Equivalence of Hidden Markov Models with Continuous Observations

Oscar Darwin
Department of Computer Science, Oxford University, United Kingdom

Stefan Kiefer
Department of Computer Science, Oxford University, United Kingdom

Abstract

We consider the Sequential Probability Ratio Test applied to Hidden Markov Models. Given two Hidden Markov Models and a sequence of observations generated by one of them, the Sequential Probability Ratio Test attempts to decide which model produced the sequence. We show relationships between the execution time of such an algorithm and Lyapunov exponents of random matrix systems. Further, we give complexity results about the execution time taken by the Sequential Probability Ratio Test.

2012 ACM Subject Classification Theory of computation → Random walks and Markov chains; Mathematics of computing → Stochastic processes; Theory of computation → Logic and verification

Keywords and phrases Markov chains, equivalence, probabilistic systems, verification

Digital Object Identifier 10.4230/LIPIcs.CVIT.2016.23

1 Introduction

2 Preliminaries

We write $\mathbb{N}$ for the set of non-negative integers, $\mathbb{Q}$ for the set of rationals and $\mathbb{Q}_+$ for the set of positive rationals. For $d \in \mathbb{N}$ and a finite set Q we use the notation $|Q|$ for the number of elements in Q, $[d] = \{1, \ldots, d\}$ and $[Q] = \{1, \ldots, |Q|\}$. Vectors $\mu \in \mathbb{R}^N$ are viewed as row vectors. The norm $\|\mu\|$ is assumed to be the l_1 norm: $\|\mu\| = \sum_{i=1}^{N} |\mu_i|$. A matrix $M \in \mathbb{R}^{N\times N}$ is *stochastic* if M is non-negative and $\sum_{j=1}^{N} M_{i,j} = 1$ for all $i \in [N]$.

Throughout this paper, we use Σ to denote a finite set of *observations*. The set Σ^n is the set of words over Σ of length n and $\Sigma^* = \bigcup_{n=0}^{\infty} \Sigma^n$.

▶ **Definition 1.** *A* Hidden Markov Model *(HMM) is a triple (Q, Σ, Ψ) where Q is a finite set of states, Σ is a set of observations, and the* observation mass matrix *$\Psi : \Sigma \to [0,\infty)^{|Q|\times|Q|}$ specifies the transitions such that $\sum_{a\in\Sigma} \Psi(a)$ is a stochastic matrix.*

▶ **Example 2.** The HMM from the introduction is the triple $(\{q_1, q_2\}, \mathbb{R}, \Psi)$ with

$$\Psi(x) \;= TBC?!? \tag{1}$$

We extend Ψ to the mapping $\Psi : \Sigma^* \to [0,\infty)^{|Q|\times|Q|}$ with $\Psi(x_1 \cdots x_n) = \Psi(x_1) \times \cdots \times \Psi(x_n)$ for $x_1, \ldots, x_n \in \Sigma$. For the empty word ϵ we define $\Psi(\epsilon)$ to be the identity matrix. We say that $A \subseteq \Sigma^n$ is a *cylinder set* if $A = A_1 \times \cdots \times A_n$ and $A_i \in \mathcal{G}$ for $i \in [n]$. For every n there is an induced measure space $(\Sigma^n, \mathcal{G}^n, \lambda^n)$ where $\mathcal{G}^n$ is the smallest σ-algebra containing all cylinder sets in Σ^n and $\lambda^n(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n} \lambda(A_i)$ for any cylinder set $A_1 \times \cdots \times A_n$. Let $A \subset \Sigma^n$ and write $A\Sigma^\omega$ for the set of infinite words over Σ where the first n observations fall in the set A. Given a HMM (Q, Σ, Ψ) and initial distribution π on Q viewed as vector $\pi \in \mathbb{R}^{|Q|}$, there is an induced probability space $(\Sigma^\omega, \mathcal{G}^*, \mathbb{P}_\pi)$ where Σ^ω is the set of infinite words over Σ, and $\mathcal{G}^*$ is the smallest σ-algebra containing (for all $n \in \mathbb{N}$) all sets $A\Sigma^\omega$ where $A \subseteq \Sigma^n$ is a cylinder set and $\mathbb{P}_\pi$ is the unique probability measure such that

42nd Conference on Very Important Topics (CVIT 2016).
Editors: John Q. Open and Joan R. Access; Article No. 23; pp. 23:1–23:24
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

$\mathbb{P}_\pi(A\Sigma^\omega) = \sum_{w\in A} \|\pi\Psi(w)\|$ for any cylinder set $A \subseteq \Sigma^n$. Let $\mathbb{1}_F$ be the indicator random variable for the set $F \in \mathcal{G}^*$, i..e $\mathbb{1}_F(w) = 1$ if $w \in F$ and $\mathbb{1}_F(w) = 0$ if $w \notin F$.

Let (Q, Σ, Ψ) be an HMM and let π_1, π_2 be two initial distributions. The *total variation distance* is

$$d(\pi_1, \pi_2) = \sup_{E\in\mathcal{G}^*} |\mathbb{P}_{\pi_1}(E) - \mathbb{P}_{\pi_2}(E)|.$$

This supremum is actually a maximum due to Hahn's decomposition theorem; i.e., there is an event $S \in \mathcal{G}^*$ such that $d(\pi_1, \pi_2) = \mathbb{P}_{\pi_1}(S) - \mathbb{P}_{\pi_2}(S)$. We call π_1 and π_2 *distinguishable* if $d(\pi_1, \pi_2) = 1$. In [2] it is shown that distinguishability is decidable in polynomial time.

One could define the distinguishability of two pairs (C_1, π_1) and (C_2, π_2) where $C_i = (Q_i, \Sigma, \Psi_i)$ are HMMs and π_i are initial distributions for $i = 1, 2$. We do not need that though, as we can define, in a natural way, a single HMM over the disjoint union of Q_1 and Q_2 and consider instead the distinguishability of π_1 and π_2 (where π_1, π_2 are appropriately padded with zeros).

– For the rest of the paper we assume that (Q, Σ, Ψ) is an HMM. L_n depending on π_i etc.

► **Definition 3.** *Let π_1 and π_2 be initial distributions. For $w \in \Sigma^\omega$ we write w_n for the length n prefix of w. For any $n \in \mathbb{N}$, the* likelihood ratio *L_n is a random variable on Σ^ω given by $L_n(w) = \frac{\|\pi_1\Psi(w_n)\|}{\|\pi_2\Psi(w_n)\|}$.*

It follows from [2, Proposition 6] that L_n is a martingale and the following lemma holds due to Doob's forward convergence theorem[7].

► **Lemma 4.** *We have $\lim_{n\to\infty} L_n$ exists $\mathbb{P}_{\pi_2}$-almost surely.*

The following result characterises when the limit is 0 and follows from [2].

► **Theorem 5.** *We have $\lim_{n\to\infty} L_n = 0$ $\mathbb{P}_{\pi_2}$-a.s. if and only if π_1 and π_2 are distinguishable.*

► **Example 6.** We will illustrate convergence of the likelihood ratio using an example from [4] where the authors use HMMs to model sleep cycles. They took measurements of 51 healthy and 51 diseased individuals and using electrodes attached to the scalp, they read electrical signal data as part of an electroencephalography (EEG) during sleep. They split the signal into 30 second intervals and mapped it onto the simplex $\Delta^3 = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}_+^4 \mid \sum_{i=1}^4 x_i = 1\}$. For each individual this results in a time series of points in Δ^3. They modelled this data using two HMMs, each with 5 states, for healthy and diseased individuals using a numerical maximum likelihood estimate. This resulted in two embedded Markov chains defined by the row-stochastic block diagonal matrices

$$T_1 = \begin{bmatrix} 0.793 & 0.099 & 0.035 & 0.064 & 0.009 \\ 0.078 & 0.769 & 0.006 & 0.144 & 0.003 \\ 0.018 & 0.004 & 0.833 & 0.134 & 0.012 \\ 0.022 & 0.094 & 0.054 & 0.827 & 0.002 \\ 0.011 & 0.005 & 0.035 & 0.005 & 0.945 \end{bmatrix}, T_2 = \begin{bmatrix} 0.641 & 0.109 & 0.031 & 0.040 & 0.015 \\ 0.202 & 0.699 & 0.008 & 0.089 & 0.003 \\ 0.026 & 0.002 & 0.823 & 0.062 & 0.035 \\ 0.123 & 0.189 & 0.114 & 0.808 & 0.016 \\ 0.007 & 0.001 & 0.024 & 0.001 & 0.931 \end{bmatrix}.$$

Their HMMs are state-labelled. For each state i, they fit a Dirichlet pdf f_i describing the distribution of observations in Δ^3 emitted at state i. The pdfs of diseased and healthy individuals were so similar that they used the same pdf for both HMMs which was estimated from the whole population. Thus the two HMMs differ only in the transition probabilities.

The f_i are continuous and since in this paper we assume finite observation alphabets, we partition the simplex into the sets

$$U_k = \{x \in \Delta^3 \mid f_k(x) \geq \sup_i f_i(x)\}$$

for $k = 1, \ldots, 5$. The set U_k contains the points in Δ^3 most likely to be produced in state k. We assign a letter a_k for each U_k, and define a set of observations $\Sigma = \{a_1, \ldots, a_5\}$. Thus, the probability of producing letter a_k from state i is given as $O_{i,k} = \int_{U_k} f_i(x)dx$. We estimated the entries of O using a numerical Monte Carlo technique. We generated 100,000 samples from all 5 Dirichlet distributions in their paper which yielded the estimate

$$O = \begin{pmatrix} 0.9172 & 0.0803 & 0 & 0.0002 & 0.0024 \\ 0.0719 & 0.8606 & 0 & 0.0665 & 0.0010 \\ 0 & 0.0007 & 0.8546 & 0.1055 & 0.0392 \\ 0.0008 & 0.0998 & 0.0663 & 0.8257 & 0.0075 \\ 0.0109 & 0.0094 & 0.1046 & 0.0334 & 0.8416 \end{pmatrix}.$$

Since we consider transition labelled HMMs, we define observation density functions Ψ_1, Ψ_2 by

$$\Psi_m(a_k)_{i,j} = \left(T_m\right)_{i,j} O_{i,k}$$

for $m = 1, 2$. Let $Q = [10]$. We construct the block diagonal HMM (Q, Σ, Ψ) where

$$\Psi(a) = \begin{pmatrix} \Psi_1(a) & 0 \\ 0 & \Psi_2(a) \end{pmatrix}$$

for each $a \in \Sigma$. Let π_1 and π_2 be the Dirac distributions on states 1 and 6 respectively. These initial distributions correspond to healthy and diseased individuals started from sleep state 1. Using the algorithm from [2] one can show that π_1 and π_2 are distinguishable.

We sampled runs of the HMM started from π_1 and π_2 and plotted the corresponding sequences of $\ln L_n$. We refer to each of these two plots as a *log-likelihood plot*; see Figure 1.

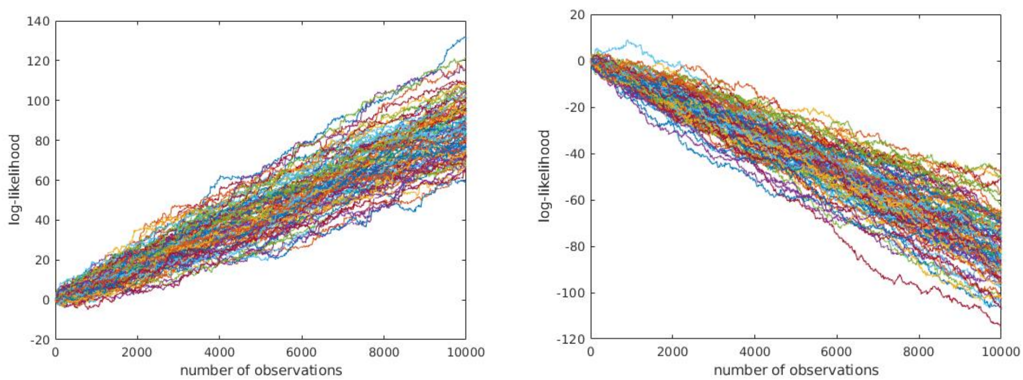


Figure 1 The two images show two log-likelihood plots of sample runs produced by π_1 and π_2, respectively.

By Theorem 5 it follows that $\ln L_n$ converges to ∞ and $-\infty$ in the case of words produced by π_1 and π_2 respectively. This is affirmed by Figure 1. Both log-likelihood plots also appear to follow a particular slope. This suggests that we can distinguish between words produced by π_1 and π_2 by tracking the value of $\ln L_n$ to see whether it crosses a lower or upper bound. This is the intuition behind the *Sequential Probability Ratio Test* (SPRT).

3 Sequential Probability Ratio Test

The SPRT runs as follows. We specify two error probabilities $\alpha, \beta > 0$ such that the SPRT gives the result π_2 from observations produced by π_1 with probability at most α and similarly the test gives the result π_1 from observations produced by π_2 with probability at most β. We then set $A = \ln \frac{\alpha}{1-\beta}$ and $B = \ln \frac{1-\alpha}{\beta}$. The algorithm continues to read observations and computes the value of $\ln L_n$ until $\ln L_n$ leaves the interval $[A, B]$. If $\ln L_n \leq A$ the test result is π_2 and if $\ln L_n \geq B$ the test result is π_1. We may express SPRT as a random variable $\mathrm{SPRT}_{\alpha,\beta} : \Sigma^\omega \to \{\pi_1, \pi_2\}$. We also define the stopping time

$$N_{\alpha,\beta} = \min\{n \in \mathbb{N} \mid \ln L_n \notin [A, B]\}.$$

We have that $N_{\alpha,\beta}$ is monotone decreasing in the sense that for $\alpha' \leq \alpha$ and $\beta' \leq \beta$ we have $N_{\alpha',\beta'} \geq N_{\alpha,\beta}$. When π_1 and π_2 are distinguishable $N_{\alpha,\beta}$ is $\mathbb{P}_{\pi_2}$-almost surely finite by Theorem 5.

