STUDYING POLICY RETENTION RATES USING MARKOV CHAINS

JOSEPH O. MARKER

Abstract

How does one measure the effect of improved policy retention on such key variables as market share and profitability?

This paper will analyze this problem by:

- *using the theory of Markov chains to model policy retention and to determine key values such as steady-state probabilities;*
- *using current spreadsheet technology to solve the key matrix equations from Markov chain theory; and*
- *applying these results to determine key business variables such as effects on profitability and market share.*

1. INTRODUCTION AND PROBLEM STATEMENT

You run an insurance company. You know that retaining policies is good business, but you want to quantify its value.[1] To simplify the analysis, you assume that all policies are written for a fixed policy term, expire at the same time, and have no mid-term activity. It turns out that the theory of Markov chains provides help with the analysis.

Markov chains assume discrete time periods and a system with "states" and "transition probabilities," the probabilities of moving from one state to another in one time period. For example, a physical system may consist of particles that move from a state to state in each discrete time period.

[1] D'Arcy and Doherty [1] discuss the "aging phenomenon." Their paper looks at this phenomenon relative to the profitability of insuring a policyholder for several periods. This paper views the same phenomenon from the aggregate financial viewpoint of an entire corporation.

The number of successes in a sequence of independent Bernoulli trials with probability of success p is a Markov chain. The system is defined to be in state k at time n if there have been exactly k successes in the first n trials. The transition probability of going from state k to state $k+1$ is p and the transition probability of staying at state k is $q = 1 - p$. In this paper, the term "Markov chain"[2] refers to a system with stationary transition probabilities. This means that if a particle is in state j at time t, then the conditional probability of going to state k at time $t+1$ does not depend on t, nor does it depend on any of the states that the particle was in prior to time t.

For the policy retention problem of this paper we replace the term "particle" by the term "customer." We say that the customer is in state k for $k = 0,1,2\ldots$ if the customer has been insured with the company for k consecutive time periods (one time period is equal to one policy term).

$k = 0$ refers to a person not currently insured with the company.

$k = 1$ refers to a policyholder in his/her first policy term.

$k = 2$ refers to a policyholder who has renewed once.

To study retention we define retention probabilities $\{r_k,\ k = 0,1,2,\ldots\}$ such that

r_k is the probability of renewing a policy that has been with the company for k time periods (that is, r_k is the probability that a customer currently in state k will pass to state $k+1$ in the next time period), and

r_0 is the probability of writing a customer who is not currently insured with the company.

We need an initial distribution $\{p_k^{(0)},\ k = 0,1,2\ldots\}$ where $p_k^{(0)}$ is the proportion of the entire population that has been insured

[2] For discussion of Markov chains see Feller [2, p. 372] or Resnick [4, p. 60]. The notation in this paper more closely follows that of Feller.

with the company for k years. Note that $1 - p_0^{(0)}$ is the company's initial "market share."

With this notation the *matrix of transition probabilities*[3] is:

$$\mathbf{A} = \begin{pmatrix} 1-r_0 & r_0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1-r_1 & 0 & r_1 & 0 & 0 & \cdots & 0 & 0 \\ 1-r_2 & 0 & 0 & r_2 & 0 & \cdots & 0 & 0 \\ 1-r_3 & 0 & 0 & 0 & r_3 & \cdots & 0 & 0 \\ 1-r_4 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1-r_{N-1} & 0 & 0 & 0 & 0 & \cdots & 0 & r_{N-1} \\ 1-r_N & 0 & 0 & 0 & 0 & \cdots & 0 & r_N \end{pmatrix}.$$

Here a_{ij} is the conditional probability of going from state i to state j in one time period. The indices i and j range from 0 to N. (See Appendix C for a discussion of chains with an infinite number of states.) The maximum value N may be set at a number of policy periods after which the retention is essentially constant.

For this retention problem only the first column (more correctly called the zeroth column) and the superdiagonal are non-zero, along with a_{NN}. This is because a customer in state j either moves to state $j+1$ (if the policy renews) or to state 0 (if the customer takes his/her business elsewhere). The retention rate is simply the probability that the policy will renew at its next expiration.

Notation Conventions

Superscripts within parentheses, such as $^{(n)}$, refer to time periods, or in the case of matrix elements, refer to n-step transi-

---

[3] Feller [2] contains an example that is mathematically equivalent to this Markov chain, except that, in his example, the number of states is infinite. He refers to state k as the "age," and says that at the next time period the system will either pass to age $k+1$ or will go back to age 0 and start afresh. See [2, pp. 382, 390, 398, 403].

tion probabilities. The n-step transition probability of going from state i to state j is the conditional probability that a customer in state i will, in n periods, be in state j.

Plain superscripts refer to exponents, except ^{t} refers to matrix transpose.

Subscripts refer to states of the system.

Vectors and matrices are in boldface.

2. GENERAL RESULTS ABOUT MARKOV CHAINS AND THE STEADY-STATE DISTRIBUTION

Given an initial distribution of states $\mathbf{p}^{(0)}$, the distribution of states in the next period is given by $\mathbf{p}^{(1)} = \mathbf{A}^t\mathbf{p}^{(0)}$, where $\mathbf{A}^t$ is the transpose of $\mathbf{A}$. This follows immediately, since $p_k^{(1)} = \sum_j a_{jk} p_j^{(0)}$. Each term on the right represents the probability that the system is in state j at time 0 and passes to state k at time 1. The summation over j then is the total probability of being in state k at time 1. The kth element of $\mathbf{p}^{(1)}$ is thus the inner product of the vector $\mathbf{p}^{(0)}$ and the kth column of matrix $\mathbf{A}$, which is the definition of multiplication on the left by the transpose.

Similarly, the conditional probability $a_{jk}^{(2)}$ of moving to state k in two steps given initial state j is given by

$$a_{jk}^{(2)} = \sum_m a_{jm} a_{mk},$$

which means that the two-step transition matrix is given by $\mathbf{A}^{(2)} = \mathbf{A}^2$. This is intuitively obvious by observing that, in order to get from j to k in two steps, one must stop at some state m at the first step. By induction, the n-step transition matrix[4] is given by $\mathbf{A}^n$. By definition, the element $a_{jk}^{(n)}$ is the probability, given state j, of being in state k n-periods later.

