$\mathcal{I}^2\mathcal{SDS}$

The Institute for Integrating Statistics in Decision Sciences

Technical Report TR-2010-5

April 10, 2010

Markov and Hidden Markov Models

Refik Soyer

Department of Decision Sciences

The George Washington University

Markov and Hidden Markov Models

Refik Soyer
Department of Decision Sciences
The George Washington University
Washington, DC 20052, USA

Abstract

In this paper we give an overview of Markov models that are used for describing dynamics of operating environments in reliability analysis. We present both continuous and discrete-time Markov chains as well as diffusion processes. We also consider Markov modulated stochastic processes that are known as hidden Markov models and discuss their properties. More specifically, we give an overview of Markov modulated Poisson and Markov modulated Bernoulli processes.

Keywords. Reliability assessment, random environment, Markov chains, Markov modulated Poisson process, Markov modulated Bernoulli process

1 Introduction and Overview

Stochastic process models are fundamental for assessing the reliability of items that operate under a dynamic environment. As pointed out by Singpurwalla [42], since the dynamic environment causes changes in the physics of failure, use of stochastic processes provides flexibility in describing failure characteristics under these environments. Earlier uses of stochastic process models include Mercer [24] who modeled item wear and Klein [21] who used a Markov chain model to describe stochastic deterioration of a system. Gaver [13] is the first who propose modeling the failure rate as a stochastic process and introduced the notion of a *randomly changing environment*.

Assessment of reliability of an item/system as a function of the stochastic process governing the environment was considered by Çinlar ([5] [6], [7]), and the notion further developed by Çinlar and Özekici [9] who considered models of stochastic dependence caused by a random environment. The model of [9] is studied further in Çinlar et al. [10]. Related developments are those by Lindley and Singpurwalla [23], and by Singpurwalla and Youngren [44]. Özekici [26] analyzed the optimal maintenance problem of a single-component device operating in a random environment. In all these cases, the environment is characterized by a stochastic process model; see Singpurwalla [42], for a survey. Recent work in stochastic hazard processes can be found in Özekici [28].

Note that a stochastic process is a collection of random variables that are indexed by a parameter space T. In our setup the parameter space T will be time which could be discrete, say $T = I = (0, 1, 2, \ldots,)$, or continuous, say $T = \mathbb{R}^+ = (t; 0 \leq t < \infty)$. In our notation, the collection of random variables $X_0, X_1, \ldots, X_n, \ldots$, will denote a discrete time stochastic process, $\{X_n; n \in I\}$, whereas the collection $\{Y_t; t \in \mathbb{R}^+\}$ denotes a continuous time stochastic process. The state space E of the stochastic process, that is, the possible set of values that either X_n or Y_t can take, can be discrete or continuous. When E is discrete, the process $\{X_n; n \in I\}$ is said to be a discrete state, discrete time stochastic process, and $\{Y_t; t \in \mathbb{R}^+\}$ is a discrete state, continuous time process. Similarly, when E is continuous the process $\{Y_t; t \in \mathbb{R}^+\}$ is referred to as a continuous state, continuous time process.

Of the several stochastic process models that have been considered by the above authors, the simplest is the Markov process, but many stochastic models used in reliability analysis belong to the Markov processes. Important class of Markov processes include discrete and continuous time Markov chains and diffusion processes. In Section 2 we will give an overview of Markov processes and discuss some of the widely used examples. Most of the material in this section can be found in Çinlar [8], Karlin and Taylor [19] or Ross [37].

Stochastic processes that are governed by Markov processes which cannot be observed, are referred to as *hidden Markov models*; see [41]. Such processes have found applications in areas such as speech recognition [36], signal processing [15], biology and medicine [14], environmental sciences [16], software engineering ([29], [11], [38]) and have been of interest to statisticians ([4], [40], [30], [3])because of the difficult inferential issues that they pose. The hidden Markov models will be discussed in Section 3. An application of the hidden Markov models to software reliability analysis is presented in Section 4.

2 Markov Processes

The stochastic process $\{Z_t;\ t \in T\}$ is said to be a Markov process (MP) if for any $t, s \in T$ and for any set of states $A \in E$,

$$P(Z_{t+s} \in A | Z_u; u \leq t) = P(Z_{t+s} \in A | Z_t). \quad (1)$$

The above relationship is referred to as the *Markov property*. It simply states that given the present the future of the process is independent of the past. If (1) holds for all values of t, then MP is said to be *time-homogeneous*.

A continuous time MP, Z_t, with continuous state space E is called a *diffusion process*. Diffusion processes are used to describe item wear and system degradation over time; see for example, Park and Padgett [35]. Brownian motion and gamma process [2] are the most commonly used diffusion processes for modeling such phenomena. An important class of MPs are processes with *independent increments*. A continuous time stochastic process Z_t will have independent increments if for all $t_1 < t_2 < \ldots < t_n$ the random variables Z_{t_1}, $Z_{t_2} - Z_{t_1}, \ldots$,

$Z_{t_n} - Z_{t_{n-1}}$ are independent; see [37]. Furthermore, the process will have *stationary increments* if the distribution of $(Z_{t+s} - Z_t)$ does not depend on t. It is easy to see that every stochastic process Z_t with independent increments is a MP.

