VU Research Portal

A note on extended uniformization for non-exponential stochastic networks

van Dijk, N.M.

1989

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
van Dijk, N. M. (1989). *A note on extended uniformization for non-exponential stochastic networks*. (Serie Research Memoranda; No. 1989-26). Faculty of Economics and Business Administration, Vrije Universiteit Amsterdam.

SERIE RESEARCH MEMORANDA

A NOTE ON

EXTENDED UNIFORMIZATION FOR

NON-EXPONENTIAL STOCHASTIC NETWORKS

Nico M. van Dijk

A NOTE ON
EXTENDED UNIFORMIZATION FOR
NON-EXPONENTIAL STOCHASTIC NETWORKS

Nico M.van Dijk
Free University, Amsterdam, The Netherlands

Abstract

The standard uniformization technique for continuous-time Markov chains is generalized to non-exponential stochastic networks.

1.   Introduction

The technique of uniformization, as initiated by results in [3], and described for instance in [7], p. 26 or [8], p. 110, is known to be a most useful tool for modelling, simulating or numerically solving continuous-time Markov chains. (cf. [8], [9]). This technique is essentially based on a Markovian or rather exponential structure. For non-exponential structures, however, an equivalent or generalized result has not been reported.

Probably the most famous present-day applications of continuous-time Markov chains are stochastic service networks which arise when modelling communication, computer or manufacturing systems under exponentiality assumptions. These assumptions, however, are rarely met in practice. Furthermore, explicit stationary distributions are usually not available due to practical phenomena, such as blocking, age dependent routing or breakdowns. Simulation or numerical computation is therefore generally required.

To this end, this note is concerned with an extended uniformization procedure that applies to non-exponential stochastic service networks. An equivalence result is obtained for the stationary distribution of the original model with state dependent jump times and a modified model with state independent jump times. This result is of practical interest as it may enable one to reduce the simulation or numerical computation of a non-exponential complex network to that of a Markov chain.

The result is intuitively appealing and may have been used already by practitioners. To the best of knowledge, however, it has not been reported or advocated as such in the literature. This note therefore is primarily meant to bring it to the attention of practitioners along with formal justification. The proof, which appears to be rather simple, is of theoretical interest in itself.

The system under study allows blocking and general state dependent characteristics such as modelling blocking and age dependent routing. For notational convenience the presentation is restricted to closed frameworks but extensions to open formulations are readily concluded. The restriction to bounded failure rates can be relaxed.

The organization is as follows. In section 2 the model descriptions and equivalence result are presented. Section 3 provides the formal proof. A discussion on extensions and computation concludes the paper.

2.   Model

Consider a stochastic network with a fixed number of M jobs. A state [L,T] with $L = (\ell_1,\ldots,\ell_M)$ and $T = (t_1,\ldots,t_M)$ denotes for each job i the current job mark ℓ_i of job i with $\ell_i \in S$, where S is some countable space of possible job marks, and t_i the time (age) after the last service completion of this job.

For example, in queueing network applications a jobmark ℓ can be of the form: $\ell = (r,j,p)$ with r some type number of the job, j the station at which it is present and p the position at this queue that it occupies,

while t is the time that the job has already been present at that station.

Original Model. The law of motion is determined by the characteristics:

$F_\ell(.)$:   distribution functions

$s_i([L,T])$:   service rates (speeds)

$p_i(\ell \mid [L,T])$:   transition probabilities

as follows. When a job changes its jobmark in ℓ it requires a random amount of service with distribution function F_ℓ. When the system is in state $[L,T]$, the service rate i.e. the amount of service per unit of time provided to job i is $s_i([L,T])$. When the system is in state $[L,T]$ and job i completes its service its jobmark is changed in ℓ' with probability $p_i(\ell' \mid [L,T])$.

Remarks

1. Note that the service rate for a particular job in a particular state can be equal to zero as naturally arising for instance in a queueing network with FCFS-service stations as in the example below.
2. Clearly, the above parametrization could have been combined in one service completion rate function. However, the present more detailed formulation is preferred as it corresponds more naturally to queueing network protocols.

Example (Queueing Network) Consider a closed queueing network with N FCFS-single server service stations and M numbered jobs. A job requires random amounts of service at the various stations, say at station j according to a distribution function G_j. The service rate at station j is $s_j(n_j,t_j)$ when n_j jobs are present while the job in service has received already t_j units of service. (This service rate thus depends stochastically also on the total amount of residual workload at this station). Upon service completion at station j a job routes to station j' with probability $p_{jj'}(\bar{n},\bar{t})$ where $\bar{n} = (n_1,\ldots,n_N)$ and $\bar{t} = (t_1,\ldots,t_M)$ denote the population sizes n_s and received amounts of service t_s at station s for all s. (Age and workload dependent routing as well as blocking are hereby involved).