[4]Feller [2, pp. 382, 383].

Let $\mathbf{B}$ = transpose of $\mathbf{A} = \mathbf{A}^t$. Then the distribution at time n is given by

$$\mathbf{p}^{(n)} = (\mathbf{A}^t)\mathbf{p}^{(n-1)} = \mathbf{B}\mathbf{p}^{(n-1)} = (\mathbf{B}^n)\mathbf{p}^{(0)} \qquad \text{for any} \quad n \geq 1.$$

The *steady-state (or invariant) probability distribution* is defined as that solution of the equation $\mathbf{p}^* = \mathbf{B}\mathbf{p}^*$ for which $\sum p_j^* = 1$. It turns out that the steady-state probabilities are very important to our original business retention problem. We will discuss later how to calculate $\mathbf{p}^*$.

A key result is that

$$\mathbf{B}^n\mathbf{p}^{(0)} \to \mathbf{p}^* \quad \text{as} \quad n \to \infty \quad \text{for any initial distribution } \mathbf{p}^{(0)}.$$

The proof is in Appendix B. This limiting result says that the ultimate distribution of customers by state (remember: "state" is the number of consecutive renewals) is independent of the initial distribution but depends only on the steady state probabilities associated with the retentions.

3. CALCULATING THE STEADY-STATE (INVARIANT) DISTRIBUTION

There are several approaches to calculating the invariant distribution for our retention problem.

3.1. Use the Definition Directly

Recall that the matrix $\mathbf{B}$ for the retention problem is given by

$$\mathbf{B} = \mathbf{A}^t = \begin{pmatrix} 1-r_0 & 1-r_1 & 1-r_2 & 1-r_3 & 1-r_4 & \cdots & 1-r_{N-1} & 1-r_N \\ r_0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & r_1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & r_2 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & r_3 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & r_{N-1} & r_N \end{pmatrix},$$

where all the r_k are strictly between 0 and 1.

The defining equations for invariance are:

$$p_k = r_{k-1} p_{k-1} \qquad \text{for} \quad k = 1, 2, 3 \cdots N-1, \tag{3.1}$$

$$p_N = r_{N-1} p_{N-1} + r_N p_N, \qquad \text{and} \tag{3.2}$$

$$p_0 = (1 - r_0) p_0 + (1 - r_1) p_1 + \cdots + (1 - r_N) p_N. \tag{3.3}$$

From Equation 3.1 we obtain

$$p_k = r_0 r_1 r_2 \cdots r_{k-1} p_0 \qquad \text{for} \quad k = 1, 2, 3, \ldots N-1. \tag{3.4}$$

From Equation 3.4 we can see that the terms 0 through $N-1$ on the right-hand side of Equation 3.3 add to $p_0 - r_{N-1} p_{N-1}$. From Equation 3.2 the last term on the right-hand side of Equation 3.3 equals $r_{N-1} p_{N-1}$. Thus, we can choose an arbitrary value for p_0, define the remaining p_k by Equations 3.1 and 3.2, and Equation 3.3 will be automatically satisfied. Once all the p_k are calculated, just rescale them so they add to 1 and these values are the invariant probabilities.

Thus the retention problem has a particularly simple form of transition matrix that allows the steady-state probabilities to be calculated directly from the definition.

3.2. A Simple Machine-Oriented Approach[5]

The vector $\mathbf{p}^*$, whose transpose is defined by

$$(\mathbf{p}^*)^t = (1, 1, \ldots 1, 1)(\mathbf{I} - \mathbf{A} + \mathbf{ONE})^{-1},$$

defines an invariant distribution. Here $\mathbf{I}$ is the identity matrix and **ONE** is the square matrix all of whose entries are 1. Resnick [4] proves this handy proposition. This result requires that $\mathbf{A}$ be irreducible, which we prove in Appendix C.

[5] Resnick [4, p. 138].

3.3. Use a Spreadsheet "Solver"

A spreadsheet "solver"[6] can solve for the steady-state probabilities. A typical spreadsheet solver

(a) maximizes, minimizes, or sets a target cell to a specific value

(b) subject to constraint equations or inequalities

(c) by changing a set of "decision cells."

The use of the target cell is optional. The solver can be used to simply produce values of the decision cells that satisfy the given constraints.

Recall that steady state probability vector is simply the solution $\mathbf{p}^*$ of the matrix equation $\mathbf{Bx} = \mathbf{x}$, for which $\sum x_j = 1$, where $\mathbf{B} = \mathbf{A}^t$. This equation can be rewritten as $\mathbf{Cx} = 0$, where $\mathbf{C} = \mathbf{B} - \mathbf{I}$ and $\mathbf{I}$ is the identity matrix.

Now setting up the solver is simple:

1. Set up the matrix $\mathbf{C}$, which is a function of the transition probabilities $\mathbf{A}$.
2. Set up a vector $\mathbf{x}$, the vector of decision variables that are allowed to change when the solver is run.
3. Set up a vector $\mathbf{z}$ as the matrix product $\mathbf{Cx}$.
4. Run the solver with the following constraints:

$$\mathbf{z} = 0 \qquad \text{and} \qquad \sum x_j = 1.$$

The resultant vector $\mathbf{x}$ is the steady-state probability vector $\mathbf{p}^*$.

We present this solution using the solver because solvers are being commonly used to handle problems involving maximizing, minimizing, and satisfying constraints, and a solver for our

[6] The particular solver used in this paper is that from the Microsoft Excel spreadsheet.

retention problem does not require the same linear algebra skills that other solutions entail.

4. SPREADSHEET EXAMPLE TO MODEL THE RETENTION PROBLEM

Recall that we have translated the retention problem into Markov chain terms and have reviewed some characteristics of Markov chains. Appendix A displays printouts from a spreadsheet set up to analyze retentions. The spreadsheet is documented, but here are some of the highlights.

The Basic Data section asks us to input the retention probabilities $\{r_k,\ k = 0,1,2,\ldots N\}$ and initial probability distribution $\{p_k^{(0)}, k = 0,1,2\ldots N\}$. Recall that r_k is the probability that a policyholder that has been insured for k policy periods will renew when his/her policy expires, and r_0 is the probability that the company will capture as new business a customer not currently insured with the company. The end of the Basic Data section translates these retention probabilities into the matrix $\mathbf{A}$ of one-step transition probabilities.