▶ **Theorem 7.** *Suppose π_1 and π_2 are distinguishable. Let α, β be inputs to the SPRT. Then by choosing $A = \ln \frac{\alpha}{1-\beta}$ and $B = \ln \frac{1-\alpha}{\beta}$, $\mathbb{P}_{\pi_1}(\mathrm{SPRT}_{\alpha,\beta} = \pi_2) \leq \alpha$ and $\mathbb{P}_{\pi_2}(\mathrm{SPRT}_{\alpha,\beta} = \pi_1) \leq \beta$.*

In the following we consider the SPRT with respect to the measure $\mathbb{P}_{\pi_2}$. This is without loss of generality as there is a dual version of the SPRT, say $\overline{\mathrm{SPRT}}$ with $\overline{L}_n = 1/L_n$ instead of L_n, such that $\overline{\mathrm{SPRT}}_{\beta,\alpha} = \mathrm{SPRT}_{\alpha,\beta}$.

3.1 Expectation of $N_{\alpha,\beta}$

Consider the two single state HMMs where $p_1 \neq p_2$.

(The Dirac distributions on) s_1 and s_2 are distinguishable. Further, the increments $\ln L_{n+1} - \ln L_n$ are i.i.d. and $0 > \mathbb{E}_{s_1}[\ln L_1 - \ln L_0] = p_2 \ln \frac{p_1}{p_2} + (1-p_2) \ln \frac{1-p_1}{1-p_2} =: \ell$. Intuitively as ℓ gets more negative, the HMMs become more different.[1] Indeed, Wald [6] shows that the expected stopping time $\mathbb{E}_{s_2}[N_{\alpha,\beta}]$ and ℓ are inversely proportional:

$$\mathbb{E}_{s_2}[N_{\alpha,\beta}] = \frac{\beta \ln \frac{1-\alpha}{\beta} + (1-\beta) \ln \frac{\alpha}{1-\beta}}{\ell}. \tag{2}$$

This Wald formula cannot hold general (multi-state) HMMs. The increments $\ln L_{n+1} - \ln L_n$ need not be independent and $\mathbb{E}_{s_1}[\ln L_{n+1} - \ln L_n]$ can be different for different n. Further, $|\ln L_{n+1} - \ln L_n|$ can be unbounded –REF– .

Nevertheless, in Figure 1 we observed that $\ln L_n$ appears to decrease linearly (on the π_2 plot). Indeed we show, in Theorem 11 below, that the limit $\lim_{n\to\infty} \frac{1}{n} \ln L_n$ exists $\mathbb{P}_{\pi_2}$-almost surely. Intuitively it corresponds to the average slope of the log-likelihood plot for π_2. In the two state case, there is a simple proof of this using the law of large numbers:

$$\lim_{n\to\infty} \frac{1}{n} \ln L_n = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \ln L_{i+1} - \ln L_i = \mathbb{E}_{\pi_2}[\ln L_1 - \ln L_0] = \ell \quad \mathbb{P}_{\pi_2}\text{-a.s.}$$

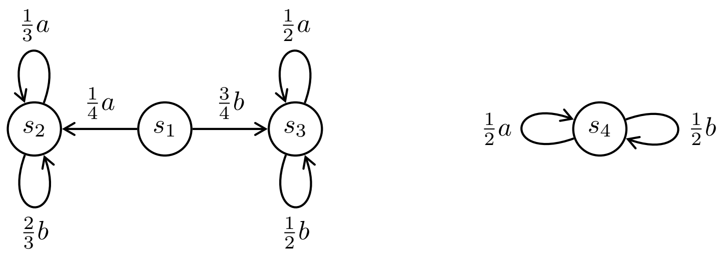
The number ℓ is called a likelihood exponent, as defined generally in the following definition.

[1] In fact, ℓ is the *KL-divergence* of the distributions f_1, f_2 where $f_i(a) = p_i$ and $f_i(b) = 1 - p_i$ for $i = 1, 2$.

▶ **Definition 8.** *A number ℓ is a* likelihood exponent *for initial distributions π_1 and π_2 if* $\mathbb{P}_{\pi_2}(\lim_{n\to\infty} \frac{1}{n} \ln L_n = \ell) > 0$.

▶ **Example 9.** In the case of Example 6 we show in the appendix that $\Lambda_{\pi_1,\pi_2} = \{\ell\}$. The slope of the right hand side of Figure 1 suggests that $\ell \approx -\frac{80}{10000} = -0.008$.

▶ **Example 10.** Even for fixed π_1, π_2 there may be multiple likelihood exponents. Consider the following pair of HMMs.



We observe two different likelihood exponents depending on the first letter produced. If the first letter is a then $L_1 = \frac{1}{2}$ and if the first letter is b then $L_1 = \frac{3}{2}$. If the first letter is a then $\ln L_{n+1} - \ln L_n$ are i.i.d. for $n > 1$ and $\lim_{n\to\infty} \frac{1}{n} \ln L_n = \frac{1}{2} \ln \frac{2}{3} + \frac{1}{2} \ln \frac{4}{3} = \frac{1}{2} \ln \frac{8}{9}$ like the two state example above. If the first letter is b then $L_n = \frac{3}{2}$ for all $n > 1$ and $\lim_{n\to\infty} \frac{1}{n} \ln L_n = 0$.

We denote the set of likelihood exponents for initial distributions π_1 and π_2 by Λ_{π_1,π_2}. Further, we write $\Lambda = \bigcup_{\pi_1,\pi_2} \Lambda_{\pi_1,\pi_2}$. In addition, for $\ell \in \Lambda$ we define the event $E_\ell = \{\lim_{n\to\infty} \frac{1}{n} \ln L_n = \ell\}$. In Example 10, we have $\Lambda_{e_1,e_4} = \{\frac{1}{2} \ln \frac{8}{9}, 0\}$. Writing $\ell := \frac{1}{2} \ln \frac{8}{9}$ it follows that $\mathbb{P}_{s_4}(E_\ell) = \mathbb{P}_{s_4}(E_0) = \frac{1}{2}$.

▶ **Theorem 11.** *The set of likelihood exponents satisfies $\Lambda \subset [-\infty, 0]$ and $|\Lambda| \le |Q|^2$. Further $\lim_{n\to\infty} \frac{1}{n} \ln L_n$ exists (and by definition is in Λ) $\mathbb{P}_{\pi_2}$-almost surely for any π_1, π_2.*

We investigate how $\lim_{n\to\infty} \frac{1}{n} \ln L_n$ influences the performance of the SPRT when α and β are small. Intuitively we expect a steeper slope in the likelihood plot (such as in Figure 1) to lead to faster termination. In the two state case, Wald's formula (2) becomes

$$\mathbb{E}_{s_2}[N_{\alpha,\beta}] = \frac{\beta \ln \frac{1-\alpha}{\beta} + (1-\beta) \ln \frac{\alpha}{1-\beta}}{\ell} \sim \frac{\ln \alpha}{\ell} \quad (\text{as } \alpha, \beta \to 0), \tag{3}$$

where we use the asymptotic notation $\sim$ defined in the preliminaries. In Theorem 12 below we generalise Equation (3) to arbitrary HMMs. Indeed a very similar asymptotic identity holds. In the case that $\Lambda = \{\ell\}$ and $\ell \in (-\infty, 0)$ then $\mathbb{E}_{s_2}[N_{\alpha,\beta}] \sim \frac{\ln \alpha}{\ell}$ as $\alpha, \beta \to 0$. If $|\Lambda| > 1$ then we condition our expectation on $\lim_{n\to\infty} \frac{1}{n} \ln L_n$.

▶ **Theorem 12.** *[Generalised Wald Formula] Let ℓ be a likelihood exponent and let π_1 and π_2 be initial distributions. If $\ell \in (-\infty, 0)$ then*

$$\mathbb{E}_{\pi_2}\Big[N_{\alpha,\beta} \mid E_\ell\Big] \sim \frac{\ln \alpha}{\ell} \quad (as\ \alpha, \beta \to 0).$$

If $\ell = 0$ then there exist $\alpha, \beta > 0$ such that

$$\mathbb{E}_{\pi_2}\Big[N_{\alpha,\beta} \mid E_\ell\Big] = \infty. \tag{4}$$

If $\ell = -\infty$ then

$$\sup_{\alpha,\beta} \mathbb{E}_{\pi_2}\Big[N_{\alpha,\beta} \mid E_\ell\Big] < \infty.$$

The theorem above pertains to $\mathbb{E}_{\pi_2}[N_{\alpha,\beta}]$. In the next subsection we give additional information about the distribution of $N_{\alpha,\beta}$.

3.2 Distribution of $N_{\alpha,\beta}$

3.2.1 Likelihood Exponent 0

▶ **Example 13.** We continue with Example 10 to illustrate the second case in Theorem 12. By picking $\alpha = \frac{1}{4}, \beta = \frac{1}{4}$ the bounds for the SPRT are $A = \ln \frac{1}{3}$ and $B = \ln 3$. If the first letter is b then $\ln L_n = \ln \frac{3}{2}$ for all $n > 1$, thus never crosses the SPRT bounds and $\lim_{n\to\infty} \frac{1}{n} \ln L_n = 0$. Hence with probability $\frac{1}{2}$ the SPRT fails to terminate and $N_{\alpha,\beta} = \infty$. It follows that $\mathbb{P}_{\pi_2}(E_0) = \frac{1}{2}$ and $E_{\pi_2}[N_{\alpha,\beta} \mid E_0] = \infty$. Also, $E_{\pi_2}[N_{\alpha,\beta}] = \infty$.

The second part of Theorem 12 says that the expectation of $N_{\alpha,\beta}$ conditioned under E_0 is infinite. The following proposition strengthens this statement. Conditioning under E_0, the probability that $N_{\alpha,\beta}$ is infinite converges to 1 as the errors $\alpha, \beta \to 0$. Recall that $N_{\alpha,\beta}$ is monotone decreasing. It follows that $\{N_{\alpha',\beta'} = \infty\} \subset \{N_{\alpha,\beta} = \infty\}$ if $\alpha' \le \alpha$ and $\beta' \le \beta$.

▶ **Proposition 14.** *The following equalities hold up to $\mathbb{P}_{\pi_2}$-null sets:*

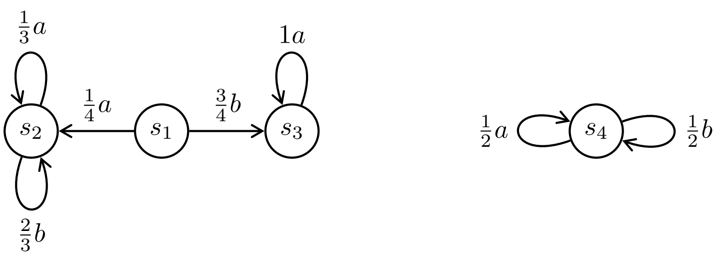
$$E_0 \; = \; \left\{ \lim_{n\to 0} L_n > 0 \right\} \; = \; \bigcup_{\alpha,\beta} \{N_{\alpha,\beta} = \infty\}.$$

Thus, $\lim_{\alpha,\beta\to 0} \mathbb{P}_{\pi_2}(N_{\alpha,\beta} = \infty) = \mathbb{P}_{\pi_2}(E_0)$.

It follows with Theorem 5 that π_1 and π_2 are distinguishable if and only if $\mathbb{P}_{\pi_2}(E_0) = 0$ if and only if $\mathbb{P}_{\pi_2}(N_{\alpha,\beta} < \infty) = 1$ for all α, β.

3.2.2 Likelihood Exponent $-\infty$

▶ **Example 15.** Consider now a modification of Example 10 where state s_3 has the b loop removed.



The likelihood exponents are $-\infty$ and $\ell := \frac{1}{2} \ln \frac{8}{9}$ so that $\Lambda = \{-\infty, \ell\}$. Also, $\mathbb{P}_{s_4}(E_{-\infty}) = \mathbb{P}_{s_4}(E_\ell) = \frac{1}{2}$. Up to $\mathbb{P}_{s_4}$-null sets the events $E_{-\infty}$, $b\Sigma^\omega$ and $ba^*b\Sigma^\omega$ are equal. The event $ba^*b\Sigma^\omega$ represents the right chain producing an observation which the left chain cannot produce, causing the SPRT to terminate for any α, β. Therefore conditioned on $E_{-\infty}$, the random variable $N_{\alpha,\beta} - 1$ is bounded by a geometric random variable with parameter $\frac{1}{2}$. Hence $\sup_{\alpha,\beta} \mathbb{E}_{\pi_2} \left[N_{\alpha,\beta} \mid E_{-\infty} \right] \le 1 + 2$.

We define the stopping time $N_\perp = \min\{n \in \mathbb{N} \mid L_n = 0\}$. Note that $\sup_{\alpha,\beta} N_{\alpha,\beta} \le N_\perp$ since $\{L_n = 0\} \subseteq \{L_n \le \frac{\alpha}{1-\beta}\}$ for all α, β. The following proposition states that the reverse inequality also holds.

▶ **Proposition 16.** *We have* $\sup_{\alpha,\beta} N_{\alpha,\beta} = N_{\perp}$. *Thus,* $\lim_{\alpha,\beta\to 0} \mathbb{P}_{\pi_2}(N_{\alpha,\beta} < \infty) = \mathbb{P}_{\pi_2}(E_{-\infty})$.

Applying this to Example 15, we have $\sup_{\alpha,\beta} \mathbb{E}_{\pi_2}\left[N_{\alpha,\beta} \mid E_{-\infty}\right] = 3$.

3.2.3 Likelihood Exponent in $(-\infty, 0)$

Conditioned on E_ℓ where $\ell \in (-\infty, 0)$, Theorem 12 shows that $N_{\alpha,\beta}$ scales with $\frac{\ln\alpha}{\ell}$ in expectation. The following result shows that this relationship also holds $\mathbb{P}_{\pi_2}$-almost surely.

▶ **Proposition 17.** *Let* $\ell \in \Lambda$ *and assume* $\ell \in (-\infty, 0)$. *We have*

$$\mathbb{P}_{\pi_2}\left(N_{\alpha,\beta} \sim \frac{\ln\alpha}{\ell} \quad (\textit{as } \alpha,\beta \to 0) \;\middle|\; E_\ell\right) \;=\; 1.$$

In fact, we prove the first part of using Proposition 17. If there is a bound $M \in \mathbb{N}$ such that $\mathbb{P}_{\pi_2}$-almost surely $\frac{N_{\alpha,\beta}}{-\ln\alpha} \le M$, the first part of Theorem 12 would follow from Proposition 17 by the Dominated convergence theorem. However this is not the case in general. Instead we show the following lemma and use Vitali's convergence theorem.

