BIVARIATE MARKOV CHAINS CONTAINING A FAILURE PROCESS

BY NANCY FLOURNOY

The American University

This paper is directed toward the challenge to model dependencies among discrete-state processes. In an earlier motivating application, we used proportional hazards regression models with time-dependent covariates to examine the relationship between relapse following treatment for leukemia, internal biological processes fighting the leukemia, and interventions intended to fix defects in these processes or to stimulate them to behave more aggressively. Complexities in this application led to the introduction of new dependency measures derived from extending Kolmogorov's differential equations. In this paper these dependency measures are interpreted for a bivariate Markov chain where one process is a failure process and another evolves concurrently. Likelihood construction using these dependency measures and others currently used in the context of a failure process are discussed.

1. Introduction. Our interest in measures of association that evolve in time has its genesis in an application involving multiple parallel, non-independent, discrete-state random processes evolving in time. A motivating application involving certain random events and the relationship between these events is described briefly in Section 2 to provide a context for the theoretical development. The principal issue is how to define and model the dependency of a failure process on discrete-state continuous-time random covariates. In Flournoy (1990) we develop measures of dependency among multivariate discrete state stochastic processes by extending the concept of intensity functions using multivariate extensions of Kolmogorov's differential equations. Now these dependency measures are made explicit and interpreted for the situation where one process is a failure process and another (possibly vector-valued) process evolves concurrently. Also we describe

AMS subject classifications. 60J27; 15A51.

Key words and phrases. Intensity functions, Kolmogorov's differential equations, point processes, proportional hazards regression, time-dependent covariates.

The author wishes to acknowledge the comments and suggestions of the reviewer that lead to an improved manuscript.

the incorporation of these dependency measures into the likelihood function for purposes of estimation.

It is important to note that there may not be a unique natural multivariate and conditional extension of Markovian processes or of univariate intensity functions. Indeed a variety of extensions can be created depending on different needs and different criteria. There are two approaches to incorporating the Markov property of lack of memory. One approach assumes that each process is marginally memoryless and characterizes multivariate processes with this property. However, the resulting multivariate processes are not memoryless. This approach was used, for example, by Yadin and Syski (1979) to explore the randomization of intensities in a Markov chain, and by Cogburn (1980) in describing random environments. In contrast to this approach, we examine measures of dependency based on the assumption that two processes are jointly memoryless, yet inhomogeneous.

The rationale for a model that is jointly memoryless, yet inhomogeneous, arises from medical applications such as the one described in the next section. Section 3 provides the necessary notation and a discussion of likelihood construction for one general class of continuous time Markov chains, and for failure processes in particular. In Section 4 a bivariate derivation of intensity functions is obtained that is analogous to Kolmogorov's (1931) univariate derivation. The likelihood is given for the situation in which one process is a failure process. In Section 5 dependencies within the bivariate chain are described and incorporated into the joint likelihood. These dependencies are motivated by one proposed by Cox (1972).

2. A Motivating Scientific Problem. A description of the original motivating application should serve to provide a context to our development. The comparison of several proportional hazards regression models analyzing this application is reported by Weiden, Flournoy, Thomas, Fefer, and Storb (1981) with conclusions reported earlier by Weiden, Flournoy, Thomas, Prentice, Fefer, Buckner, and Storb (1979). To be brief and retain focus, the present description is restricted to the situation surrounding two primary events of interest without regard to competing causes of failure or other covariates.

A bone marrow transplant is preceded by a dose of radiation that destroys the bone marrow that produces leukemic cells. Originally, the effect of this treatment was hypothesized to be dual: the first hypothesis has been demonstrated effectively, namely, that extremely high doses of radiation can be given, increasing the destruction of leukemic cells with the transplant rescuing the patient; the second hypothesis is the subject of inquiry, that if the body's failure to mount an effective fight against the leukemic cells results from a defect in its ability to identify the leukemic cells as foreign objects, and if a healthy immune system is transplanted, one that matches the patient's own immune system so closely that it will accept the patient as itself, the transplanted immune system will recognize that the leukemic cells are strangers and wage an aggressive battle against any that survived the radiation treatment.

Assume that two random processes initiate realizations on the day a patient receives a bone marrow transplant to treat leukemia. One process governs the recurrence of the leukemia, the relapse rate. It is this process that we seek to control and model. A second process governs the emergence of side-effects caused by the bone marrow graft that resemble lupus, an auto-immune disease. If the second hypothesis holds, then an appearance of this side-effect, called graft-versus-host disease, indicates that the transplanted immune system recognizes the foreign cells, and is engaging in a battle that reduces relapse probabilities. However, these battles are frequently fatal. If the hypothesis is true, the battle between the transplanted immune system and its new host, the patient, should be encouraged rather than stifled. But accepting this second hypothesis implies risking the lives of patients believed to be at high risk of later relapse in order to learn how to control the transplant's attack on residual leukemia cells without killing the patient.

The second hypothesis is that the onset of graft-versus-host disease reduces the probability of relapse. Yet if graft-versus-host disease does affect relapse, it is hard to imagine that its effect is constant or even proportional. The need to model inhomogeneous effects led us to consider the Chapman-Kolmogorov equations as a tool for developing multivariate measures of dependency.

3. Notation and the Likelihood. Now we present the necessary notation in the context of a single discrete state process. Then we specialize to a two-state failure process such as is commonly called a death process (see, e.g., Karlin and Taylor (1981)), and establish a correspondence with the notation commonly used in survival analysis. The likelihood is given for the situation in which independent censoring mechanisms terminate the observation of sample paths. A method now used to incorporate dependencies into failure models is described and used to motivate alternative approaches. One such approach in which dependencies are defined by straightforward extensions of classical univariate Markov chain theory is described in subsequent sections. Connections to the theory of counting processes are not pursued at this time.