For example, in Appendix A the company's initial market share is 10%, since the proportion $p_0^{(0)}$ of the market not insured by the company is 90%. Since $N = 9$ and $p_N^{(0)} = .043$, 4.3% of the market has been insured with the company nine or more policy periods. At the next renewal cycle $r_0 = 1.0\%$ of the population not insured by the company will be captured as new business.

The section labeled "Distribution At Time n" shows how the distribution $\mathbf{p}^{(n)}$ changes after n time periods. Recall that $\mathbf{p}^{(n)} = \mathbf{B}\mathbf{p}^{(n-1)}$, and that $\mathbf{p}^{(n)} = \mathbf{B}^n\mathbf{p}^{(0)}$, where $\mathbf{B} = \mathbf{A}^t$. We have shown that $\mathbf{p}^{(n)} \rightarrow \mathbf{p}^*$ (the steady-state probability) as $n \rightarrow \infty$.

The value of these calculations is that they allow us to get a feel for how fast the limit is approached. In the real world, a company does not have an infinite time horizon to wait for the limiting behavior to be realized. The "Distribution At Time n"

explanation also allows us to restrict the model to a finite planning horizon. Too many managerial changes during the convergence period could invalidate the Markov chain assumption that the transition matrix is stationary over time. The n-step transition matrices in Appendix A converge to a matrix which has the steady-state distribution vector in each of its columns.

5. RETURN TO THE ORIGINAL RETENTION PROBLEM

When any action affects retention, it changes the transition matrix **A**. Improved retention means larger superdiagonal elements (probabilities of renewal) and smaller elements in the first column (probabilities of non-renewal). In this section, we will study our original set of retention assumptions and their effect on key business variables. The spreadsheet with those results is shown as Appendix A. Then we will see how a shift in retention (Appendix B) may change the results.

We have used the theory of Markov chains, along with spreadsheet tools, to compute steady-state probabilities for a given set of retention rates. We have shown that the distribution of states of the system (recall that the "state" of an individual customer is the number of consecutive policy renewals for that customer) approaches the steady-state probabilities, as time goes on, regardless of the initial distribution.

The spreadsheet in Appendix A gives us a sense for how quickly this convergence takes place. It is easy to calculate the distribution at time n, because the n-step transition matrix is just the nth power of the one-step transition matrix. Mahler's paper, "A Markov Chain Model of Shifting Risk Parameters," provides a mathematical treatment of the rate of convergence [3].

Thus we have a wealth of tools that give us information about the probability distribution of states throughout time. These probabilities are not in themselves of much interest to management. However, there are functions of these probabilities that are of great interest. For example, the projected market share is of keen interest. Because we have included state 0 in our definition of states (the customer is in state 0 if he/she is not currently insured with the company), the market share at time n is given by $1 - p_0^{(n)}$.

Loss ratios, expense ratios, and combined ratios greatly interest management. Most observers would agree that renewing an existing policy is much less expensive than writing a new policy. It follows that increasing the retention rate will improve the expense ratio. Most would also agree that the loss ratio for a customer who has been on the company books for a long period of time will be lower than for a new or recent customer. Actions that improve retention should improve the loss ratio.

The last page of Appendix B, Combined Ratio Differential, illustrates how to estimate this effect. To estimate the effect of retention on combined ratio one needs a sense of how loss/expense ratios vary by state (the number of consecutive policy renewals). The phrase "needs a sense of" is intentionally vague. It could mean that we have data on loss or expense ratios by state. More likely it means that we have some information that would enable us to make an assumption about how the loss or expense ratio varies by state. For instance we may be able to say that a new policyholder has a 10% worse loss ratio than a long-standing policyholder. Or it could mean that we accept a management estimate of this differential and use the model to check the effect of retention under different estimates.

Once we have made a reasonable assumption about these differentials by state, we are ready to estimate the effect of improved retention. This is simply a matter of:

1. entering the initial distribution and retention probabilities in to the spreadsheet;
2. running the spreadsheet to determine the steady-state probabilities and how quickly the system approaches those limiting probabilities; and

TABLE 1

ORIGINAL RETENTION ASSUMPTIONS

	State k									
	0	1	2	3	4	5	6	7	8	9
Retention	.0100	.8500	.9000	.9000	.9000	.9000	.9000	.9000	.9000	.9000
Steady-State Probability	.9132	.0091	.0077	.0069	.0063	.0057	.0051	.0046	.0041	.0371
Combined Ratio Differential	.2000	.1000	.0800	.0600	.0400	.0200	0	0	0	0

3. applying the differentials in loss/expense ratios to the various probabilities to arrive at an "average differential" or "average loss/expense ratio."

Then we make the same calculation using the "improved retentions" in Appendix B and compare the results to estimate the effect of the change in retention.

The "Combined Ratio Differential" section in Appendix A shows a calculation of this nature for the original retention probabilities. Here we externally determined (or hypothesized) various combined ratio differentials by state relative to the combined ratio for a long-standing (i.e., seven term or longer) policyholder. The results are summarized in Table 1. The retention and combined ratio differentials are inputs to the calculation. The steady-state probabilities and the average differential are calculated. From Appendix A the average differential is +.0446 using the steady-state probabilities as weights. That is, on average the book of business will have a 4.5% higher average combined ratio than if the book consisted entirely of long-term customers.

Now suppose that the company takes some action that improves retention. Such an action might be a new billing option, more advertising, etc. The number of such actions is limited only

TABLE 2

IMPROVED RETENTION ASSUMPTIONS

	State k									
	0	1	2	3	4	5	6	7	8	9
Retention	.0120	.8700	.9200	.9200	.9200	.9200	.9200	.9200	.9200	.9200
Steady-State Probability	.8753	.0105	.0091	.0084	.0077	.0071	0066	.0060	.0055	.0637
Combined Ratio Differential	.2000	.1000	.0800	.0600	.0400	.0200	0	0	0	0

by the creativity of the sales or marketing manager. In the example in Appendix B, our improved retention assumption is that r_k increases by .02 for $k >= 1$ and r_0 increases by .002 (recall r_0 is the probability that the company writes a new customer). We can use the same spreadsheet with the revised retention and obtain the results shown in Table 2, assuming that the differentials have not changed.