Both Brownian motion and gamma process have the independent increments property. As noted by Singpurwalla [42] and Park and Padgett [35], the Gaussian assumption of increments in the Brownian motion makes it an unattractive choice for modeling wear and degradation since the process is not monotonically increasing. One way to alleviate this is to use a geometric Brownian motion or a gamma process as suggested in [35]. As pointed by the authors, the gamma process is more desirable than the geometric Brownian motion since it is always positive and increasing.

The original use of gamma process for descibing deterioration goes back to Abdel-Hameed [1]. A gamma process Z_t is a diffusion process with the following properties:

(i) $Z_0 = 0$;

(ii) Z_t has independent increments;

(iii) $(Z_t - Z_s)$ has a gamma distribution $G[(a_t - a_s), b]$ with shape parameter $(a_t - a_s) > 0$ and scale parameter $b > 0$ for all $t > s$.

An excellent review of gamma processes and their use in maintenance is given in van Noortwijk [45].

2.1 Markov Chains

A MP with discrete state space E is called a *Markov chain*. If the process is a discrete time process, that is, if $T = I$ then the process is called a discrete time Markov chain or more commonly a Markov chain. More specifically, a stochastic process $\{X_n; n \in I\}$ is said to be a Markov chain, with state space E, if for all $n, m \in I$, and $(x_1, x_2, \ldots,) \in E$,

$$P(X_{n+m} = x_{n+m}|X_n = x_n, \ldots, X_0 = x_0) = P(X_{n+m} = x_{n+m}|X_n = x_n). \tag{2}$$

The Markov chain (MC) is said to be time homogeneous, if the above relationship is independent of n, that is,

$$P(X_{n+m} = j|\ X_n = i) = P_m(i, j),$$

where $P_m(i, j)$ is called the *m-step transition function* of the MC. $P(i, j) = P_1(i, j)$ is called the transition probability from state i to state j of the time homogeneous MC, that is,

$$P(X_{n+1} = j|\ X_n = i, \ldots, X_0) = P(X_{n+1} = j|\ X_n = i) = P(i, j).$$

The matrix $\mathbf{P}$, whose $(i,j)^{th}$ element is $P(i,j)$ is called the transition probability matrix of the MC. Note that $P(i,j) \geq 0$, for any $i, j \in E$ and $\sum_{j \in E} P(i,j) = 1$.

A continuous time discrete state space MP, Y_t, is called a continuous time Markov chain. Our discussion of continuous time Markov chains in the next section is based on [41].

2.2 Continuous Time Markov Chains

The stochastic process $\{Y_t; t \in \mathbb{R}^+\}$ is said to be a continuous time Markov chain (CTMC) with discrete state space E, if for any $j \in E$, and $t, s \in \mathbb{R}^+$

$$P(Y_{t+s} = j | Y_u; u \leq t) = P(Y_{t+s} = j | Y_t). \tag{3}$$

If the above relationship holds for all values of t, then the CTMC is said to be *time-homogeneous*, and we write

$$P(Y_{t+s} = j | Y_t = i) = P_s(i,j).$$

where $P_s(i,j)$ is called the transition function of the CTMC and it has the following properties:

(i) $P_s(i,j) \geq 0$

(ii) $\sum_{j \in E} P_s(i,j) = 1$

(iii) $P_{r+s}(i,j) = \sum_{k \in E} P_r(i,k) P_s(k,j)$.

The relationship (iii) above is called the *Chapman-Kolmogorov Equation.*

A CTMC with

$$P(Y_{t+s} = j | Y_t = i) = \begin{cases} P_s(i,j) = 0, & \text{if } \ j < i \\ \frac{e^{-\lambda s} (\lambda s)^{j-i}}{(j-i)!} \ , & \text{if } \ j \geq i, \end{cases} \tag{4}$$

is known as a homogeneous *Poisson process* for some constant $\lambda > 0$.

Associated with any CTMC, is a discrete parameter, discrete state stochastic process $\{X_n; n \in I\}$, where X_n can be constructed from the process $\{Y_t; t \in \mathbb{R}^+\}$ as follows:

$$X_n = Y_t \text{ , for } T_n \leq t < T_{n+1} \text{ , } n = 0, 1, 2, \ \ldots \ , \tag{5}$$

where $0 \equiv T_0 < T_1 < \cdots < T_n < T_{n+1} < \cdots$ are the times at which the CTMC changes states. The process $\{X_n; n \in I\}$ keeps track of the various states that the process $\{Y_t; t \in \mathbb{R}^+\}$ takes over time starting with the initial state Y_{T_0} at time 0. The process $\{X_n; n \in I\}$ is said to be *engendered* by the process $\{Y_t; t \in \mathbb{R}^+\}$ and has certain properties.

Result 1: Suppose that $\{Y_t; t \in \mathbb{R}^+\}$ is a CTMC with state space E. Let $\{X_n; n \in I\}$ be the corresponding discrete parameter process that is engendered by the above CTMC via the tracking scheme mentioned before. Then, $\{X_n; n \in I\}$ is a MC with some transition matrix $\mathbf{P}$, where $P(i,j) > 0$, for $i \neq j$, $P(i,i) = 0$, and $\sum_{j \in E} P(i,j) = 1$.

