Journal of Mathematical Psychology 59 (2014) 65–81

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Reformulating Markovian processes for learning and memory from a hazard function framework

Richard A. Chechile*, Lara N. Sloboda

Tufts University, United States

HIGHLIGHTS

- Link hazard functions to the probabilities in a Markov model.
- Show that a Markov chain is linked to a Weibull model with a shape parameter $c = 1$.
- Show that a general Markovian model rather than a Markov chain is needed.
- Develop a Markovian model for the Chechile et al. (2012) IES model.

ARTICLE INFO

Article history:
Available online 17 October 2013

Keywords:
Markov models
Hazard functions
Learning models
Forgetting functions
Explicit versus implicit memory

ABSTRACT

With the development of stimulus sampling theory (SST), William K. Estes demonstrated the importance of Markov chains for capturing many important features of learning. In this paper, learning and memory retention are reexamined from a hazard function framework and linked to the stochastic transition matrices of a Markov model. The probabilities in the transition matrix are shown to be discrete hazard values. In order for the stochastic matrix to be a homogeneous Markov chain, there is a requirement that the transition matrix values remain constant. Yet for some learning and memory retention applications, there is evidence that the transition matrix probabilities are dynamically changing. For list learning, the change in hazard is attributed in part to differences in the learning rate of individual items within the list. Even on an individual basis, any variability in item difficulty whatsoever is enough to induce a change in hazard with training. Another analysis was done to delineate the hazard function for memory loss. Evidence is again provided that the hazard associated with the loss of memory is systematically changing. A Markov chain is not a suitable model when there are dynamic changes in the hazard. However, for both the learning and memory applications, a general Markovian model can be used, where transition probabilities are a function of trial number or interpolated event number. Finally, a more complex, four-state application is considered. This application is based on the Chechile, Sloboda, and Chamberland (2012) multinomial processing tree model called the IES model. The IES model obtains probability estimates for the representation of target information in memory in terms of four possible states—explicit memory, implicit memory, fractional memory, and non-storage. Stochastic matrices for the IES model are provided and are shown to yield new insights about implicit memory.

Preface to the paper

As one of the originators of Mathematical Psychology, William K. Estes had a keen interest in the careers of scientists who worked in this field. For many in mathematical psychology, Bill was an intellectual father figure or grandfather figure. He read and commented on our papers. He faithfully listened to our conference presentations. He encouraged us to keep asking good questions. Also for 20 years, Bill was a close neighbor to the first author of this paper; Harvard University and Tufts University are about three miles apart. During that period there were numerous visits to share ideas. This paper is a belated outgrowth from several conversations with Bill about how hazard functions are a powerful alternative framework for understanding stochastic processes and from that perspective there were some new insights that can be gleaned about some classic problems. In this paper, hazard functions are used to reexamine Markovian processes for learning and memory.

1. Introduction

The utilization of stochastic models to study learning and memory should be ranked very high among the many technical and

* Correspondence to: Psychology Department, Tufts University Medford, MA 02155, United States.
E-mail address: Richard.Chechile@Tufts.edu (R.A. Chechile).

0022-2496/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmp.2013.09.004

conceptual contributions of William K. Estes. Stimulus sampling theory (SST) was one of the earliest models in this area (Atkinson & Estes, 1963; Estes, 1959a,b; Estes & Burke, 1953; and Neimark & Estes, 1967). In the Estes and Suppes (1974) paper the axiomatic foundations for SST were delineated. Some of those axioms made explicit the properties that link SST to finite state Markov chain models, i.e., "... we turn to the proof of what is probably the most important general theorem of stimulus sampling theory, namely, that under very broad conditions an appropriately chosen sequence of events is a Markov chain" (Estes & Suppes, 1974, p. 174.). Of course general Markov processes and Markov chains were already well established topics in probability theory prior to their use in psychology. The topic of Markov processes was developed in the first decade of the 20th century, and Markov models were prominently featured in the Feller (1950) textbook that had a strong influence on the early work in mathematical psychology. Markov models remain a fundamental technique in the conceptual toolbox of mathematical psychologists (Brainerd, 1985). Markov models are discussed in virtually all the mathematical psychology textbooks (viz., Busemeyer & Diederich, 2010; Coombs, Dawes, & Tversky, 1970; Laming, 1973; Levine & Burke, 1972; Restle & Greeno, 1970; Roberts, 1976; Wickens, 1982).[1]