Now what has been the effect of the management action to improve the retention? The ultimate market share increases from 8.7% to 12.5%. The ultimate loss ratio decreases by 0.8% (i.e., the combined ratio differential drops from 4.4% to 3.6%). Now remember that these are "ultimate" results and we know that, for Markov chains, it may take quite a few renewal cycles to approach these limiting results!

The insurer must weigh these benefits against the costs. For example, if an improved billing system produces the increased retention, then the improved market share and loss ratio must overcome the cost of maintaining and building the billing system. If instead, a rate decrease is used to improve retention then it is likely that the overall combined ratio itself will increase and wipe out the benefits from the retention improvement. The exact effect will depend on the price elasticity of demand for the product.

5. COMPARISON TO SINGLE POLICYHOLDER APPROACH

The approach used in this paper examines the financial effects of retaining policies on the entire company's book of business. Starting with an initial distribution of business by policy age and a set of transition probabilities, we use Markov chain theory to model the distribution over time. Because one of the states of the system (i.e., state 0) consists of potential customers not insured by the company, the model produces estimates of total growth as well as distribution by policy age. We then hypothesize differences in loss ratio by policy age to examine changes in profitability over time. The Markov chain approach enables us to examine the effects on growth and profitability of changes in the transition probabilities.

This entire approach is an aggregate approach in that it looks at the growth and profitability of a company's entire book of business over time. In contrast, D'Arcy and Doherty [1] approach the "aging phenomenon" by tracing the profitability of a single insured over time. They start with a new customer (corresponding to state 1 in this paper) and calculate the profitability of that customer's policies from the initial date through the last renewal, discounting all calculations to the initial policy inception date. D'Arcy and Doherty hypothesize differing levels of profitability by policy age. In their model the probability of renewal is constant over time.

D'Arcy and Doherty study the price that will optimize present and future profits from a customer added to the books. How do the approach of this paper and D'Arcy-Doherty relate? We can express the D'Arcy-Doherty models in Markov chain terms as follows:

The initial distribution $\mathbf{p}^0$ consists of a probability of 1.0 of being a new policyholder.

The retention probabilities r_k are constant (called W in [1]).

The state 0 becomes an "absorbing state." That is, there is no more action for the individual policyholder once he/she non-renews.

Using Markov chains to study the aging phenomenon in [1] is not useful because the transition probabilities are so simple that Markov chain theory is not needed. D'Arcy and Doherty concentrate on a single policyholder and the transition matrix does not satisfy the criteria for using the theorems about invariant distributions.

D'Arcy and Doherty concentrate on the single policyholder and are sophisticated in treating differing loss ratios and the time value of money in arriving at proper prices. Their analysis could be used as an input to this paper's aggregate model. We could use the models in [1] to enable us to calculate the expected present value of profit for each policy renewal (i.e. for each state k). This gives us a set of expected profits corresponding to the various states in the retention model. There is no need to sum these discounted present values for all the renewals of a single customer as is done in [1]. We can then hypothesize an initial distribution and use the transition matrix as was done in this paper. Our Markov chain model determines the distribution of states of the system over time. With this information and the expected profits by state, we can determine the company's expected profit over time. The Markov chain model allows us to easily vary the retention rate by state of the system.[7]

Both papers refer to optimizing profitability over time. In [1] this is done by calculating the present value of expected profits over the life of an individual policyholder as a function of price. The renewal rate W is adversely affected by increasing price, so that there is an optimum price above which the profits begin decreasing.

---

[7]The possibility of renewal rates changing by policy age is mentioned briefly in [1, p. 38].

In this paper, the expected profits for the entire corporation are calculated using Markov chains. Increasing the price for policies increases the profit at each renewal. However, price increases lower the renewal probabilities r_k. This decreases both the market share and the number of customers in the higher states (i.e., long-term policyholders). At some point raising the price adversely affects renewal probabilities so much that total profit is adversely affected.

Both papers mention elasticity of demand (with respect to price) as critical values. Basically, the more elastic the demand, the more difficult it is to increase overall profit through price increases.

In general, we could use D'Arcy and Doherty [1] to establish the expected profit by age of policy. Then we could plug this information into the Markov chain model to determine aggregate profitability over time and growth for the company.

6. OBSERVATIONS AND CAVEATS

Many companies have neither very good retention information nor very good ideas of loss ratio differentials by retention. The Markov chain model is useful even in these circumstances. To illustrate, some company managements have wildly inflated ideas of the benefits of improved retention on market share and profitability. Let's assume that the actuary can persuade management to "guess" the improvement both in retention rate and in combined ratio differentials by state. The company can then use the model to produce profitability and market share change estimates that are more realistic than management's original "feeling." As the company obtains better data, some of the hypothesizing can be replaced with observations. There is a high probability that retention data will improve because it is of universal interest among top management.

In using this type of modeling one must be careful not to compound too many assumed improvements. For example, suppose the retention on long-standing business is 95%. A new billing plan claims to increase this by 2% (additively). A few months later the ability to "account sell" increases the number again by 2%. Then a fancy new endorsement produces another 2% increase. Now the implied retention rate is 101%, which is absurd. This sounds ridiculous, but companies do act this way when the actions are separated in time and the company loses its memory due to management changes.

A better way to view this is to express these increases as reductions in the non-renewal or lapse rate and then compound them properly. For example, we might say that each of the three actions above reduces the lapse rate by 40% (i.e., reduces it from .05 to .03), so that the resultant retention from this series of actions becomes:

$$1 - (.05 \times .60 \times .60 \times .60) = .989.$$

The assumption that the policy renewal process is a Markov chain is a simplification of the real world. Recall that the Markov property says that the probability of passing to a given future state depends on the current state but does not depend on any prior history. This implies, for example, that the probability r_0 of capturing a new customer is the same whether or not that customer has ever been previously insured with the company. This is probably not an accurate assumption.

We can attempt to get around this assumption by defining two "0" states: state "0a" for potential customers who have never been with the company, and state "0b" for potential policyholders who had been previously insured. With this formulation the transition matrix is such that current policyholders (state 1 or higher) can never get to state 0a. It turns out that the invariant distribution assigns probability 0 to state 0a (that is, everyone eventually becomes a policyholder or former policyholder of the company).

