Theoretical Computer Science 410 (2009) 4067–4084

Contents lists available at ScienceDirect

Theoretical Computer Science

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

An analysis of the exponential decay principle in probabilistic trust models

Ehab ElSalamouny [a], Karl Tikjøb Krukow [b], Vladimiro Sassone [a,*]

[a] ECS, University of Southampton, UK
[b] Trifork, Aarhus, Denmark

ARTICLE INFO

This paper is dedicated to Professor Mogens Nielsen on the occasion of his 60th birthday, who along the years shared with us research and much more. *To your sharpness of mind and kindness of spirit.*

Keywords:
Trust
Computational trust
Probabilistic trust models
Hidden Markov Models
Decay principle
Beta distribution

ABSTRACT

Research in models for experience-based trust management has either ignored the problem of modelling and reasoning about dynamically changing principal behaviour, or provided ad hoc solutions to it. Probability theory provides a foundation for addressing this and many other issues in a rigorous and mathematically sound manner. Using Hidden Markov Models to represent principal behaviours, we focus on computational trust frameworks based on the 'beta' probability distribution and the principle of exponential decay, and derive a precise analytical formula for the estimation error they induce. This allows potential adopters of beta-based computational trust frameworks and algorithms to better understand the implications of their choice.

1. Introduction

This paper concerns *experience-based* trust management systems. The term 'trust management' is usually associated with the traditional *credential-based* trust management systems in which trust is established primarily as a function of available credentials (see e.g. [2]). In experience-based systems, trust in a principal is represented as a function of *information* about past principal behaviour. This encompasses also reputation information, i.e., information about principal behaviour obtained not by direct observation but from other sources (e.g., ratings made by other principals). There are several approaches to experience-based trust management; this paper is concerned with the *probabilistic* approach, which can broadly be characterised as aiming to build probabilistic models upon which to base predictions about principals' future behaviours. Due to space limitations, we shall assume the reader to be familiar with experience-based trust management (the interested reader will find a comprehensive overview in [11]) and with the basics of probability theory.

Many systems for probabilistic trust management assume, sometimes implicitly, the following scenario. There is a collection of principals $(p_i \mid i \in I)$, for some finite index set I, which at various points in time can choose to interact in a pair-wise manner; each interaction can result in one of a predefined set of outcomes, $O = \{o_1, \ldots, o_m\}$. For simplicity, and without loss of generality, in this exposition we shall limit the outcomes to two: s (for success) and f (for failure). Typically, outcomes are determined by behaviours: when principal p_i interacts with principal p_j, the behaviour of p_j relative to the protocol used for interaction defines the outcome. Specifically, compliant behaviours represent successful interactions, whilst behaviour which diverge from the interaction protocol determine failure. Hence, the most important component in the framework is the *behaviour model*. In many existing frameworks the so-called *Beta model* [10] is chosen. According

* Corresponding author.
E-mail address: vs@ecs.soton.ac.uk (V. Sassone).

0304-3975/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2009.06.011

to the Beta model, each principal p_j is associated with a fixed real number $0 \leq \Theta_j \leq 1$, to indicate the assumption that an interaction involving p_j will yield success with probability Θ_j. This is a *static* model in the precise sense that the behaviour of principal p_j is assumed to be representable by a fixed probability distribution over outcomes, invariantly in time. This simple model gives rise to trust computation algorithms that attempt to 'guess' p_j's behaviour by approximating the unknown parameter Θ_j from the history of interactions with p_j (cf., e.g., [17]).

There are several examples in the literature where the Beta model is used, either implicitly or explicitly, including Jøsang and Ismail's Beta reputation system [10], the systems of Mui et al. [13] and of Buchegger [4], the Dirichlet reputation systems [9], TRAVOS [18], and the SECURE trust model [5]. Recently, in a line of research largely inspired by Mogens Nielsen's pioneering ideas, the Beta model and its extension to interactions with multiple outcomes have been used to provide a first formal framework for the analysis and comparison of computational trust algorithms [17,14,12]. In practice, these systems have found space in different applications of trust, e.g., online auctioning, peer-to-peer filesharing, mobile ad hoc routing and online multiplayer gaming.

All the existing systems use the family of beta probability density functions (pdfs) or some generalisation thereof, as e.g. the Dirichlet family — again for the sake of simplicity and with no loss of generality, in this paper we shall confine ourselves to beta. The choice of beta is a reasonable one, as a history of interactions h with principal p_j can be summarised compactly by a beta function with parameters α and β, in symbols $\mathbf{B}(\alpha, \beta)$, where $\alpha = \#_{\mathrm{s}}(h) + 1$ (resp. $\beta = \#_{\mathrm{f}}(h) + 1$) is the number of successful (resp. unsuccessful) interactions in h plus one. This claim can be made mathematically precise in the language of Bayesian Theory, as the family of beta distributions is a *conjugate prior* to the family of Bernoulli trials (cf. Section 2 and [17] for a full explanation). An important consequence of this representation is that it allows us to estimate the so-called *predictive probability*, i.e., the probability of a success in the next interaction with p_j given history h. Such an estimate is given by the expected value of Θ_j according to $\mathbf{B}(\alpha, \beta)$[1]:

$$P(\mathrm{s} \mid h\,\mathbf{B}) = \mathbf{E}_{\mathbf{B}(\alpha,\beta)}(\Theta_j) = \frac{\alpha}{\alpha + \beta}.$$

So in this simple and popular model, the predictive probability depends only on the number of past successful interactions and the number of past failures.

Many researchers have recognised that the assumption of a fixed distribution to represent principals is a serious limitation for the Beta model [10,4,19]. Just consider, e.g., the example of an agent which can autonomously switch between two internal states, a normal 'on-service' mode and a 'do-not-disturb' mode. This is why several papers have used a '*decay*' principle to favour recent events over information about older ones. The decay principle can be implemented in many different ways, e.g., by a using a finite 'buffer' to remember only the most recent n events, or linear and exponential decay functions, which scale according the parameters α and β of the beta pdf associated with a principal. This paper will focus on exponential decay.

Whilst decay-based techniques have proved useful in some applications, to our knowledge, there is as yet no formal understanding of which applications may benefit from them. Indeed, the lack of foundational understanding and theoretical justification leaves application developers alone when confronting the vexing question of which technique to deploy in their applications. In the recent past, two of the present authors in joint work with Mogens Nielsen have proposed that rather than attempting to lift too simplistic assumptions (viz., the Beta model) with apparently effective, yet ad hoc solutions (viz., decay), one should develop models sufficiently sophisticated to encompass all the required advanced features (e.g., dynamic principal behaviours), and then derive analysis and algorithms from such models. In this way, one would provide a foundational framework suitable formulate and compare different analyses and algorithms and, therefore, to underpin their deployment in real-world applications. It is our contention that such an encompassing model is afforded by *Hidden Markov Models* [1] (an enjoyable tutorial is [16]). In the present work we elect to use HMMs as a reference model for probabilistic, stateful behaviour.

Original contribution of the paper. Aiming to address the issue of whether and when exponential decay may be an optimal technique for reasoning about dynamic behaviour in computational trust, we use a simple probabilistic analysis to derive some of its properties. In particular, we study the error induced on the predictive probability by using the Beta model enhanced with exponential decay in comparison with (the so-to-say 'ideal') Hidden Markov Models (HMMs). Under mild conditions on principals' behaviours, namely that their probabilistic state transition matrices are *ergodic*, we derive an analytic formula for the error, which provides a first formal tool to assess precision and usefulness of the decay technique. Also, we illustrate our results by deploying it in an experimental setting to show how system *stability* has a dramatic impact on the precision of the model.

Structure of the paper. We first recall the basic ideas of Bayesian analysis, beta distributions, decay and HMMs, in Sections 2–4; Section 3 also proves some simple, yet interesting properties of exponential decay. Then, Section 5 contains the derivation of our error formula, whilst Section 6 analyses exponential decay in terms of an original notion of system stability.

1 In our probability notation, juxtaposition means logical conjunction, as in Jaynes [8] whose notations we follow. So, $P(\mathrm{s} \mid h\,\mathbf{B})$ reads as 'the probability of success conditional to history h and the assumptions of the Beta model $\mathbf{B}$.'

2. Bayesian analysis and beta distributions

Bayesian analysis consists of formulating hypotheses on real-world phenomena of interest, running experiments to test such hypotheses, and thereafter updating the hypotheses – if necessary – to provide a better explanation of the experimental results, a better fit of the hypotheses to the observed behaviours. In terms of conditional probabilities on the space of interest and under underlying assumptions $\boldsymbol{\lambda}$, this procedure is expressed succinctly by Bayes' Theorem:

$$P(\boldsymbol{\Theta} \mid \boldsymbol{h}\,\boldsymbol{\lambda}) \propto P(\boldsymbol{h} \mid \boldsymbol{\Theta}\,\boldsymbol{\lambda}) \cdot P(\boldsymbol{\Theta} \mid \boldsymbol{\lambda}).$$

Reading from left to right, the formula is interpreted as saying: the probability of the hypotheses $\boldsymbol{\Theta}$ *posterior* to the outcome of experiment $\boldsymbol{h}$ is *proportional* to the *likelihood* of such an outcome under the hypotheses multiplied by the probability of the hypotheses *prior* to the experiment. In the context of computational trust described in the Introduction, the prior $\boldsymbol{\Theta}$ will be an estimate of the probability of each potential outcome in our next interaction with principal p, whilst the posterior will be our amended estimate after some such interactions took place with outcome $\boldsymbol{h}$.

In the case of binary outcomes $\{\mathsf{s}, \mathsf{f}\}$ discussed above, $\boldsymbol{\Theta}$ can be represented by a single probability Θ_p, the probability that an interaction with principal p will be successful. In this case, a sequence of n experiments $\boldsymbol{h} = o_1 \cdots o_n$ is a sequence of binomial (Bernoulli) trials, and is modelled by a binomial distribution

$$P(\boldsymbol{h} \text{ consists of exactly } k \text{ successes}) = \binom{n}{k} \Theta_p^k (1 - \Theta_p)^{n-k}.$$

It then turns out that if the prior $\boldsymbol{\Theta}$ follows a beta distribution, say

$$\mathbf{B}(\alpha, \beta) \propto \Theta_p^{\alpha-1} (1 - \Theta_p)^{\beta-1}$$

of parameters α and β, then so does the posterior: viz., if $\boldsymbol{h}$ is an n-sequence of exactly k successes, $P(\boldsymbol{\Theta} \mid \boldsymbol{h}\,\boldsymbol{\lambda})$ is $\mathbf{B}(\alpha + k, \beta + n - k)$, the beta distribution of parameters $\alpha + k$ and $\beta + n - k$. This is a particularly happy event when it comes to applying Bayes' Theorem, because it makes it straightforward to compute the posterior distribution and its expected value from the prior and the observations. In fact, the focus here is not to compute $P(\boldsymbol{\Theta} \mid \boldsymbol{h}\,\boldsymbol{\lambda})$ for any particular value of $\boldsymbol{\Theta}$ but, as $\boldsymbol{\Theta}$ is the unknown in our problem, rather to derive a symbolic knowledge of the entire distribution in order to compute its expected value and use it as our next estimate for $\boldsymbol{\Theta}$. A relationship as the one between binomial trials and the beta distributions is very useful in this respect; indeed it is widely known and studied in the literature as the condition that the family of beta distributions is a *conjugate prior* for the binomial trials.