▶ **Lemma 18.** *The set of random variables* $\{\frac{N_{\alpha,\beta}}{-\ln\alpha} \mid 0 < \alpha,\beta \le \frac{1}{2}\}$ *is uniformly integrable with respect to the measure* $\mathbb{P}_{\pi_2}$*; i.e.*

$$\lim_{K\to\infty} \sup_{\alpha,\beta} \mathbb{E}_{\pi_2}\left[-\frac{N_{\alpha,\beta}}{\ln\alpha}\mathbb{1}_{\frac{N_{\alpha,\beta}}{-\ln\alpha} \ge -K}\right] = 0.$$

▶ **Example 19.** Recall Example 6 concerning sleep cycle HMMs with $\Lambda = \{\ell\}$. Figure 2 demonstrates the asymptotic relationship in Proposition 17. Each of the 50 lines correspond to a sample run and we record the value of $N_{\alpha,\beta}$ for $0 \le -\ln\alpha = -\ln\beta \le 1000$. From the figure we estimate $-\frac{1}{\ell}$ as $\frac{10^5}{800} = 125$. This coincides with the estimate in Example 9.

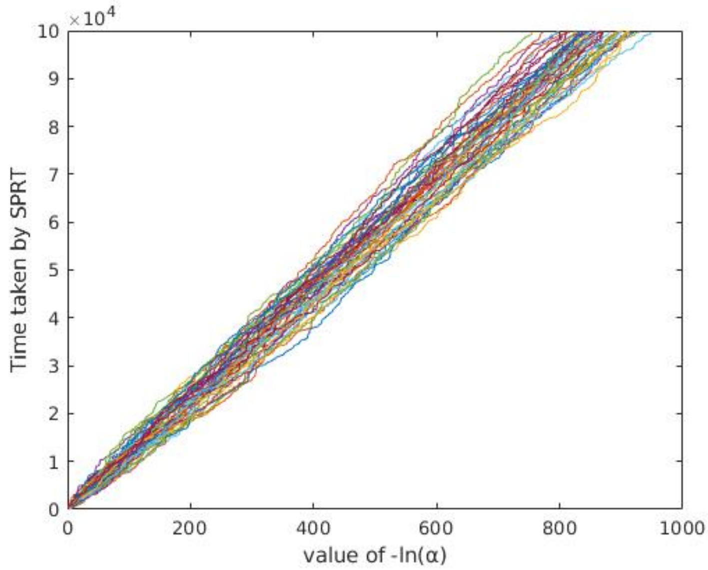


Figure 2 The time taken by the SPRT for $0 \le -\ln\alpha = -\ln\beta \le 1000$.

Proof of Proposition 17. Since $\Psi_{\min}^n \le L_n \le \Psi_{\min}^{-n}$ it follows that

$$N_{\alpha,\beta} \ge \frac{\min\{\ln \frac{\alpha}{1-\beta}, \ln \frac{\beta}{1-\alpha}\}}{\ln \Psi_{\min}}$$

Hence $N_{\alpha,\beta} \to \infty$ $\mathbb{P}_{\pi_2}$-almost surely as $\alpha, \beta \to 0$. Consider the case $\ell_k \in (-\infty, 0)$. Let $U_{\alpha,\beta} = \{w \in \Sigma^\omega \mid \ln L_{N_\alpha} \le \ln \frac{\alpha}{1-\beta}\}$. The set $\bigcap_{\alpha,\beta\in(0,1]} U^c_{\alpha,\beta} \subseteq \{L_n \text{ is unbounded}\}$. Hence, $\lim_{\alpha,\beta\to 0} \mathbb{1}_{U_{\alpha,\beta}} = 1$ $\mathbb{P}_{\pi_2}$-almost surely. Conditioned on V_k it follows that

$$0 \le \mathbb{1}_{U_{\alpha,\beta}} \frac{\ln \frac{\alpha}{1-\beta} - \ln L_{N_{\alpha,\beta}}}{N_\alpha} \le \mathbb{1}_{U_{\alpha,\beta}} \frac{\ln L_{N_{\alpha,\beta}-1} - \ln L_{N_{\alpha,\beta}}}{N_{\alpha,\beta}} \to 0 \quad \text{as } \alpha \to 0.$$

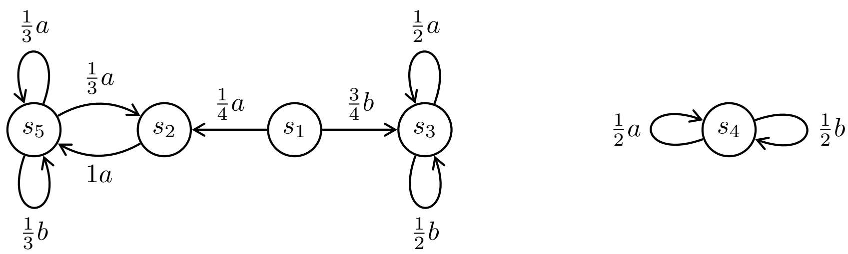
And so

$$\lim_{\alpha,\beta\to 0} \frac{\ln \alpha}{N_{\alpha,\beta}} = \lim_{\alpha\to 0} \frac{\ln \frac{\alpha}{1-\beta}}{N_{\alpha,\beta}} = \lim_{\alpha\to 0} \frac{\ln L_{N_{\alpha,\beta}}}{N_{\alpha,\beta}} = \ell_k.$$

◀

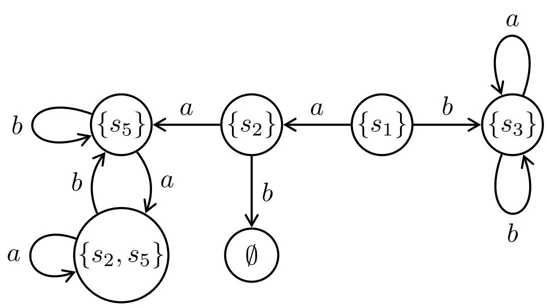
3.3 Qualitative Problems

▶ **Example 20.** Consider another adaption of Example 10.



If the first letter produced by s_4 is a b, then $L_n = \frac{3}{2}$ for all $n \in \mathbb{N}$. If the first two letters are ab then $L_1 = \frac{1}{2}$ and $L_n = 0$ for $n \ge 2$. If the first two letters are aa then $s_5 \in \operatorname{supp} \mathrm{e}_1 \Psi(aaw)$ for all words $w \in \Sigma^*$. Therefore in this case, up to a $\mathbb{P}_{s_4}$-null set, $L_n > 0$ for all $n \in \mathbb{N}$ and we show in the appendix that there is $\ell \in (-\infty, 0)$ such that $\lim_{n\to\infty} \frac{1}{n} \ln L_n = \ell$. Thus, $\Lambda_{s_1,s_4} = \{-\infty, \ell, 0\}$.

The likelihood ratio $L_n = 0$ if and only if $\operatorname{supp} \pi_1 \Psi(w_n) = \emptyset$. In order to track the support of $\pi_1 \Psi(w)$, we consider the left part of the HMM as an NFA with s_1 as the initial state and its determinisation as shown in the DFA below.



Almost surely, s_4 produces a word that drives this DFA into a bottom SCC. This bottom SCC determines $\lim_{n\to\infty} \frac{1}{n} \ln L_n$.

In general, s_4 is not a single state so the change of support of $\pi_1\Psi(w_n)$ depends also on the state of the producing HMM started from π_2. We construct a Markov chain that takes into account both supp $\pi_1\Psi(w_n)$ and also the state of the producing chain.

Let $\mathcal{S} = \{(\mu, q) \mid \mu \in 2^Q, q \in Q\}$ be a set of states and let

$$P_{(\mu_1,s_1),(\mu_2,s_2)} = \sum_{a\in\Sigma} \Psi(a)_{i_1,i_2} \mathbb{1}_{\text{supp } \mu_1\Psi(a)=\text{supp } \mu_2}$$

be a transition matrix. The pair $(\mathcal{S}, P)$ is a Markov chain –REQUIRES PROOF?. We then define the initial distribution

$$(u_{\pi_1,\pi_2})_{(\mu,q)} = \begin{cases} (\pi_2)_q & \text{supp } \mu = \text{supp } \pi_1 \\ 0 & \text{else.} \end{cases}$$

▶ **Lemma 21.** *Consider the set* $\{(\vec{0}, q) \mid q \in Q\}$

▶ **Lemma 22.** *Let* $C \subset \mathcal{S}$ *be a bottom SCC of* $(\mathcal{S}, P)$*. there exists an* $\ell \in (-\infty, 0]$ *such that for all* π_1 *and for all states* $q \in Q$ *such that* $(\text{supp } \pi_1, q) \in C$ *we have* $\Lambda_{\pi_1,\delta_q} = \{\ell\}$*.*

▶ **Lemma 23.** *Let* $C \subset \mathcal{S}$ *be a bottom SCC of* $(\mathcal{S}, P)$*. For any* $(\mu, q) \in C$*, there is a partition of* μ *written* $\mathcal{U}(\mu, q)$ *such that the mapping* $\phi : \mathcal{U}(\mu, q) \to \mathcal{P}(\mathcal{G})$ *given by*

$$\phi(\rho) = \{(s, q) \mid s \in \rho\}$$

is injective and for all $\rho \in \mathcal{U}(\mu, q)$*,* $\phi(\rho) \subseteq C$ *for some right bottom SCC of* $\mathcal{G}$*.*

the set of likelihood exponents $\Lambda_{\mu/\|\mu\|,\delta_q}$ *contains the single element*

$$\sup \{\ell \in \Lambda_{\rho,\delta_q} \mid (p, q) \textit{ in right bottom SCC of } G \textit{ and } , \text{supp } \rho \subseteq \text{supp } \mu\}.$$

Proof. – Prove the set is non-empty and $\ell > -\infty$

Let the set be maximised on (p, q). Then, there is $(\rho, q) \in \mathcal{S}$ such that $p \in \text{supp } \rho$.

We have

$$\begin{aligned} -\infty &< \lim_{n\to\infty} \frac{1}{n} \ln \|\rho\Psi(w)\| \\ &= \lim_{n\to\infty} \frac{1}{n} \ln \sum_{s\in\text{supp } \rho} \|\delta_s\Psi(w)\| \\ &= \sup_{s\in\text{supp } \rho} \lim_{n\to\infty} \frac{1}{n} \ln \|\delta_s\Psi(w)\| \\ &\leq \lim_{n\to\infty} \frac{1}{n} \ln \|\rho\Psi(w)\|. \end{aligned}$$

We have $\sup_{s\in\text{supp } \rho} \lim_{n\to\infty} \frac{1}{n} \ln \|\delta_s\Psi(w)\| = \sup_{s\in\text{supp } \rho} \Lambda_{\delta_s,\delta_q}$

◀

▶ **Lemma 24.** *There is a mapping* $\phi : \mathcal{S} \to \Lambda$ *such that for any* $\ell \in \Lambda$*, the events* $\{\textit{hitting } \phi^{-1}(\ell)\}$ *and* $\{\lim_{n\to\infty} \frac{1}{n} \ln L_n = \ell\}$ *are equal.*

– each bottom SCC is associated with a likelihood exponent. We can find those bottom SCCs that are associated with $-\infty$ and the zero SCCs.

Let (p_2, p_5) be the state distributions on s_2 and s_5 respectively. We have $\|(p_2, p_5)\Psi(a)\| = \|(\frac{1}{3}p_5 + p_2, \frac{1}{3}p_5)\| \leq \|(p_2, p_5)\|$ and $\|(p_2, p_5)\Psi(b)\| = \|(\frac{1}{3}p_5, 0)\| \leq \frac{1}{3}\|(p_2, p_5)\|$.

▶ **Theorem 25.** *Given an HMM and initial distributions* π_1, π_2*,*

1. *one can compute $\mathbb{P}_{\pi_2}(E_{-\infty})$ and $\mathbb{P}_{\pi_2}(E_0)$ in PSPACE;*
2. *one can decide whether $\mathbb{P}_{\pi_2}(E_0) = 0$ (i.e., $0 \notin \Lambda_{\pi_1,\pi_2}$) in polynomial time;*
3. *deciding whether $\mathbb{P}_{\pi_2}(E_0) = 1$, whether $\mathbb{P}_{\pi_2}(E_{-\infty}) = 0$, and whether $\mathbb{P}_{\pi_2}(E_{-\infty}) = 1$ are all PSPACE-complete problems.*

the *mortality* problem, which asks, given a finite set of states Q, a finite alphabet Σ, and a function $\Phi : \Sigma \to \{0,1\}^{Q \times Q}$, whether there exists a word $w \in \Sigma^*$ such that $\Phi(w)$ is the zero matrix. The mortality problem can be viewed as a special case of the NFA universality problem (given an NFA, does it accept every word?). Like NFA universality, the mortality problem is PSPACE-complete [CITE SHALLIT].

Concerning $\mathbb{P}_{\pi_2}(E_{-\infty})$ (cf. Theorem 25.3), we actually show a stronger result, namely that any nontrivial approximation of $\mathbb{P}_{\pi_2}(E_{-\infty})$ is PSPACE-hard. The proof is also based on the mortality problem.

▶ **Proposition 26.** *There is a polynomial-time computable function that maps any instance of the mortality problem to an HMM and initial distributions π_1, π_2 so that if the instance is positive then $\mathbb{P}_{\pi_2}(E_{-\infty}) = 1$ and if the instance is negative then $\mathbb{P}_{\pi_2}(E_{-\infty}) = 0$. Thus, any nontrivial approximation of $\mathbb{P}_{\pi_2}(E_{-\infty})$ is PSPACE-hard.*

Proof sketch. Given an instance Φ of the mortality problem, construct an HMM (Q, Σ, Ψ) so that Φ and Ψ have the same graph [DEFINED?] and $\sum_{a \in \Sigma} \Psi(a)$ is stochastic. Define π_1 as a uniform distribution on Q. Define π_2 as a Dirac distribution on a fresh state that emits letters from Σ uniformly at random. Thus, if Φ is a positive instance of the mortality problem then $\mathbb{P}_{\pi_2}(E_{-\infty}) = 1$, and if Φ is a negative instance then $\mathbb{P}_{\pi_2}(E_{-\infty}) = 0$. ◀

4 Stefan

4.1 Representing Likelihood Exponents

In the following we show that, although it is hard to *compute* likelihood exponents, one can efficiently represent them in terms of *Lyapunov exponents*. The definition of Lyapunov exponents is based on the following definition.