3.1. The Likelihood for a Univariate Markov Chain. To introduce notation and elementary definitions, consider a continuous time Markov chain $U(t)$, or simply U, with a finite discrete-state space $\theta = 1, 2, \ldots, \Omega_u$ and continuous time parameter $t \in T = [0, \infty)$. Throughout, lower case Roman letters are used to denote times and Greek letters are used to denote states. When considering functions of a process U and the context is clear, we omit the 'U'. Let $H^{t|s} \equiv H^{t|s}(U)$ be a $\Omega_u \times \Omega_u$ matrix of *transition probabilities* with elements

$$h^{t|s}_{\theta|\sigma} \equiv h^{t|s}_{\theta|\sigma}(U) \equiv P\{U(t) = \theta | U(s) = \sigma\}, \qquad \sigma, \theta = 1, 2, \ldots, \Omega_u, \tag{1}$$

each of which is a probability that the process is in state θ at time t given it was in state σ at time s. Note that in this notation the current time t and state θ are given before the vertical line, and the prior time s and state σ are given after the vertical line. Of course, some transitions may have probability zero.

We adopt the assumption that the process U cannot return to a state once it has been there and left, which implies that $H^{t|s}$ is an upper triangular matrix for all $0 < s \leq t$. Also note that, in general, $H^{r|s}$ and $H^{t|r}$ do not commute. These features are significant for solving matrix versions of Kolmogorov's differential equations. Furthermore, these features hold for the motivating application where graft-versus-host disease (the event), relapse of leukemia (the failure), and death from causes other than leukemia together with termination of observation (the censorings) are not reversible.

Assuming the limits exist, let Q^t be a $\Omega_u \times \Omega_u$ dimensional matrix with elements the univariate *intensity functions* $q^t_{\theta|\sigma}(U)$ defined as by Kolmogorov (1931) to be the derivatives of inhomogeneous transition probabilities with time s evaluated at the later time t. The derivative of $H^{t|s}$ taken elementwise yields the upper triangular matrix

$$Q^t \equiv \frac{\partial}{\partial t} H^{t|s}(U)|_{s=t}. \tag{2}$$

Since the diagonal elements of $H^{t|s}$ are positive except in degenerate cases, the inverse $(H^{t|s})^{-1}$ generally exists and so will solutions to Kolmogorov's differential equations

$$\frac{\partial}{\partial t} H^{t|s} = H^{r|s} \frac{\partial}{\partial t} H^{t|r}|_{r=t} \equiv H^{t|s} Q^t \tag{3}$$

for $s \leq r \leq t$.

As pointed out by Chang and Yang (1990), standard reference books on stochastic processes (e.g. Chang (1980)) discuss solutions only of homogeneous Markov transition probabilities, and solutions for inhomogeneous Markov transition probabilities (e.g. for (3)) remain very much application dependent. This remains true although Feller (1940) proved existence and uniqueness theorems for the solutions of inhomogeneous Markov chains; and although more recently the solutions of inhomogeneous Markov chains have been further characterized by Getz (1976) and Hartfiel (1985). Therefore, it is of practical significance that maximum likelihood estimates can be obtained without requiring a general solution to (3). This is the case in the present model.

In fact, solutions are required only for the diagonal elements of $H^{t|s}$ as we now explain. Since $H^{t|s}$ is an upper triangular matrix, both the derivative of $H^{t|s}$ and its inverse are upper triangular matrices, and consequently, the product $(H^{t|s})^{-1}\partial H^{t|s}/\partial t$ is upper triangular. In particular, each diagonal element $q^t_{\theta|\theta}$, $\theta = 1, 2, \ldots, \Omega_u$ in the solution of (3) equals the product of the respective diagonal elements in the matrices $(H^{t|s})^{-1}$ and $\partial H^{t|s}/\partial t$ yielding a system of Ω_u differential equations for each s and t with solutions

$$h^{t|s}_{\theta|\theta} = \exp\{\int_s^t q^r_{\theta|\theta} \mathrm{dr}\}, \quad \theta = 1, 2, \ldots, \Omega_u. \tag{4}$$

Given $n = 1, \ldots, N$ independent realizations of $U(t)$, a likelihood can be constructed using probability laws for sample paths. For the nth realization, let τ_{nm} be the time of the mth transition for $m = 1, \ldots, M_n$ and let $\{\theta_{n0}, \theta_{n1}, \ldots, \theta_{nM_n}\}$ denote the distinct states visited. For notational convenience, we assume that the realizations are of finite duration so that the M_nth transition is to an absorbing state. Karlin and Taylor (1981, pp. 145–149) describe the process of determining probability laws for sample paths from homogeneous Markov chains and point out that the process is analogous for sample paths from general continuous time Markov chains. Following their argument, let k be fixed and $K > 0$ be an arbitrary positive integer. Then for the nth realization $[U_n(t),\ 0 \le t \le \tau_{M_n}]$,

$$h_{\theta_{nm}|\theta_{nm}}^{\tau_{n(m+1-(k/K))}|\tau_{nm}} = P\left\{U_n(\tau) = \theta_{nm},\ \tau_{nm} \le \tau \le \tau_{n(m+1-(k/K))}\right\}$$

and (assuming $U(t)$ is separable),

$$\lim_{K\to\infty} h_{\theta_{nm}|\theta_{nm}}^{\tau_{n(m+1-(k/K))}|\tau_{nm}}$$

may be considered as just

$$h_{\theta_{nm}|\theta_{nm}}^{\tau_{n(m+1)}|\tau_{nm}} = P\left\{U_n(\tau) = \theta_{nm},\ \tau_{nm} < \tau < \tau_{n(m+1)}\right\} = \exp\left\{\int_{\tau_{nm}}^{\tau_{n(m+1)}} q_{\theta|\theta}^{r}\,\mathrm{d}r\right\},$$

where the last equality follows from (4). Also the probability that $U(t)$ remains at θ_{nm} from τ_{nm} to $\tau_{n(m+1)}$ and then jumps to $\theta_{n(m+1)} \neq \theta_{nm}$ in time $\mathrm{d}\tau$ is

$$h_{\theta_{nm}|\theta_{nm}}^{\tau_{n(m+1)}|\tau_{nm}}\, q_{\theta_{n(m+1)}|\theta_{nm}}^{\tau_{n(m+1)}}\,\mathrm{d}\tau.$$