Thus, the Markov property of the $\{Y_t; t \in \mathbb{R}^+\}$ process is inherited by its engendered $\{X_n; n \in I\}$ process. The process $\{X_n; n \in I\}$ is called the *embedded chain* of the CTMC.

Result 2: For any $u \geq 0$, and a constant $\mu(i) > 0$,

$$P(T_{n+1} - T_n \geq u \mid X_{n+1} = j, X_n = i) \ = e^{-\mu(i)u}.$$

Thus the sojourn times in a CTMC have an exponential distribution whose scale parameter depends only on the current state i; it does not depend on the state to which X_n is going to make a transition to j. Thus μ need only be indexed by i.

Result 3: For any $u_i \geq 0$, $i = 1, 2, \ldots$, and $\mu(i_j) > 0$, $j = 0, 1, 2, \ldots$.

$$\begin{aligned} P(T_1 - T_0 \geq u_1, T_2 - T_1 \geq u_2, \ldots, T_n - T_{n-1} \geq u_n \mid X_0 = i_0, \ldots, X_n = i_n) \\ = \exp(-\mu(i_0)u_1)\ \exp(-\mu(i_1)u_2)\ \cdots\ \exp(-\mu(i_{n-1})u_n)\ . \end{aligned}$$

That is, the sojourn times of a CTMC are, conditional on $X_0, \ldots, X_n$, independently, but not identically exponentially distributed. The scale parameters of the exponential distributions depend only on the current states of the X_i's.

The next result pertains to the joint distribution of the sojourn time in a state, and the state to which the CTMC is going to make a transition to. Specially, for any i, $j \in E$, and any $u \geq 0$.

Result 4:

$$\begin{aligned} P(X_{n+1} = j,\ T_{n+1} - T_n \geq u \mid X_0, \ldots, X_n, T_0, \ldots, T_n) \\ = P(X_{n+1} = j,\ T_{n+1} - T_n \geq u \mid X_n = i) = P(i,j)e^{-\mu(i)u}\ . \end{aligned}$$

Thus for a CTMC, the said joint distribution depends only on the current state of the process i, via the $\mu(i)$, and the transition probability matrix $\mathbf{P}$ of the embedded chain. Furthermore, the transition to a particular state is independent of the time spent in the current state.

In principle, we should be able to obtain results parallel to those of the four claims for processes other than the Markov. However, the attractive properties of inheriting the Markovian feature, the exponentiality and the (conditional) independence of the sojourn times, and the (conditional) independence between the sojourn time and the state transitioned to, may be lost; see Limnios [22] for *semi-Markov processes*. Finally, since homogeneous Poisson processes with the parameter $\lambda > 0$ is a special case of the CTMC, our Claims 2, 3, and 4 simplify with all the $\mu(\cdot)$'s replaced by a common λ. Consequently, here, conditioning on the current state does not matter. Furthermore, in the case of a Poisson process $P(i,j) = 1$, only if $j = i + 1$; otherwise it is zero.

2.3 The Generator of CTMC and Markovian Analysis

Analogous to the notion of a transition probability matrix of a MC, is the notion of a transition probability matrix of a CTMC. Specifically, the matrix $\mathbf{P}_s$, whose $(i,j)^{th}$ element element is $P_s(i,j)$, is of relevance for studying several characteristics of a CTMC. Recall that $P_s(i,j)$ is the probability that the CTMC transitions from state i to state j in a time interval of length s.

Result 5: For any $i, j \in E$, $P_s(i,j)$ is differentiable, and $\frac{d}{ds}P_s(i,j) = A(i,j) = \mu(i)[P(i,j) - I(i,j)]$, where $I(i,j) = 1$, if $j = i$, and is zero otherwise. Thus

$$A(i,j) = \begin{cases} -\ \mu(i), \text{ if } j = i \\ \mu(i)P(i,j), \text{ if } \ j \neq i \ . \end{cases}$$

Consequently, if $P_s(i,j)$ is known for every $i, j \in E$, then $A(i,j)$ is known, and from there we know $\mu(i)$and $P(i,j)$. Thus a knowledge of the transition probability matrix $\mathbf{P}_s$ of the CTMC enables us to obtain the transition probability matrix $\mathbf{P}$ of the embedded chain, and also, the $\mu(i)$'s, the scale parameters of the distributions of the sojourn times of the process. Conversely, if E is a finite set of say m elements, then a knowledge of $\mu(1), \ldots, \mu(m)$, and $\mathbf{P}$ enables us to obtain $\mathbf{A}$, the matrix whose $(i,j)^{th}$ element is $A(i,j)$. Once $\mathbf{A}$ is known, we may obtain the matrix $\mathbf{P}_s$ via the *matrix-geometric* relationship

$$\mathbf{P}_s = e^{s\mathbf{A}} = \sum_{k=0}^{\infty} \frac{s^k \mathbf{A}^k}{k!} \ ;$$

the term "geometric" is motivated by the fact that the summation above is a consequence of the geometric series expansion of $e^{s\mathbf{A}}$. The above development is known as a *Markovian analysis.* The matrix $\mathbf{A}$ is called the *generator* of the CTMC.

