Change Point Detection in Hidden Markov Models

Abstract

The problem of quick detection, with low false-alarm rates, of abrupt changes in stochastic dynamic systems arises in a variety of applications, including industrial quality control, segmentation of signals, financial engineering, biomedical signal processing, edge detection in images, and the diagnosis of faults in the elements of computer communication networks. Typical of such a problem in segmentation of signals is that of using an automatic segmentation of the signal as the first processing step, and a segmentation algorithm that splits the signal into homogenous segments, the lengths of which are adapted to the local characteristics of the analyzed signal. The main desired properties of a segmentation algorithm are few false alarms and missed detections, and low detection delay.

A hidden Markov model is a doubly stochastic process with an underlying stochastic process that is not directly observable but can be observed only through another set of stochastic processes that produce the sequence of observations. The model, when applied properly, has been shown to work very well in practice for several important application areas, such as speech recognition, molecular biology, ion channel, economics, digital communications over unknown channels, and bioinformatics. In the standard formulation of the change point detection problem, there is a sequence of observations whose distribution changes at some unknown time, and the goal is to detect this change as soon as possible under false alarm constraints. In this paper, we investigate the performance of the Shiryayev-Roberts-Pollak (SRP) rule for change point detection in the dynamic system of hidden Markov models. First, we present a second-order asymptotic approximation for the expected value of such stopping scheme. Next, we show that the SRP procedure is asymptotically optimal.

Introduction

A hidden Markov model (HMM) is a doubly stochastic process with an underlying stochastic process that is not directly observable (it is hidden) but can be observed only through another set of stochastic processes that produce the sequence of observations. There are two reasons why hidden Markov modeling have become increasingly popular. First the model is very rich in mathematical structure and hence can form the theoretical basis for use in a wide range of applications. The model, when applied properly, has been shown to work very well in practice for several important application areas, such as speech recognition, molecular biology, ion channel, economics, digital communications over unknown channels, and bioinformatics.

The problem of quick detection, with low false-alarm rates, of abrupt changes in stochastic dynamic systems arises in a variety of applications, including industrial quality control, segmentation of signals, financial engineering, biomedical signal processing, edge detection in images, and the diagnosis of faults in the elements of computer communication networks. In the standard formulation of the change point detection problem, there is a sequence of observations whose distribution changes at some unknown time, and the goal is to detect this change as soon as possible under false alarm constraints.

One of the pioneering works in change point detection is Shewhart's "3-sigma control charts" for controlling manufacturing processes. It is simple and yet an important tool for monitoring the stationarity of a process. More efficient procedures are the Page's cumulative sum procedure (CUSUM) and the Shiryayev-Roberts-Pollak (SRP) algorithm. When the observations ξ_n are independent with a common density function f^{θ_0} before the change and with another common density function f^{θ_1} after the change, a Bayesian formulation was proposed by Shiryayev (1963), in which the goal is to minimize the expected delay subject to an upper bound on false alarm probability. Roberts (1966) considered the non-Bayesian setting, and studied by simulation the average run length of this rule, and found it to be very good. Pollak (1985) showed that the (modified) Shiryayev-Roberts rule is asymptotically optimal under a modified formulation.

At present there is a great deal of literature on detection algorithms in complex systems but relatively little on the statistical properties and optimality theory of detection procedures beyond very simple models. The primary goal of this study was to investigate theoretical aspect of the SRP change point detection rule in hidden Markov models. First, we present a second-order asymptotic approximation for the expected stopping time of such stopping scheme. Next, we show that the SRP procedure is asymptotically minimax in the sense of Pollak (1985).

Change point detection in HMM

For each $\theta \in \Theta \subset \mathbf{R}$, the unknown parameter, we consider $\mathbf{X}=\{X_n, n \geq 0\}$ as an ergodic Markov chain on a finite state space $D=\{1, 2, \ldots, d\}$, with transition probability matrix $P(\theta)=[P_{xy}(\theta)]_{xy=1,\ldots,d}$ and stationary distribution $\pi(\theta)=(\pi_x(\theta))_{x=1,\ldots,d}$ Suppose that an additive component ξ_n, taking values in $\mathbf{R}$, is adjoined to the chain such that $\{(X_n, \xi_n), n \geq 0\}$ is a Markov chain on $D \times \mathbf{R}$, satisfying $P^\theta\{X_1 \in A \mid X_0 = x, \xi_0 = s\} = P^\theta\{X_1 \in A \mid X_0 = x\}$ for $A \subset D$. And conditioning on the full $\mathbf{X}$ sequence, ξ_n is a Markov chain with probability

$$P^\theta\{\xi_{n+1} \in B \mid X_0, X_1, \ldots; \xi_0, \xi_1, \ldots, \xi_n\} = P^\theta\{\xi_{n+1} \in B \mid X_{n+1}; \xi_n\} \quad a.s. \tag{2.1}$$