▶ **Definition 27.** *A* matrix system *is a triple $\mathcal{M} = (Q, \Sigma, \Psi)$ where Q is a finite set of states, Σ is a finite set of observations, and $\Psi : \Sigma \to \mathbb{R}_{\geq 0}^{Q \times Q}$ specifies the transitions. (Note that an HMM is a matrix system.) A* Lyapunov system *is a pair $\mathcal{S} = (\mathcal{M}, \rho)$ where $\mathcal{M} = (Q, \Sigma, \Psi)$ is a matrix system and $\rho : \Sigma \to (0, 1]$ is a probability distribution on Σ with full support [PROBABLY INTRODUCE A NOTATION FOR A PROBABILITY DISTRIBUTION IN THE PRELIMINARIES.], such that the directed graph (Q, E) with $E = \{(q, r) \mid \sum_{a \in \Sigma} \Psi_{q,r}(a) > 0\}$ is strongly connected.*

We can identify the probability distribution ρ from this definition with the single-state HMM $(\{s\}, \Sigma, \Psi_\rho)$ where $\Psi_\rho(a)_{s,s} = \rho(a)$ for all $a \in \Sigma$. In this way, ρ produces a random infinite word from Σ^ω. For $w \in \Sigma^\omega$ we write w_n for the length-n prefix of w. The following lemma is known from [CITE].

▶ **Lemma 28** ([CITE])**.** *Let $((Q, \Sigma, \Psi), \rho)$ be a Lyapunov system. Then there is $\ell \in \mathbb{R}$ such that, for all $q \in Q$, $\mathbb{P}_\rho$-a.s., either $e_q \Psi(w_n) = \vec{0}$ [NOTATION DEFINED?] for some $n \in \mathbb{N}$ or the limit $\lim_{n \to \infty} \frac{1}{n} \ln \|e_q \Psi(w_n)\|$ exists and equals ℓ.*

For a Lyapunov system $\mathcal{S}$ we call $\ell(\mathcal{S}) = \ell$ from the lemma the *Lyapunov exponent* defined by $\mathcal{S}$. We prove the following theorem, which implies Theorem 11. [CHECK $|Q|^2$ vs. $|Q|^2 + 1$]

▶ **Theorem 29.** *Given an HMM (Q, Σ, Ψ) we can compute in polynomial time $2K \leq 2|Q|^2$ Lyapunov systems $\mathcal{S}_1^1, \mathcal{S}_1^2, \mathcal{S}_2^1, \mathcal{S}_2^2, \ldots, \mathcal{S}_K^1, \mathcal{S}_K^2$ such that for any initial distributions π_1, π_2 the limit $\lim_{n\to\infty} \frac{1}{n} \ln L_n$ exists $\mathbb{P}_{\pi_2}$-a.s. and lies in*

$$\Lambda \;\subseteq\; \{-\infty\} \cup \{\ell(\mathcal{S}_1^1) - \ell(\mathcal{S}_1^2), \ldots, \ell(\mathcal{S}_K^1) - \ell(\mathcal{S}_K^2)\}\,.$$

In particular, the HMM has at most $|Q|^2 + 1$ likelihood exponents.

In the rest of the section we provide more details on the construction underlying Theorem 29. As an intermediate concept (between the given HMM and the Lyapunov systems from Theorem 29) we define *generalized Lyapunov systems.*

First, for two matrix systems $\mathcal{M}_1 = (Q_1, \Sigma, \Psi_1)$ and $\mathcal{M}_2 = (Q_2, \Sigma, \Psi_2)$ with finite Q_1, Q_2, Σ and transitions $\Psi_1, \Psi_2 : \Sigma \to \mathbb{R}_{\geq 0}^{Q\times Q}$ we define the directed graph $G_{\mathcal{M}_1,\mathcal{M}_2} = (Q_1 \times Q_2, E)$ such that there is an edge from (q_1, q_2) to (r_1, r_2) if there is $a \in \Sigma$ with $\Psi_1(a)_{q_1,r_1} > 0$ and $\Psi_2(a)_{q_2,r_2} > 0$.

A *generalized Lyapunov system* is a triple $\mathcal{S} = (\mathcal{M}, \mathcal{H}, C)$ where $\mathcal{M} = (Q_1, \Sigma, \Psi_1)$ is matrix system and $\mathcal{H} = (Q_2, \Sigma, \Psi_2)$ is a strongly connected [DEFINED?] HMM and $C \subseteq Q_1 \times Q_2$ is a bottom SCC of $G_{\mathcal{M},\mathcal{H}}$. Given a generalized Lyapunov system, one can efficiently compute an "equivalent" Lyapunov system:

▶ **Lemma 30.** *Let $\mathcal{S} = ((Q_1, \Sigma, \Psi_1), (Q_2, \Sigma, \Psi_2), C)$ be a generalized Lyapunov system.*

1. *There is $\ell(\mathcal{S}) \in \mathbb{R}$ such that, for all $\pi_1 \in \mathbb{R}_{\geq 0}^{Q_1}$ and all probability distributions [NOTATION?] $\pi_2 \in [0,1]^{Q_2}$ with $\{q \mid \pi_1(q) > 0\} \times \mathrm{supp}(\pi_2) \subseteq C$, we have $\mathbb{P}_{\pi_2}$-a.s. that either $\pi_1\Psi_1(w_n) = \vec{0}$ [NOTATION?] for some $n \in \mathbb{N}$ or the limit $\lim_{n\to\infty} \frac{1}{n} \ln \|\pi_1\Psi_1(w_n)\|$ exists and equals $\ell(\mathcal{S})$.*
2. *One can compute in polynomial time a Lyapunov system $\mathcal{S}'$ such that $\ell(\mathcal{S}) = \ell(\mathcal{S}')$.*

Let $\mathcal{H} = (Q, \Sigma, \Psi)$ be an HMM. Let $R \subseteq Q \times Q$ be a (not necessarily bottom) SCC of the graph $G_{\mathcal{H},\mathcal{H}}$ such that $Q_2 := \{q_2 \in Q \mid \exists q_1 \in Q : (q_1, q_2) \in R\}$ is a bottom SCC of the graph of $\sum_{a\in\Sigma} \Psi(a)$. We call such R a *right-bottom* SCC. Clearly there are at most $|Q|^2$ right-bottom SCCs. Towards Theorem 29 we want to define, for each right-bottom SCC R, two generalized Lyapunov systems $\mathcal{S}_R^1, \mathcal{S}_R^2$.

For a function of the form $\Phi : \Sigma \to \mathbb{R}^{Q\times Q}$ and $P \subseteq Q$ we write $\Phi_{|P} : \Sigma \to \mathbb{R}^{P\times P}$ for the function with $\Phi_{|P}(a)(q, r) = \Phi(a)(q, r)$ for all $a \in \Sigma$ and $q, r \in P$; i.e., $\Phi_{|P}(a)$ denotes the principal submatrix obtained from $\Phi(a)$ by restricting it to the rows and columns indexed by P.

Define $\Psi'(a, r)_{q,r} := \Psi(a)_{q,r}$ for all $a \in \Sigma$ and $q, r \in Q$. Then $(Q, \Sigma \times Q, \Psi')$ is an HMM, which is similar to $\mathcal{H}$, but which emits, in addition to an observation from Σ, also the next state. Since Q_2 is a bottom SCC of the graph of $\sum_{a\in\Sigma} \Psi(a)$, the HMM $\mathcal{H}_2 := (Q_2, \Sigma \times Q_2, \Psi'_{|Q_2})$ is strongly connected. This HMM $\mathcal{H}_2$ will be used both in $\mathcal{S}_R^1$ and in $\mathcal{S}_R^2$.

Next, define $\overline{\Psi} : (\Sigma \times Q) \to [0,1]^{(Q\times Q)\times(Q\times Q)}$ by

$$\overline{\Psi}(a, r_2)_{(q_1,q_2),(r_1,r_2)} \;:=\; \Psi(a)_{q_1,r_1} \quad \text{for all } a \in \Sigma \;\text{ and } q_1, q_2, r_1, r_2 \in Q\,.$$

Now define $\mathcal{S}_R^1 := (\mathcal{M}^1, \mathcal{H}_2, C^1)$, where $\mathcal{M}^1 := (R, \Sigma \times Q_2, \overline{\Psi}_{|R})$ and $C^1 := \{((q_1, q_2), q_2) \mid (q_1, q_2) \in R\}$. Finally, denoting by $R' \subseteq Q_2 \times Q_2$ the SCC of the graph $G_{\mathcal{H},\mathcal{H}}$ that contains the "diagonal" vertices $(q, q) \in Q_2 \times Q_2$, define $\mathcal{S}_R^2 := (\mathcal{M}^2, \mathcal{H}_2, C^2)$, where $\mathcal{M}^2 := (R', \Sigma \times Q_2, \overline{\Psi}_{|R'})$ and $C^2 := \{((q_1, q_2), q_2) \mid (q_1, q_2) \in R'\}$.

For sets $U, V \subseteq Q \times Q$ write $U \longrightarrow^*_{G_{\mathcal{H},\mathcal{H}}} V$ denote that there are $u \in U$ and $v \in V$ such that v is reachable from u in $G_{\mathcal{H},\mathcal{H}}$. We are ready to state the following key technical lemma:

▶ **Lemma 31.** *Given an HMM (Q, Σ, Ψ), let $\mathcal{R} \subseteq 2^{Q\times Q}$ be the set of its right-bottom SCCs, and, for $R \in \mathcal{R}$, let $\mathcal{S}_R^1, \mathcal{S}_R^2$ be the generalized Lyapunov systems defined above. Then,*

$$\Lambda \subseteq \{-\infty\} \cup \{\ell(\mathcal{S}_R^1) - \ell(\mathcal{S}_R^2) \mid R \in \mathcal{R}\}.$$

NEW AND STRONGER VERSION:

▶ **Lemma 32.** *Given an HMM (Q, Σ, Ψ), let $\mathcal{R} \subseteq 2^{Q\times Q}$ be the set of its right-bottom SCCs, and, for $R \in \mathcal{R}$, let $\mathcal{S}_R^1, \mathcal{S}_R^2$ be the generalized Lyapunov systems defined above. Then, for any initial distributions π_1, π_2, the limit $\lim_{n\to\infty} \frac{1}{n} \ln L_n$ exists $\mathbb{P}_{\pi_2}$-a.s. and lies in*

$$\{-\infty\} \cup \{\ell(\mathcal{S}_R^1) - \ell(\mathcal{S}_R^2) \mid R \in \mathcal{R},\ \mathrm{supp}(\pi_1) \times \mathrm{supp}(\pi_2) \longrightarrow^*_{G_{\mathcal{H},\mathcal{H}}} R\}.$$

Thus, $\Lambda_{\pi_1,\pi_2} \subseteq \{-\infty\} \cup \{\ell(\mathcal{S}_R^1) - \ell(\mathcal{S}_R^2) \mid R \in \mathcal{R},\ \mathrm{supp}(\pi_1) \times \mathrm{supp}(\pi_2) \longrightarrow^*_{G_{\mathcal{H},\mathcal{H}}} R\}$.

Proof sketch. Let π_1, π_2 be initial distributions. Very loosely speaking, we show in the appendix that on $\mathbb{P}_{\pi_2}$-almost every run w there is a right-bottom SCC R which "traps" "most" of the mass of $\pi_1\Psi(w_n)$ and $\pi_2\Psi(w_n)$. This can be made meaningful and formal using (the cross-product systems) $\mathcal{S}_R^1, \mathcal{S}_R^2$. We then show that on $\mathbb{P}_{\pi_2}$-almost every such run w, for both $i = 1, 2$, the limit $\lim_{n\to\infty} \frac{1}{n} \ln \|\pi_i \Psi(w_n)\|$ exists and equals $\ell(\mathcal{S}_R^i)$ (or $\pi_1\Psi(w_n) = \vec{0}$ for some n). It follows that

$$\lim_{n\to\infty} \frac{1}{n} \ln L_n \;=\; \lim_{n\to\infty} \frac{1}{n} \ln \frac{\|\pi_1\Psi(w_n)\|}{\|\pi_2\Psi(w_n)\|} \;=\; \ell(\mathcal{S}_R^1) - \ell(\mathcal{S}_R^2). \qquad ◀$$

With Lemma 32 at hand, the proof of Theorem 29 is easy:

Proof of Theorem 29. As argued before, the set $\mathcal{R}$ of right-bottom SCCs of the given HMM has at most $|Q|^2$ elements. These right-bottom SCCs R and the associated generalized Lyapunov systems $\mathcal{S}_R^1, \mathcal{S}_R^2$ can be computed in polynomial time. By Lemma 32 we have $\Lambda = \bigcup_{\pi_1,\pi_2} \Lambda_{\pi_1,\pi_2} \subseteq \{-\infty\} \cup \{\ell(\mathcal{S}_R^1) - \ell(\mathcal{S}_R^2) \mid R \in \mathcal{R}\}$. By Lemma 30.2, for each $R \in \mathcal{R}$ one can compute in polynomial time an equivalent Lyapunov system. ◀

5 Computational Aspects

– We may define a mapping on this Markov chain onto a set of at most $|Q|^2$ HMMs such that
– In the case of deterministic HMM, this exponential markov chain has at most $|Q|^2$ reachable states.