Therefore, assuming the probability of more than one jump in the time span from $\tau_{n(m+1)}$ to $\tau_{n(m+1)} + \mathrm{d}\tau$ goes to zero as $\mathrm{d}\tau \to 0^+$ and given the initial states $\{\theta_{n0},\ n = 1, \ldots, N\}$, the likelihood of $[U_n(t),\ 0 \le t \le \tau_{M_n}]$ is

$$\begin{aligned}(5)\qquad \ell_n &= \left[\prod_{m=0}^{M_n-1} h_{\theta_{nm}|\theta_{nm}}^{\tau_{n(m+1)}|\tau_{nm}}\, q_{\theta_{n(m+1)}|\theta_{nm}}^{\tau_{n(m+1)}}\right] \\ &= \prod_{m=0}^{M_n-1} \exp\left\{\int_{\tau_{nm}}^{\tau_{n(m+1)}} q_{\theta_{nm}|\theta_{nm}}^{r}\,\mathrm{d}r\right\} q_{\theta_{n(m+1)}|\theta_{nm}}^{\tau_{n(m+1)}},\end{aligned}$$

where the last equality follows from (4). Note in the first expression in (5) that the likelihood depends on diagonal elements from $H^{t|s}$ that correspond to the time spent in those states that were actually visited and off-diagonal elements from Q^t that correspond to the actual state changes at observed transition times. The second expression in (5) shows that, by solving for only the diagonal elements of $H^{t|s}$, the likelihood can be written strictly in terms of elements of Q^t. This fact facilitates the modeling dependencies on failure processes.

3.2. The Likelihood of a Failure Process. A simple *failure process*, namely, a two state process with one absorbing state is now described, and its likelihood is derived assuming that independent censoring mechanisms terminate observation of the sample paths. Let V be a failure process with state space $\Omega_v = \{0, 1\}$, where state 1 denotes failure. Then the matrix $H^{t|s}$ of transition probabilities simplifies because the transition matrix depends on only one element, i.e.,

$$H^{t|s}(V) = \begin{bmatrix} h_{0|0}^{t|s} & h_{1|0}^{t|s} \\ h_{0|1}^{t|s} & h_{1|1}^{t|s} \end{bmatrix} = \begin{bmatrix} 1 - h_{1|0}^{t|s} & h_{1|0}^{t|s} \\ 0 & 1 \end{bmatrix}, \tag{6}$$

with the consequence that Q^t also simplifies:

$$Q = \begin{bmatrix} -q_{1|0}^t & q_{1|0}^t \\ 0 & 0 \end{bmatrix}. \tag{7}$$

Clearly, the failure process is completely specified by $q_{1|0}^t$ and the system of differential equations (3) reduces to the single one-to-one correspondence familiar in survival analysis with the solution:

$$S(t) = \exp\left\{-\int_0^t q_{1|0}^t \mathrm{d}r\right\}, \tag{8}$$

where $S(t) \equiv h_{0|0}^{t|s}|_{s=0}$, the probability that no state changes occur up to time t, is called the *survival function* and the *hazard function*, commonly denoted by $\lambda(t)$, is the intensity of failure $q_{1|0}^t$.

The failure process we consider is extended to a three state process to accommodate censoring; the initial state θ_{n0} is survival, and transitions are possible to the absorbing states, θ_{n1} and θ_{n2}, of failure and censoring, respectively. Since only the first transition is observable, realizations are described by the observations $\{(\tau_n, \theta_n),\ n = 1, \ldots, N\}$, where $\tau_n \equiv \min\{\tau_{n1}, \tau_{n2}\}$ is the minimum of the failure and censoring times for the nth observation, and θ_n equals one if the failure state θ_{n1} is observed and zero if the censoring state θ_{n2} is observed. Using (7) and assuming an independent censoring mechanism, Kalbfleisch and Prentice (1980) show that the likelihood (5) for the nth observation is

$$\ell_n = \left(q_{1|0}^{\tau_n}\right)^{\theta_n} \exp\left\{-\int_0^{\tau_n} q_{1|0}^r \mathrm{d}r\right\}. \tag{9}$$

One approach for incorporating dependencies into a Poisson survival model is becoming popular in applications. If $\mathbf{U}_n(t)$ is a row vector of covariate processes for the nth observation and $\boldsymbol{\beta}$ is a row vector of unknown parameters, then a model

$$q_{[1|0]_n}^t = h(\mathbf{U}_n(t), \boldsymbol{\beta}) \tag{10}$$

relating the nth observed hazard function $q_{[1|0]_n}^t$ to covariates $\mathbf{U}_n(t)$ is selected. Then dependencies are evaluated by replacing $q_{[1|0]_n}^t$ with $h(\mathbf{U}_n(t), \boldsymbol{\beta})$ in (9) or in Cox's (1972) partial likelihood and evaluating $\boldsymbol{\beta}$ by maximum likelihood methods. The most famous example of this approach to evaluating the dependency of a covariate process $U(t)$ on a failure process $V(t)$ is the use of Cox's (1972) proportional hazards regression model

$$q^t_{[1|0]_n} = h(t) \exp \left\{ \mathbf{U}_n(t) \boldsymbol{\beta}^T \right\},$$

where $h(t) \geq 0$ is an arbitrary function of time and each covariate acts proportionally on the failure intensity $q^t_{1|0}$, in Cox's (1975) partial likelihood. The partial likelihood ℓ_{pn} for the nth observation is

$$\ell_{pn} = \left[\left(q^{\tau_n}_{[1|0]_n} \right) \Big/ \sum_{j=1}^{R_n} q^{\tau_n}_{[1|0]_j} \right]^{\theta_n}, \tag{11}$$

where R_n is the set $\{j : \tau_j \geq \tau_n\}$ of realizations that have not failed or been censored at the time τ_n that the nth observation fails. Heuristically, $\theta_n = 1$ for each failure, and the corresponding ratio in the partial likelihood is the probability that, of all the realizations at risk at τ_n, it was the nth realization that failed.

When $U(t)$ is deterministic, the insertion of (10) into (9) or (11) can be justified using conditionality arguments. However, when $U(t)$ is a random function of time, insertion of (10) into (9) or into (11) does not take the Jacobian of the transformation into account and clearly their probabilistic content is altered (see Flournoy (1980) and Yashin and Arjas (1988)). The increasing number of applications that use (10) in (9) or (11) when covariates $U(t)$ are random is one motivation for exploring alternative approaches.

Another motivation is to establish a framework which extends the types of dependencies that can be modeled beyond those that can be expressed in the form of (10). Note that in (10) as written, the covariate process(es) maps into the failure intensity at parallel times t. Meaningful dependencies between $U(t)$ and $V(t)$ may involve lag or lead times, or it may be other functions such as the rate at which states change in $U(t)$ that most directly affect $V(t)$. An obvious extension of (10) is $q^t_{[1|0]_n} = h(\mathbf{U}_n(s), \boldsymbol{\beta},\ 0 \leq s \leq t)$. But this extension is so general that forms of $h(\cdot)$ that are meaningful in applications must be determined for the extension to be useful.