The Bayesian approach has proved as successful in computational trust as in any of its several applications (cf., e.g., [8]), yet it is fundamentally based on the assumption that a principal p can be represented precisely enough by a single, immutable Θ_p. As the latter is patently an unacceptable limitation in several real-world applications, in the next section we will illustrate a simple idea to address it.

3. The exponential decay principle

One purpose of the *exponential decay* principle is to improve the responsiveness of the Beta model to principals exhibiting dynamic behaviours. The idea is to scale by a constant $0 < r < 1$ the information about past behaviour, viz., $\#_o(h) \mapsto r \cdot \#_o(h)$, each time a new observation is made. This yields an exponential decay of the weight of past observations, as in fact the contribution of an event n steps in the past will be scaled down by a factor r^n. Qualitatively, this means that picking a reasonably small r will make the model respond quickly to behavioural changes. Suppose for instance that a sequence of five positive and no negative events has occurred. The unmodified Beta model would yield a beta pdf with parameters $\alpha = 6$ and $\beta = 1$, predict the next event to be positive with probability higher than 0.85. In contrast, choosing $r = 0.5$, the Beta model with exponential decay would set $\alpha = 1 + 31/16$ and $\beta = 1$. This assigns probability 0.75 to the event that the next interaction is positive, as a reflection of the fact that some of the weight of early positive events has significantly decayed. Suppose however that a single negative event occurs next. Then, in the unmodified Beta model the parameters are updated to $\alpha = 6$ and $\beta = 2$, which still assign a probability 0.75 to 'positive' events, reflecting the relative unresponsiveness of the model to change. On the contrary, the model with decay assigns $63/32$ to α and 2 to β, which yields a probability just above 0.5 that the next event is *again* negative. So despite having observed five positive events and a negative one, the model with decay yields an approximately uniform distribution, i.e., it considers positive and negative almost equally likely in the next interaction.

Of course, this may or may not be appropriate depending on the application and the hypotheses made on principals behaviours. If on the one hand the specifics of the application are such to suggest that principals do indeed behave according to a single, immutable probability Θ, then discounting the past is clearly not the right thing to do. If otherwise one assumes that principals may behave according to different Θs as they switch their internal state, then exponential decay for a suitable r may make prediction more accurate. Our assumption in this paper is precisely the latter, and our main objective is to analyse properties and qualities of the Beta model with exponential decay in dynamic applications, by contrasting it with Hidden Markov Model, the '*par excellence*' stochastic model which includes state at the outset.

We conclude this section by observing formally that with exponential decay, strong certainty can be impossible. To this end, we study below the expression $\alpha + \beta$, as that is related to the width of $\mathbf{B}(\alpha, \beta)$ and, in turn, to the variance of the distribution, and represents in a precise sense the confidence one can put in the predictions afforded by the pdf. Consider e.g. the case of sequences of n positive events, for increasing $n \geq 0$. For each n, the pdf obtained has parameters $\alpha_n = 1 + \sum_{i=0}^{n} r^i$ and $\beta_n = 1 + r^n$. Since $0 < r < 1$, the sequence converges to the limit $\alpha = 1 + (1-r)^{-1}$ and $\beta = 1$. In fact, we can prove the following general proposition.

Proposition 1. *Assume exponential decay with factor* $0 < r < 1$ *starting from the prior beta distribution with parameters* $\alpha = 1$ *and* $\beta = 1$. *Then, for all* $n \geq 0$,

$$\begin{aligned} 2 + \frac{1}{1-r} \leq \alpha_n + \beta_n + 2r^n \leq 4 \quad & \textit{if } r \leq \frac{1}{2} \\ 4 \leq \alpha_n + \beta_n + 2r^n \leq 2 + \frac{1}{1-r} \quad & \textit{if } r \geq \frac{1}{2}. \end{aligned}$$

Proof. The proof is a straightforward induction, which we exemplify in the case $r \geq \frac{1}{2}$. The base case is obvious, as $\alpha_0 + \beta_0 = 2$. Assume inductively that the proposition holds for n, and observe first that $\alpha_{n+1} + \beta_{n+1} = 2 + 1 + r(\alpha_n + \beta_n - 2)$. Then

$$\alpha_{n+1} + \beta_{n+1} + 2r^{n+1} = 2 + 1 + r\,(\alpha_n + \beta_n - 2 + 2r^n) \leq 2 + 1 + r\left(2 + \frac{1}{1-r} - 2\right) = 2 + \frac{1}{1-r}.$$

Similarly, as $r \cdot (\alpha_n + \beta_n + 2r^n - 2) \geq r \cdot 2 \geq 1$, it follows that $\alpha_{n+1} + \beta_{n+1} + 2r^{n+1} \geq 4$, which ends the proof. □

Proposition 1 gives a bound on $\alpha_n + \beta_n$ as a function on r, which in fact is a bound on the Bayesian inference via beta function. For instance, it means that a principal using exponential decay with $r = 1/2$ can never achieve higher confidence in the derived pdf than they initially had in the uniform distribution. In order to assess the speed of convergence of α and β, let us define $D_n = \left|\alpha_n + \beta_n + 2r^n - 2 - \frac{1}{1-r}\right|$. We then have the following.

Proposition 2. *Assume exponential decay with factor* $0 < r < 1$ *starting from the prior beta distribution with parameters* $\alpha = 1$ *and* $\beta = 1$. *For every* $n \geq 0$, *we have*

$$D_{n+1} = r \cdot D_n$$

Proof. Let $n \geq 0$. Then

$$\begin{aligned} D_{n+1} &= \left| 2 + 1 + r\,(\alpha_n + \beta_n - 2) + 2r^{n+1} - 2 - \frac{1}{1-r} \right| \\ &= r \cdot \left| \alpha_n + \beta_n + 2r^n - 2 - \frac{1}{1-r} \right| = r \cdot D_n. \quad \square \end{aligned}$$

One may of course argue that the choice of r is the key to bypass issues like these. In any case, the current literature provides no underpinning for assessing what a sensible value for r should be. Buchegger et al. [4] suggest that $r = 1 - 1/m$, where m is "*... is the order of magnitude of the number of observations over which we believe it makes sense to assume stationary behaviour*;" however, no techniques are indicated to estimate such a number. Similarly, Jøsang and Ismail [10] present simulations with $r = 0.9$ and $r = 0.7$. Interestingly, even at $r = 0.9$, the sum $\alpha_n + \beta_n$ is uniformly bound by ten; this means that the model yields at best a degree of certainty in its estimates comparable to that obtained with the unmodified Beta model after only eight observations. Again, whether or not this is the appropriate behaviour for a model depends entirely on the application at hand.

4. Hidden Markov Models

In order to assess merits and problems of the exponential decay model, and in which scenario it may or may not be considered suitable, we clearly need a probabilistic model which captures the notion of dynamic behaviour primitively, to function so-to-say as an unconstraining testbed for comparisons.

In ongoing joint work with Mogens Nielsen, we are investigating Hidden Markov Models (HMM) as a general model for stateful computational trust [12,14,17]. These are a well-established probabilistic model essentially based on a notion of system state. Indeed, underlying each HMM there is a Markov chain modelling (probabilistically) the system's state transitions. HMMs provide the computational trust community with several obvious advantages: they are widely used in scientific applications, and come equipped with efficient algorithms for computing the probabilities of events and for parameter estimation (cf. [16]), the chief problem for probabilistic trust management.

Definition 1 (*Hidden Markov Model*). A (discrete) *hidden Markov model* (HMM) is a tuple $\lambda = (Q, \pi, A, O, s)$ where Q is a finite set of *states*; π is a distribution on Q, the *initial distribution*; $A : Q \times Q \to [0, 1]$ is the *transition matrix*, with $\sum_{j \in Q} A_{ij} = 1$; finite set O is the set of possible *observations*; and where $s : Q \times O \to [0, 1]$, the *signal*, assigns to each state $j \in Q$, a distribution s_j on observations, i.e., $\sum_{o \in O} s_j(o) = 1$.

It is worth noticing how natural a generalisation the models illustrated in the preceding sections HMMs provide. Indeed, in the context of computational trust, representing a principal p by a HMM λ_p affords us a different distribution s_j on O for each possible state j of p. In particular, one could think of the states of λ_p as a collection of independent Beta models, the

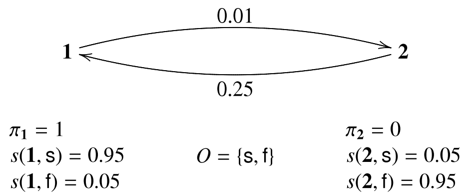

Fig. 1. Example Hidden Markov Model.

transitions between which are governed by the Markov chain formed by π and A, as principal p switches its internal state. The outcome of interactions with p, given a (possibly empty) past history $h = o_1 o_2 \cdots o_{k-1}$, is then modelled according to HMMs rules, for $o \in O$:

$$P(o \mid h\ \lambda_p) = \sum_{q_1,\ldots,q_k \in Q} \pi(q_1) \cdot s_{q_1}(o_1) \cdot A_{q_1 q_2} \cdot s_{q_2}(o_2) \cdots A_{q_{k-1} q_k} \cdot s_{q_k}(o).$$

This makes the intention quite explicit that in HMMs, principals' states are '*hidden*' from the observer: we can only observe the result of interacting with p, not its state, and the probability of outcomes is always computed in the ignorance of the path through Q the system actually traced.

Example 1. Fig. 1 shows a two-state HMM over our usual set of possible observations $\{\mathsf{s}, \mathsf{f}\}$. State **1** is relatively stable, i.e., there is only 0.01 probability mass attached to the transition $\mathbf{1} \mapsto \mathbf{2}$. Also, in state **1** the output s is much more likely than f. In contrast, state **2** is not nearly as stable and it signals f with probability 0.95. So, intuitively, the likely observation sequences from this HMM are long sequences consisting mostly of s, followed by somewhat shorter sequences of mostly f; this pattern is likely to repeat itself indefinitely. □

We conclude this section by recalling some fundamental properties of finite Markov chains which we shall be using in the rest of the paper to analyse HMMs. For a fuller treatment of these notions the reader is referred to, e.g., [6,15,3]. On the contrary, the reader not interested in the details can safely jump to the next section.

The key property we rely on is the *irreducibility* of the discrete Markov chain (DMC) underlying a HMM λ. Intuitively, a DMC is irreducible if at any time, from each state i, there is a positive probability to eventually reach each state j. Denoting by A^m_{ij} the (i, j)-entry of the mth power of matrix A, we can then express the condition formally as follows.

Definition 2 (*Irreducibility*). For A a DMC, we say that state i reaches state j, written $i \mapsto j$, whenever $A^m_{ij} > 0$, for some m, and that A is irreducible if $i \mapsto j$, for all i and j.

A state i of a DMC can be classified as either *recurrent* or *transient*, according to whether or not starting from i one is guaranteed to eventually return to i. Recurrent states can be *positive* or *null* recurrent, according to whether or not they have an 'average return time.' In the following, we shall write $q_k = i$ to indicate that i is the kth state visited by a DMC in a given run $q_0 q_1 q_2 \cdots$.