Letting $\ell = (i,j,p)$ denote the job-number i, the station number j and the position p at this station, where p=1 is the head of the queue, and reading $1\{A\} = 1$ if an event A is satisfied and $1\{A\} = 0$ if not, the above parametrization applies with

$F_\ell = G_j$

$s_i([L,T]) = s_i(n_j,t_j) \, 1\{p=1\}$

$p_i(\ell' \mid [L,T]) = p_{jj'}(\bar{n},\bar{t}) \, 1\{\ell'=(i,j',0)\}$

                                                                □

We will now make some assumptions in order to define a related so-called uniformized model.

Assumptions

1. For all ℓ, the function $F_\ell(t)$ is absolute continuous for $t \in (0,\infty)$ with density function $f_\ell(t)$. Hence, its failure rate is well-defined by $f_\ell(t)/[1-F_\ell(t)]$ for all $t \in (0,\infty)$. We introduce the notation:

(1) $d_i([L,T]) = s_i([L,T]) \ f_{\ell_i}(t_i)/[1-F_{\ell_i}(t_i)]$

2. For some constant $B < \infty$ and all $[L,T]$:

(2) $d([L,T]) = \Sigma_i \ d_i \ ([L,T]) \leq B$

Uniformized model. The law of motion is now defined as follows. At exponential times with some parameter $Q \geq B$ the network is inspected. Suppose that directly after an inspection the system is in state $[L,T]$. Then at the next inspection, say after time t, with probability

(3) $d_i([L,T]) \ p_i(\ell' | [L,T])/Q$

the state is changed in $[L',T']$ with $\ell_j'=\ell_j$ and $t_j'=t_j + t$ for $j \neq i$ but $\ell_i'=\ell'$ and $t_i'=0$, for all $i = 1,\ldots,M$, while with probability

(4) $1 - d([L,T])/Q$

the state is merely updated, i.e. changed in $[L,T']$ with $t_i'=t_i+t$ for all $i=1,\ldots,M$.

Remarks

1. Note that (3) is indeed a probability with (3) summed over i and (4) together equal to one.
2. The formulation (3) and (4) reduces to the standard uniformization technique (e.g. [7], p. 26 or [8] p. 110) when all distributions $F_\ell(.)$ are exponential.

Without restriction of generality, now assume that both the original and uniformized model have a unique stationary density function at one and the same irreducible set S which we denote by $\pi_1(L,T)$ and $\pi_2(L,T)$ respectively. The following result will then be proven in section 3.

Theorem

(5) $\pi_1(L,T) = \pi_2(L,T)$ $(L,T) \in S$

Remark

Recently, in [6], a new elegant approximation method for continuous-time Markov chains has been introduced by actually inspecting the original chain at exponential times. Numerical results turned out to be amazingly accurate. In contrast, note that the above uniformization by exponential inspection concerns a different model but leads to an equality result.

3. Proof of the theorem

We need to verify that the global balance (or more precisely, stationary forward Kolmogorov) equations of both models have the same solution, where these solutions are assumed to be unique and to be continuously differentiable in its argument t_i for all i.

To this end, for a state [L,T] we write $[L_i,T_i] + (\ell',t')$ for the same state with the job-i specification (ℓ_i,t_i) replaced by (ℓ',t'). We write (L,T-s) for the same state with t_i replaced by t_i-s for all i. Further, a symbol 0^+ is used to indicate a right hand limit at 0 and $o(\Delta s)$ for an expression such that $o(\Delta s)/\Delta s \to 0$ as $\Delta s \to 0$.

Let a state [L,T] be fixed. Then, in the standard manner (e.g. [1], [4]) of first considering a point of time τ and conditioning upon time $\tau-\Delta t$, these global balance equations (GBE) require:

GBE for original model

(6) $\pi_1(L,T) =$

$\pi_1(L,T-\Delta s) \prod_i [1-d_i([L,T-\Delta s])\Delta s] + o(\Delta s)$

provided t_i-$\Delta s > 0$ for all i, while for any i such that $t_i < \Delta s$:

(7) $\pi_1([L_i,T_i] + (\ell_i,t_i)) =$

$\sum_{\ell'} \int_0^\infty dt' \{\pi_1([L_i,T_i] + (\ell',t')) \times$

$\Delta s\, d_i([L_i,T_i] + (\ell',t'))p_i(\ell_i|[L_i,T_i] + (\ell',t')) \times$

$\prod_{j \neq i} [1-d_j([L_i,T_i] + (\ell',t'))\Delta s]\} + o(\Delta s)$

GBE for uniformized model.