4. A Failure Process in a Bivariate Markov Chain. A common practice in modeling a Markov chain with a multivariate state space is to reparameterize to a univariate Markov chain by expressing the state space as a list of all possible points in the multidimensional space. Such an approach yields a parameter for every possible transition which makes the model intractable for most realistic applications and almost certainly intractable if the processes are taken to be inhomogeneous. One way to reduce the complexity of inhomogeneous Markov chains is to assume that functional dependencies exist between the transitions. This is the role of the proportional hazards assumption in Cox's (1972) regression model for failure data. We characterize a multivariate representation of the Markov chain in order to facilitate the conceptualization of other types of functional dependencies within a more general probability framework.

Let $[U(t),\ V(t)]$ be a two-dimensional row vector of discrete-state stochastic processes, where $V(t)$ is a univariate failure process and $U(t)$, without loss of generality, might be vector-valued. Note that the elements of $[U(t),\ V(t)]$ are jointly indexed by a single continuous time parameter $t \in T = [0, \infty)$. When 't' is not critical to the context, we write $[U, V] \equiv [U(t),\ V(t)]$.

Denote the bivariate state space of $[U(t),\ V(t)]$ by $\Omega = \Omega_u \times \Omega_v$. For some t, the probability of being in certain states may be zero. For example, when a failure in $V(t)$ is fatal to $U(t)$, the combined sample space might degenerate to a single point representing the absorbing state of the entire system. Let $[\theta_{ui}^t,\ \theta_{vj}^t]$ denote an element in the bivariate state space Ω at time t. Rather than maintain the superscript 't' in the state space notation, we indicate the states at different times by another letter σ. That is, $[\theta_{ui},\ \theta_{vj}]$ and $[\sigma_{ui},\ \sigma_{vj}]$ denote the ith state of process U and the jth state of process V at two different times.

The state space Ω_u of U is determined by the application, whereas there are only two states for the failure process V. Let θ_{v1} or '$V = 1$' denote the state of failure while θ_{v0} or '$V = 0$' denotes no failure. Thus in the bivariate state space, $[\theta_{ui},\ \theta_{v0}] \equiv [\theta_{ui},\ 0]$ and $[\theta_{ui},\ \theta_{v1}] \equiv [\theta_{ui},\ 1]$ denote $U(t) = \theta_i$ jointly with survival or failure, respectively. When it is not necessary to reference a specific state, we let $(\theta_u,\ \sigma_u)$ and $(\theta_v,\ \sigma_v)$ denote arbitrary states of U in Ω_u and of V in Ω_v at times t, and s, respectively. The nth realization of U and V is denoted by U_n and V_n, respectively, and the mth state visited by the nth realization is denoted by $[\theta_u,\ \theta_v]_{nm}$.

DEFINITION 1. Now let $H^{t|s} \equiv H^{t|s}(U, V)$ denote a $2\Omega_u \times 2\Omega_u$ dimensional matrix of *joint transition probabilities*, where $H^{t|s}$ can be partitioned into four $\Omega_u \times \Omega_u$ upper triangular submatrices,

$$H^{t|s} = \begin{bmatrix} H_{0|0}^{t|s} & H_{1|0}^{t|s} \\ 0 & H_{1|1}^{t|s} \end{bmatrix}, \tag{12}$$

where the 0 submatrix reflects the assumption that the state of failure is absorbing. The matrix $H_{1|0}^{t|s}$ is upper triangular with elements defined by the probability of transition in U from state i at time s to state j at time t $(i, j \in \Omega_u)$ jointly with the transition of V from survival (0) into failure (1):

$$\begin{aligned} h_{\theta_u 1|\sigma_u 0} &\equiv h_{\theta_u 1|\sigma_u 0}^{t|s}(U, V) \\ &\equiv P\{U(t) = \theta_{uj},\ V(t) = 1 | U(s) = \sigma_{ui},\ V(s) = 0\}; \end{aligned} \tag{13}$$

$H_{0|0}^{t|s}$ and $H_{1|1}^{t|s}$ are also upper triangular matrices with elements defined to be probability transitions function of U joint with continued survival $(V = 0)$ from s to

t and with continued failure ($V = 1$), respectively. In referencing the transition probabilities (13), we omit the notation '(U, V)' for simplicity when it does not play a role. The superscript '$t|s$' is also omitted for notational simplicity whenever it is not essential to the context.

The nature of a failure's impact on U is expressed through $H_{1|1}$. For example, $H_{1|1}$ may be the identity matrix I_{Ω_u} implying $U(t)$ is frozen in its current state at the time of failure, or $U(t)$ may enter a single absorbing state regardless of its current state at the time of failure. Often when we focus on how failures depend on a coprocess, it is natural to treat $H_{1|1}$ as ancillary and restrict consideration of joint transition probabilities to the upper submatrices in (12), namely $H_{1|0}$, in which failure occurs, and $H_{0|0}$, in which failure does not occur. $H_{1|1}$ is ancillary when it does not contain parameters in common with $H_{0|0}$ or $H_{1|0}$ which are conditioned on prior survival.

DEFINITION 2. Two discrete state continuous time random processes U and V form a *bivariate Markov chain* if $[U, V]$ have joint loss of memory, that is, if for any s, r, and t such that $0 \leq s \leq r \leq t < \infty$,

$$\begin{aligned} P\{[U(t), V(t)] = [\theta_u, \theta_v] | [U(s), V(s)], \forall s : s \in [0, r]\} \\ = P\{[U(t), V(t)] = [\theta_u, \theta_v] | [U(r), V(r)]\}. \end{aligned} \tag{14}$$

A bivariate extension of the Chapman-Kolmogorov equations follows directly:

$$H^{t|s}(U, V) = H^{r|s}(U, V) H^{t|r}(U, V). \tag{15}$$

Note that the bivariate Chapman-Kolmogorov equation (15) implies that $[U, V]$ is a bivariate Markov chain and does not imply that either marginal process U or V is a Markov chain, that is, (15) does not imply that the univariate Chapman-Kolmogorov equations hold for U or V. Indeed, the bivariate loss of memory property (15) will hold also for marginal processes only under very restrictive conditions (see Yadin and Syski (1979)).