Definition 3 (*Classification of States*). For A a DMC and i a state, we say that i is:

$$\begin{array}{ll} \textit{recurrent} & \text{if} \quad P(\, q_k = i, \ \text{for some } q_0 \cdots q_k \mid q_0 = i\,) = 1; \\ \textit{transient} & \text{otherwise.} \end{array}$$

It can be proved that j is recurrent if and only if $\sum_m^\infty A^m_{jj} = \infty$, and this characterisation has important corollaries. Firstly, it follows easily that $\sum_m^\infty A^m_{ij} = \infty$, for all i such that $i \mapsto j$. Thus, if $i \mapsto j$ and $j \mapsto i$, then i and j are either both transient or both recurrent. It is then an immediate observation that in an irreducible chain either all states are transient, or they all are recurrent. We can also conclude that if j is transient, then $A^m_{ij} \to 0$ as $m \to \infty$, for all i. From this last observation it follows easily that if A is *finite*, as it is in our case, then at least one state must be recurrent and, therefore, all states must be recurrent if A is also irreducible. In fact, if all states were transient, we would have that $\lim_{m\to\infty} \sum_{j\in Q} A^m_{ij} = 0$, which is incompatible with the fact that $\sum_{j\in Q} A^m_{ij} = 1$ for each m, since each A^m is a stochastic matrix.

Let us define T_i as the random variable yielding the *time of first visit* to state i, namely $\min\{\, k \geq 1 \mid q_k = i\,\}$. Exploiting the independence property of DMC (homogeneity), we can define the *average return time* of state i as

$$\mu_i = \mathbf{E}[T_i \mid q_0 = i].$$

Definition 4 (*Classification of Recurrent States*). For A a DMC and i a recurrent state, we say that i is:

$$\begin{array}{rl} \textit{null} & \text{if} \quad \mu_i = \infty; \\ \textit{positive} & \text{if} \quad \mu_i < \infty. \end{array}$$

Similarly to above, one can prove that a recurrent state is *null* if and only if $A^m_{ij} \to 0$ as $m \to \infty$. Then, for the same reasons as above, one concludes that a finite DMC has no null recurrent states. Moreover, if A is *finite* and *irreducible*, then *all* its states are positive recurrent, which leads us to state the following fundamental theorem.

Definition 5 (*Stationary Distribution*). A vector $\pi = (\pi_j \mid j \in Q)$ is a stationary distribution on Q if

$$\pi_j \geq 0 \quad \text{for all } j, \quad \text{and} \quad \sum_{j \in Q} \pi_j = 1;$$

$$\pi A = \pi .$$

So a stationary distribution characterises the limiting behaviour of a chain beyond the typical fluctuations of stochastic behaviours by the existence of an initial distribution which remains invariant in time for the DMC. In an irreducible chain, the average time to return determines such invariant.

Theorem 1 (*Existence of Stationary Distribution*). *An irreducible Markov chain has a stationary distribution π if and only if all its states are positive recurrent. In this case, π is the unique stationary distribution and is given by $\pi_i = \mu_i^{-1}$.*

The existence of a stationary distribution goes a long way to describe the asymptotic behaviour of a DMC yet, as it turns out, it is not sufficient. Indeed, if one wants to guarantee convergence to the stationary distribution regardless of λ's initial distribution, one needs to add the condition of aperiodicity.

Definition 6 (*Aperiodicity*). For A a DMC, the period of i is $d(i) = \gcd\{\, m \mid A^m_{ii} > 0 \,\}$. State i is aperiodic if $d(i) = 1$; and A is *aperiodic* if all its states are such.

Theorem 2 (*Convergence to Stationary Distribution*). *For A an irreducible aperiodic Markov chain, $\lim_{m\to\infty} A^m_{ij} = \mu_j^{-1}$, for all i and j.*

Observe that as $P(q_m = j) = \sum_i P(q_0 = i) \cdot A^m_{ij}$, it follows that $P(q_m = j) \to \mu_j^{-1}$ when $m \to \infty$, regardless of the initial distribution.

In the rest of this paper we shall assume each HMM to have an *irreducible* and *aperiodic* underlying A, so as to be able to rely on Theorems 1 and 2. As A is indeed *positive recurrent* (because finite), this is the same as requiring that A is ***ergodic***.

5. Estimation error of Beta model with a decay scheme

This section presents a comparative analysis of the Beta model with exponential decay. More precisely, we set out to derive an analytic expression for the error incurred by approximating a principal exhibiting dynamic behaviour by the Beta model enhanced with a decay scheme. As explained before, for the sake of this comparison we select HMMs as a 'state-based' probabilistic model sufficiently precise to be confused with the principal under analysis; therefore fix a generic HMM λ which we refer to as the *real model*. Following the results from the theory of Markov chains recalled in Section 4, we shall work under the hypothesis that λ is *ergodic*. This corresponds to demanding that all the states of λ remain 'live' (i.e., probabilistically possible) at all times, and does seem as a standard and reasonably mild condition. It then follows by general reasons that λ admits a *stationary probability distribution* over its states Q_λ (cf. Theorem 1); we denote it by the row vector

$$\Pi_\lambda = \begin{bmatrix} \pi_1 & \pi_2 & \dots & \pi_n \end{bmatrix},$$

where π_q denotes the stationary probability of the state q. If A_λ is the stochastic state transition matrix representing the Markov chain underlying λ, vector Π_λ satisfies the stationary equation

$$\Pi_\lambda = \Pi_\lambda A_\lambda . \tag{1}$$

As we are only interested λ's *steady-state* behaviour, and as the state distribution of the process is guaranteed to converge to Π_λ after a transient period (cf. Theorem 2), without loss of generality in the following we shall assume that Π_λ is indeed λ's initial distribution. Observe too that as λ is finite and irreducible, all components of Π_λ are *strictly* positive and can be computed easily from matrix A_λ.

For simplicity, we maintain here the restriction to binary outcomes (s or f), yet our derivation of the estimation error can be generalised to multiple outcomes cases (e.g., replacing beta with Dirichlet pdfs, cf. [14]).

It is worth noticing that HMMs can themselves be used to support Bayesian analysis and/or supplant the Beta model, as indicated, e.g., in [17]. That is however a matter for another paper and another line of work. We remark again that our focus here remains the analysis of the decay principle, in which HMMs' sole role is to provide us with a suitable model for principals and a meaningful testbed for comparisons.

Beta model with a decay factor

We consider observation sequences $h_\ell = o_0 o_1 \cdots o_{\ell-1}$ of arbitrary length ℓ, where o_0 and $o_{\ell-1}$ are respectively the least and the most recent observed outcomes. Then, for r a decay factor ($0 < r < 1$), the beta estimate for the probability distribution on the next outcomes $\{\, \mathsf{s},\ \mathsf{f} \,\}$ is given by $(B_r(\mathsf{s} \mid h_\ell),\ B_r(\mathsf{f} \mid h_\ell))$, where

$$\begin{aligned} B_r(\mathsf{s} \mid h_\ell) &= \frac{m_r(h_\ell) + 1}{m_r(h_\ell) + n_r(h_\ell) + 2} \\ B_r(\mathsf{f} \mid h_\ell) &= \frac{n_r(h_\ell) + 1}{m_r(h_\ell) + n_r(h_\ell) + 2} \end{aligned} \tag{2}$$

and

$$m_r(h_\ell) = \sum_{i=0}^{\ell-1} r^i \delta_{\ell-i-1}(\mathrm{s}) \qquad n_r(h_\ell) = \sum_{i=0}^{\ell-1} r^i \delta_{\ell-i-1}(\mathrm{f}) \tag{3}$$

for

$$\delta_i(X) = \begin{cases} 1 & \text{if } o_i = X \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Under these conditions, from Eqs. (3) and (4), the sum $m_r(h_\ell) + n_r(h_\ell)$ forms a geometric series, and therefore

$$m_r(h_\ell) + n_r(h_\ell) = \frac{1 - r^\ell}{1 - r}. \tag{5}$$

The error function

We call the real probability that the next outcome will be s the *real* predictive probability, and denote it by σ. In contrast, we call the estimated probability that the next outcome will be s the *estimated* predictive probability. We define the estimation error as the expected squared difference between the real and estimated predictive probabilities. Observe that whilst the real predictive probability σ depends on λ, the chosen representation of principal's behaviour, and its current state, the estimated predictive probability $B_r(\mathrm{s} \mid h_\ell)$ depends on the interaction history h_ℓ and the decay parameter r. Here we derive an expression for the estimation error parametric in ℓ as a step towards computing its limit for $\ell \to \infty$, and thus obtain the required formula for the *asymptotic* estimation error. Here we start by expressing the estimation error as a function of the behaviour model λ and the decay r. Formally,

$$Error_\ell(\lambda, r) = \mathbf{E}\left[(B_r(\mathrm{s} \mid h_\ell) - \sigma)^2\right]. \tag{6}$$

Using the definition in (2) for $B_r(\mathrm{s} \mid h_\ell)$, and writing $a = m_r(h_\ell) + n_r(h_\ell) + 2$ for brevity, we rewrite the error function as:

$$Error_\ell(\lambda, r) = \mathbf{E}\left[\left(\frac{m_r(h_\ell) + 1}{a} - \sigma\right)^2\right] = \mathbf{E}\left[\frac{1}{a^2}\left(m_r(h_\ell)^2 + 2m_r(h_\ell) + 1\right) - \frac{2\sigma}{a}\left(1 + m_r(h_\ell)\right) + \sigma^2\right]. \tag{7}$$

Using (5), we obtain

$$a = \frac{3 - 2r - r^\ell}{1 - r}. \tag{8}$$

Observe now that a depends on the decay parameter r and the sequence length ℓ. Using the linearity property of expectation, we can rewrite Eq. (7) as:

$$Error_\ell(\lambda, r) = \frac{1}{a^2}\mathbf{E}\left[m_r(h_\ell)^2\right] + \frac{2}{a^2}\mathbf{E}\Big[m_r(h_\ell)\Big] + \frac{1}{a^2} - \frac{2}{a}\mathbf{E}\left[\sigma\right] - \frac{2}{a}\mathbf{E}\Big[\sigma m_r(h_\ell)\Big] + \mathbf{E}\left[\sigma^2\right]. \tag{9}$$

In order to express the above error in terms of the real model λ and the decay r, we need to express $\mathbf{E}\big[m_r(h_\ell)^2\big]$, $\mathbf{E}\big[m_r(h_\ell)\big]$, $\mathbf{E}\big[\sigma m_r(h_\ell)\big]$, $\mathbf{E}\big[\sigma\big]$, and $\mathbf{E}\big[\sigma^2\big]$ in terms of the parameters of the real model λ and r. We start with evaluating $\mathbf{E}\big[m_r(h_\ell)\big]$.

Using the definition of $m_r(h_\ell)$ given by (3) and the linearity of the expectation operator, we have

$$\mathbf{E}\left[m_r(h_\ell)\right] = \sum_{i=0}^{\ell-1} r^i \cdot \mathbf{E}\left[\delta_{\ell-i-1}(\mathrm{s})\right]. \tag{10}$$

Then, by Eq. (4), we find that

$$\mathbf{E}\left[\delta_{\ell-i-1}(\mathrm{s})\right] = P\left(\delta_{\ell-i-1}(\mathrm{s}) = 1\right). \tag{11}$$

Denoting the system state at the time of observing o_i by ν_i we have

$$\begin{aligned} P\left(\delta_{\ell-i-1}(\mathrm{s}) = 1\right) &= \sum_{x \in Q_\lambda} P\Big(\nu_{\ell-i-1} = x, \delta_{\ell-i-1}(\mathrm{s}) = 1\Big) \\ &= \sum_{x \in Q_\lambda} P\left(\nu_{\ell-i-1} = x\right) P\left(\delta_{\ell-i-1}(\mathrm{s}) = 1 \mid \nu_{\ell-i-1} = x\right) \end{aligned} \tag{12}$$

where Q_λ is the set of states in the real model λ.

