FORMULAS FOR THE VARIANCE OF THE SAMPLE MEAN IN FINITE STATE MARKOV PROCESSES

Gordon B. Hazen
Purdue University

and

A. Alan B. Pritsker
Purdue University and
Pritsker & Associates, Inc.
West Lafayette, IN

ABSTRACT

Formulas are derived for the variance of sample means associated with finite state Markov processes. Results are presented to illustrate the use of the formulas for example processes. The formulas can be used to evaluate proposed statistical methods for estimating the variance of a sample mean obtained from simulation experiments.

INTRODUCTION

Simulation of real world systems on a digital computer is now common place. When the system is a queueing system, one of the most often derived statistics is the average or sample mean number-in-the-system. The utility of this statistic obviously lies in the fact that it is presumably an unbiased estimator of the equilibrium mean number in system. However, the simulator needs to know more than just the value of such a statistic in order to make statistical statements concerning the true mean. Specifically, if the statistic is highly variable across simulations, the result from a single simulation may be misleading. The solution to this problem is the standard statistical approach: use simulation data to estimate the variance of the average-number-in-system statistic and construct a confidence interval. The approach in more detail is as follows (see Fishman [3]).

Let $N(t)$ be the number in system at time t and assume the simulation starts at time zero. Then the sample mean statistic at time t, $\overline{N}_t$, is given by

$$\overline{N}_t = \frac{1}{t} \int_0^t N(u)\, du$$

If we assume the $N(t)$ process is covariance stationary, then the autocovariance of lag s, R_s, is given by

$$R_s = \text{cov}[N(t), N(t+s)] \ .$$

Let var $\overline{N}_t$ denote the variance of $\overline{N}_t$. It is well known that the variance of a discrete average of n independent observations approaches zero as n increases. The analogous result

$$\lim_{t\to\infty} \text{var}\ \overline{N}_t = 0$$

holds, under reasonable conditions, for $\overline{N}_t$ (see Appendix). More significantly, it can be shown that

$$\lim_{t\to\infty} t \ \text{var}\ \overline{N}_t = m$$

where we define m by

$$m = \int_{-\infty}^{\infty} R_s\, ds \quad .$$

Therefore, if the simulator has an estimate $\hat{m}$ of m, he can estimate var $\overline{N}_t$ by $\hat{m}/t$. If it is reasonable to assume that $\overline{N}_t$ is normally distributed, that $\hat{m}$ has a chi-square distribution, and that $\overline{N}_t$ and $\hat{m}$ are independent, then a confidence interval for the true mean may be derived (see Fishman [3]).

Procedures that have been proposed for estimating m use: replication; subinterval sampling; spectral methods; autoregressive techniques; and regenerative methods [3,4]. Analytic methods for evaluating these estimators all rely on assumptions (large sample size, independence) which may not strictly hold in practice. Thus, an empirical evaluation of the performance of an estimator $\hat{m}$ on systems for which m is known is desirable. Such evaluations have been carried out [2,3]; however, they have been restricted to simple systems, such as the M/M/1 queue, for which the true value of m can be calculated. It is not clear if the results of such evaluations can be extended to more complex systems.

With the intent of alleviating this situation, we present in this paper formulas for the direct computation of m and t var $\overline{N}_t$ for any finite state Markov process. Proofs of theorems are included in an Appendix to the paper.

1. FINITE STATE MARKOV PROCESSES

In this section we summarize the results we will need concerning finite state Markov processes. For additional information, the reader is referred to Cinlar [1] or Parzen [5].

We consider a finite state Markov process whose state at time $t \geq 0$ is denoted by $S(t)$. Let the finite set of possible states be $s_0, s_1, \ldots, s_p$ and define the transition probability

$$P_{qr}(t) = \Pr(S(t)=s_r \mid S(0)=s_q) \ ,$$

that is, $P_{qr}(t)$ is the probability that the process is in state s_r at time t given that it is in state

s_q at time zero. Since

$$\sum_r P_{qr}(t) = 1,$$

the row vector

$$P_q(t) = [P_{q0}(t), P_{q1}(t), \ldots, P_{qp}(t)]$$

represents the conditional distribution of S(t) given that $S(0) = s_q$. As t becomes large, $P_q(t)$ tends to a limiting distribution which does not depend on q. We write

$$\pi = \lim_{t\to\infty} P_q(t) \qquad q = 0,1,\ldots,p \quad ,$$

and call π the <u>stationary</u> or <u>steady state</u> distribution. If the initial state S(0) of the process is chosen according to the distribution π we say we have <u>steady state initial conditions</u>. In this case, the unconditional distribution of S(t) for any $t \geq 0$ is also π; in other words, a process beginning in "steady state" remains in steady state.

