Moments of Transition-Additive Random Variables
Defined on Finite, Regenerative Random Processes

RICHARD SHIFFRIN AND MAYNARD THOMPSON

Indiana University

Random processes consisting of a sequence of jumps from one to another of a finite set of states are considered. Such processes are regenerative if the progress after the nth jump depends only upon the state entered at the jump. Examples include discrete and continuous Markov processes. A method is given to restrict attention to a subset of sample paths according to criteria based on the transitions allowed for a single jump. The primary concern is with transition-additive random variables: these sum for any valid sample path the values of independent random variables assigned to each jump in the path. A simple formula for finding all moments of such random variables is derived. Necessary and sufficient conditions for the existence of the solutions are demonstrated, and illustrations of the computational simplicity of the approach are provided.

I. INTRODUCTION AND OVERVIEW

In this paper we study a class of random variables defined on random, regenerative, finite-state processes. These include Markov and semi-Markov processes as special cases. A realization of the process is a path of jumps from one to another of the states of the process. The random variable under investigation sums values that are assigned to the jumps in a path. Our primary result is a theorem giving an elementary procedure for calculating all moments of such random variables. Examples and related questions have been investigated in many instances, both pure and applied (e.g., Barbour & Schassberger, 1981; Chung, 1967; Gibbon, 1971; Ginsberg, 1971; Glynn & Inglehart, 1986; Howard, 1971; Kao,

This research was supported, in part, by a grant to the first author, NIMH 12717. Reprint requests should be sent to Richard M. Shiffrin, Dept. of Psychology, Indiana University, Bloomington, IN 47405. The authors thank the reviewers for their careful and helpful comments.

1 The results presented here were reported at the Mathematical Psychology Meetings at Santa Barbara, California, in August, 1981, and at the IMACS Congress on Scientific Computation in Oslo, Norway, in August, 1985. A version of this paper was submitted to the Journal of Mathematical Psychology in January, 1983. Although that version was not accepted because it was too concise and lacked examples, the mathematical content was the same as that of the present version. We view the mathematical contribution as significant for both the generality of the setting and the ease of implementing the computational algorithm. The paper contains a number of published results as special cases, including some that have appeared since our early submission. Our results not only are quite general but also contain ideas that are likely to be of special interest and usefulness to theorists in psychology.
In psychology, Markov models have been used quite extensively (a few examples include Atkinson & Crothers, 1964; Bower & Theios, 1964; Brainerd, Desrochers, & Howe, 1981; Rapoport, Stein, & Burkheimer, 1979, Chap. 3, pp. 52–54). Most applications have involved discrete chains in which the time variable, if it is treated at all, is a function of the number of transitions. When the transitions between states are governed by distributions of time, the processes are usually termed "semi-Markov." Such processes have not generally been used as models in psychology, although a few examples exist (e.g., Fisher & Goldstein, 1983; Townsend, 1976; Meyer, Yantis, Osman, & Smith, 1975).

We show in this article that calculation methods for semi-Markov processes are little different from those for Markov processes and both involve elementary operations (solutions of simultaneous linear equations). The methods presented are widely applicable and may make semi-Markov processes readily available to theorists in psychology and other fields. Even within discrete Markov chains, the methods provide a straightforward approach to the calculation of moments that might otherwise seem inordinately complex.

Although the theorems in question involve just two equations (actually two systems of simultaneous equations: see Eqs. (6) and (8)), we think an understanding of the approach and its applicability is best achieved by an orderly progression from simpler processes and uses to the more complex cases. This introduction will therefore be devoted to a somewhat simplified and informal presentation of the basic results, along with illustrative examples. The precise formulation and technical details are given in the following sections: In Section II are the definitions and notation; in Section III are the theorems and proofs. Section IV contains an algorithm useful for speeding calculations and contains some additional examples. Section V contains concluding remarks.

A. Discrete Markov Chains

Although calculation methods for discrete Markov chains have been well studied, our exposition is clarified by beginning with this case. In addition, the approach we suggest offers increased power and simplicity even in the discrete case.

Suppose there are \( N \) states. The process consists of a series of jumps (transitions) from one to another of these states. The sequence of jumps made is termed a path. The probability that the next jump occurs to state \( j \) given that the process is currently in state \( i \) is denoted by \( p_{ij} \) (usually given as a set of entries in an \( N \times N \) matrix). This probability is independent of the trial number and of the sequence of jumps prior to the current state.

When at least one of the states (or a nonempty subset of the \( N \) states) can be entered but not exited (due to certain of the \( p_{ij} \) being zero), then the chain is said to have "absorbing" states. Our methods apply to chains whether or not they have absorbing states, but the methods apply only to sets of paths that have a definite
ending jump. To specify these, we define an indicator variable, $e_{ij}$, which is 1.0 if a jump from $i$ to $j$ ends a path, and 0.0 otherwise. We also might wish to consider only a subset of the remaining possible jumps. We therefore define an indicator variable $c_{ij}$ which is 1.0 if a jump from $i$ to $j$ continues a path and is 0.0 otherwise (obviously $e_{ij}$ and $c_{ij}$ cannot both be 1.0). We are interested in all paths that consist of a sequence of continuing jumps followed by an ending jump. These are called valid paths. It should be noted that the $c_{ij}$ and $e_{ij}$ can be chosen at the discretion of the researcher, subject only to the constraint that $c_{ij}$ and $e_{ij}$ cannot both be 1.0 (examples will be given shortly).

The ideas are illustrated in Fig. 1. A three state Markov chain is under consideration. The transition probabilities are given in the upper left matrix (labeled $(p_{ij})$) in the figure. The possible paths starting in state 2 that are three jumps long are depicted in the figure as branches in the tree. Each jump is labeled with the probability of that jump. At the termination of each path is given the probability of that three jump path. The matrices labeled $(c_{ij})$ and $(e_{ij})$ at the top of the figure give the continuing and ending jumps that define valid paths. Paths that are valid, or potentially valid, are connected by solid lines; invalid jumps are indicated by dashed lines, and all subsequent jumps in those paths are dashed as well. In addition, jumps in paths following any ending jump are dashed.

The solid horizontal lines indicate paths that have ended (validly at that point. The probability of each such valid path is given in the trapezoidal box. (The matrix labeled $(v_{ij})$ and the values given in parentheses will be discussed shortly).
The first task is to provide a method to calculate for each starting state \( i \), the probability that the process will result in a valid path. This probability is termed \( g_i \). These are given by solving the following set of simultaneous linear equations (for \( i \) such that \( g_i \neq 0 \)):

\[
g_i = \sum_{j=1}^{N} e_{ij} p_{ij} + \sum_{j=1}^{N} c_{ij} p_{ij} g_j.
\]

This equation is similar to the well-known equation used to calculate the probabilities of absorption in particular states in an absorbing Markov chain. In words, it says that the probability of a valid path starting in \( i \) is a sum over the next steps that end a valid path at once (the first sum) and the next steps that continue a valid path (the second sum); if a continuing step occurs (say into state \( j \)), the probability of moving to that state must of course be multiplied by the probability that a valid path will result afterwards (\( g_j \)).

Since, for certain starting states, \( g_i \) may be zero (e.g., when state \( i \) cannot be left, and the jump from \( i \) to itself is not an ending jump: \( p_{ii} = 1.0; e_{ii} = 0 \)), the equations we give apply only to states \( i \) for which \( g_i \neq 0 \). (Otherwise an indeterminacy results.) It can be concluded that \( g_i \neq 0 \) if there exists a path with greater than zero probability connecting state \( i \) to at least one state \( k \) from which an ending transition can be made. These notions will be formalized in Section III, but it is always easy in practice to determine which \( g_i \) equal zero.

Although the \( g_i \) are of some interest in their own right, our primary goal is the determination of moments of certain random variables defined on valid sample paths. The idea is simple; each possible jump in a path is assigned a fixed value, termed \( v_{ij} \) when the jump is from state \( i \) to state \( j \). We define a random variable \( Z_i \) that is simply the sum of the values assigned to each jump in a valid path starting in state \( i \), conditional upon the path being valid. It is the moments of \( Z_i \) that are desired. This idea also is illustrated in Fig. 1. The matrix labeled \((v_{ij})\) at the top of the figure gives the values assigned to each jump. These values are given in parentheses next to each jump in the figure, and the accumulated values for each path are in parentheses at each path termination. Some values for \( Z_2 \) are those in the trapezoidal boxes at the end of valid paths.