In this situation the distribution of states approaches the invariant distribution very slowly. In one reasonable example (where the probability of capturing a new customer is a high 5%) the limiting distribution was not approached even after sixty-four time periods. Thus, in this situation the Markov chain model is useful for finite time periods, but the study of the invariant distribution is somewhat academic.

The model in this paper assumes that time is discrete, that all customers have policies with inception dates at these discrete time periods, and that the only possible actions are renewal or non-renewal. Of course, we know that customers can cancel or purchase policies at any time, and that endorsement activity is probably more frequent than renewal activity. This would require a continuous time Markov process with a richer set of options.

In selecting actions that improve financial results through "improved retention," we must verify that the action itself does not adversely affect the profitability for each state. A classic action that violates this condition is a rate decrease. Obviously, this action would decrease the profitability of each state, even though it improves retention.

7. CONCLUSION

This paper uses the theory of Markov chains to analyze retention rates and how they affect key insurance variables. In the paper, the Markov chain state for a customer is the number of consecutive policy periods the customer has been insured with the company. Determining the ultimate, or limiting, distribution for Markov chains involves solving matrix equations of the form $\mathbf{Bx} = \mathbf{x}$.

The paper shows how to do this using spreadsheets. Finally, the paper illustrates how changing the retention rates (i.e. the transition probabilities in the Markov chain) might change key business variables such as profitability and market share. There is also a discussion of how the model interrelates with an earlier "policy age" model by D'Arcy and Doherty.

REFERENCES

[1] D'Arcy, Stephen and Neil Doherty, "The Aging Phenomenon and Insurance Prices," *PCAS* LXXVI, 1989, pp. 24–44.

[2] Feller, William, *An Introduction to Probability Theory and its Applications*, Volume I, John Wiley & Sons, Inc., 1968.

[3] Mahler, Howard, "A Markov Chain Model of Shifting Risk Parameters," *PCAS* LXXXIV, 1997, pp. 581–659.

[4] Resnick, Sidney, *Adventures in Stochastic Processes*, Birkhauser: Boston, 1992.

APPENDIX A

INITIAL RETENTION ASSUMPTIONS

INTRODUCTION

This spreadsheet carries out the calculations for the insurance retention problem. The retention problem is set up as as a Markov Chain, where a customer is in state k if he/she has been insured with the company for k consecutive periods. State 0 refers to a potential customer not currently insured with the company. Each state k has an associated "retention probability" r_k, where r_k is the probability that a customer in state k renews his/her policy. The customer non-renews, i.e. moves to state 0, with probability $1-r_k$.

The retention problem translates to a Markov chain as follows:

The states of the Markov chain are defined as in the retention problem.

The matrix of transition probabilities $\mathbf{A} = (a_{i,j})$ is defined as follows:

$$\begin{aligned} a_{k,k+1} &= r_k, \\ a_{k,0} &= 1-r_k, \qquad \text{and} \\ a_{k,j} &= 0 \qquad \text{for all other } j. \end{aligned}$$

William Feller [2, p. 382] discusses this Markov chain problem. Sidney Resnick [4] describes this Markov chain as the "Success Run Chain."

The retention problem also requires a vector $\mathbf{p}^{(0)}$, the initial probability distribution of states.

With the transition probabilities $\mathbf{A}$ and the initial distribution $\mathbf{p}^{(0)}$ specified, the spreadsheet calculates the "steady-state," or invariant, distribution of states, to which the system converges; the probability distributions at various points in time, to check for rate of convergence; and changes in market share and profitability over time.

BASIC DATA

Section 1: Calculate the distribution given the matrix of transition probabilities **A**, where $a_{i,j}$ is probability of going from state i to state j in one step. The initial distribution is $\mathbf{p}^{(0)}$. This particular example is an effort to model insurance retention. State i is the number of years the customer has been insured with the company. The first state (zero) refers to a potential customer not currently insured. The next state (one) refers to a first-year insured, etc.

Input Section

Input the *retention probabilities* of going from state i to state $i+1$. That is, the input for state 0 is the probability that someone currently insured elsewhere will be written as new business. The input for state $i>0$ is the probability of renewing a policy of someone that the company has insured for i years.

Then input the *initial distribution* $\mathbf{p}^{(0)}$ of insureds. For $i = 0$, this is the proportion of the population not currently insured with the company. For $i > 0$, this is the proportion of the population insured with the company for i consectutive policy terms. The last column is the proportion insured with the company for 9 or more consecutive terms.

	State i									
	0	1	2	3	4	5	6	7	8	9
Retention Probabilities	0.01	0.85	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
Initial Distribution $\mathbf{p}^{(0)}$	0.9	0.01	0.009	0.008	0.007	0.007	0.006	0.005	0.005	0.043

Following is the resultant matrix **A** of one-step transition probabilities:

0.99	0.01	0	0	0	0	0	0	0	0
0.15	0	0.85	0	0	0	0	0	0	0
0.10	0	0	0.90	0	0	0	0	0	0
0.10	0	0	0	0.90	0	0	0	0	0
0.10	0	0	0	0	0.90	0	0	0	0
0.10	0	0	0	0	0	0.90	0	0	0
0.10	0	0	0	0	0	0	0.90	0	0
0.10	0	0	0	0	0	0	0	0.90	0
0.10	0	0	0	0	0	0	0	0	0.90
0.10	0	0	0	0	0	0	0	0	0.90

DISTRIBUTION AT TIME *n*

This section shows how to calculate the distribution at time n, given the initial distribution $\mathbf{p}^{(0)}$ and the matrix **A**.

Note that if **p** is the distribution of states at any time, then $\mathbf{A}^t$ (the transpose of **A**) times **p** is the distribution in the next time period. That is, the probability that the system is in state m in the next time period is the mth column of matrix **A** times the distribution **p**.