3 Hidden Markov Models

As previously pointed out, hidden Markov models (HMMs) are stochastic processes that are governed by MPs which cannot be observed. Examples of HMMs include Markov modulated Bernoulli processes ([27] and [30]) where the governing or modulating process is a MC and Markov modulated Poisson processes [12] where the modulating process is a CTMC. It is important to note that here we use the term HMM or a Markov modulated process in a more general manner to include any MP as a governing process. Sometimes in the literature the term HMM is used to refer only to discrete time processes governed by MCs (see for example, [3]) whereas the term Markov modulated process is reserved for the case where the governing process is a CTMC; see for example, Rydén [39].

In reliability modeling, the HMMs are sometimes referred to as random environment models as in Özekici and Soyer [32]. In [32] the item/system in question is assumed to operate in a randomly changing environment depicted by $Y = \{Y_t; t \in T\}$ where Y_t is the state of the environment at time t. The environmental process Y is a MP with some state space E which is assumed to be discrete to simplify the notation. Note that if the environmental process is assumed to be a semi-Markov process, then the resulting processes will be hidden semi-Markov models as considered by [22] and [33].

In what follows, we assume that the environmental process Y is a MP and we consider the Markov Modulated Poisson Process (MMPP) and the Markov modulated Bernoulli process. As noted in [41], MMPP is a special case of the *Markov Renewal Process* (MRP) of Çinlar [5] as well as the D*oubly Stochastic Poisson Process* of Kingman [20] which is also known as the Cox process.

3.1 Markov Modulated Poisson Model

Let N be a modulated Poisson process such that N_t depicts the total number of arrivals until time t. The modulation is done via an environmental process Y with a discrete state space E where Y_t represents the state of the environment at time t. The rate of arrivals at time t is $\lambda(Y_t)$ for some arrival rate vector λ defined on E. We suppose that while the environment is at state i arrivals occur according to an ordinary Poisson process with rate $\lambda(i)$. To be more precise,

$$P[N_t = k|Y] = \frac{e^{-A_t} A_t^k}{k!} \tag{6}$$

where

$$A_t = \int_0^t \lambda(Y_s) ds \tag{7}$$

for all $k = 0, 1, \cdots$ and $t \geq 0$.

It follows from the above that, given Y, N is a nonstationary Poisson process with mean value function $E[N_t|Y] = A_t$. Defining T to be the arrival time process so that T_n is the time of the nth arrival, we have the conditional interarrival time distribution

$$P[T_{n+1} - T_n > t|Y, T_n] = e^{-(A_{T_n+t} - A_{T_n})}. \tag{8}$$

The modulated process reduces to the ordinary Poisson process with rate λ if the arrival rate vector is $\lambda(i) = \lambda$ independent of the environment for all i. In this case, $A_t = \lambda t$ deterministically.

The arrival process N can be studied via the additive functional A of Y. In particular, (6) and (8) directly yield

$$P[N_t = k] = E[\frac{e^{-A_t} A_t^k}{k!}] \tag{9}$$

and

$$P[T_{n+1} - T_n > t|T_n] = E[e^{-(A_{T_n+t} - A_{T_n})}]. \tag{10}$$

Therefore, the probability law of A, thus that of the environmental process Y, will play an important role in our analysis of N and T.

We assume that Y is the minimal Markov process associated with a Markov renewal process (X, S) with Markov kernel Q. In other words, X_n is the nth state visited by the environmental process and S_n is the time of this visit such that

$$Y_t = X_n \text{ whenever } S_n \leq t < S_{n+1}. \tag{11}$$

Moreover, the Markov kernel gives

$$Q(i, j, t) = P(i, j)(1 - e^{-\mu(i)t}) \tag{12}$$

where $\mathbf{P}$ is the transition matrix of the embedded Markov chain X and μ is the vector of jump rates. More precisely, if the process Y is in some state i, then it stays there for an exponentially distributed amount of time with rate $\mu(i)$ and then jumps to some other state j with probability $P(i, j)$. It also follows that

$$F_i(t) = 1 - e^{-\mu(i)t} \tag{13}$$

and the generator of the Markov process Y is given by the matrix

$$G(i, j) = \begin{cases} -\mu(i), & \text{if } j = i \\ \mu(i)P(i, j), & \text{if } j \neq i. \end{cases} \tag{14}$$

Let P_t denote the transition function of Y so that

$$P_t(i, j) = P[Y_t = j | Y_0 = i]. \tag{15}$$

Following [33], we define another process Y^λ such that

$$Y_t^\lambda = \begin{cases} Y_t, & \text{if } t < T_1 \\ \Delta, & \text{if } t \geq T_1 \end{cases} \tag{16}$$

where T_1 is the time of the first arrival. While the environment is in state i, the time of an arrival has the exponential distribution with rate $\lambda(i)$. It is clear that Y^λ is also a Markov process on the extended state space $E_\Delta = E \cup \{\Delta\}$ and it is obtained by "stopping" the Markov process Y as soon as an arrival occurs. Here, Δ is an absorbing state where the process is dumped to as soon as it is stopped. The transition matrix of the embedded Markov chain is now extended as

$$P_\lambda(i, j) = \begin{cases} \frac{\mu(i)}{\mu(i)+\lambda(i)} P(i, j), & \text{if } i, j \in E \\ \frac{\lambda(i)}{\mu(i)+\lambda(i)}, & \text{if } i \in E, j = \Delta \\ 1, & \text{if } i, j = \Delta \end{cases} \tag{17}$$

and the transition rate vector is

$$\mu_\lambda(i) = \begin{cases} \mu(i) + \lambda(i), & \text{if } i \in E \\ 0, & \text{if } i = \Delta. \end{cases} \tag{18}$$