(8) $\pi_2(L,T) =$

$\pi_2(L,T-\Delta s) [1 - Q\Delta s] +$

$\pi_2(L,T-\Delta s) Q\Delta s [1 - d([L,T-\Delta s]) \Delta s] + o(\Delta s)$

provided t_i-$\Delta s > 0$ for all i, while for any i such that $t_i < \Delta s$:

(9) $\quad \pi_2([L_i,T_i] + (\ell_i,t_i)) =$

$\quad\quad \Sigma_{\ell'} \; _0\!\int^\infty \pi_2([L_i,T_i] + (\ell',t')) \times$

$\quad\quad dt'(Q\Delta s \; d_i([L_i,T_i] + (\ell',t'))p_i(\ell_i|[L_i,T_i] + (\ell',t'))/Q) + o(\Delta s)$

Then, by dividing the left and right hand sides of equations (6)-(9) by Δs, letting Δs tend to 0 and using the continuity and differentiability assumptions in L_i, one easily concludes that both (6) and (8) lead to

(10) $\quad \Sigma_i \; \{\frac{\partial}{\partial t_i} \; \pi_p \; ([L,T]) - \pi_p \; ([L,T]) \; d_i([L,T])\} = 0$

for any $[L,T]$ and $p = 1,2$, while

(11) $\quad \pi_p([L_i,T_i] + (\ell_i,0^+)) -$

$\quad\quad \Sigma_{\ell'} \; _0\!\int^\infty dt' \{\pi_p([L_i,T_i] + (\ell',t')) \times$

$\quad\quad d_i([L_i,T_i] + (\ell',t'))p_i(\ell_i|[L_i,T_i] + (\ell',t'))\} = 0$

for any $[L_i,T_i]$, ℓ_i, i and $p=1,2$. As these equations are assumed to have a unique solution the proof is hereby completed.

4. Further remarks

1. (Discrete phase-type version; limiting approach)
Clearly, a discrete version of the uniformized model and a corresponding equivalence result can be provided also if one assumes the service requirements to be mixtures of Erlang distributions. In that case the sojourn times t_i should be read as the number of completed exponential phases after the last service completion. The present result could then be concluded by approximating the given distributions by such mixtures (in the sense of weak convergence) and applying weak convergence limiting arguments on so-called D-sample path spaces as developed in [2]. The technical details of such an approach, however, are known to be cumbersome and complex.

2. (Unbounded failure rates; finite distributions)
Assumptions A1 and A2 can be rather restrictive from a practical point of view. For example, both assumptions are violated by deterministic service requirements. However, similarly to [11], for the standard uniformization technique in the exponential case, under appropriate conditions one can extend the above uniformization to an approximate uniformization for situations with unbounded failure rates and finite distributions.

3. (Simulation/Computation).

As continuous distributions are involved, the actual simulation or computation of the transition probabilities may still lead to technical complications. In simulation, the rejection method (cf. [5] or [7]) may come in handy. In computing, a discretization or approximation either by exponential phase type distributions as mentioned under 1, or by using discrete time grids (cf. [10]) for the service times, seems most natural.

References

[1] Gavish, B. and Schweitzer P.J. (1977), The Markovian queue with bounded waiting time, Management Science 23, 1349-1357.

[2] Hordijk, A. and Schassberger, R. (1982), "Weak convergence of generalized semi-Markov processes", Stochastic Procc. Appl. 12, 271-291.

[3] Jensen, A. (1953), Markov chains as an Aid in the Study of Markov Processes, Skan. Aktuartidskr. 36, 87-91.

[4] Keilson, J. (1963), A Gambler's ruin type problem in queueing theory, Opns. Res. 570-576.

[5] Law, A.M. and Kelton, W.D. (1982), "Simulation modeling and analysis", McGraw-Hill.

[6] Ross, S.M. (1987), Approximating transition probabilities and mean occupation times in continuous-time Markov chains. Probability in the Engineering and Informational Sciences 1: 251-264.

[7] Ross, S.M. (1985), "Introduction to probability models", Academic Press.

[8] Tijms, H.C. (1986), "Stochastic Modelling and Analysis"; A computational Approach, Wiley, New York.

[9] Tijms, H.C. and Eikeboom, A.M. (1986), "A simple technique in Markovian control with applications to resource allocation", Oper. Res. Letters, 1, 25-32.

[10] Van Dijk, N.M. (1984), "Controlled Markov processes: time-discretization, CWI Tract 11, Mathematical Center, Amsterdam, The Netherlands.

[11] Van Dijk, N.M. (1988), "Approximate uniformization for continuous-time Markov chains with an application to performability analysis", Research Report, Free University, Amsterdam.