DEFINITION 3. Let two random processes U and V form a bivariate Markov chain and assume that the derivatives of (13) exist with respect to t, then their $\Omega \times \Omega$-dimensional *joint intensity matrix* is defined analogous to (2):

$$Q^t \equiv Q^t(U, V) \equiv \frac{\partial}{\partial t} H^{t|s}(U, V)|_{s=t} \equiv \frac{\partial}{\partial t} H^{t|s}|_{s=t}. \tag{16}$$

The elements of $Q^t(U, V)$ are denoted by $q^t_{\theta_u \theta_v | \sigma_u \sigma_v}$, and are called *joint intensity functions.*

Extending the forward differential equations by taking the derivative of both sides of (15) elementwise with respect to t, and evaluating the result at $r = t$, gives bivariate forward equations for all s and t analogous to (3). When $[U, V]$ is a bivariate Markov chain and V is a failure process, the forward equations are

$$(17)\qquad \frac{\partial}{\partial t}H^{t|s} \equiv \begin{bmatrix} \frac{\partial}{\partial t}H^{t|s}_{0|0} & \frac{\partial}{\partial t}H^{t|s}_{1|0} \\ 0 & \frac{\partial}{\partial t}H^{t|s}_{1|1} \end{bmatrix} = \begin{bmatrix} H^{t|s}_{0|0} & H^{t|s}_{1|0} \\ 0 & H^{t|s}_{1|1} \end{bmatrix} \begin{bmatrix} Q^{t}_{0|0} & Q^{t}_{1|0} \\ 0 & Q^{t}_{1|1} \end{bmatrix}$$
$$= \begin{bmatrix} H^{t|s}_{0|0}Q^{t}_{0|0} & H^{t|s}_{0|0}Q^{t}_{1|0} + H^{t|s}_{1|0}Q^{t}_{1|1} \\ 0 & H^{t|s}_{1|1}Q^{t}_{1|1} \end{bmatrix},$$

and since $H^{t|s}$ is upper triangular with positive diagonal elements (except in degenerate cases), the solution of (17) for the diagonal elements in $H^{t|s}$ is analogous to (4):

$$(18)\qquad h^{t|s}_{\theta_j|\theta_j} = \exp\left\{\int_s^t q^{r}_{\theta_j|\theta_j}\,\mathrm{dr}\right\}; \quad j = 0,1,;\ \theta = 1,2,\ldots,\Omega_u.$$

Recall that in the case where a failure in V freezes the coprocess U, $H_{1|1}$ is the identity matrix and hence (17) simplifies considerably because $Q^{t}_{1|1} \equiv 0$.

Assuming for notational convenience that the M_nth transition is to an absorbing state, the likelihood for the nth realization is of the same form as (5) except that the realized states $[\theta_u, \theta_v]_{nm}$ are now binary with $\theta_v = 0,1$; $m = 1,\ldots,M_n$:

$$(19)\qquad \ell_n = \prod_{m=0}^{M_n-1} h^{\tau_{n(m+1)}|\tau_{nm}}_{[\theta_u\theta_v]_{nm}|[\theta_u\theta_v]_{nm}}\, q^{\tau_{n(m+1)}}_{[\theta_u\theta_v]_{n(m+1)}|[\theta_u\theta_v]_{nm}}$$
$$= \prod_{m=0}^{M_n-1} \exp\left\{\int_{\tau_{nm}}^{\tau_{n(m+1)}} q^{r}_{[\theta_u\theta_v]_{nm}|[\theta_u\theta_v]_{nm}}\,\mathrm{dr}\right\} q^{\tau_{n(m+1)}}_{[\theta_u\theta_v]_{n(m+1)}|[\theta_u\theta_v]_{nm}},$$

where the last equality follows from (18). Hence the likelihood can be constructed using the diagonal elements of $H^{t|s}$, modeled in terms of the diagonal elements of Q^t using (18), together with the off-diagonal elements of Q^t as was the case for the univariate model. If a failure in $V(s)$ freezes $U(t)$ for $t \geq \inf\{s : V(s) = 1\}$ so that $Q^{t}_{1|1} = 0$, (19) becomes an extension of (9):

$$(20)\qquad \ell_n = \left(\prod_{m=0}^{M_n-1} \exp\left\{\int_{\tau_{nm}}^{\tau_{n(m+1)}} q^{r}_{[\theta_u 0]_m|[\theta_u 0]_m}\,\mathrm{dr}\right\}\right)\left(\prod_{m=0}^{M_n-2} q^{\tau_{n(m+1)}}_{[\theta_u 0]_{(m+1)}|[\theta_u 0]_m}\right)$$
$$\times \left(q^{\tau_{n(m+1)}}_{[\theta_u 0]M_n|[\theta_u 0]_{(M_n-1)}}\right)^{1-\delta_{nM_n}} \left(q^{\tau_{n(m+1)}}_{[\theta_u 1]M_n|[\theta_u 0]_{(M_n-1)}}\right)^{\delta_{nM_n}},$$

where $\delta_{nM_n} = 1$ if the M_nth transition is to failure and 0 otherwise.

5. Dependencies Within the Bivariate Chain. We have established the analogy between likelihood construction for univariate and bivariate discrete state

Markov chains and shown that their respective likelihoods can be modeled strictly in terms of the intensity matrix Q^t. However, the joint intensities in Q^t constitute a limited class of dependencies. We now describe two other classes of dependencies defined in Flournoy (1990) in the context of a failure process; other dependencies in time and space could also be considered.

First we consider the process $U(t)$ as it depends on prior realizations of $U(s)$ and $V(s)$ jointly. There are a variety of applications in which the dependency of $U(t)$ on $U(s)$, given $V(s) = 0$ for $0 \leq s \leq t$, is of interest. For example, consider a gross simplification of an educational application developed by Debanne, Rowland, Eielefeld, and Maw (1989) in which U has states that are school grades (i.e. freshman, sophomore, junior, senior), and V is equal to 0 if a student remains in high school and is equal to 1 if the student drops out. One dependency of interest involves the transition through grades of school among students who remain in school.