We may form the matrix $\mathbb{P}(t)$ whose rows are $P_0(t), \ldots, P_p(t)$. $\mathbb{P}(t)$ satisfies the Chapman-Kolmogorov differential equations, which in matrix form are

$$\mathbb{P}'(t) = \mathbb{P}(t)A$$

where $A = \mathbb{P}'(0)$ is the matrix of so-called transition rates, which uniquely determine the probabilistic structure of the process. If $A = (a_{qr})$ then for $q \neq r$, $a_{qr} \geq 0$ is the transition rate from s_q to s_r. The row sums of the matrix A are zero, so that

$$a_{qq} = -\sum_{\substack{r=0\\ r\neq q}}^{p} a_{qr} \quad .$$

If the process is in state s_q at any time, it remains there for a length of time T which has a negative exponential distribution with mean $-1/a_{qq}$ and then moves to state s_r with probability

$$a_{qr} \Big/ \sum_{\substack{k=0\\ k\neq q}}^{p} a_{qk} \quad .$$

The steady state distribution π may be obtained from A by solving the system of equations

$$\pi A = 0$$

$$\sum_{i=0}^{p} \pi_i = 1 \quad .$$

The matrix $\mathbb{P}(t)$ can be obtained by evaluating

$$\mathbb{P}(t) = V^{-1}\mathrm{diag}(e^{\alpha_0 t}, e^{\alpha_1 t}, \ldots, e^{\alpha_p t})V$$

provided A has linearly independent eigenvectors. Here V is the matrix whose rows are the left eigenvectors of A and $\alpha_0, \alpha_1, \ldots, \alpha_p$ are the corresponding eigenvalues. In the above and subsequently, the expression

$$\mathrm{diag}(x_0, x_1, \ldots, x_p)$$

represents the matrix with diagonal entries $x_0, x_1, \ldots, x_p$, and zeros elsewhere.

2. SUMMARY OF RESULTS

Consider the finite state Markov process on the states $s_0, s_1, \ldots, s_p$, having transition probability matrix $\mathbb{P}(t)$, a matrix A of transition rates, and a stationary distribution $\pi = [\pi_0, \pi_1, \ldots, \pi_p]$. Furthermore, let

$$\Pi = \lim_{t\to\infty} \mathbb{P}(t)$$

be the matrix where p+1 rows are all π.

Let η be a real valued function which assigns to each state s a number $\eta(s)$, and define the process N(t) by

$$N(t) = \eta(S(t)) \ .$$

N(t) could, for example, represent number-in-system at time t. It can be shown (see appendix) that under steady state initial conditions for the S(t) process, the N(t) process is covariance stationary, which means that EN(t) and cov(N(t+s), N(t)) do not depend on t. We therefore define under these conditions

$$EN = EN(t)$$

$$R_s = \mathrm{cov}(N(t+s), N(t))$$

It is easily seen that

$$EN = \sum_q \eta(s_q)\pi_q \quad .$$

As above we define

$$\overline{N}_t = \frac{1}{t}\int_0^t N(u)du \ .$$

We are interested in var $\overline{N}_t$. The proof of the following theorem is given in the appendix.

<u>Theorem 1</u>: If the matrix A of transition rates has p+1 linearly independent eigenvectors then the integral

$$m = \int_{-\infty}^{\infty} R_s ds$$

exists and is finite. Moreover,

(i) $\quad m = 2((EN)^2 - h'T(A+\Pi)^{-1}h)$

and under steady state initial conditions

(ii) $\quad t \text{ var } \overline{N}_t = m + \frac{2}{t} h'T(A+\Pi)^{-2}(\mathbb{P}(t)-I)h$

(iii) $\quad \lim_{t\to\infty} t \text{ var } \overline{N}_t = m \ ,$

where I is the identity matrix of order p+1 and

$$h' = [\eta(s_0), \eta(s_1), \ldots, \eta(s_p)]$$

$$T = \mathrm{diag}(\pi_0, \pi_1, \ldots, \pi_p) \quad .$$

It should be noted that the matrix inversions indicated are always possible, that is, the matrix $A+\Pi$ is nonsingular.