The expected value of \( Z_i \) can be obtained from the following set of simultaneous equations (for \( i \) such that \( g_i \neq 0 \)) (let \( \mu_i = E[Z_i \mid \text{valid path}] \)):

\[
\mu_i = \frac{1}{g_i} \left\{ \sum_{j=1}^{N} e_{ij} p_{ij} v_{ij} + \sum_{j=1}^{N} c_{ij} p_{ij} g_j (v_{ij} + \mu_j) \right\}.
\]

The idea here is related to that underlying Eq. (1). In words, the expected value starting in state \( i \) is a sum over valid next steps, multiplying the probability of that step by the sum of the value of that step and any expected future value to come after that step. Since all future paths might not be valid, the expected value of a continuing path must be multiplied by the probability of future validity: \( g_j \). Also, to gain a proper expectation, since the sum across all paths may have probability less
than one, the sum must be divided by $g_i$. As is the case for Eq. (1), Eq. (2) applies only for those states $i$ for which $g_i \neq 0$.

It is useful at this point to look at applications and examples. Consider how the $c_{ij}$ and $e_{ij}$ may be used. If paths terminating with the first entry in state $k$ are of interest, the column $k$ entries in the matrix $(e_{ij})$ can be set to 1.0, with others set to 0.0 ($e_{ik} = 1.0$, $i = 1, 2, ..., N$). If only paths ending with a jump from state $j$ to state $k$ are desired, then only the $e_{jk}$ entry is set to 1.0. If paths that do not go through state $k$ are desired (see Chung, 1967, p. 45) then column $k$ of the continuing matrix is set to 0.0 ($c_{jk} = 0$, $j = 1, ..., N$). These examples suffice to illustrate the general idea.

The value matrix can be used for many purposes. If all $v_{ij}$ are set to 1.0, then $Z$ is a count of all steps taken. Setting only certain $v_{ij}$ to 1.0 sums only those steps. For example, setting $v_{ij} = 1.0$ for $j = 1, ..., N$ counts each step away from state $i$. Of course if visits to certain states are worth more than others, or if certain jumps are worth more than others, then the $v_{ij}$ may be chosen to reflect those values (as illustrated in the examples).\footnote{It is probably good form to set $e_{ij}$ and $c_{ij}$ to zero when $p_{ij} = 0$, and to set $v_{ij}$ to zero when both $e_{ij} = 0$ and $c_{ij} = 0$. However, since these terms multiply each other in the equations, it is unnecessary to do so. It may even prove convenient when exploring different sets of restrictions and values on a given chain to leave certain of these cells filled with nonzero values. The only necessary constraint is that $0 \leq c_{ij} + e_{ij} \leq 1.$}

**Example 1.** Our first example is based on the Markov chain defined by the $p_{ij}$ matrix in Fig. 1. To calculate the probability of a valid path, the use of Eq. (1) gives

\[
g_1 = 0.6g_2,
\]

\[
g_2 = 0.5g_1 + 0.5g_3
\]

\[
g_3 = 0.1 + 0.6 + 0.3g_2.
\]

Solving, $g_1 = \frac{21}{55}$; $g_2 = \frac{7}{11}$; $g_3 = \frac{49}{55}$.

To calculate the expected number of steps in a valid path, replace the value matrix in Fig. 1 with a matrix filled with ones. Then Eq. (2) leads to

\[
\mu_1 = \frac{1}{g_1} \{0.6g_2(1 + \mu_2)\}
\]

\[
\mu_2 = \frac{1}{g_2} \{0.5g_1(1 + \mu_1) + 0.5g_3(1 + \mu_3)\}
\]

\[
\mu_3 = \frac{1}{g_3} \{0.1 + 0.6 + 0.3g_2(1 + \mu_2)\}.
\]

Substituting the values of $g_i$ and then solving give $\mu_1 = \frac{51}{11}$; $\mu_2 = \frac{40}{11}$; $\mu_3 = \frac{137}{77}$.
For the value matrix of Fig. 1, Eq. (2) gives

\[ \mu_1 = \frac{1}{g_1} \{0.6g_2(4 + \mu_2)\} \]

\[ \mu_2 = \frac{1}{g_2} \{0.5g_1(5 + \mu_1) + 0.5g_3(9 + \mu_3)\} \]

\[ \mu_3 = \frac{1}{g_3} \{0.1(2) + 0.6(6) + 0.3g_2(7 + \mu_2)\} \].

Substituting for the \( g_i \) and solving, \( \mu_1 \approx 27.70; \mu_2 \approx 23.70; \mu_3 \approx 10.84 \).

Consider next the calculation of higher moments of \( Z_i \). The idea is the same as that in Eq. (2), except that in this case the appropriate formula is

\[ E[Z^p_i] = \frac{1}{g_i} \left\{ \sum_{j=1}^{N} e_{ij} p_{ij} v^p_{ij} + \sum_{j=1}^{N} c_{ij} p_{ij} g_j E[(v_{ij} + Z_j)^p] \right\}. \]  

(3)

Letting \( E[Z^p_i] = \mu^p_i \), Eq. (3) can be rewritten

\[ \mu^p_i = \frac{1}{g_i} \left\{ \sum_{j=1}^{N} e_{ij} p_{ij} v^p_{ij} + \sum_{j=1}^{N} c_{ij} p_{ij} g_j \sum_{m=0}^{p} \binom{p}{m} v^m_{ij} \mu^{p-m}_j \right\}. \]  

(4)

Once again, these equations apply only to those states \( i \) for which \( g_i \neq 0 \). To solve for the \( p \)th moment, one solves the simultaneous equations for the first moments, substitutes the values obtained into the simultaneous equations for the second moments, and so forth. In each case, a set of \( N \) (or fewer) simultaneous linear equations needs to be solved.\(^3\)

**Example 2.** To illustrate the solution method, we use an example that keeps the calculations fairly simple: a five state random walk with two absorbing barriers (see Kemeny & Snell, 1976, p. 27).

The probabilities \( (p_{ij}) \) are given by a transition matrix,

\[ (p_{ij}) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \]

\(^3\) Actually, things are even easier than this. The solution of a set of simultaneous equations is equivalent to the inversion of a certain matrix. It turns out that the same inverted matrix can be used to solve for all moments and all value matrices, as shown in Section IV.
We take the natural definitions of continuing and ending transitions:

\[
(e_o) = \begin{bmatrix}
0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}, \quad (e_o) = \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}.
\]

Let us calculate the first three moments of the number of times the process is in state 2 prior to absorption. To do this, define the value matrix

\[
(v_o) = \begin{bmatrix}
0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}.
\]

It is clear that \(g_i = 0\) for \(i = 1, 5\) and \(g_i = 1\) for \(i = 2, 3, 4\) (although this could be derived from Eq. (1)). Equation (3) then gives

\[
E[Z_2^p] = \frac{1}{2}(\frac{1}{4}) (1)^p + \frac{1}{2} (\frac{3}{4}) (1) E[(1 + Z_3)^p]
\]

\[
E[Z_3^p] = \frac{1}{2}(\frac{1}{4}) (1) E[(0 + Z_2)^p] + \frac{1}{2} (\frac{3}{4}) (1) E[(0 + Z_4)^p]
\]

\[
E[Z_4^p] = \frac{1}{2}(\frac{1}{4}) (1) E[(0 + Z_3)^p].
\]

Solving these equations for \(p = 1, 2,\) and 3 sequentially, we find that the first moments of \(Z_2, Z_3,\) and \(Z_4\) are respectively \(7/5,\) \(3/5,\) and \(1/5.\) Setting \(p = 2,\) expanding where necessary, and using the first moments, we find the second moments to be \(63/25,\) \(27/25,\) and \(9/25,\) respectively. Continuing, we obtain \(E[Z_2^2] = 763/125,\) \(E[Z_3^2] = 327/125,\) and \(E[Z_4^2] = 109/125.\) Extensions to higher moments are obvious.

B. Semi-Markov Processes

The methods outlined work equally well when the transitions between states have a temporal distribution. Let the time a state \(i\) is entered be set to zero. Then let the probability that a transition to state \(j\) occurs in \([0, t]\) be \(p_{ij}(t)\). Continuing transitions at time \(t\) are defined by \(c_{ij}(t) = 1,\) and ending transitions at time \(t\) defined by \(e_{ij}(t) = 1\) (and both cannot be 1 at the same value of \(t\)). Otherwise \(e_{ij}(t)\) and \(c_{ij}(t) = 0.\) Finally, the value of a jump occurring at time \(t\) is given as \(v_{ij}(t)\).