Matrix $\mathbf{A}^t$

0.99	0.15	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
0.01	0	0	0	0	0	0	0	0	0
0	0.85	0	0	0	0	0	0	0	0
0	0	0.90	0	0	0	0	0	0	0
0	0	0	0.90	0	0	0	0	0	0
0	0	0	0	0.90	0	0	0	0	0
0	0	0	0	0	0.90	0	0	0	0
0	0	0	0	0	0	0.90	0	0	0
0	0	0	0	0	0	0	0.90	0	0
0	0	0	0	0	0	0	0	0.90	0.90

Probability Distributions $\mathbf{p}^{(n)}$ at time period n, for $n = 1, 2, 3 \ldots$

$\mathbf{p}^{(0)}$	State	1	2	3	4	5	6	7	8	9
0.900	0	0.9015	0.9028	0.9039	0.9049	0.9059	0.9067	0.9074	0.9080	0.9086
0.010	1	0.0090	0.0090	0.0090	0.0090	0.0090	0.0091	0.0091	0.0091	0.0091
0.009	2	0.0085	0.0077	0.0077	0.0077	0.0077	0.0077	0.0077	0.0077	0.0077
0.008	3	0.0081	0.0077	0.0069	0.0069	0.0069	0.0069	0.0069	0.0069	0.0069
0.007	4	0.0072	0.0073	0.0069	0.0062	0.0062	0.0062	0.0062	0.0062	0.0062
0.007	5	0.0063	0.0065	0.0066	0.0062	0.0056	0.0056	0.0056	0.0056	0.0056
0.006	6	0.0063	0.0057	0.0058	0.0059	0.0056	0.0050	0.0050	0.0050	0.0050
0.005	7	0.0054	0.0057	0.0051	0.0052	0.0053	0.0050	0.0045	0.0045	0.0045
0.005	8	0.0045	0.0049	0.0051	0.0046	0.0047	0.0048	0.0045	0.0041	0.0041
0.043	9	0.0432	0.0429	0.0430	0.0433	0.0431	0.0430	0.0430	0.0428	0.0422

Note that the n-step transition probability is given by raising matrix $\mathbf{A}$ to the nth power. The distribution at time n is given by $(\mathbf{A}^t)^n$ times $\mathbf{p}^{(0)}$.

Shown below are the transposes of some n-step transition matrices:

Two-step									
0.9816	0.2335	0.1890	0.1890	0.1890	0.1890	0.1890	0.1890	0.1890	0.1890
0.0099	0.0015	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
0.0085	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.7650	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.8100	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.8100	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.8100	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.8100	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.8100	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.8100	0.8100	0.8100

Four-step									
0.9675	0.3741	0.3388	0.3388	0.3388	0.3388	0.3388	0.3388	0.3388	0.3388
0.0097	0.0031	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027	0.0027
0.0083	0.0020	0.0016	0.0016	0.0016	0.0016	0.0016	0.0016	0.0016	0.0016
0.0076	0.0011	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008
0.0069	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.6197	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.6561	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.6561	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.6561	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.6561	0.6561	0.6561	0.6561	0.6561

STEADY-STATE PROBABILITIES

This section shows the steady-state probabilities found using the solver.

The steady-state probability $\mathbf{p}^*$ is characterized by $\mathbf{A}^t \times \mathbf{p}^* = \mathbf{p}^*$, or $(\mathbf{A}^t - \mathbf{I}) \times \mathbf{p}^* = 0$, where $\mathbf{I}$ is the identity matrix.

Use the solver to find the solution. Let $\mathbf{C} = \mathbf{A}^t - \mathbf{I}$. The steady-state probability $\mathbf{p}^*$ is the solution of the linear system $\mathbf{Cx} = 0$ for which the elements of $\mathbf{x}$ sum to 1.0. After using the solver, the vector $\mathbf{x}$ contains the steady-state probabilities, and the vector $\mathbf{z} = \mathbf{Cx}$ contains all zeros.

Matrix $\mathbf{C} = \mathbf{A}^t - \mathbf{I}$										$\mathbf{x}$	$\mathbf{z} = \mathbf{Cx}$
−0.01	0.15	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.91324	0
0.01	−1.00	0	0	0	0	0	0	0	0	0.00913	0
0	0.85	−1.00	0	0	0	0	0	0	0	0.00776	0
0	0	0.90	−1.00	0	0	0	0	0	0	0.00699	0
0	0	0	0.90	−1.00	0	0	0	0	0	0.00629	0
0	0	0	0	0.90	−1.00	0	0	0	0	0.00566	0
0	0	0	0	0	0.90	−1.00	0	0	0	0.00509	0
0	0	0	0	0	0	0.90	−1.00	0	0	0.00458	0
0	0	0	0	0	0	0	0.90	−1.00	0	0.00413	0
0	0	0	0	0	0	0	0	0.90	−0.10	0.03713	0
										1.00000	

In this application of the solver, the "target cell" for the solver is undefined, since there is no objective function to maximize or minimize.

COMBINED RATIO DIFFERENTIAL

This section illustrates how studying the retention problem helps businesses evaluate profitability and market share. Generally combined ratios are better for customers who have been retained longer, due to lower expenses and/or better loss ratios. By comparing the average combined ratios before and after improving retention, one can measure the financial effects of changing policy retention.

State i	Steady State Probability $\mathbf{x}$	Assumed Combined Ratio Differential	Assumed Base Combined Ratio	Combined Ratio
0	0.913242	N/A		
1	0.009132	20.00%		115.54%
2	0.007763	10.00%		105.54%
3	0.006986	8.00%		103.54%
4	0.006288	6.00%		101.54%
5	0.005659	4.00%		99.54%
6	0.005093	2.00%		97.54%
7	0.004584	0.00%		95.54%
8	0.004125	0.00%		95.54%
9	0.037128	0.00%	95.54%	95.54%
Market share		8.68%		
Average combined ratio differential		4.46%		
Average combined ratio		100.00%		

APPENDIX B

IMPROVED RETENTION ASSUMPTIONS

BASIC DATA

Section 1: Calculate the distribution given the matrix of transition probabilities **A**, where $a_{i,j}$ is the probability of going from state i to state j in one step. The initial distribution is $\mathbf{p}^{(0)}$. This particular example is an effort to model insurance retention. State i is the number of years a customer has been insured with the company. The first state (zero) refers to a potential customer not currently insured. The next state (one) refers to a first-year insured, etc.

Input Section

Input the retention probabilities of going from state i to state $i+1$. That is, the input for state 0 is the probability that someone currently insured elsewhere will be written as new business. The input for state $i > 0$ is the probability of renewing a policy of someone that the company has insured for i years.

Then input the initial distribution $\mathbf{p}^{(0)}$ of insureds. For $i = 0$, this is the proportion of the population not currently insured with the company. For $i > 0$, this is the proportion of the population insured with the company for i consectutive policy terms. The last column is the proportion insured with the company for nine or more consecutive terms.