A second important result can also be derived. We may emphasize the dependence of m and t var $\overline{N}_t$ on the matrix A of transition rates by writing

$$m = m(A)$$

$$t \text{ var } \overline{N}_t = v_t(A)$$

If A_1 and A_2 are two different transition matrices, we may inquire as to the relationship between $m(A_1)$ and $m(A_2)$ and between $v_t(A_1)$ and $v_t(A_2)$. If $A_2 = rA_1$ for some $r > 0$, the question can be answered as follows.

Theorem 2: Under the conditions of theorem 1, and for $r > 0$

(i) $m(rA) = \frac{1}{r} m(A)$.

Under steady state initial conditions

(ii) $v_t(rA) = \frac{1}{r} v_{rt}(A)$.

An easy convergence result follows from this theorem, namely that $v_t(rA)$ converges N times faster to $m(rA) = \frac{1}{r} m(A)$ than $v_t(A)$ does to m(A). More precisely, we have

$$\frac{v_t(rA) - m(rA)}{m(rA)} = \frac{\frac{1}{r} v_{rt}(A) - \frac{1}{r} m(A)}{\frac{1}{r} m(A)} = \frac{v_{rt}(A) - m(A)}{m(A)},$$

where the ratios indicate it takes rt time units for the system described by the matrix A to achieve the same convergence ratio as the system described by the matrix rA.

We may apply Theorem 2, for example, to the family of M/M/s/n queueing systems with arrival rate λ and service rate μ. The matrix A of transition rates is entirely determined by λ and μ (if s and n are fixed) and we obtain

$$m(r\lambda, r\mu) = \frac{1}{r} m(\lambda,\mu)$$

$$v_t(r\lambda, r\mu) = \frac{1}{r} v_{rt}(\lambda,\mu) \quad .$$

This is illustrated in section 3.

Theorem 2 may hold for countable state Markov processes as well. For example, Fishman [3] gives the following data for an M/M/1 system.

i	λ_i	μ_i	m_i
1	4.5	5.0	6840
2	4	40/9	7695

For this data,

$$r = \frac{\lambda_1}{\lambda_2} = \frac{\mu_1}{\mu_2} = \frac{9}{8}$$

and $$m_1 = \frac{1}{r} m_2$$

since $$6840 = \frac{8}{9} (7695) \ .$$

3. COMPUTATIONAL RESULTS

The results of section 2 are amenable to computer implementation. We give three examples.

Example 1. Using numerical integration techniques, Duket and Pritsker [2] have computed the value of

$$m = \sum_{s=-\infty}^{\infty} R_s$$

for the purpose of discrete data collection on an M/M/1 queueing system. We wished to compare their value of m = 361 with the results obtained by computing m for an M/M/1/n queueing system for large values of n, using the equations of theorem 1. The results are displayed in table 1, and it in fact does appear that m is approaching 361 as n increases. The values of t var $\overline{N}_t$ for several values of t are also given, to illustrate the convergence to m.

Example 2. The results of theorem 2 are illustrated in table 2 for an M/M/2/8 queueing system. All values were calculated using the equations of theorem 1. Note that the values of $v_t(2,1)$ for t = 5000,6000,7000,8000 can be deduced since the values $v_t(4,2)$ are known for t = 2500,3000,3500,4000.

Example 3. More complex systems can also be examined, although available computer storage tends to limit the size of such systems, in terms of the number of possible states.

Table 1

Values of m and t var $\overline{N}_t$ for M/M/1/n queueing system with $\lambda = 4$ and $\mu = 5$

n	m	t var $\overline{N}_t$: t=500	1000	1500	2000	2500	3000	3500	4000
10	34.91	34.71	34.83	34.85	34.87	34.88	34.88	34.88	34.89
20	182.7	180.5	181.6	182.0	182.2	182.3	182.3	182.4	182.4
40	345.3	338.1	341.7	342.9	343.5	343.9	344.1	344.3	344.4
50	357.0	349.1	353.0	354.3	355.0	355.4	355.7	355.8	356.0

Table 2

Values of m and t var $\overline{N}_t$ for an M/M/2/8 queueing system

	m	t var $\overline{N}_t$							
		t = 500	1000	1500	2000	2500	3000	3500	4000
$\lambda = 2$, $\mu = 1$	44.79	44.46	44.62	44.68	44.71	44.72	44.73	44.74	44.75
$\lambda = 4$, $\mu = 2$	22.40	22.31	22.35	22.37	22.37	22.38	22.38	22.38	22.38