In order to use the theorems, certain terms must first be calculated. Let \(q_{cij}\) be the probability of making a continuing jump to \(j\) given a start in \(i;\) let \(q_{eij}\) be the probability of making an ending jump to \(j\) given a start in \(i.\) Then

\[
q_{cij} = \int_0^\infty c_{ij}(t) \, dp_{ij}(t), \quad q_{eij} = \int_0^\infty e_{ij}(t) \, dp_{ij}(t),
\]

(5)
assuming that the integrals exist. (In (5), \( dp_{ij}(t) \) refers to Lebesgue–Stieltjes integration. In many applications \( dp_{ij}(t) \) may be replaced by \( p_{ij}(t) dt \), that is, by the density function: see Example 3 and Fig. 2.) It is then possible to solve for the \( g_i \) as in Eq. (1),

\[
g_i = \sum_{k=1}^{N} q_{eik} + \sum_{k=1}^{N} q_{cik} g_k, \tag{6}
\]

for states \( i \) such that \( g_i \neq 0 \). Next, let \( \tau^p_{cij} \) be the \( p \)th moment of the value of a continuing jump from \( i \) to \( j \), and \( \tau^p_{eij} \) be the \( p \)th moment of the value of an ending jump from \( i \) to \( j \), when \( q_{cij} \neq 0 \) and \( q_{eij} \neq 0 \):

\[
\tau^p_{cij} = \frac{1}{q_{cij}} \int_{0}^{\infty} [v_{ij}(t)]^p c_{ij} dp_{ij}(t) \tag{7A}
\]

\[
\tau^p_{eij} = \frac{1}{q_{eij}} \int_{0}^{\infty} [v_{ij}(t)]^p e_{ij}(t) dp_{ij}(t). \tag{7B}
\]

Fig. 2. An example of the use of the equations for semi-Markov processes. (See the discussion of the Land of Oz example in the text.) The chain, the valid paths, and the value functions are defined by the \( L_j(t), cV(z), e_j(t), \) and \( u_j(t) \) matrices. The various state to state jump probabilities that must be calculated in order to use the equations in the text are given in the lower portion of the figure, as are the weighted values assigned to those jumps.
Moments can then be calculated using an equation essentially identical to Eq. (4):

\[ E[Z^p_i] = \mu^p_i = \frac{1}{g_i} \left\{ \sum_{k=1}^{N} q_{eik} \tau^p_{eik} + \sum_{k=1}^{N} q_{eik} g_k \sum_{m=0}^{P} \binom{p}{m} \tau^m_{eik} \mu^p_{k-m} \right\}. \]  

As usual, these equations hold for states \( i \) for which \( g_i \neq 0 \).

Note that the basic formulas (Eqs. (1) and (4)) remain essentially unchanged in the setting of semi-Markov processes. The computations will be more complex only because certain integrals (Eqs. (5) and (7)) must be evaluated before the basic equations (Eqs. (6) and (8)) can be used. The integrals involved in Eqs. (5) and (7) can be evaluated numerically.

The roles played by the matrices \((c_{ij}(t))\), \((e_{ij}(t))\), and \((v_{ij}(t))\) remain the same as those in the discrete case, although the entries in the matrices are now functions of time. Most often, \( c_{ij}(t) \) and \( e_{ij}(t) \) will be set equal to 0.0, or 1.0, over certain intervals of time \((t_0, t_1)\). If, for instance, \( c_{ij}(t) = 0 \) within such an interval, then no valid path can have a jump from state \( i \) to state \( j \) within that interval. Likewise, if \( e_{ij}(t) = 1 \) in an interval, then any valid path jumping from state \( i \) to state \( j \) within that interval ends with that jump. Note in particular that it is possible for a transition from \( i \) to \( j \) to be continuing for some values of \( t \) and ending for other values.

The uses of the \( v_{ij}(t) \) matrix are also straightforward. Perhaps the most common usage in the semi-Markov domain will involve setting \( v_{ij}(t) = t \) for all \( i, j, \) and \( t \). Then the random variable gives the cumulative time of travel in a valid path. Obviously many other options are available. For example, the time spent only in certain states can be counted by setting \( v_{ij}(t) = t \) only for those states \( i \) in question. Setting \( v_{ij}(t) = 1 \) will count number of transitions, as in the discrete case; setting \( v_{ij}(t) = 1 \) only in certain time intervals will count number of transitions made in those time intervals. Thus if for all \( i, j, v_{ij}(t) = 0 \) for \( t \leq t_0 \), and \( v_{ij}(t) = 1 \) for \( t > t_0 \), then the random variable counts number of transitions in a valid path that take longer than \( t_0 \). Finally, note that the \( v_{ij}(t) \) assignments need not be restricted to 1 and 0, or to \( t \); any values called for by the setting of the problem may be used. For example, if a system has the property that an output occurs from state \( k \) once state \( k \) is resided in for at least \( t_0 \) sec, and the output is proportional to the square root of the additional residence time beyond \( t_0 \), then the output value could be obtained by setting \( v_{kj}(t) = 0 \) for \( t \leq t_0 \) and \( v_{kj}(t) = a(t - t_0)^{1/2} \) for \( t > t_0 \), for some constant \( a \), and all states \( j \).

**Example 3.** We use a version of the Land of Oz example of Kemeny and Snell (1976, p. 29) to illustrate the application of the method to continuous processes. Suppose that the weather in Oz changes randomly from moment to moment, among snow (S), nice (N), and rain (R), according to the matrix \((f_{ij}(t))\) of densities given in Fig. 2. In words, this matrix says that the probability density of a transition (somewhere) from a state is decreasing exponentially so that the cumulative probability of a transition by time \( t \) is \( P_i - P_i \exp\{-a,t\} \), where \( 1 - P_i \) is clearly the
probability that no transition will ever be made. If a transition is made, $b_i$ gives the conditional probability that it will be made to the first of the remaining two states (ordered R, N, S). For this example, we set $P_R = 0.9$, $a_R = 1$, $b_R = 0.4$, $P_N = 0.8$, $a_N = 0.5$, $b_N = 0.2$, $P_S = 0.7$, $a_S = 0.4$, $b_S = 0.7$.

Now suppose we wish to calculate the moments of the variable which sums up time as follows: it counts the total time spent in going from rainy to snowy weather, and from snowy to nice weather. It also counts the time going from rainy to nice weather, unless this time for any one such transition happens to be less than 1 day, in which case it counts a full day, or happens to be between 3 and 4 days, in which case it counts nothing, or happens to be greater than 4 days, in which case it counts 3 days.

We will consider only continuing weather transitions in which rainy to nice transitions take less than 4 days, rainy to snowy transitions take more than 2 days, nice to rainy transitions and snowy to nice transitions of any sort occur, and snowy to rainy transitions take less than 2 days. Finally, we will stop counting whenever a nice to snowy transition occurs, or when a rainy to nice transition takes more than 4 days, or when a rainy to snowy transition takes less than 1 day or when a snowy to rainy transition takes more than 2 days.

These assumptions are embodied in the $(v_0(t))$, $(c_0(t))$, and $(e_0(t))$ matrices of Fig. 2. To apply Eqs. (6) and (8), we must find the various $q$'s and $\tau$'s for single jumps, and the results of these calculations are shown in the lower portion of Fig. 2. From Eq. (6) we have

\[
\begin{align*}
q_R &= q_{eR} + q_{eR} g_N + q_{eR} g_S, \\
g_S &= q_{eS} + q_{eS} g_R + q_{eS} g_N. 
\end{align*}
\]

and consequently (using the values for the $q$'s) $g_S = 0.552$, $g_N = 0.744$, $g_R = 0.651$. Next we use Eq. (8) to obtain

\[
\begin{align*}
\mu_R^k &= \frac{1}{g_R} \left\{ q_{eR} \tau_{eR}^p + q_{eS} \tau_{eS}^p + q_{eR} g_N \sum_{m=0}^{p} \binom{p}{m} \tau_{eR}^m \mu_N^{p-m} \\
&+ q_{eS} g_N \sum_{m=0}^{p} \binom{p}{m} \tau_{eS}^m \mu_R^{p-m} \right\} \\
\mu_S^k &= \frac{1}{g_S} \left\{ q_{eS} \mu_R^p g_R + q_{eSN} g_N \sum_{m=0}^{p} \binom{p}{m} \tau_{eSN}^m \mu_N^{p-m} \\
&+ q_{eSR} g_R \mu_R^p \right\} \\
\mu_N^k &= \frac{1}{g_N} \left\{ q_{eNR} g_R \mu_R^p \right\}.
\end{align*}
\]

Solving for the first moments only, one finds $\mu_S = 1.08$, $\mu_N = 0.146$, and $\mu_R = 1.04$. The higher moments may be determined iteratively.