	State i									
	0	1	2	3	4	5	6	7	8	9
Retention Probabilities	0.012	0.87	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
Initial Distribution $\mathbf{p}^{(0)}$	0.90	0.01	0.009	0.008	0.007	0.007	0.006	0.005	0.005	0.043

Following is the resultant matrix **A** of one-step transition probabilities:

0.988	0.012	0	0	0	0	0	0	0	0
0.130	0	0.87	0	0	0	0	0	0	0
0.080	0	0	0.92	0	0	0	0	0	0
0.080	0	0	0	0.92	0	0	0	0	0
0.080	0	0	0	0	0.92	0	0	0	0
0.080	0	0	0	0	0	0.92	0	0	0
0.080	0	0	0	0	0	0	0.92	0	0
0.080	0	0	0	0	0	0	0	0.92	0
0.080	0	0	0	0	0	0	0	0	0.92
0.080	0	0	0	0	0	0	0	0	0.92

DISTRIBUTION AT TIME n

This section shows how to calculate the distribution at time n, given the initial distribution $\mathbf{p}^{(0)}$ and the matrix **A**.

Note that if **p** is the distribution of states at any time, then $\mathbf{A}^t$ (the transpose of **A**) times **p** is the distribution in the next time period. That is, the probability that the system is in state m in the next time period is the mth column of matrix **A** times the distribution **p**.

Matrix $\mathbf{A}^t$

0.988	0.13	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
0.012	0	0	0	0	0	0	0	0	0
0	0.87	0	0	0	0	0	0	0	0
0	0	0.92	0	0	0	0	0	0	0
0	0	0	0.92	0	0	0	0	0	0
0	0	0	0	0.92	0	0	0	0	0
0	0	0	0	0	0.92	0	0	0	0
0	0	0	0	0	0	0.92	0	0	0
0	0	0	0	0	0	0	0.92	0	0
0	0	0	0	0	0	0	0	0.92	0.92

Probability Distributions $\mathbf{p}^{(n)}$ at time period n, for $n = 1, 2, 3\ldots$

$\mathbf{p}^{(0)}$	State	1	2	3	4	5	6	7	8	9
0.900	0	0.8977	0.8957	0.8938	0.8921	0.8906	0.8892	0.8879	0.8867	0.8857
0.010	1	0.0108	0.0108	0.0107	0.0107	0.0107	0.0107	0.0107	0.0107	0.0106
0.009	2	0.0087	0.0094	0.0094	0.0094	0.0093	0.0093	0.0093	0.0093	0.0093
0.008	3	0.0083	0.0080	0.0086	0.0086	0.0086	0.0086	0.0086	0.0086	0.0085
0.007	4	0.0074	0.0076	0.0074	0.0080	0.0079	0.0079	0.0079	0.0079	0.0079
0.007	5	0.0064	0.0068	0.0070	0.0068	0.0073	0.0073	0.0073	0.0073	0.0073
0.006	6	0.0064	0.0059	0.0062	0.0064	0.0062	0.0067	0.0067	0.0067	0.0067
0.005	7	0.0055	0.0059	0.0055	0.0057	0.0059	0.0057	0.0062	0.0062	0.0062
0.005	8	0.0046	0.0051	0.0055	0.0050	0.0053	0.0055	0.0053	0.0057	0.0057
0.043	9	0.0442	0.0449	0.0459	0.0473	0.0481	0.0491	0.0502	0.0510	0.0522

Note that the n-step transition probability is given by raising matrix $\mathbf{A}$ to the nth power. The distribution at time n is given by $(\mathbf{A}^t)^n$ times $\mathbf{p}^{(0)}$.

Shown below are the transposes of some n-step transition matrices:

Two-step									
0.9777	0.1980	0.1526	0.1526	0.1526	0.1526	0.1526	0.1526	0.1526	0.1526
0.0119	0.0016	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
0.0104	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.8004	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.8464	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.8464	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.8464	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.8464	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.8464	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.8464	0.8464	0.8464

Four-step									
0.9598	0.3161	0.2786	0.2786	0.2786	0.2786	0.2786	0.2786	0.2786	0.2786
0.0116	0.0031	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026	0.0026
0.0102	0.0021	0.0016	0.0016	0.0016	0.0016	0.0016	0.0016	0.0016	0.0016
0.0095	0.0012	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008	0.0008
0.0088	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.6775	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.7164	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.7164	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.7164	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	0.0000	0.7164	0.7164	0.7164	0.7164	0.7164

STEADY-STATE PROBABILITIES

This section shows the steady-state probabilities found using the solver.

The steady-state probability $\mathbf{p}^*$ is characterized by $\mathbf{A}^t \times \mathbf{p}^* = \mathbf{p}^*$, or $(\mathbf{A}^t - \mathbf{I}) \times \mathbf{p}^* = 0$, where $\mathbf{I}$ is the identity matrix.

Use the solver to find the solution. Let $\mathbf{C} = \mathbf{A}^t - \mathbf{I}$. The steady-state probability $\mathbf{p}^*$ is the solution of the linear system $\mathbf{Cx} = 0$ for which the elements of $\mathbf{x}$ sum to 1.0. After using the solver, the vector $\mathbf{x}$ contains the steady-state probabilities, and the vector $\mathbf{z} = \mathbf{Cx}$ contains all zeros.

Matrix $\mathbf{C} = \mathbf{A}^t - \mathbf{I}$										$\mathbf{x}$	$\mathbf{z} = \mathbf{Cx}$
−0.012	0.13	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.87527	0
0.012	−1.00	0	0	0	0	0	0	0	0	0.01050	0
0	0.87	−1.00	0	0	0	0	0	0	0	0.00914	0
0	0	0.92	−1.00	0	0	0	0	0	0	0.00841	0
0	0	0	0.92	−1.00	0	0	0	0	0	0.00773	0
0	0	0	0	0.92	−1.00	0	0	0	0	0.00712	0
0	0	0	0	0	0.92	−1.00	0	0	0	0.00655	0
0	0	0	0	0	0	0.92	−1.00	0	0	0.00602	0
0	0	0	0	0	0	0	0.92	−1.00	0	0.00554	0
0	0	0	0	0	0	0	0	0.92	−0.08	0.06372	0
										1.00000	

In this application of the solver, the "target cell" for the solver is undefined, since there is no objective function to maximize or minimize.

