ABDELAZIZ NASROALLAH

Simulation of transient performance measures for stiff Markov chains

RAIRO. Recherche opérationnelle, tome 34, n° 4 (2000), p. 385-396

<http://www.numdam.org/item?id=RO_2000__34_4_385_0>
SIMULATION OF TRANSIENT PERFORMANCE MEASURES FOR STIFF MARKOV CHAINS (*)

by Abdelaziz NASROALLAH

Communicated by P. TOLLA

Abstract. We consider the simulation of transient performance measures of high reliable fault-tolerant computer systems. The most widely used mathematical tools to model the behavior of these systems are Markov processes. Here, we deal basically with the simulation of the mean time to failure (MTTF) and the reliability, R(t), of the system at time t. Some variance reduction techniques are used to reduce the simulation time. We will combine two of these techniques: Importance Sampling and Conditioning Technique. The resulting hybrid algorithm performs significant reduction of simulation time and gives stable estimations.

Keywords: Reliability, stiff Markovian models, performance measures, variance reduction, Monte-Carlo simulation.

1. INTRODUCTION

There is a growing interest in evaluating the performances of fault-tolerant computer systems, i.e. systems that are able to recover from a non-operational state after a fault. These systems are used in many fields such as general purposes computers or telephone switching. The high technology involves high reliability for these systems. The Markovian models are the most appropriate tools to study such systems (cf. Arlat 1987). A consequence of high reliability is that the Markov chains at issue are stiff, i.e. the failure and recovery rates have very different orders of magnitude. In many cases, the stiffness makes analytical resolution almost impossible. So, in this case, a Monte-Carlo simulation is appealed in order to get estimations. But, a very long time is necessary for obtaining good estimations from a standard simulation. There exist many variance reduction techniques which reduce the simulation time. Some of these techniques are discussed in

(*) Received May, 1997.

(Département de Mathématiques, Université Cadi Ayyad, Faculté des Sciences Semlalia, boulevard du Prince Moulay Abdellah, BP. 2390, Marrakech, Morocco.)

Recherche opérationnelle/Operations Research, 0399-0559/00/04/$ 7.00 © EDP Sciences, 2000
Hammersley (1979) and in Rubinstein (1981). The Importance Sampling (IS) is an adequate technique in the simulation of Markovian models. It has been applied, with success, to simulate stationary performances Markovian models, and it is well presented in Nasroallah (1991) and in Glynn et al. (1989).

In this paper, we focus on the simulation of transient performance measures of high reliable systems. We basically deal with the estimation of \( MTTF \) and the reliability of the system at time \( t \), \( R(t) \), by simulating a discrete Markov process. There is also some variance reduction techniques which are adapted to simulation of transient performances. For example Forced Transitions (FT) and Conditioning Technique (CT) are often used to reduce the effect of the parameter \( t \), and consequently reduce the simulation time (cf. Lewis et al. 1984 and Nasroallah 1991).

The main aim of this work is to combine IS and CT in order to reduce the time complexity of the simulation caused by stiffness and parameter \( t \). The resulting hybrid procedure, noted IS + CT, leads to significant reduction of simulation time and performs stable estimations.

The paper is organized as follows: in Section 2 we give a description of the Markovian model used. The \( MTTF \) and \( R(t) \) expressions are studied in Section 3. Section 4 is concerned with the approximation adopted for the confidence intervals used in simulation. Some simulation results are given in the Section 5. We close this paper by giving a brief conclusion in Section 6 and some references.

2. MARKOVIAN MODEL

Let \( X = \{X_t; t \geq 0\} \) be an irreducible continuous time Markov chain (CTMC) with finite state space \( \mathcal{E} \). We assume that \( \mathcal{E} \) can be represented as \( \mathcal{E} = \mathcal{O} \cup \mathcal{F} \) where \( \mathcal{O} \) and \( \mathcal{F} \) are the operational and unoperational state space respectively. The process \( X \) is used to study the behavior of a fault-tolerant computer system. Since the system is assumed to be highly reliable, then \( X \) is a stiff CTMC. Let \( Q = (q_{ij})_{i,j \in \mathcal{E}} \) be the generator of \( X \) and assume that all the components of the system are operational at time 0. We denote this initial state \( s_0 \).