In this section we have given an informal exposition of the results in a slightly restricted setting and have given examples. The discussion in the following section introduces a somewhat more general setting in which "time" is replaced by a variable in a more general parameter space, perhaps multidimensional. It is in this more general setting that we make the concepts and methods precise and that we prove the theorems (in Section III).
II. Definitions and Notation

We introduce a setting which is general enough to emphasize the wide applicability of the results. Consider a random process, taking values in a finite state space \( S = \{1, 2, ..., N\} \), which jumps from state to state with probabilities depending on the states and on a parameter in a parameter space \( \mathcal{Y} \). More specifically, suppose that we have an \( N \times N \) matrix of real-valued set functions defined\(^4\) on a \( \sigma \)-algebra \( \mathcal{U} \) in \( \mathcal{F} \) with the properties

\[
p_{ij}(A) \geq 0 \quad \text{for each} \quad A \in \mathcal{U}, \quad i, j = 1, 2, ..., N
\]

\[
\sum_j p_{ij}(\mathcal{F}) \leq 1, \quad i = 1, 2, ..., N
\]

and

\[
p_{ij} \left( \bigcup_n A_n \right) = \sum_n p_{ij}(A_n) \quad \text{for each countable collection of nonoverlapping sets} \ A_n \in \mathcal{U}, \ i, j = 1, 2, ..., N.
\]

For \( i, j \in S \) and a set \( \theta \in \mathcal{U} \), a set of jumps (or transitions) from state \( i \) to \( j \) with parameter in \( \theta \) will be denoted \((i, \theta, j)\) and we introduce a probability measure \( P \), conditioned on the starting state, by

\[
P[(i, \theta, j) | i] = p_{ij}(\theta).
\]

In words, conditioned upon the starting state being \( i \), \( p_{ij}(\theta) \) gives the joint probability that a jump is made to \( j \) and that the parameter will take a value in \( \theta \). For each \( t \in \theta \) we write \((i, t, j)\) as that element of \((i, \theta, j)\) with parameter \( t \). In the jump \((i, t, j)\) we refer to \( i \) as the beginning state and \( j \) as the ending state. A sample path is a sequence of jumps with the beginning state of the \((n+1)\)st jump coinciding with the ending state of the \(n\)th jump. A sample path \( \gamma \) can be denoted by

\[
\gamma = (k_0, t_1, k_1, t_2, k_2, ...),
\]

where \( k_0, k_m \in S, t_m \in \mathcal{F} \) for \( m = 1, 2, ... \). The set of sample paths beginning in state \( k_0 \) and passing (in order) through the states \( k_1, k_2, ... \), and with \( t_m \) in a set \( \theta_m \in \mathcal{U} \) for \( m = 1, 2, ... \), will be denoted by \( \Gamma \). Equivalently \( \Gamma = \{(k_0, \theta_1, k_1, \theta_2, k_2, ...)\} \). Note that the case \( k_{m-1} = k_m \) is not excluded. The transition \((k_{m-1}, t_n, k_m)\) will be referred to as the \(n\)th jump and \( t_n \) as the parameter for this jump. The probability of the set of sample paths \( \Gamma \) is defined to be

\[
P(\Gamma) = \prod_{m \geq 1} p_{k_{m-1}k_m}(\theta_m).
\]

\(^4\) A \( \sigma \)-algebra is a collection of subsets which is closed under countable union and complementation. Such a class of subsets provides the appropriate setting for the study of probabilistic problems of the type considered here (and almost all others).
This construction guarantees the regenerative character of our process: there is a probability $p_{ij}(\theta)$ of entering $j$ from $i$ with parameter value in the set $\theta$ which is independent of the initial part of the sample path. In particular, this probability is independent of when in the sequence of states $i$ is entered.

In the special case when $\mathcal{F} = R^+ = \{x \in R : x \geq 0\}$ and the sets $\theta$ are intervals $[0, t]$ we use the notation $p_{ij}(t)$ instead of $p_{ij}(\theta)$. In this case we define $p_{ij}(t)$ to be the conditional probability that if there is a jump into state $i$ at $t = 0$, then the next jump is into state $j$ and it takes place in the interval $(0, t]$.

Note that we allow for the possibility that for some states $i$,

$$\sum_{j=1}^{N} p_{ij}(\mathcal{F}) < 1.$$

When $\mathcal{F}$ is time, we may interpret such a defective case by saying that the process may reach some state and with positive probability never make another jump. This is illustrated in one of the versions of Example 4 in Section IV.

Initial segments of sample paths will be denoted by

$$\phi = (k_0, t_1, k_1, t_2, k_2, \ldots, k_{s-1}, t_s, k_s).$$

A segment $\phi$ of the form (*) will be said to be of length $s$.

Let $C$ (continuing) and $E$ (ending) be disjoint sets of jumps with indicator functions

$$c_{ij}(t) = \begin{cases} 1 & \text{if } (i, t, j) \in C, \\ 0 & \text{otherwise} \end{cases}, \quad e_{ij}(t) = \begin{cases} 1 & \text{if } (i, t, j) \in E, \\ 0 & \text{otherwise} \end{cases},$$

$i, j = 1, 2, \ldots, N$, $t \in \mathcal{F}$. Define $\Delta_i$, $i = 1, 2, \ldots, N$, to be the set of all finite initial segments of sample paths of the form (*) with $s \geq 1$ and

$$k_0 = i,$$

$$(k_{j-1}, t_j, k_j) \in C, \quad j = 1, 2, \ldots, s - 1,$$

$$(k_{s-1}, t_s, k_s) \in E.$$

(These concepts are related to the taboo states of Chung [1967, Sect. 1.9]).

To facilitate our discussion it will be convenient to use $\phi_y$ to represent a jump $(i, t, j) \in C$, and $\phi_j$ to represent a segment $(j, t_1, k_1, \ldots, t_s, k_s) \in \Delta_j$. With this notation we define $\phi_y + \phi_j$ to be the segment $\phi \in \Delta_i$:

$$\phi = (i, t, j, t_1, k_1, \ldots, t_s, k_s).$$

Thus the symbol $+$ in $\phi_y + \phi_j$ indicates the concatenation of $\phi_y$ and $\phi_j$. We write $(k_1, t, k_2) \in \phi$ to mean that there is a jump $(k_1, t, k_2)$ somewhere in the segment $\phi$. With this notation the following assertion is obvious.
Every $\phi \in A_i$ is exactly one of:

(a) a single jump, in which case $\phi \in E$, or

(b) a sequence of two or more jumps, in which case it can be written as $\phi_{ij} + \phi_j$ with $\phi_{ij} \in C$ and $\phi_j \in A_j$.

Let $(v_{ij}(t))$ be an $N \times N$ matrix of finite real valued functions defined on $\mathcal{T}$. Define a random variable $Z_i$ on $A_i$ by setting

$$Z_i(\phi) = \sum_{j=1}^{s} v_{k_{i-1}k_j}(t_j)$$

for $\phi \in A_i$ of length $s$. Since $Z_i$ is a sum of values associated with one step jumps or transitions, we refer to such random variables as transition-additive.

In what follows it will be useful to have notation for specifying the value assigned to a single jump, continuing or ending. To this end

for $\phi = (i, t, k) \in C$ set $T_{cik}(\phi) = v_{ik}(t)$

and

for $\phi = (i, t, k) \in E$ set $T_{eik}(\phi) = v_{ik}(t)$.

Finally, we need notation for the probabilities of certain sets of valid sample paths or segments of such paths. For a given state $i$, let $g_i$ be the probability of the set of valid sample paths starting in state $i$ where the universal set is all sample paths starting in state $i$. Similarly, for $i, j \in S$ set

$$q_{cij} = \int_{\mathcal{T}} c_{ij}(t) \, dp_{ij}(t), \quad q_{ej} = \int_{\mathcal{T}} e_{ij}(t) \, dp_{ij}(t).$$

That is, $q_{cij}$ is the probability of making a continuing jump to state $j$ given a start in state $i$, and $q_{ej}$ is the probability of making an ending jump to state $j$ given a start in state $i$.

Our interest is in the moments of the random variable $Z_i$, and we next introduce notation for these moments. For $i, j \in S$ and $p = 1, 2, ..., \mu_i^p,$ we set

$$E[Z_i^p \mid \phi \in A_i] = \mu_i^p,$$

$$E[T_{cij}^p (i, t, j) \in C] = \tau_{cij}^p, \quad E[T_{eij}^p (i, t, j) \in E] = \tau_{eij}^p.$$
when \( q_{ei} \neq 0 \) and \( q_{ej} \neq 0 \).

The states \( i \) for which \( g_i > 0 \) can be determined by forming an \( (N + 1) \times (N + 1) \) matrix \( H \) with

\[
h_{ij} = \begin{cases} 
1 & \text{if } q_{eik} > 0, \ i = 1, \ldots, N; \ j = 1, \ldots, N \\
1 & \text{if } \sum_{k=1}^{N} q_{eik} > 0, \ i = 1, \ldots, N; \ j = N + 1 \\
1 & \text{if } i = N + 1; \ j = N + 1 \\
0 & \text{otherwise}
\end{cases}
\]

and computing \( H^N \). Any state \( i \) for which the \( (i, N + 1) \) element of \( H^N \) is positive has \( g_i > 0 \) (see Kemeny & Snell, 1976, Chap. 2).