The second class of dependency we consider is one in which the failure process $V(t)$ at time t depends on itself $V(s)$ at a prior time s together with the coprocess at both times, namely $U(s)$ and $U(t)$. One might wish to model this type of dependency, for example, to study the way in which failure ($V(t) = 1$) depends on transitions in the coprocess ($U(s)$ to $U(t)$) given prior survival ($V(s) = 0$). In the next section, we show how these two classes of dependencies relate to each other and to the joint process and thereby their role in the likelihood.

5.1. Conditional Dependencies. Let $H^{t|s}(U|[U,V]) \equiv H(U(t)|[U(s),V(s)])$ be a $\Omega_u \times 2\Omega_u$ matrix of *conditional transition probabilities* with U at time t conditioned on itself together with V at time s, that is, with elements

$$(21) \qquad \begin{aligned} h^{t|s}_{\theta_u|\sigma_u\sigma_v} &\equiv P\{U(t) = \theta_u|[U(s),V(s)] = [\sigma_u,\sigma_v]\}, \\ &\theta_u,\ \sigma_u = 1,\ldots,\Omega_u;\ \sigma_v = 0,1. \end{aligned}$$

The matrix of conditional transition probabilities is partitioned with respect to $V(s)$ into two upper triangular submatrices in order to focus on the conditional transition probabilities at time t given $U(s)$ joint with survival ($V(s) = 0$) separately from conditional transition probabilities given $U(s)$ joint with failure ($V(s) = 1$):

$$(22) \qquad \begin{aligned} H^{t|s}(U|[U,V]) &\equiv [H(U|[U,0]),H(U|[U,1])] \\ &\equiv \left[H^{t|s}(U|[U,0]),H^{t|s}(U|[U,1])\right]. \end{aligned}$$

Similar to the simplification that is possible in the joint transition matrix, $H^{t|s}(U|[U,1])$ may equal I_{Ω_u} or $U(t)$ may enter a single absorbing state when a failure (i.e. $V(t) = 1$) is fatal to the entire system.

The *conditional intensity matrix* is defined, analogous to (2), through the derivatives of (21) to be

$$(23) \qquad Q^t(U|[U,V]) \equiv \left[\frac{\partial}{\partial t} H^{t|s}(U|[U,V])\right]_{s=t},$$

with elements $q^t_{\theta_u|\sigma_u\sigma_v}$. Partition $Q^t(U|[U,V])$ in (23) compatibly with $H^{t|s}(U|[U,V])$ in (22) to obtain:

$$(24) \qquad \begin{aligned} Q^t(U|[U,V]) &\equiv \left[\frac{\partial}{\partial t} H^{t|s}(U|[U,0]),\ \frac{\partial}{\partial t} H^{t|s}(U|[U,1])\right]_{s=t} \\ &\equiv \left[Q^t(U|[U,0]),\ Q^t(U|[U,1])\right]. \end{aligned}$$

The conditional dependencies are aggregations of the joint transition probabilities with

$$(25) \qquad \begin{aligned} H(U|[U,0]) &= H_{0|0} + H_{1|0}, \qquad H(U|[U,1]) = H_{1|1}, \\ Q(U|[U,0]) &= Q_{0|0} + Q_{1|0}, \qquad Q(U|[U,1]) = Q_{1|1}. \end{aligned}$$

Using (18) in (25), the diagonal elements of $H(U|[U,0])$ are

$$h^{t|s}_{\theta_u|\theta_u\sigma_v} = \exp\left\{\int_s^t q^r_{\theta_u 0|\theta_u 0}\mathrm{dr}\right\} + \exp\left\{\int_s^t q^r_{\theta_u 1|\theta_u 0}\mathrm{dr}\right\}.$$

In the situation where $Q^t_{1|0} = -Q^t_{0|0}$, the number of parameters to be modeled in the joint likelihood (20) is reduced and the analogy between (20) and the univariate likelihood (9) for a failure process is strengthened. This simplification occurs when $Q^t(U|[U,0]) = 0$, that is, when

$$(26) \qquad \frac{\partial}{\partial t} P\left\{U(t) = j | U(s) = i,\ V(s) = 0\right\}|_{s=t} = 0, \qquad i,j = 1,\ldots,\Omega_u.$$

Equation (26) is attained when the conditional transition probabilities are constant functions of t for each i. An important special case in which (26) is attained is the case in which $H^{t|s}(U|[U,0]) = I_{\Omega_u}$, that is, the case in which $U(r)$ is constant for $s \le r \le t$ almost everywhere. In many applications, there is no scientific basis for the assumption that $Q_{0|0} = -Q_{1|0}$. Consequently, the reduction of parameters in the likelihood (9) that resulted from such an equality (see 7) for the univariate failure process does not apply generally to the likelihood (20) for a bivariate chain that contains a failure process.

We now introduce a second class of dependencies before discussing the relationship of both classes to the bivariate likelihood function.

5.2. Cross-conditional Dependencies. Let $H(V|U,[U,V]) \equiv H^{t|t,s}(V|U,[U,V])$ be a $2\Omega_u \times 2\Omega_u$ dimensional matrix with elements

$$h^{t|t,s}_{\theta_v|\theta_u\sigma_u\sigma_v} \equiv P\{V(t) = \theta_v | U(t) = \theta_u,\ [U(s), V(s)] = [\sigma_u, \sigma_v]\}, \quad \theta_u, \sigma_u = 1, \ldots, \Omega_u;\ \sigma_u, \sigma_v = 0, 1, \tag{27}$$

that are called *cross-transition probabilities.* Cross-transition probabilities are dependencies between the failure process V at time t conditioned on the coprocess at the same time t as well as on the joint process $[U, V]$ at a prior time s. Thus, for example, the elements of $H^{t|t,s}(1|U, [U, 0])$ are the probabilities of having failed by time t given survival at time s and given the state of process U at s and at t with states $[\sigma_u, 0]$ and $[\theta_u, 1]$:

$$h^{t|t,s}_{\theta_v|\theta_u\sigma_u\sigma_v}\Big|_{\substack{V(s)=0\\V(t)=1}} = h^{t|t,s}_{1|\theta_u\sigma_u 0} = 1 - h^{t|t,s}_{0|\theta_u\sigma_u 0}, \quad \theta_u, \sigma_u = 1, \ldots, \Omega_u. \tag{28}$$