It is easy to see that all the hypothesis assumed here are reasonable and realistic. Since the Markovian process \( X \) is irreducible and \( \mathcal{E} \) is finite, then \( X \) is a regenerative process and each state can be considered as a regeneration point of \( X \). Let \( \tau_A \) denote the first return time of \( X \) to the subset \( A \) of \( \mathcal{E} \), after leaving the state \( s_0 \).
It is mentioned in Hordijk et al. (1976) that it is more convenient to simulate a discrete process as approximation to the continuous one. In our case, we simulate the embedded discrete time Markov chain (DTMC) \( Z = \{Z_n; n \in \mathbb{N}\} \) as approximation to \( X \). The process \( Z \) is characterized by its transition matrix \( P = (p_{ij})_{i,j \in E} \) where:

\[
p_{ij} = \frac{-q_{ij}}{q_{ii}} \mathbb{1}_{\{i \neq j\}}, \quad i, j \in E
\]

where \( \mathbb{1}_A \) is the indicator of the set \( A \) and \( q_{ii} = \sum_{j \neq i} q_{ij} \).

Some other discretization methods are discussed in Hordijk et al. (1976).

3. TRANSIENT PERFORMANCE MEASURES

A transient performance measure of a system is a measure which depends on the initial state of the system or on the time \( t \). In this paper, we'll focus on \( MTTF \) and \( R(t) \). The first measure depends on the initial state of the system, the second depends on the time \( t \).

3.1. Mean time to failure: \( MTTF \)

The mean time to failure \( MTTF \) is defined by \( MTTF = \mathbb{E}[\tau_F] \), where \( \mathbb{E}[.] \) denotes the expectation operator. This measure can be expressed as a ratio of expectations. The following two propositions point out two representations of \( MTTF \): an approximated and an exact one.

3.1.1. Proposition

Let \( \alpha = \Pr[\tau_F < \tau_{\{s_0\}} \mid X_0 = s_0] \), if the Markovian model \( X \) is stiff, then the probability distribution of the random variable \( \alpha \tau_F \) can be approximated by an exponential one with parameter \( 1/\mathbb{E}[\tau_{s_0}] \).

In the following, we give a sketch of the proof of this proposition, for more details and relevant ideas (cf. Nasroallah 1991 and Keilson 1979).

Proof. Assume that \( X_0 = s_0 \) and let \( T_1, T_2, \ldots \) be the times between the successive returns to \( s_0 \). Let \( N \) be the index of the first cycle (interval between successive returns to \( s_0 \)) when \( X \) hits \( F \). Remark that, given \( N \), the times \( T_1, T_2, \ldots, T_{N-1} \) are independent identically distributed (i.i.d) and are independent of \( T_N \). Now,

\[
\tau_F = \sum_{i=1}^{N} T_i
\]
where $\Pr[N = n] = (1 - \alpha)^{n-1}\alpha$ for $n \geq 1$. Hence

$$
\mathbb{E}[e^{-s\tau_F}] = \sum_{n=1}^{+\infty} (1 - \alpha)^{n-1}\alpha\mathbb{E}\left[e^{-s(T_1 + \cdots + T_{n-1})}e^{-sT_n}\right]
$$

$$
= \sum_{n=1}^{+\infty} (1 - \alpha)^{n-1}\alpha(\psi(s))^{n-1}\phi(s)
$$

$$
= \frac{\alpha\phi(s)}{1 - (1 - \alpha)\psi(s)}
$$

where $\psi(s)$ is the characteristic function of a cycle that does not hit $F$ and $\phi(s)$ is the characteristic function of a cycle that hits $F$.

Replacing $s$ by $\alpha s$, we find after some calculations

$$
\mathbb{E}[e^{-\alpha s\tau_F}] = \frac{1}{1 - s\psi'(0)} + O(\alpha),
$$

this shows that $\alpha\tau_F$ is approximately exponentially distributed when $\alpha$ is small (i.e. the model is stiff).