In what follows we suppose that the probabilities \( g_i \), \( q_{eij} \), and \( q_{ej} \) exist, an assumption which imposes mild restrictions on the applicability of the results. Since the sets \( C \) and \( E \) are determined by specifying the class of admissible sample paths, the existence of \( q_{eij} \) and \( q_{ej} \) depends upon the relations between the constraints of admissibility (the \( c_{ij} \) and \( e_{ij} \)) and the measure \( p_{ij} \).

### III. Determination of the Moments

Our primary goal (Theorem 1) is to derive a formula for the \( p \)th moments of transition-additive random variables. This result is supplemented by an existence theorem (Theorem 2) which shows that if the one step moments exist, then moments for valid paths of arbitrary length do also, and by a theorem (Theorem 3) which shows that the formula of Theorem 1 can be used to determine the \( p \)th moments recursively. The notation used in the statements of the theorems was introduced in Section 2, and readers interested primarily in applications will find the statements of Theorems 1, 3, and 4 self-contained. We emphasize that all probabilities in this section are understood to be conditioned on the path starting in a particular state, which we refer to as \( i \).

**Theorem 1.** For those \( i \) with \( g_i > 0 \) and any \( p \) for which the moments exist,

\[
\mu_i^p = \frac{1}{g_i} \left\{ \sum_{k=1}^{N} q_{eik} \tau_{ek}^p \right\} + \sum_{k=1}^{N} q_{eik} g_k \left( \sum_{m=0}^{p} \binom{p}{m} \tau_{eik}^m \mu_i^{p-m} \right).
\]
Remarks. (i) It is suggestive to write the formula of Theorem 1 in the form

\[ E[Z_i^p] = \frac{1}{g_i} \left\{ \sum_{k=1}^{N} q_{eik} E[T_{eik}^p] + \sum_{k=1}^{N} q_{cik} g_k E[(T_{cik} + Z_k)^p] \right\}. \]

This formulation emphasizes the basic idea of viewing each sample path as a first step followed by a (possibly missing) subsequent.

(ii) The special case of this result for first entrance times conditioned on the avoidance of specified states appears in Chung (1967, p. 62).

(iii) The probabilities \( g_i \) can be obtained by using a well-known relation involving absorption probabilities in discrete Markov chains. The details are provided in Theorem 4.

Proof of Theorem 1. Partition the set \( A_j \) into \( N + 1 \) subsets: one consisting of \( A_j \cap E \), and \( N \) other subsets of \( A_j \), each consisting of sample segments of length greater than 1 whose first transition is to a specific state. Each sample path in one of these \( N \) latter subsets can be written as \( \phi = \phi_{ik} + \phi_k, \phi_{ik} \in C, \phi_k \in A_k \). It may be that one or more of these subsets is empty.

On \( A_j \), the random variable \( Z_i \) is a probabilistic mixture of \( N + 1 \) random variables, namely \( Z_i \) restricted to each subset of \( A_j \), each weighted according to the probability of the appropriate subset. It follows that \( Z_i^p \) can be written as a sum of the weighted expectations of the restrictions of \( Z_i \), where a weight is the probability of the appropriate subset:

\[
\mu_i^p = \sum_{k=1}^{N} P[\phi_{ik} \in E | A_j] E[T_{eik}^p | (i, t, k) \in E] + \sum_{k=1}^{N} P[\phi = \phi_{ik} + \phi_k | A_j] E[(T_{cik} + Z_k)^p | \phi_{ik} + \phi_k \in A_i] \]

\[
= \frac{1}{g_i} \left\{ \sum_{k=1}^{N} q_{eik} E[T_{eik}^p | \phi_{ik} \in E] + \sum_{k=1}^{N} q_{cik} g_k E[(T_{cik} + Z_k)^p | \phi_{ik} \in C, \phi_k \in A_k] \right\}. \]

The formula of Theorem 1 is obtained by expanding the last term on the right hand side in the expression for \( \mu_i^p \) above and noting that \( T_{eik} \) and \( Z_k \) are independent.

The hypothesis of Theorem 1 that the moments \( \mu_i^p \) exist can be replaced by the (apparently simpler) assumption that the one step moments exist. This is the content of the next result.

**Theorem 2.** If the probabilities \( g_i, q_{eik}, q_{cik}, k \in S \), and the moments \( \tau_{eik}^h, \tau_{cik}^h, k \in S, \text{ and } h = 1, 2, \ldots, p \) exist, then the moments of \( Z_i \) of order up to and including \( p \) also exist.
We begin by showing that there is a discrete absorbing Markov chain for which the first passage time distribution is closely related to our process.

**Lemma 1.** There is a discrete absorbing Markov chain with one absorbing state and \( n \) (\( \leq N \)) transient states with the following property: if \( S_i \) is a random variable defined on the discrete chain which assigns to each sample path starting in state \( i \) the number of transitions to absorption, and if \( A_i^s \) is the subset of \( A_i \) which contains sample paths with exactly \( s \) jumps (\( s - 1 \) continuing and 1 ending), then

\[
P[A_i^s | A_i] = \frac{1}{g_i} P[A_i^s] = P[S_i = s]
\]

for \( i \in S \) with \( g_i > 0 \) and \( s = 1, 2, \ldots \).

**Proof.** Introduce as transient states all states \( r \in S \) for which \( g_r > 0 \). Without loss of generality we suppose these states to be numbered 1, 2, \ldots, \( n \). Form an \( (n + 1) \times (n + 1) \) transition matrix for a discrete Markov chain as

\[
p_{r,j} = \begin{cases} 
\frac{1}{g_r} (q_{rj} g_j), & j = 1, 2, \ldots, n; r = 1, 2, \ldots, n \\
\frac{1}{g_r} \left( \sum_{k=1}^N q_{rk} \right), & j = n + 1; r = 1, 2, \ldots, n 
\end{cases}
\]

\[
p_{n+1,j} = \begin{cases} 
0, & j = 1, 2, \ldots, n \\
1, & j = n + 1.
\end{cases}
\]

Since the row sums are 1, this clearly defines a legitimate transition matrix. Also, for each \( r \), since \( g_r > 0 \), there is a positive probability of reaching state \( n + 1 \) in \( n \) steps. Thus we have an absorbing Markov chain. Note that in essence the absorbing state has been obtained by collecting all ending jumps together.

Next, fix \( k_1, k_2, \ldots, k_{s-1} \) and consider the set of all sample segments which begin in state \( i \), are of length \( s \) (consist of \( s \) jumps), and pass successively through \( k_1, k_2, \ldots, k_{s-1} \) and end in state \( k_s \). Denote this set by \( A_i^s(k_1, k_2, \ldots, k_s) \). Letting \( i = k_0 \), it follows that

\[
P \left[ \bigcup_{k_s=1}^N A_i^s(k_1, \ldots, k_s) \bigg| A_i \right] = \frac{1}{g_i} P \left[ \bigcup_{k_s=1}^N A_i^s(k_1, \ldots, k_s) \right]
\]

\[
= \frac{1}{g_i} \prod_{j=1}^{s-1} q_{ck_{j-1}k_j} \sum_{m=1}^N q_{ck_s,k_{s+1}},
\]

(9)
and that

\[
\left( \prod_{j=1}^{s-1} p_{k_{j-1}, k_j} \right) p_{k_{s-1}, n+1} = \left( \prod_{j=1}^{s-1} q_{k_{j-1}, k_j} g_{k_j} \right) \sum_{m=1}^{N} q_{k_{s-1}, m} g_{k_{s-1}}
\]

\[
= \frac{1}{g_{k_s}} \prod_{j=1}^{s-1} q_{k_{j-1}, k_j} \sum_{m=1}^{N} q_{k_{s-1}, m}.
\]

Here (9) applies to the original process and (10) to the discrete analogue, and since the right hand sides are equal, the left hand sides must be equal also. Thus the probabilities of each subset of sample paths for the processes are equal. Also, the possible sets of intermediate states are identical, and the conclusion of the lemma follows.

**Lemma 2.** For \( i = 1, 2, \ldots, n \) and each \( p, E[S_i^p] < \infty \).

**Proof.** For states \( i, j \) in the discrete chain, let \( p_{ij}(m) \) denote the probability of going from \( i \) to \( j \) in \( m \) steps. There are (Kemeny & Snell, 1976, p. 43) numbers \( b > 0 \) and \( 0 < c < 1 \) such that \( p_{ij}(m) \leq bc^m \) for any transient states \( i \) and \( j \). Consequently, the probability that the process begins in state \( i \) and reaches the absorbing state in exactly \( s \) steps satisfies

\[
\sum_{i=1}^{n} p_{ij}(s-1) p_{j, n+1} \leq nb \sum_{s=1}^{\infty} s^p c^{s-1},
\]

Therefore

\[
E[S_i^p] = \sum_{s=1}^{\infty} s^p P[S_i = s] \leq nb \sum_{s=1}^{\infty} s^p c^{s-1},
\]

which is clearly finite.