Through the derivatives of (27), define a *cross-intensity matrix* with V at time t conditioned on U at time t and itself together with U at time s to be

$$Q^t(V|U, [U, V]) \equiv \left[\frac{\partial}{\partial t} H^{t|t,s}(V|U, [U, V])\right]_{s=t}, \tag{29}$$

with elements $q^t_{\theta_v|\theta_u\sigma_u\sigma_v}$ that are called *cross-intensity functions.*

Now $H(V|U, [U, V])$ and $Q(V|U, [U, V])$ each can be partitioned into four $\Omega_u \times \Omega_u$ dimensional submatrices. But these matrices can be defined in terms of one submatrix as is seen from an expanded representation of $H(V|U, [U, V])$:

$$H(V|U, [U, V]) = \begin{bmatrix} H(0|U, [U, 0]) & J - H(0|U, [U, 0]) \\ 0 & J \end{bmatrix}, \tag{30}$$

where J is a $\Omega_u \times \Omega_u$ dimensional matrix with each element identically one reflecting the absorbing state of the failure process. Note that, for convenience, the configuration of the cross-transition functions in (30) deviates from the conventional univariate representation of transition probabilities in which the row probabilities sum to one. From (30) we have, by definition (29), that

$$Q(0|U, [U, 0]) = -Q(1|U, [U, 0]) \tag{31}$$

analogous to (7).

5.3. Dependency Measures and the Likelihood. Theorem 1 states that the joint likelihood can be written in terms of conditional and cross-conditional intensity functions. Let $\delta_{M_n} = 1$ if the M_nth transition of the nth realization is to failure and 0 otherwise.

Theorem 1. If $Q^t_{1|1} = 0$, then the likelihood (20) for the nth realization of a bivariate Markov chain $[U, V]$ in which V is a failure process can be rewritten as

$$(32)\ \ell_n = \left[\prod_{m=0}^{M_n-1} \exp\left\{\int_{\tau_{nm}}^{\tau_{n(m+1)}} \left(q^r_{[\theta_u]_{n(m+1)}\|[\sigma_u 0]_{nm}} - \delta^r_{\sigma_u\theta_u} q^r_{[1]_{n(m+1)}\|[\theta_u\sigma_u 0]_{nm}}\right) \mathrm{dr}\right\}\right]$$
$$\prod_{m=0}^{M_n-2} \left(q^{\tau_{n(m+1)}}_{[\theta_u]_{n(m+1)}\|[\sigma_u 0]_{nm}} - \delta^{\tau_{n(m+1)}}_{\sigma_u\theta_u} q^{\tau_{n(m+1)}}_{[1]_{n(m+1)}\|[\theta_u\sigma_u 0]_{nm}}\right)$$
$$\times \left(q^{\tau_{nM_n}}_{[\theta_u]_{nM_n}\|[\sigma_u 0]_{n(M_n-1)}} - \delta^{\tau_{n(m+1)}}_{\sigma_u\theta_u} q^{\tau_{nM_n}}_{[1]_{nM_n}\|[\theta_u\sigma_u 0]_{n(M_n-1)}}\right)^{1-\delta_{nM_n}}$$
$$\times \left(q^{\tau_{nM_n}}_{[1]_{nM_n}\|[\theta_u\sigma_u 0]_{n(M_n-1)}}\right)^{\delta_{M_n}},$$

where $\delta^t_{\sigma_u\theta_u}$ is the Kronecker delta function.

PROOF. The joint transition probabilities factor into two components of corresponding elements from $H(U|[U,V])$ and $H(V|U,[U,V])$:

$$(33)\quad \begin{aligned} h^{t|s}_{\theta_u\theta_v|\sigma_u\sigma_v} &= P(V(t)=\theta_v|U(t)=\theta_u,\ V(s)=\sigma_v,\ U(s)=\sigma_u) \\ &\quad \times\frac{P(U(t)=\theta_u, V(s)=\sigma_v, U(s)=\sigma_u)}{P(V(s)=\sigma_v, U(s)=\sigma_u)}, \\ &= h^{t|t,s}_{\theta_v|\theta_u\sigma_u\sigma_v} h^{t|s}_{\theta_u|\sigma_u\sigma_v}. \end{aligned}$$

Apply the chain rule to (33) in taking the derivative with respect to t to obtain

$$(34)\quad \frac{\partial}{\partial t} h^{t|s}_{\theta_u\theta_v|\sigma_u\sigma_v} = \left[h^{t|t,s}_{\theta_v|\theta_u\sigma_u\sigma_v}\right]\left[\frac{\partial}{\partial t} h^{t|s}_{\theta_u|\sigma_u\sigma_v}\right] + \left[h^{t|s}_{\theta_u|\sigma_u\sigma_v}\right]\left[\frac{\partial}{\partial t} h^{t|t,s}_{\theta_v|\theta_u\sigma_u\sigma_v}\right].$$

Note that

$$(35)\quad \begin{aligned} h^{t|t,s}_{\theta_v|\theta_u\sigma_u\sigma_v} &= P(V(t)=\theta_v|U(t)=\theta_u,\ U(s)=\sigma_u,\ V(s)=\sigma_u) \\ &= \frac{P(V(t)=\theta_v, V(s)=\sigma_v|U(t)=\theta_u, U(s)=\sigma_u)}{P(V(s)=\sigma_v|U(t)=\theta_u, U(s)=\sigma_u)} \end{aligned}$$

and (35) evaluated at $s=t$ is equal to 0 if $\sigma_v \neq \theta_v$ and is equal to 1 if $\sigma_v = \theta_v$. Similarly,

$$(36)\quad \begin{aligned} h^{t|s}_{\theta_u|\sigma_u\sigma_v}|_{s=t} &= P(U(t)=\theta_u|U(s)=\sigma_u,\ V(s)=\sigma_v)|_{s=t} \\ &= \frac{P(U(t)=\theta_u, U(s)=\sigma_u|V(s)=\sigma_v)}{P(U(s)=\theta_u|V(s)=\sigma_v)}|_{s=t} = \delta^t_{\sigma_u\theta_u}. \end{aligned}$$

Evaluating (34) at $s=t$, using (35) and (36), yields a joint intensity function that is a weighted sum of a conditional and a cross-intensity function:

$$q^t_{\theta_u\theta_v|\sigma_u\sigma_v} = \delta^t_{\theta_v\sigma_v} q^t_{\theta_u|\sigma_u\sigma_v} + \delta^t_{\sigma_u\theta_u} q^t_{\theta_v|\theta_u\sigma_u\sigma_v}. \tag{37}$$