Consider a single server, capacity one system with Poisson arrivals and Erlang-p service times ($M/E_p/1/1$). Such a system is a Markov process with transition diagram as shown in Figure 1. In other words, the service times are represented as a sum of p independent exponential service stages, each with rate μ, and the interarrival times are exponential with rate λ. When the system state is zero, the system is empty, and when it is $k > 0$, the system is full with server in stage k. Thus, the η function representing number-in-system is given by

$$\eta(s) = \begin{cases} 0 & s = 0 \\ 1 & 0 < s \geq p \end{cases}$$

The mean service time τ is given by

$$\tau = \sum_{i=1}^{p} \frac{1}{\mu} = \frac{p}{\mu}$$

and the service time variance, σ^2, is

$$\sigma^2 = \sum_{i=1}^{p} \frac{1}{\mu^2} = \frac{p}{\mu^2} \quad .$$

Considering τ and λ as fixed values and p as a variable, we can write

$$\mu = \frac{p}{\tau}$$

$$\sigma^2 = \frac{p}{\mu^2} = \frac{p}{p^2/\tau^2} = \frac{\tau^2}{p} \quad .$$

Thus, as p becomes large, service time variance decreases, and in the limit as $p \to \infty$, service times approach a fixed length τ.

The behavior of $m = m_p$ as a function of p was investigated (see Table 3). When p = 1 it is not difficult to prove that

$$m_1 = \frac{2\lambda\mu}{(\lambda+\mu)^3} \quad ,$$

and this result was used to verify the values for p = 1. It was discovered empirically that for each λ and τ in Table 3 we have

$$m_p = \frac{p+1}{2p} m_1$$

although this result was not verified analytically. If valid, this result would indicate that the value m_c for constant service times is given by

$$m_c = \lim_{p \to \infty} m_p = \frac{1}{2} m_1$$

regardless of the arrival rate λ or mean service time τ.

The effect of "staging" service times on m was briefly investigated for other systems (see Table 4). For these, the limiting behavior of m_p as $p \to \infty$ was unclear. One problem is that the number of states in the process quickly becomes unmanageable as the number of service stages increases.

4. CONCLUDING REMARKS

This paper has presented formulas for the calculation of m and t var $\overline{N}_t$ under steady state initial conditions for finite state Markov processes. The formulas can be used to evaluate procedures for estimating the variance of the sample mean from simulation studies. The formula for m has been used in a study of startup policies [6].

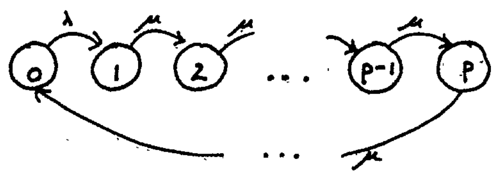

Figure 1. Transition Diagram for an $M/E_p/1/1$ Queueing System

Table 3. Values for m in an $M/E_p/1/1$ queueing system as a function of p.

$\lambda = 1;\ \tau = 1$		$\lambda = 2;\ \tau = 1$		$\lambda = 3;\ \tau = 2$	
p	m_p	p	m_p	p	m_p
1	1/4	1	4/27	1	24/343
2	.1875	2	.11111	5	.04198
3	.1667	3	.09877		
4	.1563	4	.09259		
5	.1500	5	.08889		

Table 4. Values of m as a function of p for two queueing systems

$M/E_p/1/2$ ($\lambda = 1,\ \tau = 1$)		$M/E_p/2/8$ ($\lambda = 2,\ \tau = 1$)	
p	m	p	m
1	1.333	1	44.79
2	1.038	2	43.37
3	.9377	3	43.18
4	.8868		
5	.8560		

LIST OF REFERENCES

1. Cinlar, E., Introduction to Stochastic Processes, Prentice-Hall, Inc., Englewood Cliffs, N.J. (1975).

2. Steven D. Suket and A. Alan B. Pritsker, Spectral Methods for Simulation Output, Pritsker & Associates, Inc., University Square, West Lafayette, Indiana, 47906 (1975)

3. George S. Fishman, Concepts and Methods in Discrete Event Digital Simulation, John Wiley & Sons, New York (1973).

4. Jack P. C. Kleijnen, Statistical Techniques in Simulation: Part II, Marcel Dekker, Inc., New York (1975)

5. Emanuel Parzen, Stochastic Processes, Holden Day, Inc., San Francisco (1962)

6. James R. Wilson and A. Alan B. Pritsker, "A Procedure for Evaluating Startup Policies in Simulation Experiments," Proc. 1977 Summer Computer Simulation Conference, Chicago, Ill. (1977) pp. 83-89.