**Proof of Theorem 2.** With \( A_i^s \) as defined in Lemma 1, decompose \( A_i \) into disjoint subsets \( A_i^s \) so that \( A_i = \bigcup_{j \geq 1} A_i^s \). Using this decomposition we have

\[
P[Z_i \leq u | A_i] = \frac{1}{g_i} P[Z_i \leq u]
\]

\[
= \frac{1}{g_i} \sum_{j \geq 1} P[\phi: Z_i(\phi) \leq u \text{ and } \phi \in A_i^s]
\]

\[
= \frac{1}{g_i} \sum_{j \geq 1} P[\phi: Z_i(\phi) \leq u | \phi \in A_i^s] P[A_i^s].
\]

Next,

\[
P[\phi: Z_i(\phi) \leq u | \phi \in A_i^s]
\]

\[
= \sum_{k_1, \ldots, k_s \text{ all paths}} P[\phi: Z_i(\phi) \leq u | \phi \in A_i^s(k_1, \ldots, k_s)] P[\phi \in A_i(k_1, \ldots, k_s) | \phi \in A_i^s].
\]
Also

\[ \int u^p \, dP[\phi: Z_i(\phi) \leq u | \phi \in A_i^p(k_1, ..., k_s)] \]

\[ = E[Z_i^p | A_i^p(k_1, ..., k_s)] \]

\[ - E[(T_{ci_1} + T_{ci_1k_2} + \cdots + T_{ek_s,...,k_s})^p | A_i^p(k_1, ..., k_s)] \]

\[ = \sum_{r_1, r_2, ..., r_s} \left( \frac{p}{r_1 + r_2 + \cdots + r_s = p} \right) E[T_{ci_1}^{r_1}] E[T_{ci_1k_2}^{r_2}] \cdots E[T_{ek_s,...,k_s}^{r_s}], \quad (13) \]

where there are at most \( p \) moments in the product on the right with nonzero order. Thus, if \( M^p \geq 1 \) is a bound for \( E[T^k] \) for all \( T \)'s and \( k = 1, \ldots, p \), then the right hand side of Eq. (13) is no larger than

\[ M^p \sum_{r_1, r_2, ..., r_s} \left( \frac{p}{r_1 + r_2 + \cdots + r_s = p} \right) = M^p s^p. \quad (14) \]

Using Eq. (12), Eq. (13), and the estimate Eq. (14),

\[ \int u^p \, dP[\phi: Z_i(\phi) \leq u | \phi \in A_i^p] \]

\[ = \sum_{k_1, ..., k_s} \frac{1}{P[A_i^p]} \left\{ \prod_{j=1}^{s-1} q_{ek_{j-1}k_j} \right\} q_{ek_{s-1}k_s} \]

\[ \times \int u^p \, dP[\phi: Z_i(\phi) \leq u | \phi \in A_i^p(k_1, ..., k_s)] \]

\[ \leq M^p s^p. \]

The proof is now completed by using this estimate and Eq. (11),

\[ \int u^p \, dP[Z_i \leq u | A_i] \]

\[ = \frac{1}{g_{i,s}} \sum_{i,s} \int u^p \, d\{P[Z_i \leq u | A_i^p] P[A_i^p] \} \leq \frac{1}{g_{i,s}} \sum_{i,s} P[A_i^p] M^p s^p. \]

which is finite by Lemmas 1 and 2.\(^5\) \( \square \)

\(^5\) The idea behind the proof of Lemma 2 and Theorem 2 can be used to obtain the following result for random sums. Let \( S \) be a finite set of random variables, each of which has finite moments of orders up to and including \( p \), and let \( N \) be an integer valued random variable with finite moments of all orders. Define a random variable \( T = T_1 + T_2 + \cdots + T_N \) where each \( T_i \in S \). Then \( T \) has finite moments of order up to and including \( p \).
We turn next to the use of the formula of Theorem 1. In order to determine the moments one first calculates the probabilities \( g_i \) (see Theorem 4 below), then \( \mu_i^2 \), \( i \in S \), then \( \mu_i^3 \), \( i \in S \), and so forth. For each \( p \), the determination of \( \mu_i^p \), \( i \in S \), requires the solution of no more than \( N \) simultaneous linear equations. The fact that these equations are always solvable is the content of the next result.

**Theorem 3.** If the probabilities \( g_i, q_{cik}, q_{eik}, i, k \in S \), and the moments \( \tau_{cik}^h, \tau_{eik}^h, i, k \in S, h = 1, 2, \ldots, p \), exist, then the equations of Theorem 1 can be solved recursively to yield \( \mu_i^h, i \in S, h = 1, 2, \ldots, p \).

**Proof.** Suppose that the states are labeled as in Lemma 1; i.e., states 1, 2, \ldots, \( n \) (\( n \leq N \)) have \( g_i > 0 \). For \( p \geq 1 \) suppose that the moments \( \mu_i^h, i \in S, h = 1, 2, \ldots, p - 1 \), have been determined and that the one step moments of order \( p \) exist. From Theorem 1 we have

\[
\mu_i^p - \sum_{k=1}^{n} \frac{1}{g_i} (q_{cik} g_k) \mu_k^p
\]

\[
= \frac{1}{g_i} \left\{ \sum_{k=1}^{N} q_{cik} \tau_{eik}^p + \sum_{k=1}^{n} q_{cik} g_k \sum_{m=1}^{p} \binom{p}{m} \tau_{eik}^m \mu_k^{p-m} \right\}, \tag{15}
\]

\( i = 1, 2, \ldots, n \). Denote by \( a_i(p) \) the right hand side of Eq. (15) and by \( a(p) \) the \( n \)-vector whose \( i \)-th coordinate is \( a_i(p) \). Under the hypotheses of the theorem, \( a(p) \) contains only known quantities. Also, let \( m(p) \) be an \( n \)-vector whose \( i \)-th coordinate is \( \mu_i^p \). Then Eq. (15) can be written as

\[
(\mathbb{1} - Q) m(p) = a(p),
\]

where \( \mathbb{1} \) is the \( n \times n \) identity matrix and \( Q \) is the portion of the transition matrix introduced in Lemma 2 which involves transitions between transient states; \( Q = ((1/g_i) q_{cij} g_j) \). Since \( \mathbb{1} - Q \) is invertible (Kemeny & Snell, 1976), Eq. (15) can be solved. If \( (\mathbb{1} - Q)^{-1} - N \) (the fundamental matrix for the absorbing Markov chain of Lemma 1), then we have

\[
m(p) = N a(p). \tag{16}
\]

In order to establish the formula for the probabilities of a valid path given a starting state, we need the following lemma.

**Lemma 3.** There is a discrete absorbing Markov chain with two absorbing states and \( n \leq N \) transient states with the following property: \( P[A_i | i] = p_{e_i n+1} \), where \( p_{i, n+1} \) is the probability of absorption in the first of the two absorbing states of the discrete chain.

**Proof.** Introduce as transient states all states \( r \in S \) for which \( g_r > 0 \). Without loss of generality we suppose these states to be numbered 1, 2, \ldots, \( n \). The transition
probabilities between these states are the $q_{cij}$. The first absorbing state corresponds to having made an ending transition and the second to never making a transition at all. More precisely, we form an $(n + 2) \times (n + 2)$ transition matrix for a discrete chain as

$$
p_{r,j} = \begin{cases} 
q_{cij}, & r = 1, 2, \ldots, n; j = 1, 2, \ldots, n \\
\sum_{k=1}^{n} q_{erk}, & r = 1, 2, \ldots, n; j = n + 1 \\
1 \sum_{k=1}^{n} (q_{erk} + q_{erk}), & r = 1, 2, \ldots, n; j = n + 2
\end{cases}
$$

The row sums are clearly one, so this defines a legitimate transition matrix. Using the terminology of Lemma 1, it is obvious that

$$
P^L(j_{1};(k_{1} \ldots k_{s}) = \sum_{k=1}^{N} q_{ek_{1} \ldots k_{s}} \sum_{m=1}^{N} q_{ek_{1} \ldots k_{s} \ldots m}
$$

which is the probability of a path in $A_i$ of length $s$ through intermediate states $k_1, k_2, \ldots$. The right hand side of this equation is the probability of the corresponding path for the discrete chain. The correspondence holds for all sample paths of all lengths, and $A_i = \cup A_i^j$, so the conclusion of the lemma follows.

