(\mu_1+\lambda_1)(\mu_1+\lambda_2)}{\lambda_1^2\mu_1} x_2,$$

$$x_2 = \frac{\lambda_1\mu_1(\lambda_1+\mu_2)(\lambda_2+\mu_2) - \lambda_1^2\lambda_2\mu_1}{(\lambda_1+\mu_1)(\lambda_1+\mu_2)(\lambda_2+\mu_1)(\lambda_2+\mu_2) - \lambda_1^2\lambda_2^2},$$

$$x_3 = \frac{\lambda_2}{\mu_1} x_2,$$

$$x_4 = 1 - \frac{(\lambda_1+\mu_1)(\lambda_2+\mu_1)}{\lambda_1\mu_1} x_2.$$

In our test problems, we chose $T(1)$, $T(2)$, $S(1)$ and $S(2)$ to be irreducible matrices. We computed their eigenvalues of maximum real part and corresponding left eigenvectors, chosen so as to make the PH-distributions exponential. These eigenvalues, respectively $-\lambda_1$, $-\lambda_2$, $-\mu_1$, and $-\mu_2$, were used to evaluate x from the preceding formulas. After debugging of the program, the solutions by the general algorithm were found to be in excellent agreement with those by the particular formulas.

The same idea may be used to construct accuracy checks for the other items, whose computation we have discussed.

The algorithm has a large number of other accuracy checks. The obvious ones need not be discussed. For example, we also interchanged the matrices $T(1)$ and $S(1)$ with $T(2)$ and $S(2)$, to verify that the numerical results exhibited the symmetry which is to be expected.

4. Related Models

Many of the small-scale reliability models, which have been discussed in the literature, may be analyzed in the same manner as the model treated above. The specific structure of the matrix Q is different, but the formal manipulations leading to a feasible algorithm are entirely similar. Two related models will now be briefly described.

4.1. Two Units – One Active, the Other in Cold Standby

The time-to-failure and the repair time distribution are PH with representations (α, T) and (β, S) respectively.

The sets of states *0*, *1* and *2* now have the following significance.

0: neither unit has failed,

1: standby unit in use, the other unit in repair,

2: both units failed. The first to fail is in repair, the other is waiting.

The matrix Q is now given by

$$\begin{array}{c|ccc} & \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{2} \\ \boldsymbol{0} & T & T^{\circ}\alpha \otimes \beta & 0 \\ \boldsymbol{1} & I \otimes S^{\circ} & T \oplus S & T^{\circ} \otimes I \\ \boldsymbol{2} & 0 & \alpha \oplus S^{\circ}\beta & S \end{array}$$

The number and ordering of the states in *0*, *1* and *2* is clear from the display.

4.2. Two Units with One in Cold Standby. Unit 1 is Used Whenever it is Operative

This is a more involved example. Unit 1 is preferred and is in use whenever it is operative. Unit 2 is a back-up unit. It can be used only when Unit 1 is in repair; its wear during such periods is accumulated.

The following PH-distributions are parameters of the model.

	Time-to-failure	*order*	*Repair time*	*order*
Unit 1	$[\alpha(1), T(1)]$	m_1	$[\beta(1), S(1)]$	n_1
Unit 2	$[\alpha(2), T(2)]$	m_2	$[\beta(2), S(2)]$	n_2

As in Sect. 2, sets of states *0*, *1*, *2*, *3* are considered. They have the following significance.

0: Unit 1 in use, Unit 2 on standby,

1: Unit 1 in repair, Unit 2 in use,

2: Unit 1 in use, Unit 2 in repair,

3: Unit 1 in repair, Unit 2 waiting for repair.

The matrix Q is now given by

$$Q = \begin{array}{c|cccc} & \boldsymbol{0} & \boldsymbol{1} & \boldsymbol{2} & \boldsymbol{3} \\ \boldsymbol{0} & T(1) \otimes I & T^{\circ}(1)\beta(1) \otimes I & 0 & 0 \\ \boldsymbol{1} & S^{\circ}(1)\alpha(1) \otimes I & S(1) \oplus T(2) & 0 & 0 \\ \boldsymbol{2} & I \otimes S^{\circ}(2)\alpha(2) & 0 & T(1) \oplus S(2) & I \otimes T^{\circ}(2)\beta(2) \\ \boldsymbol{3} & 0 & 0 & S^{\circ}(1)\alpha(1) \otimes I & T^{\circ}(1)\beta(1) \otimes I \\ & & & & S(1) \otimes I \end{array}$$

We see that this model differs from that treated above only in the priority of Unit 1, both for use and for repair. The numerical computations for both model proceed along entirely similar lines and some portions of the two algorithms are identical. This may be put to use to compare *numerically* the merits of the two operating policies. Such a comparison is not analytically feasible.

Acknowledgements. This paper had its origin in conversations which one of us (M. F. N.) had with Professors A. Chaudhuri of the Indian Institute of Management, Calcutta, B. Epstein of Technion, Haifa and Dr. N. Ravichandran of the Indian Institute of Technology, Madras. Their interest and comments are acknowledged with appreciation.

References

1. Bellman R (1960) Introduction to matrix analysis. McGraw-Hill, New York
2. Epstein B, Hosford J (1960) Reliability of some two-unit redundant systems. Proc. 6th National Symp. on Reliability and Quality Control, Washington, DC, pp 469–476
3. Gaver DP Jr (1963) Time to failure and availability of paralleled systems with repair. IEEE Trans Reliability R-12:30–38
4. Marcus M, Minc H (1964) A survey of matrix theory and matrix inequalities. Allyn and Bacon, Boston, MA
5. Neuts MF (1975) Probability distributions of phase type. Liber Amicorum Prof. Emeritus H. Florin, Dept. Math. Univ. Louvain, Belgium, pp 173–206
6. Neuts MF (1981) Matrix-geometric solutions in stochastic models. An algorithmic approach. The Johns Hopkins University Press, Baltimore, MD
7. Ohi F, Kodama M, Nishida T (1978) Several results on the dependent two-unit system with single repair facility. Trans. IECE Jpn E 61:710–716
8. Runnenburg JT (1965) On the use of the method of collective marks in queueing theory. Proc. Symp. on Congestion Theory, 1964. University of North Carolina Press, Chapel Hill, NC, pp 399–438