▶ **Lemma 33.** *Let (Q, Σ, Ψ) be a strongly connected HMM with initial distribution π_1 and let $(\{q\}, \Sigma, \rho)$ be an HMM with a single state with initial distribution e_q. Then the likelihood exponent $\lim_{n\to\infty} \frac{1}{n} \ln L_n$ exists and $\mathbb{P}_{e_q}$-almost surely.*

$$\lim_{n\to\infty} \frac{1}{n} \ln L_n = \lim \frac{1}{n} \ln \frac{\|\pi_1 \times \pi_2 \overline{\Psi}\|}{\|\pi_2 \times \pi_2 \overline{\Psi}\|} = \lim \frac{1}{n} \ln \frac{\|\pi_1 \times \pi_2 \overline{\Psi}\|}{\|e_q \rho\|} - \lim \frac{1}{n} \ln \frac{\|\pi_2 \times \pi_2 \overline{\Psi}\|}{\|e_q \rho\|} \tag{5}$$

For the general HMMs we simulate the joint state distribution using a new HMM $(\overline{Q}, \overline{\Sigma}, \overline{\Psi})$. This HMM has state space $\overline{Q} = (Q \cup \{\perp\}) \times Q$ where the states $(\perp, q)$ are fresh states for each $q \in Q$. The set of observations is $\overline{\Sigma} = \Sigma$. Let $a \in \Sigma$ and let $l_1, l_2 \in Q \cup \{\perp\}$ and $r_1, r_2 \in Q$. We define

$$\overline{\Psi}(a, r)_{(l_1,r_1),(l_2,r_2)} = \begin{cases} \Psi(a)_{l_1,l_2} & l_1, l_2 \in Q,\ r = r_2,\ \Psi(a)_{r_1,r_2} > 0 \\ 1 & l_1 \in Q,\ l_2 = \perp,\ r = r_2,\ \sum_{s\in Q} \Psi(a)_{l_1,s} = 0 \\ 1 & l_1 = l_2 = \perp,\ r = r_2 \\ 0 & \text{else.} \end{cases}$$

▶ **Lemma 34.** *The triple* $(\overline{Q}, \overline{\Sigma}, \overline{\Psi})$ *is an HMM and* $\|\pi_1 \times \pi_2 \overline{\Psi}\Big((a_1, r_1), \ldots, (a_n, r_n)\Big)\| = \|\pi_1 \Psi(a_1 \ldots a_n)\|$

Proof. In the case that $l_1 \in Q$ and $\sum_{s \in Q} \Psi(a)_{l_1,s} > 0$ we have

$$\begin{aligned}\sum_{a\in\Sigma}\sum_{r\in Q}\sum_{l_2\in Q\cup\{\perp\}}\sum_{r_2\in Q}\overline{\Psi}(a,r)_{(l_1,r_1),(l_2,r_2)} &= \sum_{a\in\Sigma}\sum_{r\in Q}\sum_{l_2\in Q}\sum_{r_2\in Q}\delta_{r=r_1}\Psi(a)_{l_1,l_2}\sum_{b\in\Sigma}\Psi(b)_{r_1,r_2}\\ &= \sum_{a\in\Sigma}\sum_{l_2\in Q}\Psi(a)_{l_1,l_2}\sum_{b\in\Sigma}\sum_{r_2\in Q}\Psi(b)_{r_1,r_2}\\ &= 1.\end{aligned}$$

In the case that $l_1 \in Q$ and $\sum_{s \in Q} \Psi(a)_{l_1,s} = 0$ or in the case that $l_1 = \perp$,

$$\begin{aligned}\sum_{a\in\Sigma}\sum_{r\in Q}\sum_{l_2\in Q\cup\{\perp\}}\sum_{r_2\in Q}\overline{\Psi}(a,r)_{(l_1,r_1),(l_2,r_2)} &= \sum_{a\in\Sigma}\sum_{r\in Q}\sum_{r_2\in Q}\delta_{r=r_2}\Psi(a)_{r_1,r_2}\\ &= \sum_{a\in\Sigma}\sum_{r_2\in Q}\Psi(a)_{r_1,r_2}\\ &= 1.\end{aligned}$$

Now consider a word $(a_1, r_2), (a_2, r_3), \ldots, (a_n, r_{n+1}) \in (\Sigma \times Q)^n$. We have ◀

▶ **Lemma 35.**

Write $O : (Q \cup \{\perp\}) \times Q \to Q$ for the function $O(l, r) = r$. Let $\mathcal{R}$ be the set of strongly connected components of $(\overline{Q}, \overline{\Sigma}, \overline{\Psi})$. The image $O(\mathcal{R})$ is a set of strongly connected components of (Q, Σ, Ψ). We define $\mathcal{R}' = \{R \in \mathcal{R} \mid O(R)$ is a bottom connected component of $(Q, \Sigma, \Psi)\}$.

For $R \in \mathcal{R}'$ we may define an associated HMM (R, Σ', Ψ') where $\Sigma' = \Sigma \times Q \cup \{\$\}$ and

$$\Psi'(a)_{(l_1,r_1),(l_2,r_2)} = \{ \tag{6}$$

▶ **Proposition 36.** *Whether* $0 \in \Lambda$ *can be decided in polynomial time. If* $0 \in \Lambda$ *then* $\sup_{\alpha,\beta} \mathbb{P}_{\pi_2}(N_{\alpha,\beta} = \infty) = \mathbb{P}_{\pi_2}(E_0)$ *can be computed in PSPACE. It is PSPACE-complete to decide whether* $\mathbb{P}_{\pi_2}(E_0) = 1$.

$\pi_1 \times \pi_2 = \pi_1^T \pi_2$ we have $\|\pi_1 \times \pi_2 \overline{\Psi}(a, r_2)\| = \|\pi_1 \Psi(a)\|$

▶ **Lemma 37.** *It is PSPACE-hard to decide whether* $\mathbb{P}_{\pi_2}(E_0) = 1$.

Proof. We reduce from the universality decision problem in NFAs. Let $\mathcal{N} = (Q, \Sigma, \delta, q_0, F)$ be an NFA. Define the HMM (Q', Σ', Ψ) where $Q' = Q \cup \{q_\$, q_1\}$ and $\Sigma' = \Sigma \cup \{\$\}$. Then, we may define Ψ as

$$\Psi(a)_{i,j} = \begin{cases} \delta(a)_{i,j} / \sum_{k=1}^{|Q|} \delta(a)_{i,k} & a \in \Sigma,\ i \in Q/F,\ j \in Q \\ \delta(a)_{i,j} / (1 + \sum_{k=1}^{|Q|} \delta(a)_{i,k}) & a \in \Sigma',\ i \in F,\ j \in Q \\ 1/(1 + \sum_{k=1}^{|Q|} \delta(a)_{i,k}) & a \in \Sigma',\ i \in F,\ j = s_\$ \\ 1/|\Sigma'| & a \in \Sigma,\ i = j = q_1 \\ 1/|\Sigma'| & a = \$,\ i = q_1,\ j = q_\$ \\ 1 & i = j = q_\$ \\ 0 & \text{else.} \end{cases}$$

Next, let $\pi_1 = e_{q_0}$ and $\pi_2 = e_{q_1}$ be initial distributions. If every word is accepted by $\mathcal{N}$ then for every $w \in (\Sigma)^*$, the event $w\$\Sigma^\omega = E_0$ since $\lim_{n\to\infty} L_n > 0$. But up to a $\mathbb{P}_{\pi_2}$-null set, $\Sigma^*\$(\Sigma')^\omega = \Sigma^*(\Sigma')^\omega$ which is the set of words that can be generated starting the HMM at π_2. it follows that $\mathbb{P}_{\pi_2}(E_0) = 1$. If there is a word $w \in (\Sigma)^*$ which is not final, then the word $w\$$ can be generated with non-zero probability by π_2 but is a mortal word for π_1. Hence $\mathbb{P}_{\pi_2}(E_0) < 1$. ◀

▶ **Proposition 38.** *Deciding whether $-\infty \in \Lambda$ is PSPACE-complete. If $-\infty \in \Lambda$ then $\mathbb{E}_{\pi_2}[N_\perp | E_{-\infty}] < \infty$ and both $\mathbb{P}_{\pi_2}(E_{-\infty})$ and $\mathbb{E}_{\pi_2}[N_\perp | E_{-\infty}]$ are computable in PSPACE. It is PSPACE-complete to decide whether $\mathbb{P}_{\pi_2}(E_{-\infty}) = 1$.*

Proof. It follows that $\mathbb{P}_{\pi_2}(N = n) = u_{\pi_1,\pi_2} P^n \chi_{\mathcal{K}}$ since

$$\mathbb{P}_{\pi_2}(N = n) = \mathbb{P}_{\pi_2}(L_n(w) = 0) \iff \mathbb{P}_{\pi_2}(\|\pi_1\Psi(w)\| = 0) \iff \mathbb{P}_{\pi_2}(\mathrm{supp}(\pi_1\Psi(w)) = \mathbf{0}).$$

Therefore $\mathbb{E}_{\pi_2}[N]$ is equal to the expected hitting time of $\mathcal{K}$ in the Markov chain $(\mathcal{S}, P)$ which has $|Q| \times 2^{|Q|}$ states.

–system can be computed by a PSPACE transducer. – Kemeny, Snell book citation ◀

▶ **Lemma 39.** *Deciding whether $-\infty \in \Lambda$ is PSPACE-hard.*

Proof. Let the matrices $M_1, \ldots, M_D \in \{0,1\}^{N\times N}$ be an instance of the mortality problem. If the ith row of $\sum_{d=1}^D M_i$ is a zero row, we may define $M_1', \ldots, M_D'$ as the set of matrices $M_1, \ldots, M_D$ but with the ith row and column removed. $M_1', \ldots, M_D'$ is mortal if and only if $M_1, \ldots, M_D$ is mortal.

– if there is remaining state after the product M' then add an additional letter.

Therefore, without loss of generality, we may assume $\sum_{d=1}^D M_d$ contains no zero rows. The matrices $\overline{M}_1, \ldots, \overline{M}_D$ defined by $\overline{M}_{i,j} = M_{i,j} / \sum_{k=1}^N M_{i,k}$ are mortal if and only if $M_1, \ldots, M_D$ are mortal. Finally, define the HMM (Q, Σ, Ψ) where $Q = [N+1]$, $\Sigma = \{a_1, \ldots, a_d\}$ and Ψ is defined in blocks

$$\Psi(a_i) = \begin{pmatrix} \overline{M}_i & 0 \\ 0 & 1/|\Sigma| \end{pmatrix}.$$

Let $\pi_1 = (1/N, \ldots, 1/N, 0)$ and $\pi_2 = (0, \ldots, 0, 1)$. These initial distributions lead to a set of likelihood exponents Λ and It follows that $-\infty \in \Lambda$ if and only if $M_1, \ldots, M_d$ are mortal. ◀

Define an directed graph as follows. The vertex set is $V = (Q \cup \{\perp\}) \times Q$ where $\perp$ is a fresh state. There is an edge from $(q_1, q_2) \in V$ to $(r_1, r_2) \in V$ if any of the following three conditions hold.

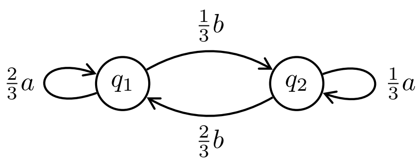
We say an HMM (Q, Σ, Ψ) is strongly connected if the Markov chain $\sum_{a\in\Sigma} \Psi(a)$ is strongly connected.

▶ **Theorem 40.** *We may compute in polynomial time a set of at most $|Q|^2$ strongly connected HMMs (Q_k, Σ, Ψ_k) and pairs of initial states (q_k, r_k) with corresponding likelihood ratios L_n^k such that for all $\ell \in \Lambda$ there is L_n^k such that $\lim_{n\to\infty} \frac{1}{n} \ln L_n^k = \ell$.*

Markov chain has states labelled (q_i, q_j) for $i, j = 1, \ldots, |Q|$ and also a special absorbing state $\perp$. The transition probabilities are taken from the outgoing transitions from q_j. It may be that a letter produced from q_j cannot be produced from q_i in which case the Markov chain transitions to $\perp$. The full definition is

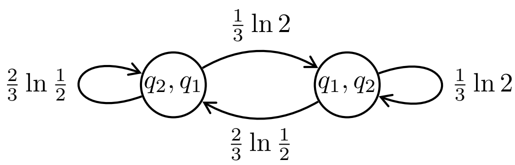
5.1 Deterministic Chains

A HMM (Q, Σ, Ψ) is deterministic if for all $a \in \Sigma$, all rows of $\Psi(a)$ contain exactly one non-zero entry. If q is a starting state, for any word in $w \in \Sigma$ the vector $\delta_q \Psi(w)$ contains at most one non-zero entry. A simple example of a deterministic chain is the following example.



.

Suppose two copies of the chain are started from states q_1 and q_2 respectively, if the chains emit the same word, they will be in opposite states. Let the chain started from q_1 emit word $aab \in \Sigma^3$. The chain will now be in state q_2. Suppose the next letter emitted is b. The value of $L_4 = L_3 \times \frac{2}{3}/\frac{1}{3} = 2L_3$ and $\ln L_4 = \ln L_3 + \ln 2$. In general deterministic chains have the property that $\ln L_n = R_n + \ln L_{n-1}$ where R_n is a random variable depending only the transition chosen. In the example above, we can build a Markov chain to characterise the joint distribution of states. We are interested in $\lim_{n\to\infty} \frac{1}{n} \ln L_n$ with words produced from starting state q_2 so the transition probabilities are lifted from the second chain. In the diagram below we also include the value of R_n on each transition.



.

The Markov chain has stationary distribution $(\frac{2}{3}, \frac{1}{3})$. By the strong ergodic theorem for Markov chains, $\lim_{n\to\infty} \frac{1}{n} \ln L_n = \frac{1}{3}\ln 2 + \frac{2}{3}\ln\frac{1}{2} = -\frac{1}{3}\ln 2$.