By the above proposition, it is easy to see that

$$
MTTF \sim \frac{\mathbb{E}[\tau(s_0)]}{\mathbb{E}[1_B]}
$$

where $B = \{\tau_F < \tau(s_0)\}$. The following proposition gives an exact ratio expression of $MTTF$.

3.1.2. PROPOSITION

$$
MTTF = \frac{\min(\tau_F, \tau(s_0))}{\mathbb{E}[1_B]}.
$$

Proof. $\tau_F$ can be written as

$$
\tau_F = \tau_F 1_B + (\tau(s_0) + (\tau_F - \tau(s_0))) 1_{B^c},
$$

where $B^c = \{\tau_F \geq \tau(s_0)\}$. This equation is equivalent to

$$
\tau_F = \min(\tau_F, \tau(s_0)) + (\tau_F - \tau(s_0)) 1_{B^c}.
$$

This implies that

$$
\mathbb{E}[\tau_F] = \mathbb{E}\left[\min(\tau_F, \tau(s_0))\right] + \mathbb{E}\left[(\tau_F - \tau(s_0)) 1_{B^c}\right]. \quad (1)
$$

Recherche opérationnelle/Operations Research
But we have

\[ E \left( \left( \tau_F - \tau_{\{s_0\}} \right) 1_{B^c} \right) = E \left( \left( \tau_F - \tau_{\{s_0\}} \right) 1_{B^c} \mid B^c \right) \Pr[B^c] \]

\[ = E[\tau_F - \tau_{\{s_0\}} \mid B^c] \Pr[B^c]. \tag{2} \]

Now by the strong Markov property we have

\[ E[\tau_F - \tau_{\{s_0\}} \mid B^c] = E[\tau_F]. \tag{3} \]

The equations (1, 2) and (3) make up the proof. \(\Box\)

Remarks:

- Since MTTF can be represented by a ratio of expectations, then its simulation by IS algorithm is similar to the simulation of the stationary performance measures (cf. Nasroallah 1991 and Nasroallah 1997).
- In Nasroallah (1991), are given some simulation results of MTTF for a stiff Markovian model with 14 states. Two algorithms are confronted: a standard Monte-Carlo simulation and a newly proposed algorithm called T-Dist (an algorithm based on a distance technique). The T-Dist algorithm is better because of stable estimations and important improvement factor (greater than 200). For more details cf. Nasroallah (1991).

Now we focus on the simulation of the reliability of stiff Markovian models. We'll deal basically with the high reliable Markovian models.

3.2. Reliability at time \(t\): \(R(t)\)

Let \([0, t]\) be an interval of time. The reliability of a system at time \(t\), \(R(t)\), is the probability that this system is operational during the whole observation period \([0, t]\). Thus, \(R(t)\) is a transient performance measure depending on time \(t\). It can be written in the form \(R(t) = \Pr[\tau_F > t]\). On simulating such measures, one must take account of the stiffness and the effect of the parameter \(t\). This time \(t\) must be compared to the holding time of sample states. In general, for stiff models, the probability that the holding time in the initial state is large, is near one. So for moderate time \(t\), a standard simulation stays a very long time in the initial state and consequently, we can’t get acceptable estimations after a reasonable period of simulation time. For transient performances, as in the stationary case, there is some variance reduction techniques that allow reduction in simulation time like FT and CT. The FT method is applied to the holding time in \(s_0\). Its philosophy is to force the next component failure to occur before time \(t\); i.e. the transition of
Z is forced by sampling the holding time in $s_0$ from a new distribution that favors leaving the state $s_0$ (cf. Lewis et al. 1984). The second procedure CT consists on conditioning on the holding time of $s_0$.

In the following, we will combine IS and CT in order to reduce the complexity of the simulation time caused by stiffness and by the parameter $t$. The IS method is well presented in Nasroallah (1991). The following lemma is the main result for CT.