We define the *state success probabilities vector*, Θ_λ, as the column vector

$$\Theta_\lambda = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix} \tag{13}$$

where θ_q is the probability of observing s given the system is in state q. Notice that these probabilities are given together with λ, viz., $s_q(\mathrm{s})$ from Definition 1. As we focus on steady-state behaviours, exploiting the properties of the stationary distribution Π_λ, we can rewrite Eq. (12) as the scalar product of Π_λ and Θ_λ:

$$P\left(\delta_{\ell-i-1}(\mathrm{s})=1\right)=\sum_{x\in Q_\lambda}\pi_x\theta_x=\Pi_\lambda\Theta_\lambda. \tag{14}$$

Substituting in Eq. (11), we get

$$\mathbf{E}\left[\delta_{\ell-i-1}(\mathrm{s})\right]=\Pi_\lambda\Theta_\lambda \tag{15}$$

and substituting in (10) we get

$$\mathbf{E}\left[m_r(h_\ell)\right]=\sum_{i=0}^{\ell-1}r^i\cdot\Pi_\lambda\Theta_\lambda. \tag{16}$$

Since $\Pi_\lambda\Theta_\lambda$ is independent of r, we use the *geometric series* summation rule to evaluate the sum in the above equation, and obtain:

$$\mathbf{E}\left[m_r(h_\ell)\right]=\left(\frac{1-r^\ell}{1-r}\right)\Pi_\lambda\Theta_\lambda. \tag{17}$$

Isolating the dependency on ℓ, we write the above equation as follows

$$\mathbf{E}\left[m_r(h_\ell)\right]=\frac{\Pi_\lambda\Theta_\lambda}{1-r}+\epsilon_1(\ell) \tag{18}$$

where

$$\epsilon_1(\ell)=-r^\ell\,\frac{\Pi_\lambda\Theta_\lambda}{1-r}. \tag{19}$$

We now move on to simplify $\mathbf{E}\left[m_r(h_\ell)^2\right]$, the next term in *Error*(λ,r). By the definition of $m_r(h_\ell)$ in Eq. (3), and using the linearity of expectation, we have

$$\begin{aligned}\mathbf{E}\left[m_r(h_\ell)^2\right]&=\mathbf{E}\left[\left(\sum_{i=0}^{\ell-1}r^i\delta_{\ell-i-1}(\mathrm{s})\right)^2\right]\\&=\mathbf{E}\left[\sum_{i_1=0}^{\ell-1}\sum_{i_2=0}^{\ell-1}r^{i_1+i_2}\,\delta_{\ell-i_1-1}(\mathrm{s})\,\delta_{\ell-i_2-1}(\mathrm{s})\right]\\&=\sum_{i_1=0}^{\ell-1}\sum_{i_2=0}^{\ell-1}r^{i_1+i_2}\cdot\mathbf{E}\left[\delta_{\ell-i_1-1}(\mathrm{s})\,\delta_{\ell-i_2-1}(\mathrm{s})\right].\end{aligned} \tag{20}$$

In fact, from the definition of $\delta_i(\mathrm{s})$ given by (4) above, it is obvious that

$$\mathbf{E}\left[\delta_{\ell-i_1-1}(\mathrm{s})\,\delta_{\ell-i_2-1}(\mathrm{s})\right]=P\left(\delta_{\ell-i_1-1}(\mathrm{s})=1,\delta_{\ell-i_2-1}(\mathrm{s})=1\right).$$

Substituting in Eq. (20) we get

$$\begin{aligned}\mathbf{E}\left[m_r(h_\ell)^2\right]&=\sum_{i_1=0}^{\ell-1}\sum_{i_2=0}^{\ell-1}r^{i_1+i_2}P\left(\delta_{\ell-i_1-1}(\mathrm{s})=1,\delta_{\ell-i_2-1}(\mathrm{s})=1\right)\\&=\sum_{i=0}^{\ell-1}r^{2i}P\left(\delta_{\ell-i-1}(\mathrm{s})=1\right)+2\sum_{i_1=0}^{\ell-2}\sum_{i_2=i_1+1}^{\ell-1}r^{i_1+i_2}P\left(\delta_{\ell-i_1-1}(\mathrm{s})=1,\delta_{\ell-i_2-1}(\mathrm{s})=1\right)\\&=\sum_{i=0}^{\ell-1}r^{2i}P\left(\delta_{\ell-i-1}(\mathrm{s})=1\right)+2\sum_{i=0}^{\ell-2}\sum_{k=1}^{\ell-1-i}r^{2i+k}P\left(\delta_{\ell-i-1}(\mathrm{s})=1,\delta_{\ell-(i+k)-1}(\mathrm{s})=1\right)\\&=\sum_{i=0}^{\ell-1}r^{2i}P\left(\delta_{\ell-i-1}(\mathrm{s})=1\right)+2\sum_{i=0}^{\ell-2}r^{2i}\sum_{k=1}^{\ell-1-i}r^kP\left(\delta_{\ell-i-1}(\mathrm{s})=1,\delta_{\ell-i-1-k}(\mathrm{s})=1\right).\end{aligned} \tag{21}$$

We use the notation $\hat{\imath}=\ell-i-1$, and write the above equation as follows,

$$\mathbf{E}\left[m_r(h_\ell)^2\right]=\sum_{i=0}^{\ell-1}r^{2i}P\left(\delta_{\hat{\imath}}(\mathrm{s})=1\right)+2\sum_{i=0}^{\ell-2}r^{2i}\sum_{k=1}^{\ell-1-i}r^kP\left(\delta_{\hat{\imath}}(\mathrm{s})=1,\delta_{\hat{\imath}-k}(\mathrm{s})=1\right). \tag{22}$$

Note now that $P(\delta_{\hat{\imath}}(\mathrm{s})=1, \delta_{\hat{\imath}-k}(\mathrm{s})=1)$ is the joint probability of observing s at times $\hat{\imath}$ and $\hat{\imath}-k$. This probability can be expressed as

$$\begin{aligned} P(\delta_{\hat{\imath}}(\mathrm{s})=1, \delta_{\hat{\imath}-k}(\mathrm{s})=1) &= \sum_{x\in Q_\lambda}\sum_{y\in Q_\lambda} P\big(\nu_{\hat{\imath}}=x, \delta_{\hat{\imath}}(\mathrm{s})=1, \nu_{\hat{\imath}-k}=y, \delta_{\hat{\imath}-k}(\mathrm{s})=1\big) \\ &= \sum_{x\in Q_\lambda} P\big(\nu_{\hat{\imath}}=x\big)\, P(\delta_{\hat{\imath}}(\mathrm{s})=1 \mid \nu_{\hat{\imath}}=x) \\ &\quad\times \sum_{y\in Q_\lambda} P\big(\nu_{\hat{\imath}-k}=y \mid \nu_{\hat{\imath}}=x\big)\, P(\delta_{\hat{\imath}-k}(\mathrm{s})=1 \mid \nu_{\hat{\imath}-k}=y). \end{aligned} \tag{23}$$

We can rewrite (23) in terms of the state stationary probabilities vector Π_λ and the state success probabilities vector Θ_λ, given by Eqs. (1) and (13), respectively.

$$P(\delta_{\hat{\imath}}(\mathrm{s})=1,\ \delta_{\hat{\imath}-k}(\mathrm{s})=1) = \sum_{x\in Q_\lambda} \pi_x\theta_x \sum_{y\in Q_\lambda} P\big(\nu_{\hat{\imath}-k}=y \mid \nu_{\hat{\imath}}=x\big)\,\theta_y. \tag{24}$$

We can simplify this further by making use of the *time reversal* model of λ (cf. [3,15] which, informally speaking, represents the same model λ when time runs 'backwards.' If λ's state transition probability matrix is $A_\lambda = (A_{ij} \mid i,j=1,\dots,n)$ then λ's *reverse state transition* probability matrix is:

$$A'_\lambda = \begin{bmatrix} A'_{11} & A'_{12} & \dots & \dots \\ A'_{21} & \ddots & \dots & \dots \\ \vdots & \dots & A'_{xy} & \vdots \\ \dots & \dots & \dots & A'_{nn} \end{bmatrix} \tag{25}$$

where A'_{xy} is the probability that the previous state is y given that current state is x. Clearly, A'_λ is derived from A_λ by the identity:

$$A'_{xy} = \frac{\pi_y}{\pi_x} A_{yx} \tag{26}$$

which exist as by the irreducibility of λ all π_x are strictly positive. It is easy to prove that A'_λ is a stochastic matrix, and is irreducible when A_λ is such. Now, observing that $P(\nu_{\hat{\imath}-k}=y \mid \nu_{\hat{\imath}}=x)$ is the probability that the kth previous state is y given that the current state is x, we can rewrite (24) in terms of Π_λ, Θ_λ and A'_λ:

$$P(\delta_{\hat{\imath}}(\mathrm{s})=1, \delta_{\hat{\imath}-k}(\mathrm{s})=1) \;=\; \big(\Pi_\lambda \times \Theta_\lambda^{\mathrm{T}}\big)\, {A'_\lambda}^{k}\, \Theta_\lambda \tag{27}$$

where we use symbol × to denote the '*entry-wise*' product of matrices. Let us now return to Eq. (22) and replace $P(\delta_{\hat{\imath}}(\mathrm{s})=1)$ and $P(\delta_{\hat{\imath}}(\mathrm{s})=1,\ \delta_{\hat{\imath}-k}(\mathrm{s})=1)$ in it using expressions (14) and (27), respectively.

$$\mathbf{E}\left[m_r(h_\ell)^2\right] = \sum_{i=0}^{\ell-1} r^{2i}\Pi_\lambda\Theta_\lambda + 2\sum_{i=0}^{\ell-2} r^{2i} \sum_{k=1}^{\ell-i-1} \big(\Pi_\lambda \times \Theta_\lambda^{\mathrm{T}}\big)\big(rA'_\lambda\big)^k\ \Theta_\lambda. \tag{28}$$

Using the summation rule for geometric series, Eq. (28) can be simplified to the following expression.

$$\mathbf{E}\big[m_r(h_\ell)^2\big] = \left(\frac{1-r^{2\ell}}{1-r^2}\right)\Pi_\lambda\Theta_\lambda + 2\sum_{i=0}^{\ell-2} r^{2i}\big(\Pi_\lambda\times\Theta_\lambda^{\mathrm{T}}\big)\big(rA'_\lambda - (rA'_\lambda)^{\ell-i}\big)\big(I-rA'_\lambda\big)^{-1}\Theta_\lambda \tag{29}$$

where I is the identity matrix of size n. Applying the geometric series rule again, the above equation can be rewritten as,

$$\begin{aligned} \mathbf{E}\big[m_r(h_\ell)^2\big] \;&=\; \left(\frac{1-r^{2\ell}}{1-r^2}\right)\Pi_\lambda\Theta_\lambda + 2r\left(\frac{1-r^{2\ell-2}}{1-r^2}\right)\big(\Pi_\lambda\times\Theta_\lambda^{\mathrm{T}}\big)A'_\lambda\big(I-rA'_\lambda\big)^{-1}\Theta_\lambda \\ &\quad - 2r^\ell\sum_{i=0}^{\ell-2} r^i\big(\Pi_\lambda\times\Theta_\lambda^{\mathrm{T}}\big)({A'_\lambda}^{\ell-i})(I-rA'_\lambda)^{-1}\Theta_\lambda. \end{aligned} \tag{30}$$