In general there may be a finite number of possible limits of $\lim_{n\to\infty} \frac{1}{n} \ln L_n$. Moreover, these limits may not be rational numbers but expressible as linear combinations of natural logarithms of rational numbers. We say a possible limit of $\lim_{n\to\infty} \frac{1}{n} \ln L_n$ is ln-symbolically computable if we can compute a set of $2M$ rational numbers where $M \le |Q|^2$, $\alpha_1, \dots, \alpha_M$ and $\beta_1, \dots, \beta_M$ such that

$$\lim_{n\to\infty} \frac{1}{n} \ln L_n = \sum_{l=1}^{M} \alpha_m \ln \beta_m.$$

For the general deterministic HMM we simulate the joint state distribution using a Markov chain M. This Markov chain has states labelled (q_i, q_j) for $i, j = 1, \dots, |Q|$ and also a special absorbing state $\perp$. The transition probabilities are taken from the outgoing transitions from q_j. It may be that a letter produced from q_j cannot be produced from q_i in which case the Markov chain transitions to $\perp$. The full definition is

$$M_{(i,j),(i^*,j^*)} = \sum_{a\in\Sigma} \Psi(a)_{j,j^*}\delta_{\Psi(a)_{i,i^*}>0}$$
$$M_{(i,j),\perp} = 1 - \sum_{(i^*,j^*)\in[Q\times Q]} M_{(i,j),(i^*,j^*)}$$
$$M_{\perp,\perp} = 1$$
$$M_{\perp,(i^*,j^*)} = 0.$$

We write T for the set of transitions of M. Let $C_0 = \{\perp\}$ and let $C_1, \dots, C_K$ be the remaining bottom connected components of M. It follows that the sets $C_0, \dots, C_K$ are disjoint.

▶ **Theorem 41.** *Let (Q, Σ, Ψ) be a deterministic HMM and let i_0, j_0 be the indices of starting states then we may compute a set of symbollic numbers $\{\lambda_1, \dots, \lambda_K\} \subseteq [-\infty, 0]$ and a probability distribution over these numbers given as $p_1, \dots, p_K$ in $O(|Q|^8)$ time such that,*

$$\mathbb{P}_{\delta_{j_0}}\Big(\lim_{n\to\infty} \frac{1}{n} \ln L_n = \lambda_k\Big) = p_k.$$

for all $k = 1, \dots, K \leq |Q|^2$.

Proof. First assume that the pair (i_0, j_0) is in the bottom connected component C_k with $k \geq 1$ of M. Fix $n \in \mathbb{N}$ then for any $a_1 \cdots a_n \in \Sigma^n$, since Ψ is deterministic, there are unique sets of states indexed by $i_1, \dots, i_n$ and $j_1, \dots, j_n$ such that

$$\frac{1}{n} \ln L_n = \frac{1}{n} \ln \frac{\Psi(a_1)_{i_0,i_1}\Psi(a_2)_{i_1,i_2} \dots \Psi(a_n)_{i_{n-1},i_n}}{\Psi(a_1)_{j_0,j_1}\Psi(a_2)_{j_1,j_2} \dots \Psi(a_n)_{j_{n-1},j_n}} = \frac{1}{n}\sum_{m=1}^{n} \ln \frac{\Psi(a_m)_{(i_{m-1},i_m)}}{\Psi(a_m)_{(j_{m-1},j_m)}}.$$

The state pair (i_m, j_m) has the same distribution under $\mathbb{P}_{\delta_{j_0}}$ as the state distribution of the sub-Markov chain M started from (i_0, j_0) after m transitions. For each transition in M there is a mapping $l : T \to (-\infty, 0]$ given by

$$l\big((i,j) \to (i^*, j^*)\big) = \ln \frac{\Psi(a)_{(i,i^*)}}{\Psi(a)_{(j,j^*)}}.$$

l is well defined because Ψ is deterministic and so a is unique to each transition. Let μ^k be the unique stationary distribution of the BCC C_k. Then by the strong Ergodic theorem, $\lim_{n\to\infty} \frac{1}{n} \ln L_n$ converges to

$$\sum_{(i,j)\to(i^*,j^*)\in T} \mu^k_{(i,j)} M_{(i,j),(i^*,j^*)} l((i,j) \to (i^*, j^*)).$$

If the starting state (i_0, j_0) is in C_0, then for all possible transitions from j_0 in Ψ, there are no transitions from i_0 and so it follows that $\lim_{n\to\infty} \frac{1}{n} \ln L_n = -\infty$. Computing the bottom connected components can be done using Tarjan's algorithm in $O(|Q|^2)$ time. Then, computing the stationary distributions μ^k for $k = 1, \dots, K$ takes $O(|Q|^6)$ time. Therefore, computing the possible $\lambda_1, \dots, \lambda_K$ takes $O(|Q|^6)$ time.

We may write $-\infty = \lambda_0, \dots, \lambda_K$ for the possible limits of $\frac{1}{n} \ln L_n$ starting in $C_0, \dots, C_K$ respectively. We write $C^{-1}(\lambda_k)$ for the union of bottom connected components associated with the likelihood exponent λ_k. In general, for starting states (i, j), $\lim_{n\to\infty} \frac{1}{n} \ln L_n = \lambda_k$ when the Markov chain M hits any of the states in $C^{-1}(\lambda_k)$. Therefore computing $p_1, \dots, p_k$ is equivalent to computing a hitting probability which can be done in $O(|Q|^6)$ because M has dimension at most $|Q|^2$. This makes the whole algorithm $O(|Q|^8)$ since $K \leq |Q|^2$. ◀

5.2 Non-Deterministic Chains

- show that it's atleast PSPACE hard to approximate and computing them is as hard as computing lyapunov exponents (known to be a hard problem).

By introducing non-determinism, the likelihood ratio can no longer be split into a simple product of random variables depending only on the transitions in a Markov chain. It follows that we may not use the Strong ergodic theorem for Markov chains to show convergence of $\lim_{n\to\infty}\frac{1}{n}\ln L_n$. We instead prove convergence using work by Protasov. The following theorem is proven in the appendix.

There is no known algorithm for computing each limit $\lim_{n\to\infty}\frac{1}{n}\ln L_n^k$. Moreover, even if we know the values of $\lim_{n\to\infty}\frac{1}{n}\ln L_n^k$, we show that computing the distribution of $\lim_{n\to\infty}\frac{1}{n}\ln L_n$ over the possible limits $\{\lim_{n\to\infty}\frac{1}{n}\ln L_n^k \mid k = 1,\dots,K\}$ is PSPACE-complete. We therefore consider consider a series of problems leading up to the final result.

The starting point is language inclusion for *non-deterministic finite automata* (NFA). An NFA is a quintuple (Q,Σ,Δ,q_0,F) such that The starting point is language inclusion for *non-deterministic finite automata* (NFA). An NFA is a quintuple (Q,Σ,Δ,q_0,F) such that

- a finite set of states Q.
- a finite set of input symbols Σ.
- a transition function $\Delta : Q\times\Sigma\to\mathcal{P}(Q)$.
- an initial state q_0
- a set of final states $F\subseteq Q$.

The language accepted by M, $L(M)$ is the set of $a_1\cdots a_n\in\Sigma^n$ such that there exists a sequence of states $q_1,\dots,q_n$ such that $q_i\in\Delta(q_{i-1},a_i)$ for $i=1,\dots,n$ and $q_n\in F$.

References

1 V.I. Bogachev. *Measure Theory*. Number v. 1 in Measure Theory. Springer Berlin Heidelberg, 2007. URL: `https://books.google.co.uk/books?id=CoSIe7h5mTsC`.

2 Taolue Chen and Stefan Kiefer. On the total variation distance of labelled Markov chains. In *Proceedings of the Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*, pages 33:1–33:10, Vienna, Austria, 2014.

3 Stefan Kiefer and A. Prasad Sistla. Distinguishing hidden Markov chains. In *Proceedings of the 31st Annual Symposium on Logic in Computer Science (LICS)*, pages 66–75, New York, USA, 2016. ACM.

4 Roland Langrock, Bruce Swihart, Brian Caffo, Naresh Punjabi, and Ciprian Crainiceanu. Combining hidden markov models for comparing the dynamics of multiple sleep electroencephalograms. *Statistics in medicine*, 32, 08 2013. `doi:10.1002/sim.5747`.

5 Vladimir Yu. Protasov. Asymptotics of products of nonnegative random matrices. *Functional Analysis and Its Applications*, 47:138–147, 2013.

6 A. Wald. Sequential Tests of Statistical Hypotheses. *The Annals of Mathematical Statistics*, 16(2):117 – 186, 1945. URL: `https://doi.org/10.1214/aoms/1177731118`, `doi:10.1214/aoms/1177731118`.

7 David Williams. *Probability with Martingales*. Cambridge University Press, 1991. `doi:10.1017/CBO9780511813658`.

▶ **Theorem 5.** *We have* $\lim_{n\to\infty} L_n = 0$ $\mathbb{P}_{\pi_2}$*-a.s. if and only if* π_1 *and* π_2 *are distinguishable.*

Proof. By proposition 6 in [2] the limit L_n exists $\mathbb{P}_{\pi_2}$-a.s. and the following equalities hold.

$$\begin{aligned}
1 - d(\pi_1, \pi_2) &= \lim_{n\to\infty} \sum_{w\in\Sigma^n} \|\pi_1\Psi(w)\| \wedge \|\pi_2\Psi(w)\| && \text{Theorem 7 of [2]} \\
&= \lim_{n\to\infty} \sum_{w\in\Sigma^n} (L_n(w) \wedge 1)\|\pi_2\Psi(w)\| \\
&= \lim_{n\to\infty} \mathbb{E}_{\pi_2}\big[L_n \wedge 1\big] \\
&= \mathbb{E}_{\pi_2}\big[\lim_{n\to\infty} L_n \wedge 1\big] && \text{because } 0 \le L_n(w) \wedge 1 \le 1.
\end{aligned}$$

Then, $\lim_{n\to\infty} L_n \wedge 1 = 0 \iff \lim_{n\to\infty} L_n = 0$. ◀

A Proofs from Section 3

▶ **Theorem 7.** *Suppose π_1 and π_2 are distinguishable. Let α, β be inputs to the SPRT. Then by choosing $A = \ln \frac{\alpha}{1-\beta}$ and $B = \ln \frac{1-\alpha}{\beta}$, $\mathbb{P}_{\pi_1}(\mathrm{SPRT}_{\alpha,\beta} = \pi_2) \le \alpha$ and $\mathbb{P}_{\pi_2}(\mathrm{SPRT}_{\alpha,\beta} = \pi_1) \le \beta$.*

Proof. We wish to control the probabilities $\mathbb{P}_{\pi_2}\big(L_N > B\big)$ and $\mathbb{P}_{\pi_1}\big(L_N < A\big)$ by choosing suitable values of A and B. Let $W_n^1 = \{w \in \Sigma^\omega \mid A \le L_m(w) \le B \ \forall m < n, L_n < A\}$ then

$$\begin{aligned}
\mathbb{P}_{\pi_1}\big(L_N < A\big) &= \sum_{n=1}^{\infty} \mathbb{P}_{\pi_1}\big(W_n^1\big) = \sum_{n=1}^{\infty} \sum_{w\in W_n^1} \pi_1\Psi(w)\mathbb{1}^T = \sum_{n=1}^{\infty} \sum_{w\in W_n^1} L_n(w)\pi_2\Psi(w)\mathbb{1}^T \\
&\le A \sum_{n=1}^{\infty} \sum_{w\in W_n^1} \pi_2\Psi(w)\mathbb{1}^T = A \sum_{n=1}^{\infty} \mathbb{P}_{\pi_2}\big(W_n^1\big) = A\mathbb{P}_{\pi_2}\big(L_N < A\big).
\end{aligned}$$

Similarly, we may derive $\mathbb{P}_{\pi_2}\big(L_N > b\big) \ge \frac{1}{b}\mathbb{P}_{\pi_1}\big(L_N > b\big)$ so it follows that

$$\begin{aligned}
A &\ge \frac{\mathbb{P}_{\pi_1}\big(L_N < A\big)}{\mathbb{P}_{\pi_2}\big(L_N < A\big)} = \frac{\mathbb{P}_{\pi_1}\big(L_N < A\big)}{1 - \mathbb{P}_{\pi_2}\big(L_N > B\big)} \\
B &\le \frac{\mathbb{P}_{\pi_1}\big(L_N > B\big)}{\mathbb{P}_{\pi_2}\big(L_N > B\big)} = \frac{1 - \mathbb{P}_{\pi_1}\big(L_N < A\big)}{\mathbb{P}_{\pi_2}\big(L_N > B\big)}
\end{aligned}$$

to guarantee the error bounds $\alpha = \mathbb{P}_{\pi_1}\big(L_{n^*} < A\big)$ and $\beta = \mathbb{P}_{\pi_2}\big(L_{n^*} > B\big)$. ◀

A.1 Proof of Theorem 40

▶ **Theorem 11.** *The set of likelihood exponents satisfies $\Lambda \subset [-\infty, 0]$ and $|\Lambda| \le |Q|^2$. Further $\lim_{n\to\infty} \frac{1}{n} \ln L_n$ exists (and by definition is in Λ) $\mathbb{P}_{\pi_2}$-almost surely for any π_1, π_2.*

The proof of Theorem 11 relies on related work in a subset of Ergodic Theory called *Lyapunov exponents*. The main papers are by Protasov [5] and Osedelets [**?**]. We first must define a similar object to a observation density matrix. A *random matrix product* is a triple (Q, Σ, Φ) where Q is a set of states, Σ is a set of letters and $\Phi : \Sigma \to [0,1]^{|Q|\times|Q|}$ is a non-negative matrix valued function. We may extend Φ to Σ^* in the same way as an observation density matrix. we use the shorthand $\Phi_n : \Sigma^n \to [0,1]^{|Q|\times|Q|}$ for Φ restricted to Σ^n. (Q, Σ, Φ) is strongly connected if for all $i, j \in [Q]$ there exists a $w \in \Sigma^*$ such that $\Phi(w)_{i,j} > 0$. (Q, Σ, Φ) is mortal if there exists a $w \in \Sigma^*$ such that $\Phi(w) = 0$. The main theorem by Protasov is stated below.