If we let the matrix $G_\lambda(i,j) = \mu_\lambda(i)(P_\lambda(i,j) - I(i,j))$ denote the generator of Y^λ, then it is well known that the transition function $P_t^\lambda(i,j) = P[Y_t^\lambda = j|Y_0^\lambda = i]$ for all $i,j \in E_\Delta$ is given by the matrix-exponential solution

$$P_t^\lambda = e^{G_\lambda t} = \sum_{n=0}^{+\infty} \frac{t^n}{n!} G_\lambda^n. \tag{19}$$

A further simplification is obtained by noting that

$$G_\lambda(i,j) = G(i,j) - \Lambda(i,j) \tag{20}$$

for all $i,j \in E$ where Λ is a diagonal matrix defined as

$$\Lambda(i,j) = \begin{cases} \lambda(i), & \text{if } j = i \\ 0, & \text{if } j \neq i. \end{cases} \tag{21}$$

Since $G_\lambda(\Delta, j) = 0$ and $G_\lambda(i,\Delta) = \lambda(i)$ for all $i \in E$ and $j \in E_\Delta$, we can rewrite (19) as

$$P_t^\lambda(i,j) = e^{G_\lambda t}(i,j) = e^{(G-\Lambda)t}(i,j) \tag{22}$$

for all $i,j \in E$.

Now, note that our construction of Y^λ implies

$$T_1 = \inf\{t \geq 0; Y_t^\lambda = \Delta\} \tag{23}$$

and T_1 is the first-passage-time to the absorbing state $\mathbf{\Delta}$. So, it has a phase-type distribution and, in particular,

$$P_i[T_1 > t] = P_i[Y_t^\lambda \in E] = \sum_{j \in E} P_t^\lambda(i,j) = \sum_{j \in E} e^{(G-\Lambda)t}(i,j). \tag{24}$$

Note that in reliability applications where a device fails exponentially with a failure rate that depends on the randomly changing environment, (24) gives the survival function. In this case, the mean time to failure is another quantity of interest. Using the Markov property, it can be computed by solving the system of linear equations

$$E_i[T_1] = \mu_\lambda(i)^{-1} + \sum_{j \in E} P_\lambda(i,j) E_j[T_1] \tag{25}$$

for $i \in E$ so that the explicit solution is

$$E_i[T_1] = \sum_{j \in E} \left[I - P_\lambda\right]^{-1}(i,j)\mu_\lambda(j)^{-1}. \tag{26}$$

Following [33] an ergodic analysis can be developed. Suppose that both X and Y are ergodic processes with limiting distributions $\nu(j) = \lim_{n \to +\infty} P[X_n = j]$ and $\pi(j) = \lim_{t \to +\infty} P[Y_t = j]$. This implies that ν is the unique solution of $\nu = \nu\mathbf{P}$ with the normalizing condition $\sum_{i \in E} \nu(i) = 1$. It can be shown that

$$\pi(j) = \frac{\nu(j)/\mu(j)}{\sum_{k\in E}(\nu(k)/\mu(k))}. \tag{27}$$

Using Theorem 1 of [33], we can obtain

$$\lim_{t\to+\infty} E_i[N_t - \hat{\lambda}t] = \sum_{j\in E}\pi(j)\left[\frac{\hat{\lambda}-\lambda(j)}{\mu(j)}\right] \tag{28}$$

and

$$\lim_{t\to+\infty}\frac{E_i[N_t]}{t} = \hat{\lambda} = \sum_{j\in E}\pi(j)\lambda(j). \tag{29}$$

It follows from Fischer and Meier-Hellstern [12], the expected number of arrivals until time t is

$$E_i[N_t] = \hat{\lambda}t +_{j\in E}\left(\left[e^{Gt}-I\right]\left[G+\Pi\right]^{-1}\right)(i,j)\lambda(j) \tag{30}$$

where $\Pi(i,j) = \pi(j)$.

3.2 Markov Modulated Bernoulli Process

We now consider discrete-time models for systems observed periodically at discrete time points. The system survives each period with a probability that depends on the state of the prevailing environment in that period. Since each period ends with a failure or survival, one can model this system as a Bernoulli process where the success probability is modulated by the environmental process. Using this setup with a Markovian environmental process, Özekici [27] focuses on probabilistic modeling and provides a complete transient and ergodic analysis. We suppose throughout the following discussion the sequence of environmental states $Y = \{Y_t; t = 1, 2, \cdots\}$ is a MC with some transition matrix $\mathbf{P}$ on a discrete state space E.