Theorem 4 now follows directly from the standard result for absorption probabilities in finite chains (Kemeny & Snell 1976) and provides a means of computing the probabilities $g_i$.

**Theorem 4.**

$$
g_i = \sum_{k=1}^{N} q_{ek_{1} \ldots k_{s} \ldots m} \sum_{k=1}^{N} q_{ek_{1} \ldots k_{s} \ldots m} g_{k}, \quad i \in S, \quad g_i > 0.
$$

**Remarks.** (i) The case of a jump from a state to itself is permitted in order to include the possibility that the process spontaneously “forgets” and the parameter values are “reset” even though there is no change in state. Also, the option of having a jump from a state to itself may be useful to theorists in cases where jumps are observable events. Of course such transitions can be excluded if they present difficulties in the design or interpretation of experiments.

(ii) Without loss of generality, $c_{ij}(t)$ and $e_{ij}(t)$ may take on values between zero and one, whose sum is less than or equal to one. In this case, the values may be given probabilistic interpretations. We illustrate in Example 5 in the next section.
(iii) Without loss of generality, the fixed value \( v_{ij}(t) \) may be replaced by a distribution of values. In such a case, Theorem 1 and Eq. (8) are correct as given, but Eqs. (7A) and (7B) must be modified so as to obtain the correct expectation. Also the versions of these equations contained in Eqs. (3) and (4) would require replacement of the terms \( v_{ij}^p \) with \( E[v_{ij}^p] \).

(iv) The proofs of the theorems in a context in which time has been replaced by a point in an arbitrary parameter space allows for some potentially useful processes. For example, a multistate process could be defined over both time and space; in this case the parameter of the process for each state would be a point in a four dimensional space.

IV. AN ALGORITHM FOR CALCULATING MOMENTS AND EXAMPLES

In Section III we developed a matrix formulation of our results. This formulation provides a convenient way to give an explicit method for determining the moments recursively.

ALGORITHM FOR CALCULATING MOMENTS

1. Determine the matrices \((p_{ij}), (c_{ij}), (e_{ij}),\) and \((v_{ij})\). (In general these matrices will depend on choices made by the theorist.)

2. Compute the probabilities \((q_{cij})\) and \((q_{ej})\) using the definitions

\[
q_{cij} = \int_{\mathcal{F}} c_{ij}(t) \, dp_{ij}(t), \quad q_{ej} = \int_{\mathcal{F}} e_{ij}(t) \, dp_{ij}(t),
\]

and the probabilities \(g_i\) (for those \(i\) for which \(g_i > 0\)) using the equation

\[
g_i = \sum_{j=1}^{N} q_{ej} + \sum_{j=1}^{N} q_{cij} g_j.
\]

3. Compute the one step moments using the equations

\[
\tau_{cij}^p = \frac{1}{q_{cij}} \int_{\mathcal{F}} [v_{ij}(t)]^p \, c_{ij}(t) \, dp_{ij}(t)
\]

\[
\tau_{ej}^p = \frac{1}{q_{ej}} \int_{\mathcal{F}} [v_{ij}(t)]^p \, e_{ij}(t) \, dp_{ij}(t)
\]

5. Following the development of Section III, suppose that the states numbered 1 through \(n\) are those with \(g_i > 0\), and the states numbered \(n+1\) through \(N\) each have \(g_i = 0\) although they may have valid sample paths ending there with positive probability. Define an \(n \times n\) matrix \(Q\) by setting \(Q = [(1/g_i) q_{cij} g_j]\) and
set \( N = (I - Q)^{-1} \). Let \( m(p) \) be an \( n \)-vector whose \( i \)th coordinate is \( \mu^p_i \), and let \( r(p) \) be an \( n \)-vector whose \( i \)th coordinate is

\[
\frac{1}{g_i} \left\{ \sum_{k=1}^{N} q_wk T^p_{kik} \right\}.
\]

Also, let \( \tau^{(p)}_e \) be the \( n \times n \) matrix \( (\tau^p_{e_k}) \), and let \( M^{(p)}_e \) be the \( n \times n \) matrix whose diagonal entries are those of \( m(p) \) and whose off diagonal entries are 0. Set \( M^{(0)}_e = I \).

Finally, for \( n \times n \) matrices \( A \) and \( B \) define \( A \star B \) to be an \( n \)-vector whose \( i \)th coordinate is \( \sum_{j=1}^{n} a_{ij} b_{ij} \). With this notation the formula for determining the moments recursively is

\[
m(p) = N \left[ r(p) + \sum_{m=1}^{P} \begin{pmatrix} p \\ m \end{pmatrix} Q \star \left( [T^{(m)}_e] M^{(p-m)}_e \right) \right]. \tag{17}
\]

Formula (17) holds under the conditions of Theorem 3.

Note that the matrix \( N \) in formula (17) is independent of the value matrix \((v_{ij}(t))\). Consequently, all moments of any transition additive random variable can be determined with the same matrix \( N \) (for given matrices \((c_{ij})\) and \((e_{ij})\)).

This algorithm is easily translated into computer code, and it provides a very efficient method of determining moments. We illustrate the simplicity and efficiency of the method with two additional examples.

**Example 4.** Using the results of this section and Eq. (17), we calculate again the first three moments of the random variable defined in Example 2:

\[
Q = \begin{bmatrix} 0 & \frac{2}{3} & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & \frac{1}{3} & 0 \end{bmatrix}, \quad N = \begin{bmatrix} \frac{3}{5} & \frac{2}{5} & \frac{4}{5} \\ \frac{3}{5} & \frac{2}{5} & \frac{4}{5} \\ \frac{3}{5} & \frac{2}{5} & \frac{4}{5} \end{bmatrix}
\]

\[
r(m) = \begin{bmatrix} \frac{1}{3} \\ 0 \\ 0 \end{bmatrix}, \quad T^{(m)}_e = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad m = 1, 2, 3.
\]

From this it follows that

\[
m(1) = N \cdot \left[ \begin{bmatrix} \frac{1}{3} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{2}{3} \\ 0 \\ 0 \end{bmatrix} \right] = \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{2}{3} \end{bmatrix}
\]

\[
m(2) = N \cdot \left[ \begin{bmatrix} \frac{1}{3} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} \frac{2}{3} \\ 0 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} \frac{2}{3} \\ 0 \\ 0 \end{bmatrix} \right] = \begin{bmatrix} \frac{10}{27} \\ \frac{10}{27} \\ \frac{10}{27} \end{bmatrix}
\]
One of the advantages of this method is that the task of computing moments for several random variables with the same conditions (same \((p_{ij}), (c_{ij}), (e_{ij})\)) is simplified. For instance, we can construct another version of this example by defining a random variable as follows: count a visit to state 2 as of value 1, a visit to state 3 as a value of \(-2\), and a visit to state 4 as of value 5 if the next state is 5 and 0 otherwise. We have

\[
(v_{ij}) = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 \\
-2 & -2 & -2 & -2 & -2 \\
0 & 0 & 0 & 0 & 5 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix},
\]

\[
r(m) = \begin{bmatrix}
\frac{1}{3} \\
0 \\
2(5)^m/3
\end{bmatrix}, \quad m = 1, 2, 3,
\]

\[
\mathbb{P}_c^{(m)} = \begin{bmatrix}
1 & 1 & 1 \\
(-2)^m & (-2)^m & (-2)^m \\
0 & 0 & 0
\end{bmatrix}, \quad m = 1, 2, 3.
\]

Consequently

\[
m(1) = \mathbb{N} \cdot \begin{bmatrix}
\frac{1}{3} \\
0 \\
\frac{10}{3}
\end{bmatrix} + \begin{bmatrix}
\frac{2}{3} \\
-2 \\
0
\end{bmatrix} = \begin{bmatrix}
\frac{5}{3} \\
\frac{11}{3}
\end{bmatrix}
\]

\[
m(2) = \mathbb{N} \cdot \begin{bmatrix}
\frac{1}{3} \\
0 \\
\frac{50}{3}
\end{bmatrix} + \begin{bmatrix}
\frac{2}{3} \\
4 \\
0
\end{bmatrix} + 2 \begin{bmatrix}
\frac{2}{3} \\
-6 \\
0
\end{bmatrix} = \begin{bmatrix}
7 \\
19
\end{bmatrix}
\]

and

\[
m(3) = \mathbb{N} \cdot \begin{bmatrix}
\frac{1}{3} \\
0 \\
\frac{240}{3}
\end{bmatrix} + \begin{bmatrix}
\frac{2}{3} \\
-8 \\
0
\end{bmatrix} + 3 \begin{bmatrix}
\frac{14}{3} \\
-30 \\
0
\end{bmatrix} + 3 \begin{bmatrix}
\frac{2}{3} \\
12 \\
0
\end{bmatrix} = \begin{bmatrix}
\frac{244}{15} \\
\frac{214}{15}
\end{bmatrix}.
\]

