Transient Behavior of DTMCs

Natarajan Gautam
Department of Industrial and Systems Engineering
Texas A&M University
235A Zachry, College Station, TX 77843-3131
Email: gautam@tamu.edu
Phone: 979-845-5458
Fax: 979-847-9005

August 17, 2009

Abstract

The objective of this article is to describe the transient behavior of DTMCs. We first provide some notation and terminology necessary for transient analysis. Then we provide a procedure for analyzing a DTMC and predicting the state after a finite number of jumps. Finally we provide algorithms to efficiently perform transient analysis for DTMCs and obtain performance measures.

Consider a discrete-time Markov chain (DTMC) which is a special type of stochastic process where state space is discrete and the state of the process is observed at discrete time epochs. The observations possess the Markov property, i.e., if the current state is given, to predict the future states, we do not need any information about the past. In other words, if X_n is the state of a process at the n^{th} observation, then Markov property states that

$$P\{X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \ldots, X_0 = i_0\} = P\{X_{n+1} = j | X_n = i\}.$$

For the purpose of transient analysis, we also assume that the DTMC is time-homogeneous, i.e. the probability of transitioning from a state to another state does not vary with time. In other words,

$$P\{X_{n+1} = j | X_n = i\} = P\{X_1 = j | X_0 = i\}.$$

This article assumes that the reader is familiar with modeling a process as a DTMC. If that is not the case, the reader is encouraged to review the previous article from this Handbook or look through one of several texts on stochastic processes (viz. [1] and [2]). With that in mind the objective of this article is to obtain expressions for the distribution of the states of a time-homogeneous DTMC at some particular time in the near future. Specifically, we will compute, say, $P\{X_m = j\}$, $P\{X_r = k | X_n = i\}$, etc. for $0 \le m < \infty$ and $n \le r < \infty$. For that we will first describe some terminology and notation used throughout the article in Section 1. Then in Section 2 we present some of the key results of transient analysis. Finally in Section 3 we will provide some techniques to efficiently perform the transient analysis from a computational standpoint. For a more thorough treatment of computational aspects of DTMCs, please see the article "Computational Methods for DTMCs."

1 Terminology and Notation

We consider a time-homogeneous DTMC X_0, X_1, X_2, $\ldots$ which we denote as $\{X_n, n \geq 0\}$, where X_n is the state of a process at the n^{th} observation. The state X_n could be multi-dimensional and when we say $X_n = i$, then i is also assumed to be multi-dimensional. Let S be the state space of the DTMC, i.e. S is the set of all possible values that X_n can take for all n. The state space S is assumed to be a countable set, although there could be infinite elements.

Next we define p_{ij} which is the one-step transition probability of going from state i to j. Therefore, for a time-homogeneous DTMC, for $i, j \in S$ and $n \geq 0$,

$$p_{ij} = P\{X_{n+1} = j | X_n = i\}.$$

Notice that p_{ij} is not a function of n because the DTMC is time-homogeneous. Using the transition probabilities p_{ij} for every $i \in S$ and $j \in S$, we can build a square matrix $P = [p_{ij}]$. The matrix P is called *one-step transition probability matrix*. The rows of the matrix correspond to the given current state, while the columns correspond to the next state. The elements of each row sum to 1 for every $i \in S$, that is

$$\sum_{j \in S} p_{ij} = 1.$$

An important notation for transient analysis is $p_{ij}^{(m)}$ which is the m-step transition probability of going from state i to j. In other words, if the current state of a process is i, then the probability that after exactly m observations, the process will be in state j is $p_{ij}^{(m)}$. By definition,

$$p_{ij}^{(m)} = P\{X_{m+n} = j | X_n = i\}.$$

Note that $p_{ij}^{(1)} = p_{ij}$, the one-step transition probability. In matrix form we denote $P(m) = [p_{ij}^{(m)}]$, the m-step transition probability matrix with $P(1) = P$. It is crucial to point out that unlike p_{ij} which is part of the modeling, $p_{ij}^{(m)}$ can actually be computed and it forms the core of transient analysis. But before deriving an expression for $p_{ij}^{(m)}$, we present one last notation.

Let $a_j^{(m)}$ be the probability that the DTMC is in state j at the m^{th} observation, i.e.

$$a_j^{(m)} = P\{X_m = j\}.$$

The row vector of $a_j^{(m)}$ values is denoted as $\mathbf{a}^{(m)}$, i.e. $\mathbf{a}^{(m)} = [a_j^{(m)}]$. To obtain $a_j^{(m)}$ we not only require P (which is described as part of the modeling) but also the initial distribution row-vector a which is made up of $a_j = P\{X_0 = j\}$ such that $a = [a_j]$. Although a is not described in DTMC modeling, for the purpose of analysis (especially $\mathbf{a}^{(m)}$), we require knowing a. Note that $a = \mathbf{a}^{(0)}$. Having described the notation, the next section develops a technique to obtain $a_j^{(m)}$ and $p_{ij}^{(m)}$ for all $m < \infty$, $i \in S$ and $j \in S$.