Evaluate the right hand side of (37) for $q^t_{\theta_u 0|\sigma_u 0}$ and $q^t_{\theta_u 1|\sigma_u 0}$. Insert the results into (20), and by (31), replace $q^t_{[0]_{nm}|[\theta_u\sigma_u 0]_{n(m-1)}}$ with $-q^t_{[1]_{nm}|[\theta_u\sigma_u 0]_{n(m-1)}}$ for each t and m to yield the theorem. ||

Cox defined the hazard function depending on a covariate value at t to be

$$\lambda(t|U(t)) = \lim_{\Delta t\to 0} \frac{1}{\Delta t} P\left\{V(t+\Delta t) - V(t) = 1 | V(t) = 0,\ U(t)\right\}, \tag{38}$$

which is similar to, but not the same as, the corresponding limit of the cross-transition probability (27) from survival to failure:

$$\lim_{\Delta t\to 0} \frac{1}{\Delta t} h^{t+\Delta t|t+\Delta t,t}_{1|\theta_u\sigma_u 0}. \tag{39}$$

Both (38) and (39) are conditioned on survival to time t, but they differ in that $\lambda(t|U(t))$ is conditioned on only one value of U at time t, whereas (39) is conditioned on U at s and t. However, neither (38) nor (39) is equal to the conditional intensity function $q^t_{1|\theta_u\sigma_u 0}$. Note that the derivative of $H^{t|t,s}$ in (29) is taken with respect to 't' which appears in the conditioning event as well as the conditioned event. The number of parameters in the likelihood (32) can be reduced in two ways. First, models such as described in (10) can be used to provide structure to the cross-conditional intensities:

$$\begin{aligned} q^t_{[1]_{nm}|[\theta_u\sigma_u 0]_{n(m-1)}} &= h(U_n(t), \boldsymbol{\beta}), \\ m &= 0, 1, \ldots, M_n,\ \theta_u, \sigma_u = 1, \ldots, \Omega_u. \end{aligned} \tag{40}$$

Also the underlying science of an application or prior data on the covariate process U (which frequently exists conditional on survival) may suggest structural models for

$$\begin{gathered} q^{\tau_{n(m+1)}}_{[\theta_u]_{n(m+1)}|[\sigma_u 0]_{nm}}, \\ m = 0, 1, \ldots, M_n,\ \theta_u, \sigma_u = 1, \ldots, \Omega_u. \end{gathered} \tag{41}$$

We expect that when the covariate process U is nonstationary, modeling the conditional (40) and cross-conditional intensity functions (41) and using the joint likelihood (32) can lead to better estimates than can be obtained by using only the model (10) in the partial likelihood (11). Structural models to reduce the parameters in (40) and (41) will be proposed, and their performance analyzed elsewhere.

REFERENCES

CHANG, C.L. (1980). *An Introduction to Stochastic Processes in Biostatistics.* R.E. Krieger Publishing Company, New York.

CHANG, M.N. and YANG, G.L. (1990). A stochastic model for prevalence of hepatitis A antibody. *Math. Bio.* **98** 157–169.

COGBURN, R. (1980). Markov chains in random environments: The case of Markovian environments. *Ann. Prob.* **8** 908–916.

COX, D.R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc.* Ser. B **34** 187–202.

COX, D.R. (1975). Partial likelihood. *Biometrika* **64** 269–276.

DEBANNE, S.M., ROWLAND, D.Y., EIELEFELD, R.A., and MAW, C.E. (1989). Adaption of the National Center for Education Statistics survey data for Markovian predicted models. P.215–219 in: *The 1989 Proceedings of the Social Statistics Section.* American Statistical Association, Alexandria, VA.

FELLER, W. (1940). On the integrodifferential equations of purely discontinuous Markov processes. *Trans. of the Amer. Math. Soc.* **48** 488–515.

FLOURNOY, N. (1980). On the survival and intensity functions. P.75 in: *1980 Abstracts. Summaries of papers presented at the joint statistical meeting of the American Statistical Association and the Biometric Society.* American Statistical Association: Alexandria, VA.

FLOURNOY, N. (1990). Dependency in multivariate Markov chains. *Linear Algebra and Its Applications* **127** 85–106.

GETZ, W.M. (1976). Stochastic equivalents of the linear and Lotka-Volterra system of equations—a general birth–and–death process formulation. *Math. Bio.* **29** 235–257.

HARTFIEL, D.J. (1985). On the solutions to $x' = A(t)x(t)$ over all $A(t)$, where $P \leq A(t) \leq Q$. *J. Math. Anal. and Appl.* **108** 230–240.

KALBFLEISCH, J.D. and PRENTICE, R.L. (1980). *The Statistical Analysis of Failure Time Data.* John Wiley and Sons, Inc., New York.

KARLIN, S. and TAYLOR, H.M. (1981). *A Second Course in Stochastic Processes.* Academic Press, Inc., New York.

KOLMOGOROV, A.N. (1931). Über die analytischen Methaden in der Wahrscheinlichkeitsrechnung. *Mathematische Annalen* **104** 415–458.

WEIDEN, P.L., FLOURNOY, N., THOMAS, E.D., FEFER, A., and STORB, R. (1981). Antitumor effect of marrow transplantation in human recipients of syngeneic or allogeneic grafts. Pages 11–23 in: *Graft-Versus-Leukemia in Man and Animal Models,* (ed. by J. Okunewick and R. Meredith), CRC Press, Inc.: Boca Raton, Florida.

WEIDEN, P.L., FLOURNOY, N., THOMAS, E.D., PRENTICE, R., FEFER, A., BUCKNER, C.D., and STORB, R. (1979). Antileukemic effect of graft-versus-host disease in human recipients of allogeneic grafts. *New Eng. J. Med.* **300** 1068–1073.

YADIN, M. and SYSKI, R. (1979). Randomization of intensities in a Markov chain. *Adv. in Appl. Prob.* **11** 397–421.

YASHIN, A. and ARJAS, E. (1988). A note on random intensities and conditional survival functions. *J. Appl. Prob.* **25** 630–635.

DEPARTMENT OF MATHEMATICS AND STATISTICS
AMERICAN UNIVERSITY
WASHINGTON, DC 20016-8050