Consider a system observed periodically at times $t = 1, 2, \cdots$ and the state of the system at time t is described by a Bernoulli random variable

$$X_t = \begin{cases} 1, & \text{if system is not functioning at time } t \\ 0, & \text{if system is functioning at time } t. \end{cases}$$

Given that the environment is in some state i at time t, the probability of failure in the period is

$$P[X_t = 1|Y_t = i] = \pi(i) \tag{31}$$

for some $0 \le \pi(i) \le 1$. The states of the system at different points in time constitute a Bernoulli process $X = \{X_t; t = 1, 2, \cdots\}$ where the success probability is a function of the environmental process Y.

Given the environmental process Y, the random quantities $X_1, X_2, \cdots$ represent a conditionally independent sequence, that is,

$$P[X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n|Y] = \prod_{k=1}^{n} P[X_k = x_k|Y]. \tag{32}$$

In the above setup, the reliability of the system is modulated by the environmental process Y which is assumed to be a MP and thus the model is referred to as the Markov Modulated Bernoulli Process (MMBP). If the system fails in a period, then it is replaced immediately by an identical one at the beginning of the next period. It may be possible to think of the environmental process Y as a random mission process such that Y_t is the tth mission to be performed. The success and failure probabilities depend on the mission itself. If the device fails during a mission, then the next mission will be performed by a new and identical device.

If we denote the lifetime of the system by L, then the conditional life distribution is

$$P[L = m|Y] = \begin{cases} \pi(Y_1), & \text{if } m = 1 \\ \pi(Y_m)\prod_{j=1}^{m-1}(1-\pi(Y_j)) & \text{if } m \geq 2. \end{cases} \tag{33}$$

Note that if $\pi(i) = \pi$ for all $i \in E$, that is, the system reliability is independent of the environment, then (33) is simply the geometric distribution $P[L = m|Y] = \pi(1-\pi)^{m-1}$. We can also write

$$P[L > m|Y] = (1-\pi(Y_1))(1-\pi(Y_2))\cdots(1-\pi(Y_m)) \tag{34}$$

for $m \geq 1$.

We represent the initial state of the MC by Y_1, rather than Y_0, as it is customarily done in the literature, so that it represents the first environment that the system operates in. Thus, most of our analysis and results will be conditional on the initial state Y_1 of the Markov chain. Therefore, for any event A and random variable Z we set $P_i[A] = P[A|Y_1 = i]$ and $E_i[Z] = P[Z|Y_1 = i]$ to express the conditioning on the initial state.

As pointed out by [27], the life distribution satisfies the recursive expression

$$P_i[L > m+1] = (1-\pi(i))\sum_{j\in E} P(i,j)P_j[L > m] \tag{35}$$

with the obvious boundary condition $P_i[L > 0] = 1$. The survival probabilities can be explicitly computed via

$$P_i[L > m] = \sum_{j\in E} Q_0^m(i,j) \tag{36}$$

where $Q_0(i,j) = (1-\pi(i))P(i,j)$. Using (36), the conditional expected lifetime can be obtained as

$$E_i[L] = \sum_{m=0}^{+\infty}\sum_{j\in E} Q_0^m(i,j) = \sum_{j\in E} R_0(i,j) \tag{37}$$

where $R_0(i,j) = \sum_{m=0}^{+\infty} Q_0^m(i,j) = (I-Q_0)^{-1}(i,j)$ is the potential matrix corresponding to Q_0.

4 Application of HMMs in Software Reliability

The notion of an *operational profile* was introduced by Musa et al. [17]. An operation is an externally initiated task performed by a system "as built". The operational profile of any software describes how users employ the system. It is a quantitative and probabilistic characterization of how a system will be used. An operational profile is defined as a set of operations and the probabilities of their occurrence; see Musa [25].

Optimal testing problems involving operational profiles are discussed in detail in [29] and [34]. Once testing is completed, the software is released to the users. This is done in an uncontrolled setting and the sequence of operations as well as their durations are now random. This operational process or the environmental process now modulates the parameters of the reliability model and play a crucial role in software reliability assessment. Now, the environmental state Y_t at time t represents the operation performed by the user. The analysis of the software failure process obviously depend on the stochastic structure of the operational process.

In Özekici and Soyer [31], Y is assumed to be a MP. Briefly, this means that the sequence of operations performed is a MC and the amount of time spent on each operation is exponentially distributed. More precisely, we let X_n denote the nth operation that the system performs and T_n be the time at which the nth operation starts. Following our developmen X is a MC with some transition matrix

$$P(i, j) = P[X_{n+1} = j | X_n = i] \tag{38}$$

and

$$P[T_{n+1} - T_n > t | X_n = i] = e^{-\mu(i)t} \tag{39}$$

so that the duration of the nth operation is exponentially distributed with rate $\mu(i)$ if this operation is i. The probabilistic structure of the operational process is given by the generator $A(i, j) = \mu(i)(P(i, j) - I(i, j))$ where I is the identity matrix.