2 Key Results

The results in this section for $a_j^{(m)}$ and $p_{ij}^{(m)}$ are due to the Chapman-Kolmogorov equation which we derive first. We start with the definition of $p_{ij}^{(m)}$ for $m > 0$ and some $0 < r \leq m$, and work our way:

$$
\begin{aligned}
p_{ij}^{(m)} &= P\{X_{m+n} = j | X_n = i\}, \\
&= \sum_{k \in S} P\{X_{m+n} = j, X_{r+n} = k | X_n = i\}, \\
&= \sum_{k \in S} P\{X_{m+n} = j | X_{r+n} = k, X_n = i\} P\{X_{r+n} = k | X_n = i\}, && (1) \\
&= \sum_{k \in S} P\{X_{m+n} = j | X_{r+n} = k\} P\{X_{r+n} = k | X_n = i\}, && (2) \\
&= \sum_{k \in S} P\{X_{m-r+n} = j | X_n = k\} P\{X_{r+n} = k | X_n = i\}, && (3) \\
&= \sum_{k \in S} p_{kj}^{(m-r)} p_{ik}^{(r)}, \\
&= \sum_{k \in S} p_{ik}^{(r)} p_{kj}^{(m-r)}. && (4)
\end{aligned}
$$

It may be worthwhile to explain the above set of equations. Equation (1) is just a standard conditional probability argument that follows a law of total probability statement. Then, the first term in Equation (2) is due to the Markov property and Equation (3) follows from the fact that the DTMC is time-homogeneous. Finally, Equation (4) is by definition and it has implications in terms of the m-step transition probability matrix. Note that Equation (4) results in

$$P(m) = P(m - r)P(r)$$

since the ij^{th} element of $P(m)$ is the left hand side of Equation (4) and is equal to the i^{th} row of $P(m - r)$ multiplied by the j^{th} column of $P(r)$.

Using the result $P(m) = P(m - r)P(r)$ for any $m > 0$ and $r \leq m$, as well as noticing that $P(0) = I$ and $P(1) = P$, we have $P(2) = P^2$ (by writing $m = 2$ and $r = 1$). Thereby we can recursively show that

$$P(m) = P^m$$

and therefore

$$p_{ij}^{(m)} = [P^m]_{ij}.$$

Thus, starting in state i, the probability that the DTMC is in state j after m steps can be computed by raising P to the m^{th} power and taking the ij^{th} term. Next, to compute $\mathbf{a}^{(m)}$ by conditioning on the initial state, we get

$$\mathbf{a}^{(m)} = aP(m) = aP^m.$$

To get $P\{X_m = j\}$ just take the term corresponding to j, i.e. $[aP^m]_j$.

We illustrate the transient analysis by means of an example. Consider three long-distance telephone companies A, B and C. Everytime a sale is announced, users switch from one company to another. Let X_n denote the long-distance company with which a particular user named Jill is just before the n^{th} sale announcement. We have $S = \{A, B, C\}$. Based on the customers' switching in the past it is estimated that Jill's switching patterns follow the following transition probability matrix:

$$P = \begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \left[\begin{array}{ccc} 0.3 & 0.4 & 0.3 \\ 0.5 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{array} \right] \end{array}.$$

For example, if Jill is with B before a sale is announced, she would switch to A with probability 0.5 and C with probability 0.2 or stay with B with probability 0.3 as evident from the second row in P.

If at time 0, Jill is with long-distance company B, then after the third sale (i.e. $m = 3$) we wish to compute the probability that Jill is with company C. We have,

$$P^3 = \begin{bmatrix} 0.3860 & 0.2770 & 0.3370 \\ 0.3930 & 0.2760 & 0.3310 \\ 0.3870 & 0.2590 & 0.3540 \end{bmatrix}.$$

What we need is

$$P\{X_3 = C | X_0 = B\} = p_{BC}^{(3)} = [P^3]_{BC} = 0.3310$$

based on the above results.