Isolating the terms which depend on ℓ, we write the above equation as follows

$$\mathbf{E}\big[m_r(h_\ell)^2\big] = \frac{\Pi_\lambda\Theta_\lambda}{1-r^2} + \frac{2r}{1-r^2}\big(\Pi_\lambda\times\Theta_\lambda^{\mathrm{T}}\big)A'_\lambda\big(I-rA'_\lambda\big)^{-1}\ \Theta_\lambda + \epsilon_2(\ell) \tag{31}$$

where

$$\begin{aligned} \epsilon_2(\ell) \;&=\; \left(\frac{-r^{2\ell}}{1-r^2}\right)\Pi_\lambda\Theta_\lambda + 2\left(\frac{-r^{2\ell-1}}{1-r^2}\right)\big(\Pi_\lambda\times\Theta_\lambda^{\mathrm{T}}\big)\big(A'_\lambda\big)\big(I-rA'_\lambda\big)^{-1}\Theta_\lambda \\ &\quad - 2r^\ell\sum_{i=0}^{\ell-2} r^i\big(\Pi_\lambda\times\Theta_\lambda^{\mathrm{T}}\big)\Big({A'_\lambda}^{\ell-i}\Big)\big(I-rA'_\lambda\big)^{-1}\Theta_\lambda. \end{aligned} \tag{32}$$

Notice that in the formulation above we use an inverse matrix, whose existence we prove by the following lemma.

Lemma 1. *For A a stochastic matrix and $0 < r < 1$, matrix $(I - rA)$ is invertible.*

Proof. We prove equivalently that

$$Det\Big(I - rA\Big) \neq 0. \tag{33}$$

By multiplying (33) by the scalar $-r^{-1}$, we reduce it to the equivalent condition

$$-\frac{1}{r} \cdot Det\Big(I - rA\Big) = Det\Big(A - \frac{1}{r}I\Big) \neq 0.$$

Observe that $Det\big(A - r^{-1}I\big)$ is the characteristic polynomial of A evaluated on r^{-1}, which is zero if and only if r^{-1} is an eigenvalue of A. Since A has no negative entry, it follows from the Perron–Frobenius Theorem (cf., e.g., [7]) that all its eigenvalues u are such that

$$|\, u\,| \leq \max_i \sum_{k=1}^{n} A_{ik}.$$

As A is stochastic and $r^{-1} > 1$, this concludes our proof. □

We remark that the argument above can easily be adapted to prove that if A a stochastic matrix, the matrix $(I - A)$ is *not* invertible.

We now turn our attention to $\mathbf{E}\left[\sigma m_r(h_\ell)\right]$, with σ the probability that the next outcome is s. As σ depends on the current state $\nu_{\ell-1}$, expectation $\mathbf{E}\big[\sigma m_r(h_\ell)\big]$ can be expressed as

$$\mathbf{E}\left[\sigma m_r(h_\ell)\right] = \mathbf{E}\left[R(x)\right] \tag{34}$$

with $R(x)$ defined for $x \in Q_\lambda$ by

$$R(x) = \mathbf{E}\left[\sigma m_r(h_\ell) \mid \nu_{\ell-1} = x\right]. \tag{35}$$

In other words, $R(x)$ is the conditional expected value of $\sigma m_r(h_\ell)$ given that the current state is x.

We define the *state predictive success probabilities vector Φ_λ* as the following column vector.

$$\Phi_\lambda = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{bmatrix} \tag{36}$$

where ϕ_x is the probability that the next outcome after a state transition is s, given that the current state is x. The entries of Φ_λ can be computed by

$$\phi_x = \sum_{y \in Q_\lambda} A_{xy}\theta_y,$$

and therefore

$$\Phi_\lambda = A_\lambda \Theta_\lambda. \tag{37}$$

Using the above, we can rewrite Eq. (35) as

$$R(x) = \mathbf{E}\Big[\phi_x m_r(h_\ell) \;\Big|\; \nu_{\ell-1} = x\Big] \tag{38}$$

for $x \in Q_\lambda$. Substituting $m_r(h_\ell)$ with its definition in (3), we obtain

$$\begin{aligned} R(x) &= \mathbf{E}\left[\phi_x \sum_{i=0}^{\ell-1} r^i \delta_{\ell-i-1}(\mathrm{s}) \;\Big|\; \nu_{\ell-1} = x\right] \\ &= \phi_x \mathbf{E}\left[\sum_{i=0}^{\ell-1} r^i \delta_{\ell-i-1}(\mathrm{s}) \;\Big|\; \nu_{\ell-1} = x\right]. \end{aligned} \tag{39}$$

Using the linearity of expectation, we then get

$$R(x) = \phi_x \sum_{i=0}^{\ell-1} r^i \mathbf{E}\Big[\delta_{\ell-i-1}(\mathrm{s}) \;\Big|\; \nu_{\ell-1} = x\Big]. \tag{40}$$

Since the possible values of $\delta_{\ell-i-1}(\mathrm{s})$ are only 0 and 1, we have

$$\mathbf{E}\Big[\delta_{\ell-i-1}(\mathrm{s}) \;\Big|\; \nu_{\ell-1} = x\Big] = P\Big(\delta_{\ell-i-1}(\mathrm{s}) = 1 \;\Big|\; \nu_{\ell-1} = x\Big).$$

Thus Eq. (40) can be written as

$$\begin{aligned} R(x) &= \phi_x \sum_{i=0}^{\ell-1} r^i P\Big(\delta_{\ell-i-1}(\mathsf{s}) = 1 \;\Big|\; \nu_{\ell-1} = x\Big) \\ &= \phi_x \sum_{i=0}^{\ell-1} r^i \sum_{y\in Q_\lambda} P\Big(\nu_{\ell-i-1} = y \;\Big|\; \nu_{\ell-1} = x\Big)\, P\Big(\delta_{\ell-i-1}(\mathsf{s}) = 1 \;\Big|\; \nu_{\ell-i-1} = y\Big) \\ &= \phi_x \sum_{i=0}^{\ell-1} r^i \sum_{y\in Q_\lambda} P\Big(\nu_{\ell-i-1} = y \;\Big|\; \nu_{\ell-1} = x\Big)\, \theta_y. \end{aligned} \tag{41}$$

We now return to Eq. (34) which expresses $\mathbf{E}\big[\sigma m_r(h_\ell)\big]$ and, making use again of the stationary distribution, substitute the expression above for $R(x)$.

$$\begin{aligned} \mathbf{E}\big[\sigma m_r(h_\ell)\big] &= \sum_{x\in Q_\lambda} P\left(\nu_{\ell-1} = x\right) R(x) = \sum_{x\in Q_\lambda} \pi_x R(x) \\ &= \sum_{x\in Q_\lambda} \pi_x \phi_x \sum_{i=0}^{\ell-1} r^i \sum_{y\in Q_\lambda} P\left(\nu_{\ell-i-1} = y \mid \nu_{\ell-1} = x\right)\ \theta_y. \end{aligned} \tag{42}$$

Exchanging the summations in the above equation, we get,

$$\mathbf{E}\big[\sigma m_r(h_\ell)\big] = \sum_{i=0}^{\ell-1} r^i \sum_{x\in Q_\lambda} \pi_x \phi_x \sum_{y\in Q_\lambda} P\left(\nu_{\ell-i-1} = y \mid \nu_{\ell-1} = x\right)\ \theta_y. \tag{43}$$

Comparing the above with Eqs. (24) and (27), we similarly obtain

$$\begin{aligned} \mathbf{E}\big[\sigma m_r(h_\ell)\big] &= \sum_{i=0}^{\ell-1} r^i \left(\Pi_\lambda \times \Phi_\lambda^{\mathrm{T}}\right) {A'_\lambda}^{i}\, \Theta_\lambda \\ &= \left(\Pi_\lambda \times \Phi_\lambda^{\mathrm{T}}\right) \left(\sum_{i=0}^{\ell-1} (rA'_\lambda)^i\right)\ \Theta_\lambda. \end{aligned} \tag{44}$$

As before, by Lemma 1, we can simplify the above formula as

$$\mathbf{E}\big[\sigma m_r(h_\ell)\big] = \left(\Pi_\lambda \times \Phi_\lambda^{\mathrm{T}}\right) \left(I - (rA'_\lambda)^\ell\right) \left(I - rA'_\lambda\right)^{-1} \Theta_\lambda. \tag{45}$$

Isolating the term which depends on ℓ, we rewrite the above equation as follows

$$\mathbf{E}\big[\sigma m_r(h_\ell)\big] = \left(\Pi_\lambda \times \Phi_\lambda^{\mathrm{T}}\right) \left(I - rA'_\lambda\right)^{-1} \Theta_\lambda + \epsilon_3(\ell) \tag{46}$$

where

$$\epsilon_3(\ell) = -r^\ell \left(\Pi_\lambda \times \Phi_\lambda^{\mathrm{T}}\right) (A'_\lambda)^\ell \left(I - rA'_\lambda\right)^{-1} \Theta_\lambda. \tag{47}$$

Let us now consider $\mathbf{E}[\sigma]$

$$\mathbf{E}[\sigma] = \sum_{x\in Q_\lambda} P\left(\nu_{\ell-1} = x\right) \phi_x = \sum_{x\in Q_\lambda} \pi_x \phi_x = \Pi_\lambda \Phi_\lambda.$$

Substituting Φ in the above equation by its definition in (37), we get

$$\mathbf{E}[\sigma] = \Pi_\lambda A_\lambda \Theta_\lambda. \tag{48}$$

Using the eigenvector property of Π_λ in Eq. (1) we obtain

$$\mathbf{E}[\sigma] = \Pi_\lambda \Theta_\lambda. \tag{49}$$

Finally, let us evaluate $\mathbf{E}\big[\sigma^2\big]$.

$$\mathbf{E}\big[\sigma^2\big] = \sum_{x\in Q_\lambda} P\big(\nu_{\ell-1} = x\big) {\phi_x}^2 = \sum_{x\in Q_\lambda} \pi_x {\phi_x}^2 = \Pi_\lambda \left(\Phi_\lambda \times \Phi_\lambda\right). \tag{50}$$

We can now in the end return to the error formula (9) and substitute the expressions we have so derived for its various components, viz., Eqs. (31), (18), (46), (49) and (50). We therefore obtain the following formula for the Beta estimation error.