COMBINED RATIO DIFFERENTIAL

This section illustrates how changing the retention assumptions affects profitability. Generally combined ratios are better for customers who have been retained longer, due to lower expenses and/or better loss ratios. By comparing the average combined ratios before (100.0%) and after improving retention (99.2%), one can measure the financial effects of changing policy retention.

State i	Steady State Probability **x**	Assumed Combined Ratio Differential	Assumed Base Combined Ratio	Combined Ratio
0	0.875274	N/A		
1	0.010503	20.00%		115.54%
2	0.009138	10.00%		105.54%
3	0.008407	8.00%		103.54%
4	0.007734	6.00%		101.54%
5	0.007116	4.00%		99.54%
6	0.006546	2.00%		97.54%
7	0.006023	0.00%		95.54%
8	0.005541	0.00%		95.54%
9	0.063719	0.00%	95.54%	95.54%
Market share		12.47%		
Average combined ratio differential		3.66%		
Average combined ratio		99.20%		

APPENDIX C

PROOF THAT **A** IS IRREDUCIBLE

In this appendix we prove that an invariant distribution exists for the Markov chain formulation of the retention problem. Recall that the transition matrix for this problem is given by:

$$\mathbf{A} = \begin{pmatrix} 1-r_0 & r_0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1-r_1 & 0 & r_1 & 0 & 0 & \cdots & 0 & 0 \\ 1-r_2 & 0 & 0 & r_2 & 0 & \cdots & 0 & 0 \\ 1-r_3 & 0 & 0 & 0 & r_3 & \cdots & 0 & 0 \\ 1-r_4 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1-r_{N-1} & 0 & 0 & 0 & 0 & \cdots & 0 & r_{N-1} \\ 1-r_N & 0 & 0 & 0 & 0 & \cdots & 0 & r_N \end{pmatrix},$$

where all the r_k are strictly between 0 and 1.

To prove the result we need to define the terms "aperiodic" and "irreducible."

State j is defined to be "periodic" if there exists an integer $t > 1$ such that $a_{jj}^{(n)} = 0$ unless n is an integer multiple of t. Here $a_{jj}^{(n)}$ is the n-step probability of returning to state j. The matrix $\mathbf{A}$ is aperiodic if no states are periodic.

We show that this system is aperiodic. Consider any state j. For any $k > 0$, with $0 < k \leq N - j$ the system can return to state j in $j + k + 1$ steps through the sequence

$$j \to j+1 \to j+2 \to \cdots \to j+k \to 0 \to 1 \to \cdots \to j \quad \text{for} \quad k \leq N - j. \tag{C.1}$$

For $k > N - j$, the system can return to state j in $j + k + 1$ steps through the same sequence except that it "parks" at state N for

$k-(N-j)$ steps before going to state 0. For example, if $j = 1$, $N = 4$, and $k = 6$, then the system returns to state j in $k + j + 1$ ($= 8$) steps through the sequence of states:

$$1 \to 2 \to 3 \to 4 \to 4 \to 4 \to 4 \to 0 \to 1.$$

This last complication only comes about because we set N as the highest state. If we had allowed an infinite number of states then Equation C.1 holds for all $k > 0$.

Thus we have shown that a system in state j can return to state j in n steps for all $n = \{j+2, j+3, j+4, \ldots\}$.[8] This means **A** has no period; i.e., **A** is aperiodic.

A chain is defined to be "irreducible" if and only if every state can be reached from every other state. This means that, given any two states j and k, there exists an integer n such that the system can move from j to k in n steps.

The chain **A** is clearly irreducible since the system can move from state j to state k through the sequence of states:

$$j \to 0 \to 1 \to \cdots \to k$$

We have now established that **A** is aperiodic and irreducible.

We now show directly that an invariant distribution **p** exists for **A** by calculating it.

The defining equations for invariance are

$$p_k = r_{k-1} p_{k-1} \qquad \text{for} \quad k = 1, 2, 3 \cdots N-1, \tag{C.2}$$

$$p_N = r_{N-1} p_{N-1} + r_N p_N, \qquad \text{and} \tag{C.3}$$

$$p_0 = (1-r_0)p_0 + (1-r_1)p_1 + \cdots + (1-r_N)p_N. \tag{C.4}$$

From Equation C.2 we get

$$p_k = r_0 r_1 r_2 \cdots r_{k-1} p_0 \qquad \text{for} \quad k = 1, 2, 3 \cdots N-1. \tag{C.5}$$

---

[8] This is true for $n = j + 1$ also, but this is not needed for the proof.

From Equation C.5 we can see that the terms 0 through $N-1$ on the right-hand side of Equation C.4 add to $p_0 - r_{N-1}p_{N-1}$. From Equation C.3 the last term on the right-hand side of Equation C.4 equals $r_{N-1}p_{N-1}$. Thus, we can choose an arbitrary value for p_0, define the remaining p_k by Equation C.2 and C.3, and Equation C.4 will be automatically satisfied. Once all the p_k are calculated, just rescale them so they add to 1 and these values of $\mathbf{p}$ are the invariant probabilities.

The following theorem[9] will now enable us to say that the n-step distributions converge to the invariant distribution, regardless of the initial distribution.

Suppose a chain is irreducible and aperiodic and that there exist probabilities $\{p_k,\ k = 0,1,2,\ldots\}$ with all $p_k \geq 0$ that satisfy the invariant distribution conditions:

$$\mathbf{p} = \mathbf{A}^t\mathbf{p}.$$

Then

$$a_{jk}^{(n)} \to p_k \qquad \text{as} \quad n \to \infty$$

independently of the initial state j, and the chain is ergodic.

We have already shown that $\mathbf{A}$ satisfies all the conditions of the theorem. (Note the term "ergodic" means that the mean recurrence time to revisit any state j is finite). What the conclusion means is that the n-step transition matrix $\mathbf{A}^n$ ultimately approaches the matrix for which every column is the invariant distribution.

[9]Feller, [2, p. 393]. Actually, the theorem in Feller is more powerful in that it provides a converse which states that if the limits exist, then they form the invariant distribution.