3.2.1. **Lemma**

Let $u$ and $v$ be two random variables such that the first and second order moments exist, then we have

$$\text{var}(E[u \mid v]) \leq \text{var}(u)$$

where \(\text{var}(.)\) denotes the variance operator.


Now, to apply the CT to $R(t)$, we need some preliminaries: let $S_k$ be the holding time in the $(k + 1)^{th}$ visited state by $Z$. It is known that $S_k$ has an exponential probability distribution with parameter $-q_{kk}$. Let $(T_j)_{j \in \mathbb{N}}$ be the sequence defined by

$$T_0 = 0 \quad \text{and} \quad T_j = \sum_{k=0}^{j-1} S_k \quad j \geq 1.$$  

$T_j$ is the $j^{th}$ transition time of the process $Z$. The number of transitions before $t$ is

$$N(t) = \max\{n \in \mathbb{N}; T_n \leq t\}.$$  

So $N(t) + 1$ is the number of visited states before $t$. For a finite sample of states $(Z_0, Z_1, \ldots, Z_n)$, where $n$ is a fixed positive integer, we define the quantities $\gamma_n$, $\beta_n$ and $n_0(n)$ by

$$\gamma_n = \sum_{k=0}^{n} S_k \mathbb{1}_{\{Z_0\}}(Z_k)$$  

$$\beta_n = \sum_{k=0}^{n} S_k \mathbb{1}_{\{Z_k \neq Z_0\}}(Z_k)$$  

$$n_0(n) = \sum_{k=0}^{n} \mathbb{1}_{\{Z_0\}}(Z_k),$$
where \( \gamma_n \) is the holding time in \( Z_0 \), \( \beta_n \) the holding time out of \( Z_0 \) and \( n_0(n) \) the number of visits to \( Z_0 \) during \( n \) transitions of the process \( Z \). Let \( \tau(t) \) be a stopping time defined by

\[
\tau(t) = \min(\tau_F, N(t) + 1),
\]

it's the time when \( Z \) hits \( F \) or the transition time of \( Z \) hits \( t \). We denote

\[
\tilde{R}(t) = 1 - R(t) = \Pr[T_{\tau(t)+1} \leq t].
\]

Now if \( \tau(t) = n \) then

\[
\tilde{R}(t) = \Pr[T_{n+1} \leq t] = \Pr[\gamma_n + \beta_n \leq t] = \mathbb{E}[\mathbbm{1}_{\{\gamma_n + \beta_n \leq t\}}]
= \mathbb{E}[\mathbb{E}[\mathbbm{1}_{\{\gamma_n \leq t - \beta_n\}} | \beta_n, n_0(n)]].
\]

From the previous lemma, we have

\[
\text{var}(\mathbb{E}[\mathbb{E}[\mathbbm{1}_{\{\gamma_n \leq t - \beta_n\}} | \beta_n, n_0(n)]] \leq \text{var}(\mathbb{E}[\mathbbm{1}_{\{\gamma_n + \beta_n \leq t\}}]).
\]

It is clear that there is an advantage to simulate expression (5) than expression (4). Since, for stiff models, the holding time in the initial state \( s_0 \) is typically large, then conditioning on this time pushes \( Z \) to leave \( s_0 \). This allows to obtain good estimations for relatively small values of \( t \). So the Monte-Carlo simulation based on this approach will give acceptable estimation of \( R(t) \) for moderate number of transitions of \( Z \).

**Remarks:**

(i) We have assumed that \( \Pr[Z_0 = s_0] = 1 \), but one can select a reference state by a given initial probability distribution on \( E \). In this case, the definitions of \( MTTF \) and \( R(t) \) must be adjusted.

(ii) It is obvious that the algorithm IS+CT can't give answer for any value of the parameter \( t \).

(iii) The algorithm IS+CT can be applied to simulate other performance measures such as the Availability Mean Interval in \([0, t]\) (\(AMI(t)\)).