The concept of hazard functions had a later introduction in the mathematical and psychological literatures. Hazard functions were originally developed in actuarial statistics (Steffensen, 1930), but it took some time before hazard functions were discussed in general probability theory. For example, hazard functions were not discussed in either volume of the classic textbooks on probability theory by Feller (1950, 1966). Consequently, there was some delay before hazard functions were used in psychology. Townsend and Ashby (1978) first used hazard functions in psychology, but their application was in the context of information processing models for response time. Although the interest in hazard functions grew in the response time literature, the use of hazard function for the study of learning and memory developed slowly (Chechile, 1987, 2006). Consequently, hazard functions were not well known tools at the time of the development of SST. There are now many established facts about hazard functions (Chechile, 2003, 2006: Luce, 1986; and Townsend & Ashby, 1983). Hazard functions also have a connection to Markov models, and this connection is explored in the present paper. In general Markov chain models impose some constraints on the hazard function, and conversely some hazard functions are not consistent with a Markov chain model.

In the next six sections of this paper, learning and memory will be reexamined in a way that links the two different mathematical perspectives, i.e. Markovian processes and hazard functions. In Section 2 the basic framework for Markov chains and general Markov models is developed with a particular focus on applications for theories of learning and memory retention. In Section 3 a framework is provided for reexamining learning data in terms of a hazard function perspective. Also in that section, a general form of a Markovian process is reexpressed in the form of hazard functions. In learning applications, hazard is associated with the conditional likelihood of establishing a new association in memory on a particular training trial. However, time and subsequent events can damage previously established memories. Hazard in the context of forgetting deals with the conditional likelihood of a memory loss at a particular point in time. The topic of hazard and memory retention is explored in Section 4. A case is made, based on a theoretical and empirical utilization of a surrogate function, that the hazard associated with memory loss has a peaked shape. A surrogate function is a function that can be used to extract information about the empirical properties of the hazard function without directly fitting a hazard function. In Section 5 the Chechile (2006) two trace hazard (TTH) model is advanced as the appropriate description of the hazard changes caused by interference in the retention interval. This theory is contrasted with other memory models, including the Estes (1997) dual-trace model. Although Estes did not advance a Markov model for general memory loss, a Markovian model for memory loss can be nonetheless formulated. In Section 5 a Markovian version of the TTH model is developed, and in Section 6 a Markovian model is developed for the recent Implicit–Explicit Separation (IES) model by Chechile et al. (2012). The IES model is a multinomial processing tree (MPT) model that enables the estimation of four basic states for representing the knowledge that the subject has in memory associated with a target event. The four types of information storage are explicit, implicit, fractional, and non-storage. The Markovian framework for the IES model leads to a new interpretation of the relationship between explicit and implicit memory. Finally in Section 7 some concluding observations are discussed.

2. Markovian models and hazard models

There are four subsections contained in this section. The first subsection defines key Markovian concepts and illustrates these ideas with four specialized models for either learning or memory retention applications. These models will be discussed in later sections of the paper. The second subsection covers key ideas from the hazard function literature and develops a general discrete Weibull model for learning applications. The third subsection deals with linking a general Markovian learning model to hazard values. In a similar fashion, the fourth subsection links a general Markovian forgetting model to hazard functions.

2.1. The Markovian framework

The context for a Markov process is in terms of a state space $\mathcal{S}$ and a discrete, ordered, support domain. The state space can be quite general. For some applications, like that of a one-dimensional random walk, the states correspond to all the discrete outcomes between the two end points. In psychological random walk applications, only the states for the two ends are observable and the other states are not directly observable. However, for other applications, there might only be two states for the Markov process, with both states representing extremes. For example, in a memory retention application, one state could be a target storage state and another state could be for the loss of target information. As for the support for a Markov process, there are also several possibilities. For many of our applications, the support domain is in terms of discrete increments of time, i.e. $t_1 < t_2 < t_3 < \cdots$ where $t_n = n\tau$, $\tau > 0$. For other applications, the support is in terms of the number of study trials, $i = 1, 2, \ldots$. Although the support set might be countably infinite for a general Markov process, we will treat the support in this paper as finite because there are only a finite number of discrete steps that occur over the life of the subject. In general, let us denote the support domain as $\mathcal{J} = 1, 2, \ldots$. The framework for a Markov process is one where the stochastic system advances, one step at a time, over the support domain, but at each step, the stochastic system is in one of m denumerable states, i.e. $\mathcal{S} = \{S_1, S_2, \ldots, S_m\}$. The general Markovian property is when the transition of the state of the system from step n to step $n+1$ is a probabilistic process with the transition probability only being dependent on the current state of the system. Let $p_{ij}^{(n)}$ denote the transition probability to state S_i on step $n+1$ given that the system was in state S_j for step n.

[1] There are other classic books of mathematical psychology that discuss random walk models, which are a form of a Markov chain (e.g., Luce, 1986; Townsend & Ashby, 1983).