APPENDIX

In this appendix, we give the proofs of the results presented in section 2. Throughout we presume a real-valued stochastic process N(t) with integrable sample paths, or more precisely that the set of integrable sample paths of the N(t) process has probability one. This is true in particular if $N(t) = \eta(S(t))$ where S(t) is a finite state Markov process. It follows that the expression

$$\overline{N}_t = \frac{1}{t}\int_0^t N(u)du$$

is well defined.

Lemma 1: Let N(t) be a covariance stationary process with continuous autocovariance function R_t, $-\infty < \infty$. Then

$$\text{var } \overline{N}_t = \frac{2}{t^2}\int_0^t (t-u)R_u \, du \tag{1}$$

Proof:

$$\text{var } \overline{N}_t = \text{cov}\left(\frac{1}{t}\int_0^t N(x)\, dx \, , \frac{1}{t}\int_0^t N(y)\, dy\right)$$

$$= \frac{1}{t^2}\int_0^t\int_0^t \text{cov}(N(x),N(y))\, dxdy$$

$$= \frac{1}{t^2}\int_0^t\int_0^t R_{x-y}\, dxdy$$

Under the change of variable

$$u = x - y$$

$$v = y$$

this becomes

$$\text{var } \overline{N}_t = \frac{1}{t^2}\int_{-t}^{0}\int_{v=-u}^{v=t} R_u\, dvdu + \frac{1}{t^2}\int_0^t\int_{v=0}^{v=t-u} R_u\, dvdu$$

$$= \frac{1}{t^2}\int_{-t}^{0}(t+u)R_u\, du + \frac{1}{t^2}\int_0^t (t-u)R_u\, du \quad .$$

$$= \frac{2}{t^2}\int_0^t (t-u)R_u\, du$$

as was to be shown.

If, in addition, it is assumed that $R_t \geq 0$ for all t, then

$$\lim_{t\to\infty} t \text{ var } \overline{N}_t = \lim_{t\to\infty} R_t$$

$$\lim_{t\to\infty} t \text{ var } \overline{N}_t = 2\int_0^\infty R_t\, dt$$

can be derived using L'Hospital's rule. However, neither of these results is needed for the subsequent development.

Lemma 2: Let S(t) for $t \geq 0$ be a finite state Markov process and let

$$N(t) = \eta(S(t))$$

for some real valued function η. If the S(t) process possesses steady state initial conditions then the N(t) process is covariance stationary and

$$R_t = h'T(\mathbb{P}(t) - \Pi)h \tag{2}$$

where

$$h' = [\eta(s_0),\eta(s_1),\ldots,\eta(s_p)]$$

$$T = \text{diag}(\pi_0,\pi_1,\ldots,\pi_p).$$

Proof: We first evaluate the noncentral moment

$$E\, N(t+u)N(u) = E\, \eta(S(t+u))\eta(S(u))$$

$$= \sum_q E(\eta(S(t+u))\eta(S(u))\,|\,S(u)=s_q)Pr(S(u)=s_q)$$

$$= \sum_q \eta(s_q)E(\eta(S(t+u))\,|\,S(u)=s_q)Pr(S(u)=s_q)$$

$$= \sum_q \eta(s_q)\sum_r \eta(s_r)P_{qr}(t)\; Pr(S(u)=s_q)$$

Under steady state initial conditions, we have the relations

$$Pr(S(u) = s_q) = \pi_q$$

$$E\, N(t) = E\, \eta(S(t)) = \sum_q \pi_q\, \eta(s_q) = EN(t+u)$$

From the above, it follows that

$$\text{cov}(N(t+u),N(u)) = EN(t+u)N(u) - EN(t+u)EN(u)$$

$$= \sum_q \eta(s_q)\sum_r \eta(s_r)P_{qr}(t)\pi_q - (\sum_q \pi_q\eta(s_q))^2$$

$$= \sum_q\sum_r \eta(s_q)\pi_q\,(P_{qr}(t) - \pi_r)\eta(s_r)$$

$$= h'T(\mathbb{P}(t) - \Pi)h$$

Since this expression is independent of u, the lemma follows.