An overview of software failure models is presented in Singpurwalla and Soyer [43]. Perhaps the most important aspect of these models is related to the stochastic structure of the underlying failure process. This could be a "times-between-failures" model which assumes that the times between successive failures follow a specific distribution whose parameters depend on the number of faults remaining in the program after the most recent failure. One of the most celebrated failure models in this group is that of Jelinski and Moranda [18] where the basic assumption is that there are a fixed number of initial faults in the software and each fault causes failures according to a Poisson process with the same failure rate. After each failure, the fault causing the failure is detected and removed with certainty so that the total number of faults in the software is decreased by one. In the present setting, the time to failure distribution for each fault in the software is exponentially distributed with parameter $\lambda(k)$ during operation k and this results in an extension of the Jelinski-Moranda model.

In dealing with software reliability, one is interested in the number of faults N_t remaining in the software at time t. Then, N_0 is the initial number of faults and the process $N = \{N_t; t \geq 0\}$ depicts the stochastic evolution of the number of faults. If there is perfect debugging, then N decreases as time goes on, eventually to diminish to zero. Defining the bivariate process $Z_t = (Y_t, N_t)$, it follows that $Z = (Y, N)$ is a MP with discrete state space $F = E \times \{0, 1, 2, \cdots\}$. As pointed out in [43], this follows by noting that Y is a MP and N is a process that decreases by 1 after an exponential amount of time with a rate that depends only on the state of Y. In particular, if the current state of Z is (i, n) for any $n > 0$, then the next state is either (j, n) with rate $\mu(i)P(i, j)$ or $(i, n-1)$ with rate $n\lambda(i)$. If $n = 0$, then the next state is $(j, 0)$ with rate $\mu(j)$. Note that 0 is an absorbing state for N.

This implies that the sojourn in state (i, n) is exponentially distributed with rate

$$\beta(i, n) = \mu(i) + n\lambda(i) \tag{40}$$

and the generator Q of Z is

$$Q((i, n), (j, m)) = \begin{cases} -(\mu(i) + n\lambda(i)), & j = i, m = n \\ \mu(i)P(i, j), & j \neq i, m = n \\ n\lambda(i), & j = i, m = n - 1 \end{cases}. \tag{41}$$

Reliability is defined as the probability of failure free operation for a specified time. We will denote this by the function

$$R(i, n, t) = P[L > t | Y_0 = i, N_0 = n] = P[N_t = n | Y_0 = i, N_0 = n] \tag{42}$$

defined for all $(i, n) \in F$ and $t \geq 0$. Note that this is equal to the probability that there will be no arrivals until time t in a Markov modulated Poisson process with intensity function $\hat{\lambda}(t) = n\lambda(Y_t)$. Thus, following Fischer and Meier-Hellstern [12], we can obtain the explicit formula

$$R(i, n, t) = \sum_{j \in E} \left[e^{(A - n\Lambda)t} \right]_{ij} \tag{43}$$

where

$$e^{(A - n\Lambda)t} = \sum_{k=0}^{+\infty} \frac{t^k}{k!} (A - n\Lambda)^k \tag{44}$$

is the exponential matrix and $\Lambda(i, j) = \lambda(i)I(i, j)$.

References

[1] M. Abdel-Hameed, *A gamma wear process*, IEEE Transactions on Reliability **24** (1975), 152–153.

[2] M. Berman, *Inhomogeneous and modulated gamma processes*, Biometrica **68** (1981), 143–152.

[3] O. Cappe, E. Moulines and T. Ryden, *Inference in hidden Markov models*, Springer, New York, 2005.

[4] G. Celeux, C. P. Robert and J. Diebolt, *Bayesian estimation of hidden Markov chains: A stochastic implementation*, Statistics and Probability Letters **16** (1993), 77–83.

[5] E. Çınlar, *Markov renewal theory*, Advances in Applied Probability **1** (1969), 123–187.

[6] E. Çınlar, *Markov additive processes: I*, Z. Wahrscheinlichkeitstheorie verw. Geb. **24** (1972), 85–93.

[7] E. Çınlar, *Markov additive processes: II*, Z. Wahrscheinlichkeitstheorie verw. Geb. **24** (1972), 95–121.

[8] E. Çınlar, *Introduction to stochastic processes*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

[9] E. Çınlar and S. Özekici, *Reliability of complex devices in random environments*, Probability in the Engineering and Informational Sciences **1** (1987), 97–115.

[10] E. Çınlar, M. Shaked and J. G. Shanthikumar, *On lifetimes influenced by a common environment*, Stochastic Processes and their Applications **33** (1989), 347–359.

[11] J. B. Durand and O. Gaudoin, *Software reliability modelling and prediction with hidden Markov chains*, Statistical Modelling **5** (2006), 75–93.

[12] W. Fischer and K. Meier-Hellstern, *The Markov-modulated Poisson process cookbook*, Performance Evaluation **18** (1992), 149–171.

[13] D. P. Gaver, *Random hazard in reliability problems*, Technometrics **5** (1963), 211–226.

[14] P. Guttorp, D. Golinelli and J. Abkowitz, *Bayesian inference in hidden stochastic two-compartment model for feline hematopoiesis*, Mathematical Medicine and Biology **23** (2006), 153–172.

[15] M. Hock and R. Soyer, *A Bayesian approach to signal analysis of pulse trains*, Bayesian Monitoring, Control and Optimization (New York) (B. M. Colosimo and E. del Castillo, eds.), CRC, 2006, pp. 215–243.