In the following section, we present the approximation on which the determination of the confidence intervals used in the simulation is based.
4. CONFIDENCE INTERVALS

Since the measure of interest is written in the form \( r = \mathbb{E}[f(Z)] \), where \( f \) is a real valued function and \( Z \) is the DTMC defined before, then by the regenerative property of the process \( Z \), \( r \) can be written as a ratio of expectations \( r = \frac{\mathbb{E}[u]}{\mathbb{E}[v]} \) (cf. Nasroallah (1991), pp. 97-99), where \( u \) and \( v \) are two random variables depending on \( f \) and on a regenerative cycle of \( Z \). We are given a sequence of i.i.d. random vectors \( \{(u_i, v_i) : 1 \leq i \leq m\} \). Suppose we focus on a \( 100(1 - \gamma) \) percent confidence interval for \( r \), where \( 0 < \gamma < 1 \). Let the sample means be

\[ \bar{u}(m) = \frac{1}{m} \sum_{i=1}^{m} u_i \quad \text{and} \quad \bar{v}(m) = \frac{1}{m} \sum_{i=1}^{m} v_i \]

and the sample variance be

\[ s_{11}(m) = \frac{1}{m-1} \sum_{i=1}^{m} (u_i - \bar{u}(m))^2, \]
\[ s_{12}(m) = \frac{1}{m-1} \sum_{i=1}^{m} (u_i - \bar{u}(m))(v_i - \bar{v}(m)), \]
\[ s_{22}(m) = \frac{1}{m-1} \sum_{i=1}^{m} (v_i - \bar{v}(m))^2. \]

Now let \( w_i = u_i - rv_i, \ 1 \leq i \leq m \) and let \( \bar{w}(m) = (1/m) \sum_{i=1}^{m} w_i \). Observe that \( \{w_i : i \geq 1\} \) are i.i.d. and that \( \mathbb{E}[w_i] = 0 \). Let \( \sigma_1^2 = \mathbb{E}[w_i^2] \) and \( s^2(m) = s_{11}(m) - 2rs_{12}(m) + r^2s_{22}(m) \). It may be shown that \( s^2(m) \to \sigma_1^2 \) as \( m \to +\infty \) with probability one. Using this fact, and the continuous mapping theorem and the central limit theorem applied to \( w_i/s \), we see after some calculations that, if \( 0 < \sigma_1 < +\infty \), then we get the following approximate \( 100(1 - \gamma) \) percent interval for \( r \):

\[ \frac{\bar{u}(m)\bar{v}(m) - k(m)s_{12}(m) - D^{(1/2)}(m)}{\bar{v}^2(m) - k(m)s_{22}(m)} \leq r \]
\[ \leq \frac{\bar{u}(m)\bar{v}(m) - k(m)s_{12}(m) + D^{(1/2)}(m)}{\bar{v}^2(m) - k(m)s_{22}(m)} \]

where \( k(m) = [\Phi^{-1}(1 - \gamma/2)]^2/m, \)
\[ D(m) = [\bar{u}(m)\bar{v}(m) - k(m)s_{12}(m)]^2 \]
\[ - [\bar{v}^2(m) - k(m)s_{22}(m)][\bar{u}^2(m) - k(m)s_{11}(m)], \]
and \( \Phi \) is the distribution function for a normal random variable with mean zero and variance one. The approximations are considered with probability \((1 - \gamma)\) for large \( m \).

**Remark:** Based on the central limit theorem for partial sums with a random number of terms, one can show that the approximations used in this section hold when \( m \) is replaced by a random integer \( N(t) \) for a fixed run length \( t \) in the simulation.

In the following section we present some simulation results for \( R(t) \) by studying two examples.