Now if we know that $a = [0.25 \quad 0.5 \quad 0.25]$, i.e. initially Jill picks A, B or C with probability 0.25, 0.5 or 0.25 respectively. Say we are interested in computing the probability that Jill is with A after 2 sale announcements. Based on the above results we need

$$P\{X_2 = A\} = a_A^{(2)} = [aP^2]_A = 0.385.$$

Further, using the initial distribution $a = [0.25 \quad 0.5 \quad 0.25]$, if we are interested in the probability that Jill is with A after the first sale, with B after the 2nd sale and with C after the third, then we can obtain that as $P\{X_1 = A, X_2 = B, X_3 = C\} = P\{X_3 = C | X_2 = B, X_1 = A\} P\{X_2 = B | X_1 = A\} P\{X_1 = A\} = p_{BC} p_{AB} [aP]_A = 0.2 \times 0.4 \times 0.425 = 0.034$ with the first expression due to Markov property and the rest from the above results.

3 Computing Issues

Based on the previous section, clearly the computationally intensive operation is in obtaining P^m especially when P is large. For most problems, mathematical software such as MATLAB, mathematica and MAPLE have efficient routines to compute P^m, however if one were to use a more fundamental tool such as c-programming or VBA it would be a good idea to incorporate a routine that would efficiently obtain P^m. Having said that, in this section we demonstrate two other techniques to obtain P^m.

3.1 Diagonalization Method

Consider a DTMC $\{X_n, n \geq 0\}$ with transition probability matrix P which is $b \times b$ square matrix, i.e. the number of states in the state space S of the DTMC is b. Here we require that b is finite. Assume that all the eigenvalues of P are distinct. Then we can write down P as XDX^{-1} where D is a diagonal matrix of all b eigenvalues of P and X is a collection of b row vectors $b \times 1$ in size that correspond to the appropriate right eigenvectors of P. We first describe a procedure to compute D and X. Let x be a right eigen vector ($b \times 1$ column vector) and d be the corresponding eigenvalue (a scalar). Then by definition of eigenvalue and eigenvector we have $Px = dx$, i.e. $(P - dI)x = 0$. Now if we solve for the determinant $|P - dI| = 0$ we get a b-degree polynomial and upon solving we get b solutions d_1, d_2, $\ldots$, d_b which form the b eigenvalues of P. We can obtain the corresponding eigenvectors x_1, x_2, $\ldots$, x_b by carefully solving for x_i in $Px_i = d_i x_i$ for $i = 1, 2, \ldots, b$. Note that x_i is a $b \times 1$ column vector but it is not unique, so just select one arbitrarily. Once d_i and x_i are known for all $i = 1, 2, \ldots, b$, we can write down

$$X = [x_1 \;\; x_2 \;\; \ldots \;\; x_b]$$

and

$$D = \begin{bmatrix} d_1 & 0 & 0 & \ldots & 0 \\ 0 & d_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & d_b \end{bmatrix}.$$

Therefore we have

$$P = XDX^{-1}$$

and hence

$$P^m = XD^mX^{-1}.$$

Notice that

$$D^m = \begin{bmatrix} d_1^m & 0 & 0 & \ldots & 0 \\ 0 & d_2^m & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & d_b^m \end{bmatrix}.$$

The only intensive computation is in obtaining X, X^{-1} and D. It may be worthwhile commenting on X^{-1} as there are two ways of obtaining it. One of course is to invert the matrix X. The other is to realize that X^{-1} is the collection of left-eigenvalues of P, stacked top to bottom corresponding to the appropriate eigenvalue.

3.2 Generating Function Method

Consider a DTMC $\{X_n, n \geq 0\}$ with transition probability matrix P which is $b \times b$ square matrix, i.e. the number of states in the state space S of the DTMC is b. We can denote the generating function of P as $\Psi(z)$ for some complex scalar z. By definition

$$\Psi(z) = I + Pz + P^2z^2 + P^3z^3 + \ldots$$

and hence can be written as

$$\Psi(z) = I + Pz\Psi(z).$$

Solving for $\Psi(z)$ we get

$$\Psi(z) = (I - Pz)^{-1}$$

provided $|z| < 1$ and b is finite (although the analysis can be extended to the infinite case). From the right hand side of the above expression, one can typically write down $[\Psi(z)]_{ij}$ of the form $\frac{f_{ij}(z)}{g_{ij}(z)}$ where both $f_{ij}(z)$ and $g_{ij}(z)$ are polynomial functions. Now, if we write down $\frac{f_{ij}(z)}{g_{ij}(z)}$ in polynomial form, it would be equivalent to $\sum_{m=0}^{\infty} p_{ij}^{(m)} z^m$ and therefore by taking the exponent multiplying z^m in $\frac{f_{ij}(z)}{g_{ij}(z)}$, we get $p_{ij}^{(m)}$.

References

[1] V.G. Kulkarni. *Modeling and Analysis of Stochastic Systems.* Texts in Statistical Science Series. Chapman and Hall, Ltd., London, 1995.

[2] S.M. Ross. *Introdution to Probability Models.* Academic Press, San Diego, CA, 2003.