$$\begin{aligned} Error_\ell(\lambda, r) \; = \; & \frac{1}{a^2}\left(\frac{\Pi_\lambda \Theta_\lambda}{1-r^2} + \frac{2r}{1-r^2}\left(\Pi_\lambda \times \Theta_\lambda^{\mathrm{T}}\right) A'_\lambda \left(I - rA'_\lambda\right)^{-1} \Theta_\lambda\right) \\ & + \frac{2}{a^2}\left(\frac{\Pi_\lambda \Theta_\lambda}{1-r}\right) - \frac{2}{a}\left(\Pi_\lambda \times \Phi_\lambda^{\mathrm{T}}\right)\left(I - rA'_\lambda\right)^{-1}\Theta_\lambda \\ & - \frac{2}{a}\Pi_\lambda \Theta_\lambda + \Pi_\lambda\left(\Phi_\lambda \times \Phi_\lambda\right) + \frac{1}{a^2} + \frac{2}{a^2}\epsilon_1(\ell) + \frac{1}{a^2}\epsilon_2(\ell) - \frac{2}{a}\epsilon_3(\ell) \end{aligned} \tag{51}$$

where $\epsilon_1(\ell), \epsilon_2(\ell)$, and $\epsilon_3(\ell)$ are given by Eqs. (19), (32) and (47) respectively. Also a is given by (8). Now, as we are interested in the asymptotic error, we evaluate the limit of the above error when $\ell \to \infty$.

$$Error(\lambda, r) = \lim_{\ell \to \infty} Error_\ell(\lambda, r). \tag{52}$$

Since $r < 1$, it is obvious that

$$\lim_{\ell\to\infty} \epsilon_1(\ell) = \lim_{\ell\to\infty} \epsilon_2(\ell) = \lim_{\ell\to\infty} \epsilon_3(\ell) = 0$$

and

$$\lim_{\ell\to\infty} a = \frac{3-2r}{1-r}.$$

Therefore, and using a few algebraic manipulations we get our final asymptotic error formula for the beta model with exponential decay.

$$\begin{aligned} Error(\lambda, r) \; = \; & \frac{(1-r)\left(4r^2-3\right)}{(1+r)(3-2r)^2}\Pi_\lambda\Theta_\lambda + \left(\frac{1-r}{3-2r}\right)^2 + \frac{2(1-r)r}{(3-2r)^2(1+r)}\left(\Pi_\lambda \times \Theta_\lambda^{\mathrm{T}}\right) A'_\lambda\left(I - rA'_\lambda\right)^{-1}\Theta_\lambda \\ & -2\left(\frac{1-r}{3-2r}\right)\left(\Pi_\lambda \times \Phi_\lambda^{\mathrm{T}}\right)\left(I - rA'_\lambda\right)^{-1}\Theta_\lambda + \Pi_\lambda\left(\Phi_\lambda \times \Phi_\lambda\right). \end{aligned} \tag{53}$$

6. System stability

The stability of a system is, informally speaking, its tendency to remain in the same state. In this section we describe the effect of system stability on beta estimation error derived in Section 5. In particular, we show that if a system is very stable, then the Beta estimation error tends to 0 as the decay r tends to 1; as the limit of the decay model for $r \to 1$ is indeed the unmodified Beta model, this means that when systems are very stable, the unmodified Beta model achieves better prediction than any decay model.

We introduce the notion of *state stability* which we define as the probability of transition to the same state. Formally, given a HMM λ with set of states Q_λ, the stability of a state $x \in Q_\lambda$ is defined as

$$Stability(x) = P\left(q_{t+1} = x \mid q_t = x\right) = A_{xx}.$$

Building on that, we define the *system stability* of λ at time t, as

$$Stability_t(\lambda) = P\left(q_{t+1} = q_t\right),$$

that is the probability that the system remains at time $t+1$ in the same state where it has been at time t. System stability can therefore be expressed as

$$Stability_t(\lambda) = \sum_{x \in Q_\lambda} P\left(q_t = x\right) A_{xx}. \tag{54}$$

Note that the system stability depends on the diagonal elements of the transition matrix A_λ. It also depends on the probability distribution over system states at time t. Assuming as before that the system is ergodic (cf. Definitions 2 and 6), when t tends to ∞ the probability distribution over the system states converges to the stationary probability distribution Π_λ. We call the system stability when $t \to \infty$ the *asymptotic* system stability, and denote it by $Stability_\infty(\lambda)$.

$$Stability_\infty(\lambda) = \sum_{x \in Q_\lambda} \pi_x A_{xx}. \tag{55}$$

As the stationary probability distribution Π_λ over states depends on the state transition matrix A_λ – see Eq. (1) – the asymptotic system stability of λ is thus determined by the transition matrix A_λ.

Regarding the analysis of the effect of the system stability on the estimation, obviously the error formula (53) is too complex to allow an analytical study of its curve. However, given a particular system model with a specific stability, the beta estimation error can be evaluated for different values of the decay factor r, which allows us to build sound intuitions about the impact of stability on the beta estimation mechanism.

Consider the model λ with the stability s where,

$$A_\lambda = \begin{bmatrix} s & \frac{1-s}{3} & \frac{1-s}{3} & \frac{1-s}{3} \\ \frac{1-s}{3} & s & \frac{1-s}{3} & \frac{1-s}{3} \\ \frac{1-s}{3} & \frac{1-s}{3} & s & \frac{1-s}{3} \\ \frac{1-s}{3} & \frac{1-s}{3} & \frac{1-s}{3} & s \end{bmatrix}. \tag{56}$$

Given the above transition matrix, it can be easily verified that

$$\Pi_\lambda = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{bmatrix}. \tag{57}$$

Let the success probabilities vector Θ_λ be defined by

$$\Theta_\lambda = \begin{bmatrix} 1.0 \\ 0.7 \\ 0.3 \\ 0.0 \end{bmatrix}. \tag{58}$$

Fig. 2 shows Beta estimation error when the system λ is unstable ($s < 0.5$). It is obvious that the minimum error value is obtained when the decay r tends to 1. The reason for this is that an unstable system is relatively unlikely to stay in the same state, and therefore unlikely to preserve the previous distribution over observations. If the estimation uses low values for the decay, then the resulting estimate for the predictive probability distribution is close to the previous distribution; this is unlikely to be the same as in the next time instant, due to instability. On the other hand, using a decay r tending to 1 favours equally all previous observations, and according to the following lemma the resulting probability distribution is expected to be the average of the distributions exhibited by the model states. Such an average provides a better estimate for the predictive probability distribution than approximating the distribution of the most recent set of states using low decay values.

Lemma 2. *Given unbounded sequences generated by a HMM λ, the expected value of beta estimate for the predictive probability as decay $r \to 1$ is given by $\Pi_\lambda \Theta_\lambda$, where Π_λ and Θ_λ are the stationary probability distribution and success probabilities vectors of λ, respectively.*

Proof. The expected value for beta estimate with decay $r < 1$ is given by,

$$\mathbf{E}\left[B_r(\mathsf{s} \mid h_\ell)\right] = \mathbf{E}\left[\frac{m_r(h_\ell) + 1}{m_r(h_\ell) + n_r(h_\ell) + 2}\right] \tag{59}$$

using Eq. (5), the equation above can be rewritten as

$$\mathbf{E}\big[B_r(\mathsf{s} \mid h_\ell)\big] = \left(\frac{1-r}{3 - 2r - r^\ell}\right)\big(\mathbf{E}[m_r(h_\ell)] + 1\big). \tag{60}$$

Substituting $\mathbf{E}\left[m_r(h_\ell)\right]$ using Eq. (18), and taking the limit when $\ell \to \infty$, we get

$$\lim_{\ell \to \infty} \mathbf{E}\big[B_r(\mathsf{s} \mid h_\ell)\big] = \frac{\Pi_\lambda \Theta_\lambda + 1 - r}{3 - 2r}, \tag{61}$$

which converges to $\Pi_\lambda \Theta_\lambda$ when $r \to 1$. $\square$

It is worth noticing that when $s = 1/|Q_\lambda|$, the minimum expected beta error is 0, when $r \to 1$. In this case all elements of A_λ are equal and therefore the predictive probability of success is $\sum_{x \in Q_\lambda} \theta_x / |Q_\lambda|$, regardless of the current state. In other words, the whole behaviour can effectively be modelled by a single probability distribution over observations. The best approximation for this probability distribution is achieved by considering the entire history using decay $r \to 1$, because in this way the expected beta estimate converges to the correct predictive distribution according to Lemma 2.

Systems which are relatively stable (i.e., with $s > 0.5$) are more likely to stay in the same state rather than transitioning to a new state. In such a case, approximating the probability distribution of a state by observing systems interactions provides a good estimate for the predictive probability distribution. However, the quality of the approximation depends heavily on the choice of an optimum value for decay. If the decay is too small, the sequence of observation considered in the computation will prove too short to reflect the correct distribution precisely. If otherwise the decay is too large (i.e., too close to 1),

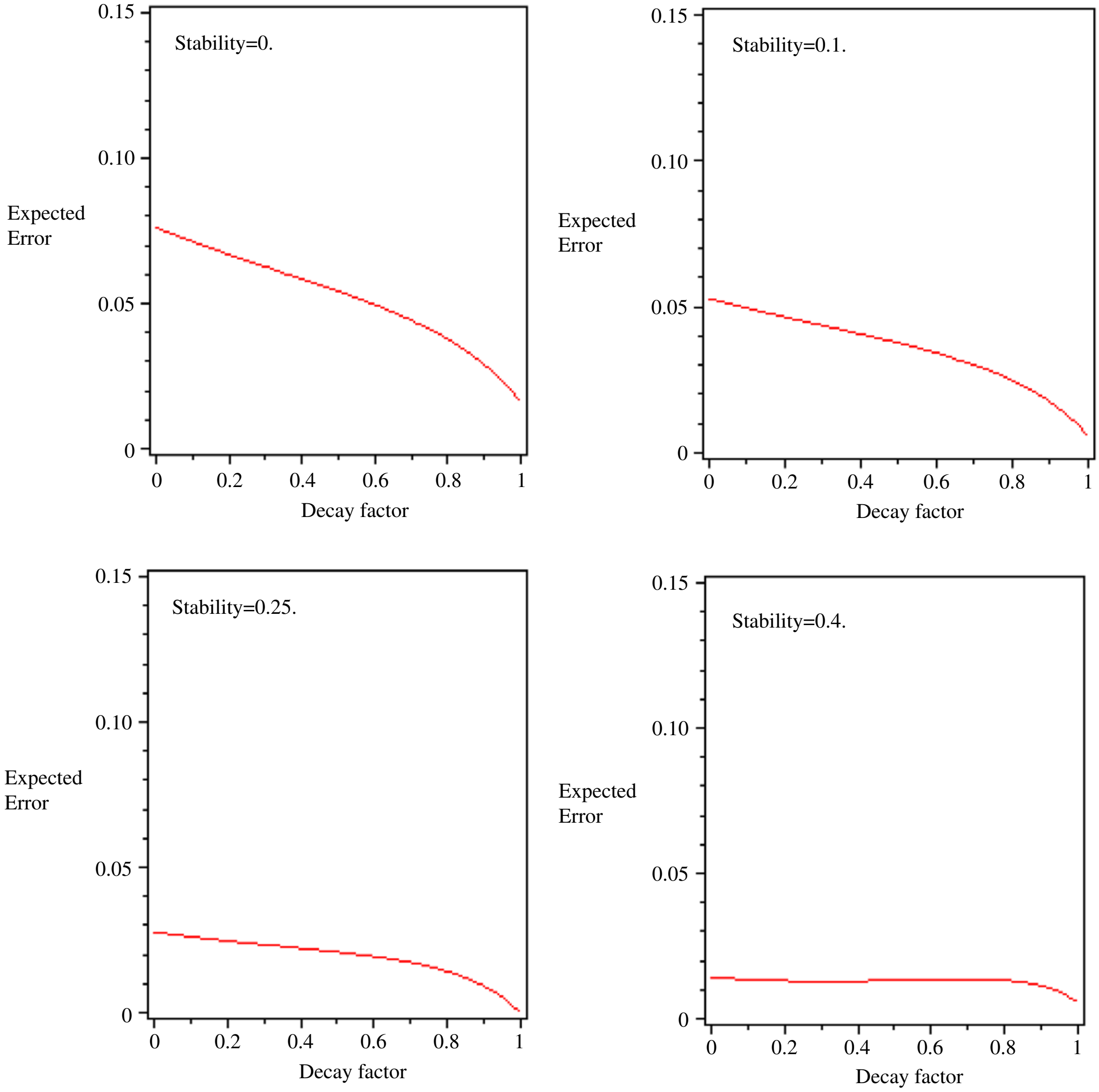

Fig. 2. Beta estimation error versus decay factor given stability <0.5.

then the resulting estimate approaches the average probability distribution as described above. Fig. 3 above shows the beta estimation error when the system λ is relatively stable.