As a final version of this example, suppose that the second version is modified so that the only ending transitions are those into state 5. State 1 is still an absorbing
state, but sample paths which terminate in state 1 are not in $A_i$, $i = 2, 3, 4$. Then $g_2$, $g_3$, $g_4$ are determined from the system

\[
\begin{align*}
g_2 &= \frac{2}{3} g_3 \\
g_3 &= \frac{1}{3} g_2 + \frac{2}{3} g_4 \\
g_4 &= \frac{2}{3} + \frac{1}{3} g_3.
\end{align*}
\]

That is, $g_2 = \frac{8}{15}$, $g_3 = \frac{4}{5}$, $g_4 = \frac{14}{15}$. We determine the first moments of the random variable defined in the second version. We have $\tau_n$, $(v_t)$, and $\mathbb{T}^{(m)}$ as in that version and

\[
r(m) = \begin{bmatrix} 0 \\ 0 \\ (5)^m + 1/7 \end{bmatrix}, \quad m = 1, 2, 3.
\]

Consequently

\[
\begin{align*}
m(1) &= \begin{bmatrix} 1.390 \\ 1.086 \\ 3.933 \end{bmatrix}, & m(2) &= \begin{bmatrix} 7.234 \\ 7.680 \\ 20.417 \end{bmatrix}, & m(3) &= \begin{bmatrix} 16.376 \\ -2.734 \\ 88.374 \end{bmatrix}.
\end{align*}
\]

It is worthwhile to reiterate a point made earlier: the $c_i(t)$ and $e_i(t)$ need not be indicator variables. Rather, they may be arbitrary (measurable) functions with values between 0 and 1 subject to $c_i(t) + e_i(t) \leq 1$. In such cases the values may be interpreted as probabilities. For example, if $c_i(t) = 0.5$ for all $t$, then the probability that a jump from state $i$ to state $j$ at time $t$ will be continuing is 0.5. None of our equations are altered by this change. We illustrate with an example.

**Example 5.** Consider a three compartment stochastic system which we model in terms of a semi-Markov process. Suppose the holding times for transition from states 1 and 2 are exponentially distributed with parameters 1 and 0.5, respectively, and suppose transitions from state 3 occur 1 time unit after it is entered. We assume that the matrix of distribution functions is

\[
(p_t) = \begin{bmatrix} 0 & 0.4(1 - e^{-t}) & 0.6(1 - e^{-t}) \\ 0.2(1 - e^{-0.5t}) & 0 & 0.8(1 - e^{-0.5t}) \\ 0.7h(t) & 0.3h(t) & 0 \end{bmatrix},
\]

where

\[
h(t) = \begin{cases} 0 & \text{for } t < 1 \\ 1 & \text{for } t \geq 1 \end{cases}.
\]

Also, suppose that valid sample paths are defined by continuing and ending transitions specified as follows: Let $T$ denote holding time. Continuing transitions are those
— from state 2 to state 1
— from state 1 to state 2 with $T < 4$
— from state 1 to state 3 with $T > 2$

and ending transitions are those
— from state 2 to state 3
— from state 1 to state 2 with $T \geq 4$
— from state 1 to state 3 with $T < 1$.

Finally, transitions from state 3 to states 1 and 2 are continuing with probability 0.5 and ending with probability 0.5. (Transitions from state 1 to state 3 with $1 \leq T \leq 2$ are not valid.)

The random variable we consider is defined as follows:
— to transitions from states 1 and 2 to state 3 assign the holding time $T$,
— to transitions from state 1 to state 2 assign the integral part of $T$ for $T < 4$ and assign the number 4 for $T \geq 4$,
— to transitions from state 3 to state 1 assign the number 1,
— to transitions from state 3 to state 2 assign the number 0.

These assumptions can be summarized as

$$c_d(t) = \begin{bmatrix}
0 & 1, T < 4 & 0, T \leq 2 \\
0 & 0, T \geq 4 & 1, T > 2 \\
0.5 & 0.5 & 0
\end{bmatrix}$$

$$e_d(t) = \begin{bmatrix}
0 & 0, T < 4 & 1, T \leq 1 \\
0 & 1, T \geq 4 & 0, T > 1 \\
0.5 & 0.5 & 0
\end{bmatrix}$$

$$v_d(t) = \begin{bmatrix}
0 & [T], T < 4 & T \\
0 & 4, T \geq 4 & T \\
1 & 0 & 0
\end{bmatrix}$$
Applying our results, we begin by determining $g_i$, $q_{ij}$, and $q_{eij}$ for $i, j = 1, 2, 3$. The nonzero $q$'s are $q_{c12} = 0.3927$, $q_{c13} = 0.0812$, $q_{c21} = 0.2000$, $q_{c31} = 0.3500$, $q_{c32} = 0.1500$, $q_{e12} = 0.0073$, $q_{e13} = 0.3793$, $q_{e23} = 0.8000$, $q_{e31} = 0.3500$, $q_{e32} = 0.1500$. Solving the system of equations of Theorem 4 for the $g_i$ we have $g_1 = 0.8434$, $g_2 = 0.9687$, $g_3 = 0.9405$.

We determine the one step moments of our random variable by using Theorem 1. We have, for $p = 1$,

\[
\begin{align*}
\tau_{c12} &= 0.5073, & \tau_{c13} &= 3, & \tau_{c31} &= 1, & \tau_{e12} &= 4, \\
\tau_{e13} &= 0.4180, & \tau_{e23} &= 2, & \tau_{e31} &= 1,
\end{align*}
\]

and for $p = 2$,

\[
\begin{align*}
\tau_{c12}^2 &= 0.8740, & \tau_{c13}^2 &= 10, & \tau_{c31}^2 &= 1, & \tau_{e12}^2 &= 16, \\
\tau_{e13}^2 &= 0.2541, & \tau_{e23}^2 &= 8, & \tau_{e31}^2 &= 1.
\end{align*}
\]

We now have the data necessary to determine the first two moments of $Z_1$, $Z_2$, and $Z_3$. Using the formula of Theorem 1 with $p = 1$ we find

\[
\mu_1 = 1.722, \quad \mu_2 = 1.951, \quad \mu_3 = 1.311.
\]

Now that we have the first moments, we can use the formula of Theorem 1 again, this time with $p = 2$, to determine the second moments. We find

\[
\mu_1^2 = 7.175, \quad \mu_2^2 = 7.856, \quad \mu_3^2 = 5.233.
\]

The technique is clear, and we could determine the moments of the random variables $Z_1$, $Z_2$, $Z_3$ of any specified order by repeatedly applying Theorem 1. Alternately, we could calculate the moments by using Eq. (17), which is essentially a matrix version of the formula of Theorem 1.

V. CONCLUDING REMARKS

Our goal in this article has not been to apply the techniques developed here to particular problems in the psychological arena, nor have we done so. We feel that specific applications might detract from the generality and simplicity of the approach: Computations for the examples given in the article, and for the moments of any transition-additive random variable, are based on just two equations, Eqs. (6) and (8), which form the content of Theorems 1 and 4.

From the standpoint of applications in psychology, we anticipate that in some situations the efficiency of the method will be of most interest. In other cases, the flexibility of defining valid sample paths or repeating the calculations easily for several random variables may be most important. However, it may well be that the greatest use will be in the calculation of moments for semi-Markov processes. This
class of processes has not been widely used, and perhaps the technical tools of this article will encourage greater use.

Finally, we remark that transition-additive random variables sum values assigned to each jump in a valid path, and there are other interesting and potentially useful random variables such as one that multiplies the values assigned to each jump in a valid path. Equations similar to Eqs. (6) and (8) can be used in such cases (and more complicated ones) but it is difficult to specify conditions under which the solutions will make sense, and under which the moments will exist. However, one such case can be handled using the present results. Suppose \( X_i \) is a random variable giving the product of values \( (v_{ij}(t)) \) along a valid path \( (v_{ij}(t) > 0) \). Define a new value matrix \( (v^*_j(t)) \) such that \( v^*_j(t) = \log v_{ij}(t) \). Define a new random variable \( Z_i \) that sum the values of \( v^* \) along a valid path. Our present methods give the moments of \( Z_i \). Since \( Z_i = \log X_i \), one can infer characteristics of the distribution of \( X_i \) from the moments of \( Z_i \). Note, however, that although the moments of \( Z_i \) may exist, the moments of \( X_i \) may not.

REFERENCES


*Received: May 19, 1987*