▶ **Theorem 42. Protasov's Theorem** *Let (Q, Σ, Φ) be a strongly connected random matrix product, let $\mathbb{P}_{ind}$ be an i.i.d probability measure on Σ^ω and let π be an initial distribution. If (Q, Σ, Φ) is mortal, then $\lim_{n\to\infty} \frac{1}{n} \ln \pi\Phi_n \mathbb{1}^T = -\infty$ $\mathbb{P}_{ind}$-a.s. otherwise there is a $\lambda \in (-\infty, 0]$ such that $\lim_{n\to\infty} \frac{1}{n} \ln \pi\Phi_n \mathbb{1}^T = \lambda$ $\mathbb{P}_{ind}$-a.s.*

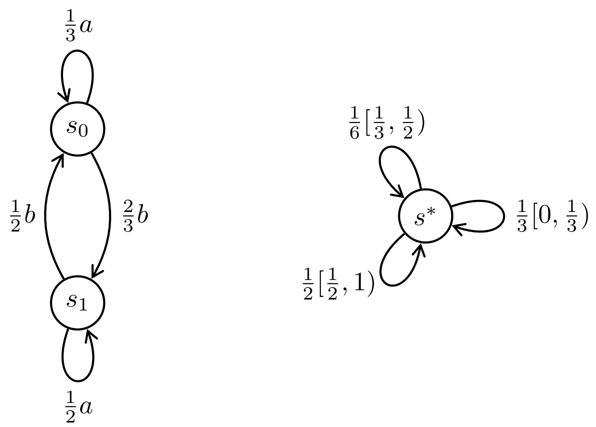
The technique we use to prove Theorem 40 involves producing an i.i.d measure $\mathbb{P}_{\mathrm{ind}}$ and random matrix product $(Q \times Q, \Sigma, \Phi)$ such that for all $n \in \mathbb{N}$ and initial distributions π,

$$\mathbb{P}_{\mathrm{ind}}\Big(\frac{1}{n} \ln \pi\Phi_n \mathbb{1}^T \in A\Big) = \mathbb{P}_{\pi_2}\Big(\frac{1}{n} \ln \pi\Psi_n \mathbb{1}^T \in A\Big).$$

Since $\lim_{n\to\infty} \frac{1}{n} \ln L_n = \lim_{n\to\infty} \frac{1}{n} \ln \pi_1\Psi_n\mathbb{1}^T - \lim_{n\to\infty} \frac{1}{n} \ln \pi_2\Psi_n\mathbb{1}^T$ such a random matrix product would imply convergence of $\lim_{n\to\infty} \frac{1}{n} \ln L_n$ due to Protasov's theorem.

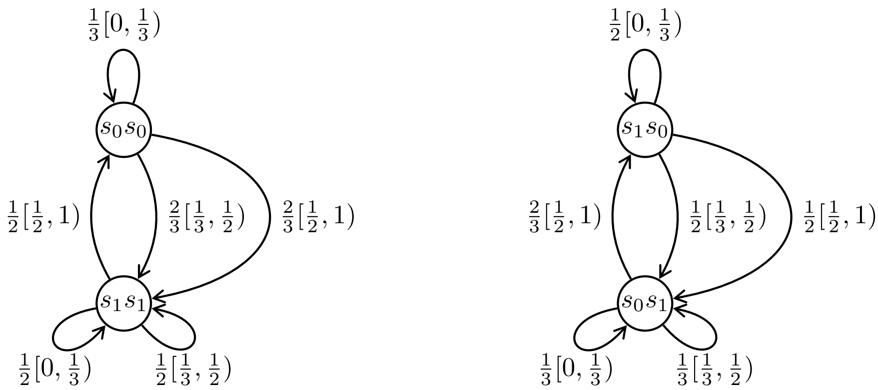
When the probability space on infinite words is derived from a general HMM, the probability of a particular letter being produced at a specific position in the infinite word depends on the state the producing HMM is in. In the constructed random matrix product, the current state of the producing HMM as started from π_2 is incorporated into the state space of the HMM started from π_1.

To accomplish this, we simulate transitions in the producing HMM (the chain started from π_2) by sampling a uniform random number in the interval $[0, 1)$. At each state, we may partition $[0, 1)$ so that each sub-interval corresponds to specific transition and the size of the sub-interval corresponds to the probability of said transition. The union over all states of these partitions has a minimal finite σ-algebra. The atoms of this σ-algebra are also a partition of $[0, 1)$ and so we may sample them independently at random with probabilities according to their size. The transformation on a simple example is demonstrated in the diagram below. On the left is the original HMM. On the right is a single state HMM representing $\mathbb{P}_{\mathrm{ind}}$. For example, with probability $\frac{1}{6}$, it produces the label $[\frac{1}{3}, \frac{1}{2})$ which corresponds to the b transition in s_0 or the a transition in s_1.



We now construct the random matrix product $(Q \times Q, \Sigma, \Phi)$ which can also be represented by a state transition system (but without any stochastic properties). In our example, $(Q \times Q, \Sigma, \Phi)$ consists of two strongly connected components. Each state is labelled with a left and right component state corresponding to the current states of the HMM started from

π_1 and π_2 respectively. The transition weights are taken from the transitions in π_1 and the transition letters are taken from the transitions in π_2 because this random matrix product simulates $\frac{1}{n}\ln \pi_1 \Psi_n \mathbb{1}^T$ under the probability measure $\mathbb{P}_{\pi_2}$.



A.1.0.1 Random Matrix Product Construction

Consider the HMM (Q,Σ,Ψ) with finite alphabet Σ. Since for each $i \in [Q]$, $\sum_{a\in\Sigma}\sum_{j=1}^{|Q|}\Psi(a)_{i,j} = 1$ it follows that we may define a function $\rho_i : [0,1) \to Q\times\Sigma$ such that for all $i \in [Q]$, $\mathcal{M}_{Leb}(\rho_i^{-1}\{(j,a)\}) = \Psi(a)_{i,j}$. Consider the minimal σ-algebra $\sigma\{\rho_i^{-1}\{(j,a)\} \mid i,j\in[Q], a\in\Sigma\}$ which is finite and has a set of atomic elements P of at most $|Q|^2|\Sigma|$ elements. P is also a partition of $[0,1)$. Let $p\in P$ then $\rho_i(x)$ is constant for all $x\in p$ so we may overload the notation and consider the function $\rho_i : P \to Q\times\Sigma$. We then extend ρ_i to $\rho_i : P^n \to (Q\times\Sigma)^n$ by iteratively defining $\rho_i(ua) = \rho_i(u)\rho_{l(\rho_i(u))}(a)$ for $u\in P^n$ and $a\in P$.

We will describe the one-state chain $(\{1\},P,\Psi_P)$ as the *singleton generator* for Ψ where $\Psi_P(p) = \mathcal{M}_{Leb}(p)$. Given an initial distribution π for (Q,Σ,Ψ), a word generated by its singleton generator uniquely defines a path of states and letters. Let $\mathbb{P}_{\text{ind}}$ be the measure on the set of infinite words P^ω generated by this HMM.

In order to incorporate the state space of the producing HMM into the state space of the chain started from π_1, we define two functions $l : Q\times\Sigma \to Q$ and $r : Q\times\Sigma\to\Sigma$ where $l(q,a) = q$ and $r(q,a) = a$. We then define the matrix valued function $\Phi : P \to [0,1]^{(Q\times Q)\times(Q\times Q)}$ as

$$\Phi(p)_{(i_1,j_1),(i_2,j_2)} = \begin{cases} \Psi(r\circ\rho_{j_1}(p))_{i_1,i_2} & l\circ\rho_{j_1}(p) = j_2 \\ 0 & \text{else.}\end{cases}$$

Further for initial distributions $\pi_1,\pi_2 \in [0,1]^{|Q|}$ we define $\pi_1\times\pi_2 = \pi_1^T\pi_2 \in [0,1]^{|Q|\times|Q|}$.

▶ **Proposition 43.** *Consider the HMM (Q,Σ,Ψ) with initial distributions π_1 and π_2 then for any measurable set $A\in[-\infty,0]$ and for all $n\in\mathbb{N}$*

$$\mathbb{P}_{\pi_2}\Big(\frac{1}{n}\ln L_n \in A\Big) = \mathbb{P}_{ind}\Big(\frac{1}{n}\ln\frac{\pi_1\times\pi_2\Phi_n\mathbb{1}}{\pi_2\times\pi_2\Phi_n\mathbb{1}} \in A\Big).$$

Proof. Let $i_0, j_0 \in [Q]$ be initial states, It follows that for a word $u_1\dots u_n \in \{\rho_{j_0}(u_1\dots u_n) = (j_1,a_1),\dots,(j_n,a_n)\}$,

$$
\begin{aligned}
\|e_{i_0}^T e_{j_0} \Phi(u)\| &= \sum_{(i_1,\ldots,i_n)\in[Q]^n} \Phi(u_1)_{(i_0,j_0),(i_1,j_1)} \ldots \Phi(u_n)_{(i_{n-1},j_{n-1}),(i_n,j_n)} \\
&= \sum_{(i_1,\ldots,i_n)\in[Q]^n} \Psi(a_1)_{i_0,i_1} \ldots \Psi(a_n)_{i_{n-1},i_n} \\
&= \|e_{i_0}\Psi(a_1\ldots a_n)\|.
\end{aligned}
$$

It follows that for any initial distributions π_1, π_2, $\|\pi_1^T \pi_2 \Psi^*(u)\| = \|\pi_1 \Psi(a_1 \ldots a_n)\|$. Then considering the word $a_1, \ldots, a_n$,

$$
\begin{aligned}
\mathbb{P}_{\pi_2}(a_1 \ldots a_n) &= \|\pi_2 \Psi(a_1 \ldots a_n)\| \\
&= \sum_{j_1,\ldots,j_n\in[Q]^n} \|\pi_2 \Psi(a_1)_{\pi_2,j_1} \ldots \Psi(a_n)_{j_{n-1},j_n}\| \\
&= \sum_{u_1,\ldots,u_n\in\{\rho_{\pi_2}(u_1,\ldots,u_n)=(j_1,a_1),\ldots,(j_n,a_n)\}} \Psi_P(u_1\ldots u_n) \\
&= \mathbb{P}_{\text{ind}}(l \circ \rho_{\pi_2}(u_1 \ldots u_n) = a_1 \ldots a_n)
\end{aligned}
$$

◀

Consider the bottom connected components $C_1, \ldots, C_k \subseteq Q \times Q$ of the Markov chain defined by $\sum_{p\in P} \Phi(p)$ and let $d : Q \times Q \to Q \times Q$ be defined as $d(s_1, s_2) = (s_2, s_2)$. The following lemmas give the $|Q|^2$ bound in Theorem 40.

▶ **Lemma 44.** *Let s be a starting state for an HMM (Q, Σ, Ψ). Then $\lim_{n\to\infty} \frac{1}{n} \ln \|(s, s)\Psi_n^*\|$ takes at most $|Q|$ values.*

Proof. Let $P_1, \ldots, P_K$ be the irreducible components of Φ. Consider states $(s_1, s_2), (r_1, r_2) \in Q \times Q$. If there is a path from (s_1, s_2) to (r_1, r_2) then there is also a path from (s_2, s_2) to (r_2, r_2). Therefore for any end component P_i it follows that the image $d(P_i) \subseteq P_j$ for some end component P_j and so we may define a function $\rho : \{P_1, \ldots, P_K\} \to \{P_1, \ldots, P_K\}$ such that $\rho(P_i) = P_j$. Suppose P_i, P_j have Lyapunov exponents λ_i and λ_j respectively. Let π_1 and π_2 be initial distributions such that the support of π_1 is in P_i and the support of π_2 is in P_j then the likelihood ratio $L_n = \frac{\|\pi_1\Psi_n\|}{\|\pi_2\Psi_n\|}$ converges to a limit in the set $[0, \infty)$ with respect to the measure $\mathbb{P}_{\pi_2}$, the same limit as $\frac{\|(\pi_1,\pi_2)\Psi_n^*\|}{\|(\pi_2,\pi_2)\Psi_n^*\|}$ with respect to $\mathbb{P}_{\text{ind}}$. Since both $\frac{1}{n} \ln \|(\pi_1, \pi_2)\Psi_n^*\|$ and $\frac{1}{n} \ln \|(\pi_2, \pi_2)\Psi_n^*\|$ converge almost surely in the set $[-\infty, 0]$ to λ_i and λ_j respectively, $\frac{1}{n} \ln \frac{\|\pi_1\Psi_n\|}{\|\pi_2\Psi_n\|}$ converges in $[-\infty, 0]$ with respect to $\mathbb{P}_{\pi_2}$. Therefore $\lambda_i \leq \lambda_j$. Now consider $\lim_{n\to\infty} \frac{1}{n} \ln \|(s, s)\Psi_n^*\|$ whose possible limits is bounded by $|Q|^2$. Suppose for some word $w \in P^n$ the support of $(s, s)\Psi_n^*$ intersects an irreducible component P_i. Then it must also intersect P_j. Since $\lambda_i \leq \lambda_j$ it follows that $\lim_{n\to\infty} \frac{1}{n} \ln \|(s, s)\Psi_n^*\| \geq \lambda_j$. Since the image $\rho\{P_1, \ldots, P_K\} \leq |Q|$ it follows that $\lim_{n\to\infty} \frac{1}{n} \ln \|(s, s)\Psi_n^*\|$ takes at most $|Q|$ values. ◀

▶ **Lemma 45.** *Consider an HMM (Q, Σ, Ψ). Let $\mathcal{E} : [0, 1]^{|Q|} \times [0, 1]^{|Q|} \to [\infty, 0]$ be a parametrised random variable on Σ^ω defined by $\mathcal{E}(\pi_1, \pi_2) = \lim_{n\to\infty} \frac{1}{n} \ln \frac{\|\pi_1\Psi(w)\|}{\|\pi_2\Psi(w)\|}$. Then $|\{\mathcal{E}(\pi_1, \pi_2) \mid \pi_1, \pi_2 \in [0, 1]^{|Q|}\}| \leq |Q|^2$ with respect to the measure $\mathbb{P}_{\pi_2}$.*

Proof. First consider the case of dirac distributions $\pi_1 = \delta_r$ and $\pi_2 = \delta_s$. We may instead consider a bound on

$$\lim_{n\to\infty} \frac{1}{n} \ln \frac{\|(\delta_r, \delta_s)\Psi_n^*\|}{\|(\delta_s, \delta_s)\Psi_n^*\|} = \lim_{n\to\infty} \frac{1}{n} \ln \|(\delta_r, \delta_s)\Psi_n^*\| - \lim_{n\to\infty} \frac{1}{n} \ln \|(\delta_s, \delta_s)\Psi_n^*\|$$

with respect to the $\mathbb{P}_{\text{ind}}$ measure. Let $C_1, \dots, C_K$ be the irreducible lethal components of Φ. For $L \le K$ without loss of generality suppose $C_1, \dots, C_L$ be the irreducible components that are also end components containing diagonal entries. Let $R_1, \dots, R_L \subseteq Q$ be disjoint and have the property that for all $q \in R_i$, $(q, q) \in C_i$.