Fig. 4 shows the beta estimation error for very stable systems, i.e., systems with $s > 0.9$. In such a case, observe that the estimation error is very sensitive to the choice of the decay value. In fact, regarded as a function of s and r, the error formula is pathological around point $(1, 1)$. Observe that the formula is undefined for $r = 1$, because in such a case all matrices $(I - rA')$ are singular. Worse than that, there is *no* limit as s and r tend to 1, as the limiting value depends on the relative speed of s and r. This is illustrated in Fig. 5, which plots $Error(\lambda, r)$ over the open unit square for our running four-state model. A simple inspection of (53), with the support of Lemma 1, shows that *Error* is continuous and well behaved on its domain, as illustrated by the top-left plot. Yes, the cusp near $(1, 1)$ – which is also noticeable in graphs of Fig. 4 – reflects its erratic behaviour in that neighborhood. The remaining three graphs of Fig. 5 show that the error function for $s \mapsto 1$ and $r \mapsto 1$ tends to different values along different lines, and therefore prove that it admits no limit at $(1, 1)$. However, if stability is actually 1, the minimum estimation error tends to 0, and the optimum decay value (which correspond to the minimum estimation error) tends to 1. The following Lemma proves this observation formally.

Lemma 3. *Let λ be a HMM. If $Stability_\infty(\lambda) = 1$, then the asymptotic beta estimation error tends to* 0 *when the decay r tends to* 1.

Proof. The asymptotic stability of a given system λ tends to 1 (i.e., a perfectly stable system) if and only if all the diagonal elements of A_λ tend to 1; this means that A_λ tends to the identity matrix I. As the latter is *not* irreducible, we first need to prove that the error formula (53) remains valid for $s = 1$. In fact, irreducibility plays its role in our assumption that

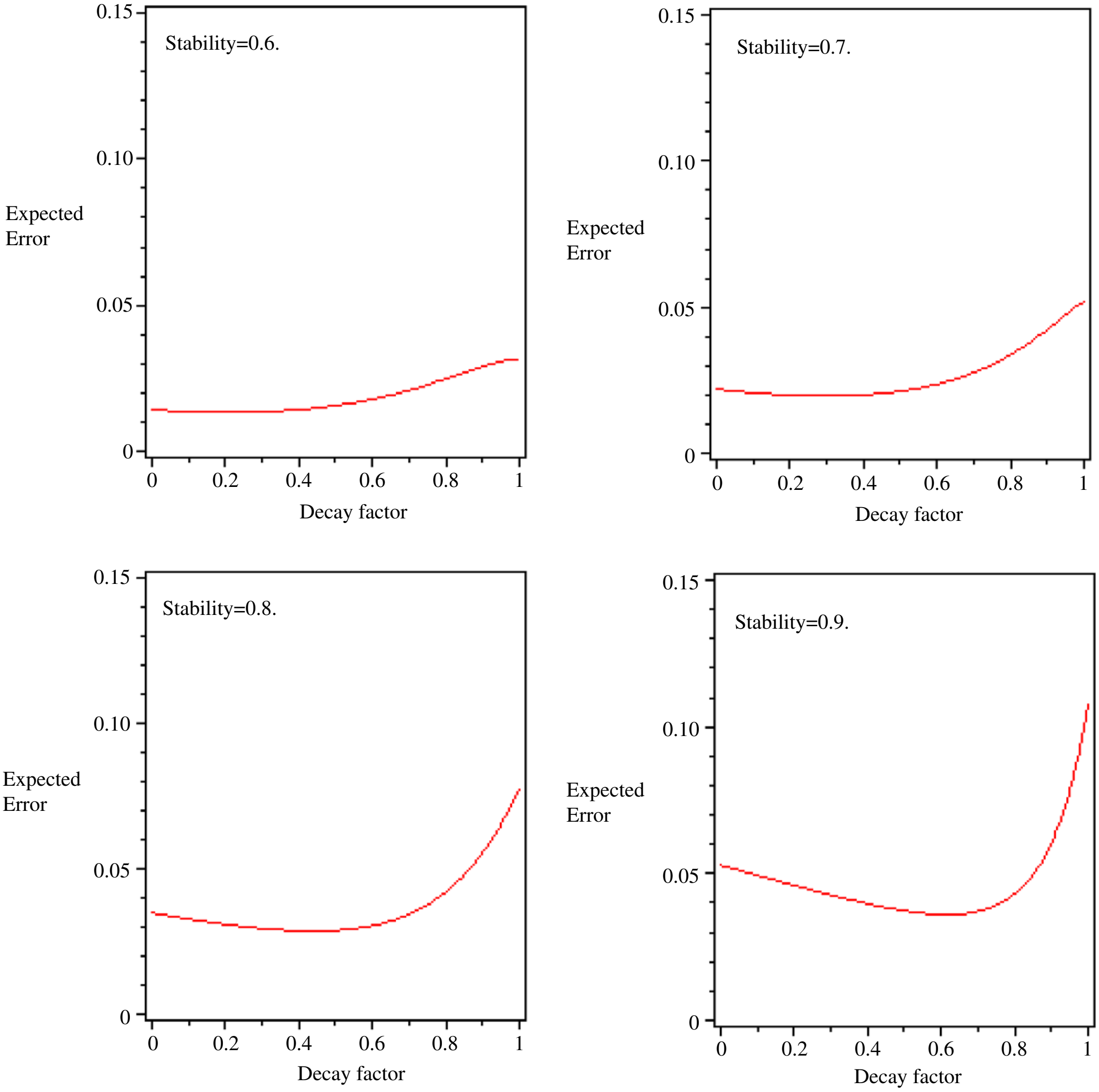

Fig. 3. Beta estimation error versus decay factor given stability >0.5.

the initial state distribution Π_λ is stable, which is obviously true in the case of I. All the steps in the derivation can then be repeated verbatim, with the exception of (26), which is undefined. Yet, it can easily be verified that I'_λ exists and is the identity matrix. We can therefore evaluate the beta estimation error in this case by replacing A'_λ by the identity matrix I in (53), while remembering that $(I - rI)^{-1} = I(1-r)^{-1}$ and $\Phi_\lambda = I\Theta_\lambda = \Theta_\lambda$. We get,

$$\begin{aligned} Error\,(\lambda, r) \; = \; & \frac{(1-r)\left(4r^2-3\right)}{(1+r)\,(3-2r)^2}\Pi_\lambda\Theta_\lambda + \left(\frac{1-r}{3-2r}\right)^2 + \frac{2\,(1-r)\,r}{(3-2r)^2\,(1+r)}\left(\Pi_\lambda \times \Theta_\lambda^{\mathrm{T}}\right)\frac{1}{1-r}\Theta_\lambda \\ & -2\left(\frac{1-r}{3-2r}\right)\left(\Pi_\lambda \times \Theta_\lambda^{\mathrm{T}}\right)\frac{1}{1-r}\Theta_\lambda + \Pi_\lambda\left(\Theta_\lambda \times \Theta_\lambda\right). \end{aligned} \tag{62}$$

Then, observing that

$$\left(\Pi_\lambda \times \Theta_\lambda^{\mathrm{T}}\right)\Theta_\lambda = \Pi_\lambda\left(\Theta_\lambda \times \Theta_\lambda\right),$$

we obtain

$$Error\,(\lambda, r) = \frac{(1-r)\left(4r^2-3\right)}{(1+r)\,(3-2r)^2}\Pi_\lambda\Theta_\lambda + \left(\frac{1-r}{3-2r}\right)^2 + \left(\frac{2r}{(3-2r)^2\,(1+r)} - \frac{2}{3-2r} + 1\right)\Pi_\lambda\left(\Theta_\lambda \times \Theta_\lambda\right) \tag{63}$$

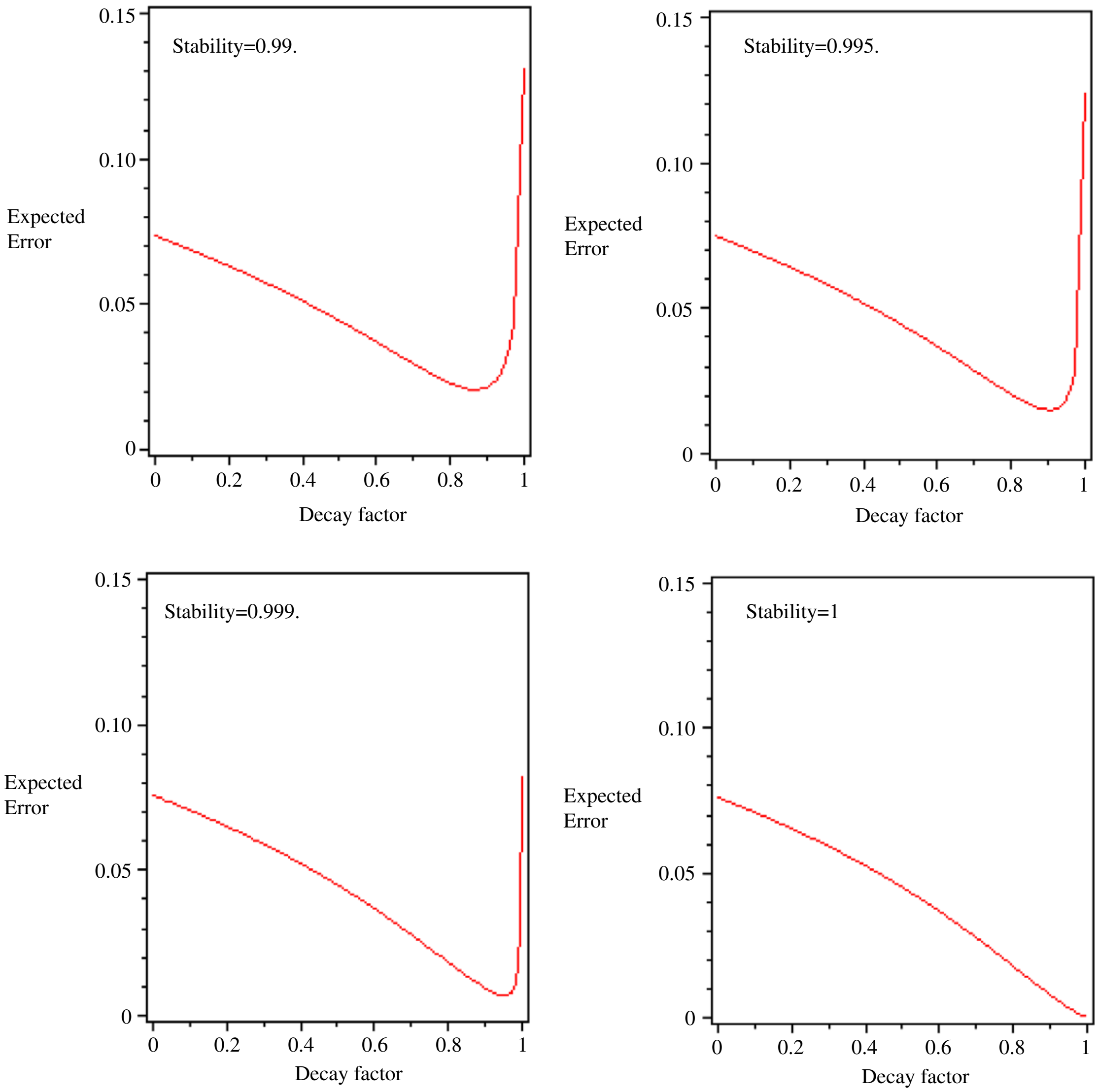

Fig. 4. Beta estimation error versus decay factor given stability >0.9.