5. SIMULATION OF \( R(t) \)

**A birth and death Markov process**

Let consider a system with two active redundant components. This system can be modelled by a birth and death Markov process \( X \) with state space \( E = \{0, 1, 2\} \). The generator of \( X \) is such that \( q_{ij} = \lambda_i \), \( q_{ji} = \mu_j \) for \( 0 \leq i < j \leq 2 \) and \( q_{02} = q_{20} = 0 \). For \( i = 1, 2 \), \( \lambda_i \) and \( \mu_i \) are failure and recovery rates respectively. We assume that \( O = \{0, 1\} \) and \( F = \{2\} \). The initial state is \( s_0 = 0 \). We take \( \lambda_0 = 2\lambda_1 \) and \( \mu_i = 1; i = 1, 2 \). In the IS algorithm we take 0.999 as the new probability to favors the rare event. This value is considered as the best one in Heidelberger *et al.* (1987). The Tables 1, 2 and 3 summarize comparisons between the hybrid algorithm IS + CT and the standard Monte-Carlo simulation algorithm SMCS. In Table 1, \( \lambda_1 = 10^{-1} \) (*i.e.* weak stiffness), \( t = 1 \) for the first three columns and \( t = 10 \) for the last three columns. For these values, we have \( R(1) \simeq 0.999 \) and \( R(10) \simeq 0.99 \). In Table 2, \( \lambda_1 = 10^{-3} \) (*i.e.* stiff model), \( t = 10 \) for the three first columns and \( t = 100 \) for the last three columns. In this case, \( R(10) \simeq 0.9999 \) and \( R(100) \simeq 0.999 \). The Tables 1 and 2 present evolution of half-width of 99% confidence interval (hwci) with respect to Central Processing Unit (CPU) time. We remark that there is an improvement factor of IS + CT with respect to SMCS, and this factor grows with stiffness. Now in Table 3, \( \lambda_1 = 10^{-5} \) (*i.e.* very stiff model) and \( t = 1000 \). The exact value of the reliability of the system is \( R(1000) = 0.9999998 \) (computed by MACSYMA, cf. Moser 1970). The Table 3 gives the evolution of the estimation \( \hat{R}(1000) \) of \( R(1000) \) and its hwci with respect to CPU time. We remark that for a CPU time near 420, the improvement factor in confidence interval (*i.e.* hwci[SMCS]/hwci[IS + CT]) is approximately 66. A second remark is that IS + CT gives more stable estimations than SMCS.

vol. 34, n° 4, 2000
A database System

We illustrate the capabilities of IS + CT algorithm by presenting a simulation of an interesting example of fault-tolerant database system. The system has two front-end systems, two databases, and two processing subsystems each one of them contains a switch, a memory, and two processors. A processing subsystem is considered operational if the memory, the switch, and one of the two processors is functioning. The entire system is...
operational if a database, a front-end, and, at least, one of the two processing subsystems is operational.

We assume that the repair and failure time distributions of all components are exponentially distributed with means 1 and 2400 hours, respectively, except for the processors which have means 1 and 120 hours, respectively. We further assume that when a processor fails, it contaminates (or fails) the database with probability \((1 - c)\), where \(c = 0.99\) is the coverage probability. There is a single repairman in the system with highest priority given to the databases and the front-ends, the next highest priority given to the memory and the switch elements, and the lowest priority to the processors. Components at the same priority level are selected at random for repair. The simulation results for this example are given in Table 4. We remark that there is an advantage of IS + CT with respect to SMCS. The average of the improvement factor is 5. This result is interesting since the Markovian model studied here is relatively stiff but large (i.e. the size of state space is \(2^{12}\)).

**Table 4**

<table>
<thead>
<tr>
<th>CPU time</th>
<th>SMCS hwci (\times 10^4)</th>
<th>IS + CT hwci (\times 10^4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>700</td>
<td>2.50</td>
<td>0.51</td>
</tr>
<tr>
<td>800</td>
<td>2.41</td>
<td>0.49</td>
</tr>
<tr>
<td>900</td>
<td>2.39</td>
<td>0.50</td>
</tr>
<tr>
<td>1050</td>
<td>2.10</td>
<td>0.48</td>
</tr>
</tbody>
</table>

6. CONCLUSION

We have presented the IS + CT algorithm which is a combination of two existent algorithm IS and CT. This allows to simulate transient performance measures depending on time such as the reliability of a system at time \(t\). For a large class of Markovian models, the stiff and large models, IS + CT reduces the simulation time and gives more stable estimations than the standard Monte-Carlo simulation algorithm SMCS.

REFERENCES