Given a state $q_i \in Q$ any letter in P defines a unique subsequent state q_j and a unique letter produced a. Therefore, projecting $(\delta_p, \delta_q)\Phi_n$ onto its right component, yields a point distribution on some state. Therefore the function $\zeta : \{(\delta_p, \delta_q)\Phi(w) \mid w \in P^*\} \to Q$ defined by $\zeta((\delta_p, \delta_q)\Phi(w)) = r$ if and only if $\operatorname{supp} \sum_{i=1}^{|Q|} ((\delta_p)_i, \delta_q)\Phi(w))_{i,j} = (\delta_r)_j$ for all j is well defined.

We may partition Σ^ω into $W_1, \dots, W_L$ such that $\zeta((\delta_r, \delta_s)\Phi_n(w))$ hits all states in R_k infinitely often for $w \in W_k$. It follows that $(\delta_s, \delta_s)\Phi_n(w)$ intersects the end component C_k and hits no other end components with diagonal entries. let $q \in C_k$ then $\frac{1}{n} \ln(\delta_s, \delta_s)\Phi_n(w)$ must converge almost surely on W_k to the Lyapunov exponent given by $\lim_{n\to\infty} \frac{1}{n} \ln(\delta_q, \delta_q)\Phi_n(w)$.

Similarly $\zeta((\delta_r, \delta_s)\Phi_n(w)) \in R_k$ and so $(\delta_r, \delta_s)\Phi_n(w)$ is contained in the set $Q \times R_k$ for $w \in W_k$. Since $\zeta((\delta_r, \delta_s)\Phi_n(w))$ hits all states in R_k infinitely often each irreducible component P_i such that $P_i \le (\delta_r, \delta_s)\Phi_n(w)$ must have the property that $|P_i| \ge |R_k|$. Therefore the total number of irreducible components hit by $(\delta_r, \delta_s)\Phi_n(w)$ where $w \in W_k$ is at most $|Q|$. Since $L \le |Q|$ the total number of possible limits for

$$\lim_{n\to\infty} \frac{1}{n} \ln \frac{\|(\delta_r, \delta_s)\Psi_n^*\|}{\|(\delta_s, \delta_s)\Psi_n^*\|}$$

is $|Q|^2$. It remains to show that $\{\mathcal{E}(\pi_1, \pi_2) \mid \pi_1, \pi_2 \in [0,1]^{|Q|}\} \subseteq \{\mathcal{E}(\delta_r, \delta_s) \mid r, s \in Q\}$. Fix π_1 and π_2 and let us consider the possible values of

$$\lim_{n\to\infty} \frac{1}{n} \ln \|(\pi_1, \delta_s)\Psi_n^*\| - \lim_{n\to\infty} \frac{1}{n} \ln \|(\pi_2, \delta_s)\Psi_n^*\|$$

where $s \in \operatorname{supp} \pi_2$. A consider again the partition of $\Sigma^\omega = \bigcup_{k=1}^{L} W_k$. For $w \in W_k$, the only end component in $C_1, \dots, C_L$ hit by $(\pi_2, \delta_s)\Phi_n(w)$ is C_k. It follows that $\lim_{n\to\infty} \frac{1}{n} \ln \|(\pi_2, \delta_s)\Psi_n^*\| = \lim_{n\to\infty} \frac{1}{n} \ln \|(\delta_s, \delta_s)\Psi_n^*\|$. Any strongly connected component hit by $(\pi_1, \delta_s)\Psi_n^*(w)$, is also hit by $(\delta_r, \delta_s)\Psi_n^*(w)$ for some $r \in \operatorname{supp}\pi_1$. It follows that we may partition Σ^ω so that on each part of the partition there is some $s, r \in Q$ such that

$$\lim_{n\to\infty} \frac{1}{n} \ln \frac{\|(\pi_1, \delta_s)\Psi_n^*\|}{\|(\pi_2, \delta_s)\Psi_n^*\|} = \lim_{n\to\infty} \frac{1}{n} \ln \frac{\|(\delta_r, \delta_s)\Psi_n^*\|}{\|(\delta_s, \delta_s)\Psi_n^*\|}.$$

◀

We may now prove Theorem 40.

Proof of Theorem 40. We may compute Φ is $O(|Q|^4|\Sigma|)$ time and a set of strongly connected components using Tarjan's algorithm $P_1, \dots, P_K \subset Q \times Q$. On each connected component, any initial distribution converges to a constant or $-\infty$ by Theorem 42 and Proposition 43. The maximum of $|Q|^2$ possible limits is a result of and Lemma 45.

for each $k \in [K]$, $(r(P_k) \cup l(P_k), \Sigma, \Psi_{\restriction l(P_k)} \bigoplus \Psi_{\restriction r(P_k)})$ is a lossy HMM. Let $(q_k, r_k) \in P_k$ then δ_{q_k} and δ_{r_k} are initial distributions such that the corresponding likelihood ratios $\lim_{n\to\infty} \frac{1}{n} \ln L_n^k$ satisfy the requirements of the second half of the theorem. ◀

A.2 Proof of the Asymptotic Wald Formula

▶ **Theorem 12.** *[Generalised Wald Formula] Let ℓ be a likelihood exponent and let π_1 and π_2 be initial distributions. If $\ell \in (-\infty, 0)$ then*

$$\mathbb{E}_{\pi_2}\Big[N_{\alpha,\beta} \mid E_\ell\Big] \sim \frac{\ln \alpha}{\ell} \quad (\textit{as } \alpha, \beta \to 0).$$

If $\ell = 0$ then there exist $\alpha, \beta > 0$ such that

$$\mathbb{E}_{\pi_2}\Big[N_{\alpha,\beta} \mid E_\ell\Big] = \infty. \tag{4}$$

If $\ell = -\infty$ then

$$\sup_{\alpha,\beta} \mathbb{E}_{\pi_2}\Big[N_{\alpha,\beta} \mid E_\ell\Big] < \infty.$$

▶ **Lemma 46.** *There is a number $c > 0$, computable in polynomial time, such that*

$$\mathbb{P}_{\pi_2}\Big(L_{2|Q|n} \geq \exp(-\frac{c^2}{36}n)\Big) \leq 4\exp\Big(-\frac{c^2}{36}n\Big).$$

Proof. By Proposition 17, $\lim_{\alpha,\beta\to 0} \frac{N_{\alpha,\beta}}{\ln \alpha}$ exists $\mathbb{P}_{\pi_2}(\cdot \mid V_k)$-almost surely. Hence, the convergence is also in $\mathbb{P}_{\pi_2}(\cdot \mid V_k)$-measure. Therefore, by the Vitali convergence theorem[1] it is sufficient to show that the set of random variables $\{\frac{N_{\alpha,\beta}}{\ln \alpha} \mid \alpha, \beta \in (0, \frac{1}{2})\}$ is uniformly integrable conditioned on V_k. In fact, because

$$\lim_{K\to\infty} \sup_{\alpha,\beta} \mathbb{E}_{\pi_2}\big[\frac{N_{\alpha,\beta}}{-\ln \alpha}\mathbb{1}_{\frac{N_{\alpha,\beta}}{-\ln\alpha} \geq -K}\big] \geq \mathbb{P}_{\pi_2}(V_k) \lim_{M\to\infty} \sup_{\alpha,\beta} \frac{1}{-\ln \alpha}\mathbb{E}_{\pi_2}\big[\frac{N_{\alpha,\beta}}{-\ln \alpha}\mathbb{1}_{\frac{N_{\alpha,\beta}}{-\ln\alpha} \geq -K} \mid V_k\big], \tag{7}$$

It is sufficient to check the uniform integrability condition without conditioning on V_k.

For fixed $M \geq \frac{144|Q|}{c^2}$, write $m_\alpha = \lfloor \frac{-M \ln \alpha}{2|Q|} \rfloor$. It follows that $\frac{2|Q|m_\alpha}{\ln \alpha} \leq M$ and $\alpha \geq \exp -\frac{c^2}{36}m_\alpha$. Further, $m_\alpha \geq \frac{M \ln 2}{2|Q|} - 1$ The following series of equalities hold

$$\begin{aligned}
\mathbb{E}_{\pi_2}\big[\frac{N_{\alpha,\beta}}{-\ln \alpha}\mathbb{1}_{\frac{N_{\alpha,\beta}}{-\ln\alpha} \geq M}\big] &= \frac{1}{-\ln \alpha}\sum_{n=0}^{\infty} \mathbb{P}_{\pi_2}\big(N_{\alpha,\beta}\mathbb{1}_{N_{\alpha,\beta} \geq 2|Q|m_\alpha}\big] \qquad \leq \frac{2|Q|}{-\ln \alpha}\Big(m_\alpha \\
&\leq M\mathbb{P}_{\pi_2}\big(L_{2|Q|m_\alpha} \geq \alpha\big) + \frac{2|Q|}{-\ln \alpha}\sum_{n=m_\alpha}^{\infty} \mathbb{P}_{\pi_2}\big(L_{2|Q|n} \geq \alpha\big) \\
&\leq M\mathbb{P}_{\pi_2}\big(L_{2|Q|m_\alpha} \geq \exp -\frac{c^2}{36}m_\alpha\big) + \frac{2|Q|}{-\ln \alpha}\sum_{n=m_\alpha}^{\infty} \mathbb{P}_{\pi_2}\big(L_{2|Q|n} \geq \exp -\frac{c^2}{36}n\big) \\
&\leq 4M\exp -\frac{c^2}{36}m_\alpha + \frac{8|Q|}{-\ln \alpha}\sum_{n=m_\alpha}^{\infty} \exp -\frac{c^2}{36}n \\
&\leq 4M\exp -\frac{c^2}{36}m_\alpha + \frac{8|Q|\exp -\frac{c^2}{36}m_\alpha}{-\ln \alpha}\,\frac{1}{1-\exp c^2/36} \\
&\leq 4M\exp -\frac{c^2}{36}\Big(\frac{M\ln 2}{2|Q|} - 1\Big) + \frac{8|Q|\exp\Big(-\frac{c^2}{36}(\frac{M\ln 2}{2|Q|} - 1)\Big)}{\ln 2}\,\frac{1}{1-\exp c^2/36} \\
&\to 0
\end{aligned}$$

as $K \to \infty$ where the fourth inequality follows by Lemma 46. Hence, Equation (7) must hold. ◀

► **Lemma 47.** *The events* $\{L_n \to 0\}$ *and* $\{\lim_{n\to\infty} \frac{1}{n} \ln L_n < 0\}$ *are equal up to a* $\mathbb{P}_{\pi_2}$*-null set.*

Proof. Fix $A < \alpha < 0$ and define the event $W_n = \{1 > L_n \geq \mathrm{e}^{n\alpha}\}$. Then

$$\begin{aligned}
\mathbb{P}_{\pi_2}(\lim_{n\to\infty} \frac{1}{n} \ln L_n > \alpha) &\leq \mathbb{P}_{\pi_2}(\liminf_n \{\frac{1}{n} \ln L_n \geq \alpha\}) \\
&\leq \liminf_n \mathbb{P}_{\pi_2}(\frac{1}{n} \ln L_n \geq \alpha) \\
&\leq \liminf_n \mathbb{P}_{\pi_2}(L_n \geq \mathrm{e}^{n\alpha}) \\
&= \liminf_n \left[\mathbb{P}_{\pi_2}(1 > L_n \geq \mathrm{e}^{n\alpha}) + \mathbb{P}_{\pi_2}(L_n \geq 1)\right] \\
&\leq \liminf_n \left[\sum_{w\in W_n} \pi_2 \Psi(w) \mathbb{1}^T + e^{An+B}\right] \\
&\leq \liminf_n \left[e^{-n\alpha} \sum_{w\in W_n} \pi_1 \Psi(w) \mathbb{1}^T\right] \\
&\leq \liminf_n \left[e^{-n\alpha} \mathbb{P}_{\pi_1}(L_n < 1)\right] \\
&\leq \liminf_n \left[e^{-n\alpha} e^{An+B}\right] \\
&= 0.
\end{aligned}$$

◄

Now suppose π_1 and π_2 are distinguishable. By Theorem 5 of [3] one may compute a $c \in (-\infty, 0)$ such that

$$\begin{aligned}
\mathbb{P}_{\pi_1}(L_n < 1) - \mathbb{P}_{\pi_2}(L_n < 1) &\geq 1 - 2e^{cn} \\
&\geq 1 - 2e^{n \max \Lambda}
\end{aligned}$$

For a given HMM (Q, Σ, Ψ), we may define a *monitor* as a function $M_n : \Sigma^n \to \{1, 2\}$. A well designed monitor reads an input word from an HMM started with either π_1 or π_2 and aims to return 1 or 2 respectively with high probability. However the following series of inequalities hold

$$\begin{aligned}
\mathbb{P}_{\pi_2}(M_n(w) = 2) - \mathbb{P}_{\pi_1}(M_n(w) = 2) &= \mathbb{P}_{\pi_2}(M_n(w) = 1) + \mathbb{P}_{\pi_1}(M_n(w) = 2) \\
&= \sum_{w\in\Sigma^n} \pi_1 \Psi(w) \mathbb{1}^T \delta_{M_n(w)=1} + \pi_2 \Psi(w) \mathbb{1}^T \delta_{M_n(w)=2} \\
&\geq \sum_{w\in\Sigma^n} \pi_1 \Psi(w) \mathbb{1}^T \wedge \pi_2 \Psi(w) \mathbb{1}^T \\
&= \mathbb{P}_{\pi_2}(L_n \leq 1) - \mathbb{P}_{\pi_1}(L_n \leq 1)
\end{aligned}$$

which means for any monitor, to guarantee an error probability bound of at most ϵ, we require atleast $\frac{\ln(\epsilon) - \ln 2}{\max \Lambda}$ observations. This bound motivates us to investigate computability properties of Λ.