and thus

$$Error\,(\lambda, r) = \frac{(1-r)\left(4r^2-3\right)}{(1+r)\,(3-2r)^2}\Pi_\lambda\Theta_\lambda + \left(\frac{1-r}{3-2r}\right)^2 + \frac{(1-r)\left(3-4r^2\right)}{(1+r)\,(3-2r)^2}\Pi_\lambda\left(\Theta_\lambda \times \Theta_\lambda\right). \tag{64}$$

By inspection of the error formula above, when $r \to 1$, the beta estimation error obviously tends to 0. That is, when the given system is stable, zero estimation error is achieved by choosing the decay r tending to 1, which is the same as saying dropping the decay altogether and using the unmodified Beta model. □

7. Conclusion

In this paper we have focussed on the exponential decay principle in the context of computational trust as a way to endow the well-known and widely-used Beta model with appropriate mechanisms to account for dynamic behaviours. Our contention is that, despite the attention the Beta model has received in the literature and its undoubted success 'on-the-ground,' the assumption that principals can be represented by a single immutable probability distribution is untenable in the real world.

Although we in general advocate fully-fledged 'stateful' models, such as the Hidden Markov Models, our purpose in this paper was to ascertain to what extent the decay principle put forward by some authors can provide the required support for principals whose behaviour changes according to their (discrete) state transitions. In doing so, we have described some

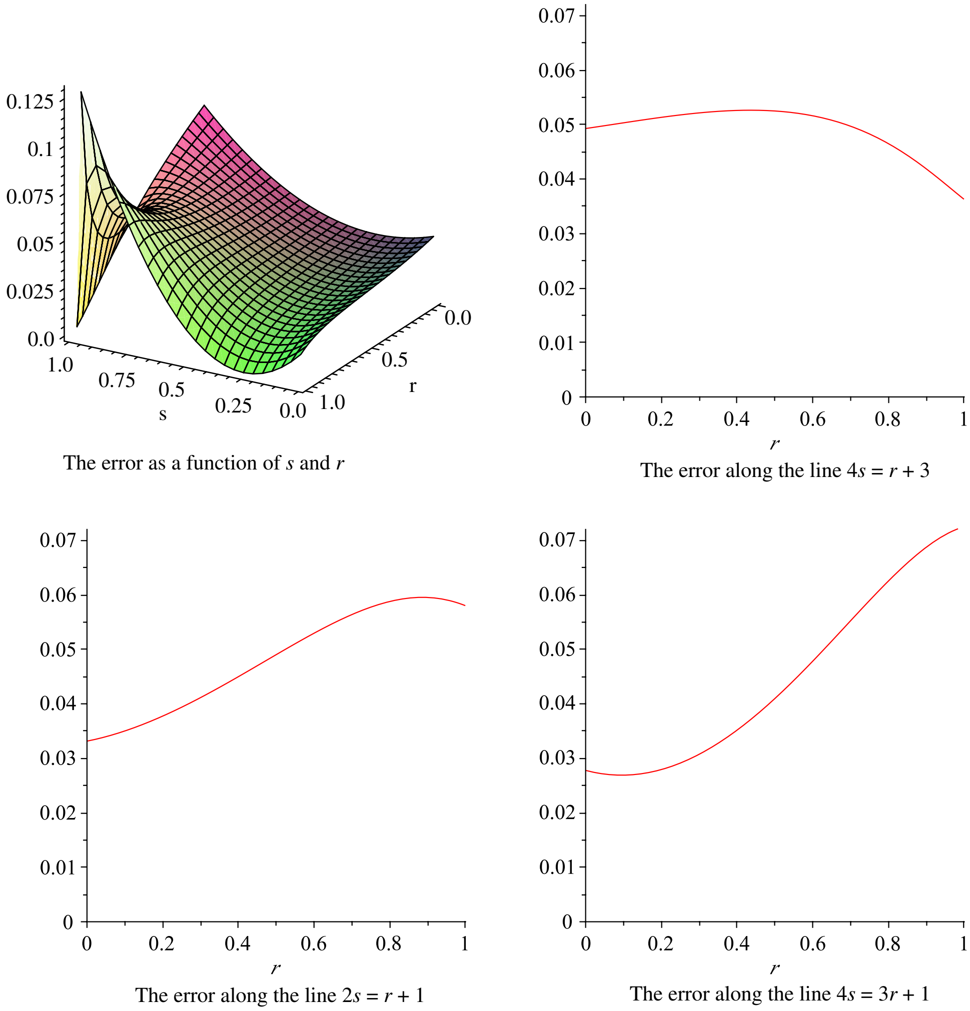

Fig. 5. Beta estimation error for the four-state model.

mathematical properties of the Beta model with exponential decay scheme, which suggest that the scheme will not be ideal in all scenarios.

We have then derived a formula for the expected error of the Beta scheme with respect to a representation of the 'real model' as a Hidden Markov Model, which can be used by algorithm developers to understand the implications of choosing a decay factor. Finally, we have exemplified one such analysis by plotting the error formula as a function of the decay parameter r according to a notion of system stability. The evidence obtained for the exercise, can be roughly summarised by saying that the choice of the 'right' parameter r remains highly sensitive and critical, and that anyway the choice of a decay scheme over the unmodified Beta model appears sensible only when systems are relatively stable, so that state changes happen rather infrequently.

Our analysis is valid under the assumption of the ergodicity of the underlying Markov chain, which in the case of finite-state systems reduces to just irreducibility and aperiodicity. Observe that the states of the model can be grouped in maximal classes – known in the literature as 'communicating' – whereby each state is reachable from any other state in the same class. By definition, reducible chains admit multiple maximal classes; every run of the system will eventually be 'trapped' in one of such classes, after which its steady-state behaviour will be described by the irreducible (sub)chain consisting of only the states in that class. As our analysis focusses on asymptotic behaviours only, this indicates that when the chain is reducible it may be sufficient to analyse each of the (sub)models determined by the maximal irreducible communicating classes in the model. The situation is more complex if the model fails to be aperiodic, as this indicates cyclic asymptotic behaviours and, potentially, causal dependencies between events, whereby a probabilistic analysis may anyway not be the best option.

Hidden Markov Models appear to be exactly the required kind of generalisation over the Beta model: they are fully probabilistic and, therefore, in principle they support all the analyses the Beta model does, whilst at the same time accounting for internal states. Our future work in this area will be dedicated to Hidden Markov Models and to other probabilistic models which, like them, embody the notion of dynamic behaviour at their core.

References

[1] L.E. Baum, T. Petrie, Statistical inference for probabilistic functions of finite-state Markov chains, Annals of Mathematical Statistics 37 (6) (1966) 1554–1563.
[2] M. Blaze, J. Feigenbaum, J. Ioannidis, A.D. Keromytis, The role of trust management in distributed systems security, in: J. Vitek, C.D. Jensen (Eds.), Secure Internet Programming: Security Issues for Mobile and Distributed Objects, in: Lecture Notes in Computer Science, vol. 1603, Springer, 1999, pp. 185–210.
[3] P. Brémaud, Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues, Springer, 1998.
[4] S. Buchegger, J.-Y. Le Boudec, A robust reputation system for peer-to-peer and mobile ad-hoc networks, in: P2PEcon 2004, 2004.
[5] V. Cahill, E. Gray, J.-M. Seigneur, C.D. Jensen, Y. Chen, B. Shand, N. Dimmock, A. Twigg, J. Bacon, C. English, W. Wagealla, S. Terzis, P. Nixon, G. di Marzo Serugendo, C. Bryce, M. Carbone, K. Krukow, M. Nielsen, Using trust for secure collaboration in uncertain environments, IEEE Pervasive Computing 2 (3) (2003) 52–61.
[6] G. Grimmet, D. Stirzaker, Probability and Random Processes, third edition, Oxford University Press, 2001.
[7] R.A. Horn, C.A. Johnson, Matrix Analysis, Cambridge University Press, Cambridge, UK, 1985.
[8] E.T. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, The Edinburgh Building, Cambridge, CB2 2RU, United Kingdom, 2003.
[9] A. Jøsang, J. Haller, Dirichlet reputation systems, in: The Second International Conference on Availability, Reliability and Security, 2007, ARES 2007, 2007, pp. 112–119.
[10] A. Jøsang, R. Ismail, The beta reputation system, in: Proceedings from the 15th Bled Conference on Electronic Commerce, Bled, 2002.
[11] K. Krukow, Towards a theory of trust for the global ubiquitous computer, Ph.D. Thesis, University of Aarhus, Denmark, 2006.
[12] K. Krukow, M. Nielsen, V. Sassone, Trust models in Ubiquitous Computing, Philosophical Transactions of the Royal Society A 366 (1881) (2008) 3781–3793.
[13] L. Mui, M. Mohtashemi, A. Halberstadt, A computational model of trust and reputation (for ebusinesses), in: Proceedings from 5th Annual Hawaii International Conference on System Sciences, HICSS'02, IEEE, 2002, p. 188.
[14] M. Nielsen, K. Krukow, V. Sassone, A Bayesian model for event-based trust, in: Festschrift in honour of Gordon Plotkin, in: Electronic Notes in Theoretical Computer Science, 2007.
[15] J.R. Norris, Markov Chains, Cambridge University Press, 1997.
[16] L.R. Rabiner, A tutorial on hidden markov models and selected applications in speech recognition, Proceedings of the IEEE 77 (2) (1989) 257–286.
[17] V. Sassone, K. Krukow, M. Nielsen, Towards a formal framework for computational trust, in: F.S. de Boer, M.M. Bonsangue, S. Graf, W.P. de Roever (Eds.), FMCO, in: Lecture Notes in Computer Science, vol. 4709, Springer, 2006, pp. 175–184.
[18] W. Teacy, J. Patel, N. Jennings, M. Luck, Travos: Trust and reputation in the context of inaccurate information sources, Autonomous Agents and Multi-Agent Systems 12 (2) (2006) 183–198.
[19] L. Xiong, L. Liu, PeerTrust: Supporting reputation-based trust for peer-to-peer electronic communities, IEEE Transactions on Knowledge and Data Engineering 16 (7) (2004) 843–857.