[16] J. P. Hughes, E. Bellone and P. Guttorp, *A hidden Markov model for relating synoptic scale patterns to precipitation amounts*, Climate Research **15** (2000), 1–12.

[17] A. Iannino, J. D. Musa and K. Okumoto, *Software reliability: Measurement, prediction, application*, McGraw-Hill, New York, NY, 1987.

[18] Z. Jelinski and P. Moranda, *Software reliability research*, Statistical Computer Performance Evaluation (W. Freiberg, ed.), Academic Press, New York, NY, 1972, pp. 465–484.

[19] S. Karlin and H. M. Taylor, *A first course in stochastic processes*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

[20] J. F. C. Kingman, *On doubly stochastic Poisson processes*, Proceedings of the Cambridge Philosophical Society **60** (1964), 923–960.

[21] M. Klein, *Inspection-maintenance-replacement schedules under markovian deterioration*, Management Science **9** (1962), 25–32.

[22] N. Limnios, *Semi-Markov processes and hidden models*, Wiley Encyclopedia of Operations Research and Management Science (New York) (J. J. Cochran, ed.), Wiley, 2010, pp. xx–xx.

[23] D. V. Lindley and N. D. Singpurwalla, *Multivariate distributions for the lifelengths of components of a system sharing a common environment*, Journal of Applied Probability **23** (1986), 418–431.

[24] A. Mercer, *Some simple wear-dependent renewal processes*, Journal of the Royal Statistical Society, Ser. B **23** (1961), 368–376.

[25] J. D. Musa, *Operational profiles in software reliability engineering*, IEEE Software **10** (1993), 14–32.

[26] S. Özekici, *Optimal maintenance policies in random environments*, European Journal of Operational Research **82** (1995), 283–294.

[27] S. Özekici, *Markov modulated Bernoulli process*, Mathematical Methods of Operations Research **45** (1997), 311–324.

[28] S. Özekici, *Stochastic hazard process*, Wiley Encyclopedia of Operations Research and Management Science (New York) (J. J. Cochran, ed.), Wiley, 2010, pp. xx–xx.

[29] S. Özekici and R. Soyer, *Bayesian testing strategies for software with an operational profile*, Naval Research Logistics **48** (2001), 747 – 763.

[30] S. Özekici and R. Soyer, *Bayesian analysis of Markov modulated Bernoulli processes*, Mathematical Methods of Operations Research **57** (2003), 125–140.

[31] S. Özekici and R. Soyer, *Reliability of software with an operational profile*, European Journal of Operational Research **149** (2003), 459–474.

[32] S. Özekici and R. Soyer, *Reliability modeling and analysis under random environments*, Mathematical Reliability: An Expository Perspective (Boston) (R. Soyer, T. A. Mazzuchi and N. D. Singpurwalla, eds.), Kluwer, 2004, pp. 249–273.

[33] S. Özekici and R. Soyer, *Semi-Markov modulated Poisson process: Probabilistic and statistical analysis*, Mathematical Methods of Operations Research **64** (2006), 125–144.

[34] S. Özekici, I. K. Altınel and S. Özçelikyürek, *Testing of software with an operational profile*, Naval Research Logistics **47** (2000), 620–634.

[35] C. Park and W. J. Padgett, *Accelerated degradation models for failure based on geometric Brownian motion and gamma processes*, Lifetime Data Analysis **11** (2005), 511–527.

[36] L. R. Rabiner and B-H. Juang, *Mixture autoregressive hidden Markov models for speech signals*, IEEE Transactions on Acoustics, Speech and Signal Processing **33** (1985), 1404–1412.

[37] S. M. Ross, *Stochastic processes*, Wiley, New York, 1996.

[38] F. Ruggeri and R. Soyer, *Advances in Bayesian software reliability modeling*, Advances in Mathematical Modeling for Reliability (Amsterdam) (T. Bedford, ed.), IOS Press, 2008, pp. 165–176.

[39] T. Rydn, *Parameter estimation for Markov modulated Poisson processes*, Stochastic Models **10** (1994), 795–829.

[40] S. L. Scott, *Bayesian methods for hidden Markov models*, Journal of the American Statistical Association **97** (2002), 337–351.

[41] N. D. Singpurwalla, T. A. Mazzuchi,S. Özekici and R. Soyer, *Stochastic process models for reliability in dynamic environments*, Handbook of Statistics: Statistics in Industry (Amsterdam) (C. R. Rao and R. Khattree, eds.), Elsevier, 2003, Vol. 22, pp. 1109– 1129.

[42] N. D. Singpurwalla, *Survival in dynamic environments*, Statistical Science **10** (1995), 86–103.

[43] N. D. Singpurwalla and R. Soyer, *Assessing the reliability of software: an overview*, Reliability and Maintenance of Complex Systems (Berlin) (S. Özekici, ed.), Springer-Verlag, 1996, NATO ASI Series Volume F154, pp. 345–367.

[44] N. D. Singpurwalla and M. A. Youngren, *Multivariate distributions induced by dynamic environments*, Scandinavian Journal of Statistics **20** (1993), 251–261.

[45] J. M. van Noortwijk, *A survey of the application of gamma processes in maintenance*, Reliability Engineering and System Safety **94** (2009), 2–21.