We now give the proofs of Theorems 1 and 2 after making some additional remarks concerning Markov processes. With terms defined as in section 1, it can be shown [1] that

$$\mathbb{P}(t) = e^{At} \tag{3}$$

is the unique solution to the Chapman-Kolmogorov equations

$$\frac{d\,\mathbb{P}(t)}{dt} = \mathbb{P}(t)A\ , \qquad \mathbb{P}(0) = I \quad .$$

If A has linearly independent eigenvectors then it can be diagonalized as

$$A = V^{-1}\,\text{diag}(\alpha_0,\alpha_1,\ldots,\alpha_p)V \tag{4}$$

where $\alpha_0,\alpha_1,\ldots,\alpha_p$ are the eigenvalues of A. (3) and (4) can be combined to give

$$\mathbb{P}(t) = V^{-1}\,\text{diag}(e^{\alpha_0 t},e^{\alpha_1 t},\ldots,e^{\alpha_p t})V \ . \tag{5}$$

Since the limit

$$\Pi = \lim_{t\to\infty}\mathbb{P}(t)$$

exists for all finite state Markov processes [5], we conclude that $Re(\alpha_i) \leq 0$ for $i=0,1,\ldots,p$. Since A is deficient by one in rank, exactly one eigenvalue, say α_0, is zero. Thus, $Re(\alpha_i) < 0$ for $i=1,\ldots,p$ and we have

$$\Pi = V^{-1}\mathrm{diag}(1,0,\ldots,0)V \qquad (6)$$

$$\mathbb{P}(t) - \Pi = V^{-1}\mathrm{diag}(0,e^{\alpha_1 t},\ldots,e^{\alpha_p t})V \qquad (7)$$

<u>Proof of Theorem 1</u>: We first show that the integral

$$\int_0^\infty R_t \, dt$$

is convergent. From (2), this will follow if the matrix integral (formed by integrating each entry of the matrix)

$$Q = \int_0^\infty (\mathbb{P}(t) - \Pi)\, dt \qquad (8)$$

converges, in which case we will have

$$m = 2h'TQh. \qquad (9)$$

From (7) it is easily seen that the integral Q does converge and that

$$Q = V^{-1}\mathrm{diag}(0,\frac{-1}{\alpha_1},\ldots,\frac{-1}{\alpha_p})V\ . \qquad (10)$$

then from (4) and (6) we obtain

$$Q = V^{-1}\mathrm{diag}(1,0,\ldots,0)V - V^{-1}\mathrm{diag}(1,\frac{1}{\alpha_1},\ldots,\frac{1}{\alpha_p})V$$
$$= \Pi - (A+\Pi)^{-1} \qquad (11)$$

Substituting (11) into (9) and noting that

$$h'T\Pi h = (EN)^2 \qquad (12)$$

we obtain the result (i).

The proof of (ii) is analogous. If we let

$$U(t) = \frac{1}{t}\int_0^t (t-u)(\mathbb{P}(u) - \Pi)\, du \qquad (13)$$

then it follows from (1) and (2) that

$$t \text{ var } \overline{N}_t = 2h'TU(t)h \qquad (14)$$

Using (7) and (13) one can obtain

$$U(t)=\frac{1}{t}V^{-1}\mathrm{diag}(0,\frac{-t}{\alpha_1}+\frac{1}{\alpha_1^2}(e^{\alpha_1 t}-1),\ldots,\frac{1}{\alpha_p^2}(e^{\alpha_p t}-1))V \qquad (15)$$

$$= Q + \frac{1}{t}(A+\Pi)^{-2}(\mathbb{P}(t) - I) \qquad (16)$$

(ii) now follows by combining (9), (14) and (16). (iii) follows directly from (ii).

Theorem 2 is a consequence of the fact that the eigenvalues of rA are $r\alpha_0,r\alpha_1,\ldots,r\alpha_p$, while the eigenvectors are the same as those of A.

<u>Proof of Theorem 2</u>: With the obvious extension in notation, we have from (15) that

$$U(t,rA) = \frac{1}{t}V^{-1}\mathrm{diag}(0,\frac{-t}{r\alpha_1}+\frac{1}{r^2\alpha_1^2}(e^{r\alpha_1 t}-1),$$
$$\ldots,\frac{-t}{r\alpha_p}+\frac{1}{r^2\alpha_p^2}(e^{r\alpha_p t}-1))V$$
$$= \frac{1}{r^2 t}V^{-1}\mathrm{diag}(0;\frac{rt}{\alpha_1}+\frac{1}{\alpha_1^2}(e^{r\alpha_1 t}-1),$$
$$\ldots,\frac{-rt}{\alpha_p}+\frac{1}{\alpha_p^2}(e^{r\alpha_p t}-1))V$$
$$= \frac{1}{r}U(rt,A)$$

(ii) is now an immediate consequence of this and (14). Part (i) of Theorem 2 follows directly from part (ii) and Theorem 1 by letting $t\to\infty$.