for each and the Borel -algebra of . Furthermore, we assume

$$P^\theta\{X_1 \in A, \xi_1 \in B \mid X_0 = x, \xi_0 = s_0\} = \sum_{y \in A} \int_B p_{xy}(\theta) f(s; \varphi_y(\theta) \mid s_0) d\mu(s), \tag{2.2}$$

where $f(\xi_k; \psi_{X_k}(\theta) \mid \xi_{k-1})$ is the transition probability density of ξ_k given ξ_{k-1}

and X_k with respect to μ, and $\varphi_y(\cdot)$ is a function defined on the parameter space Θ for each $y=1,\ldots,d$. We call a process $\{\xi_n, n\geq 0\}$ a hidden Markov model if there is a Markov chain $\{X_n, n\geq 0\}$ such that the process $\{(X_n, \xi_n), n\geq 0\}$ satisfies (2.1) and (2.2).

Let $\xi_0, \xi_1, \ldots, \xi_{\omega-1}$ be the observations from the hidden Markov model $\{\xi_n, n\geq 0\}$ with distribution P^{θ_0}, and let $\xi_\omega, \xi_{\omega+1},\ldots$ be the observations from the hidden Markov model $\{\xi_n, n\geq 0\}$ with distribution P^{θ_1}. Both θ_0 and θ_1 are given, while the change point ω is unknown. We shall use P_ω to denote such a probability (with change time ω) and use P_∞ to denote the case $\omega=\infty$ (no change point). Denote E_ω as the corresponding expectation under P_ω. A detection scheme is a stopping time on the sequence of observations and to minimize the number of post change observations. Here we want to find a stopping time N to minimize

$$\sup_{1\leq k<\infty} E_k(N-K \mid N\geq K) \tag{2.3}$$

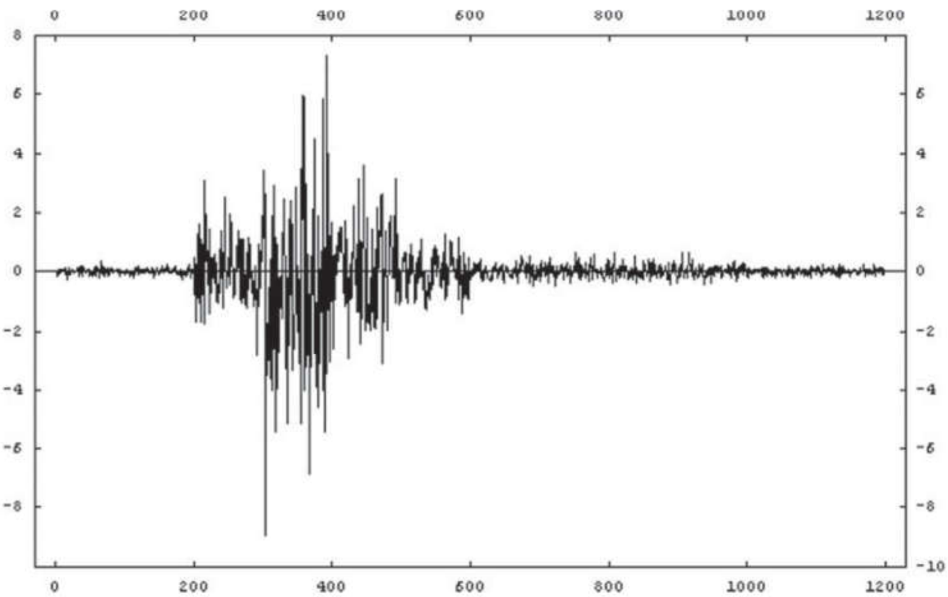
subject to $E_\infty N\geq\gamma$, for some specified (large) constant γ. A detection scheme is called asymptotically minimax, if it minimizes (2.3), within an $o(1)$ order, among all stopping rules that satisfy $E_\infty N\geq\gamma$, where $o(1)\longrightarrow$ as $\gamma\longrightarrow\infty$.

Let $\xi_0, \xi_1, \ldots, \xi_n$ be the observations given from the hidden Markov model $\{\xi_n, n\geq 0\}$. For $0\leq k\leq n$, let $LR_n^k = Pn(\xi_k, \xi_{k+1},\ldots,\xi_n;\theta_1)/P_n(\xi_k,\xi_{k+1},\ldots,\xi_n;\theta_0)$ Given an approximate threshold $B>0$ and setting $b=\log B$, the Shiryayev-Roberts scheme is defined as

$$N_b{:}=\inf\{n{:}\sum_{k=0}^{n} LR_n^k\geq B\}=\inf\{n{:}\log\sum_{k=0}^{n} LR_n^k\geq b\}. \tag{2.4}$$

A simple modification of (2.4) was given by Fuh (2004a) by adding a randomization on the initial LR_n^0.

An example of speech signal segmentation.



Asymptotic optimality of the SRP algorithm

By using the same notations as those in Fuh (2004a), for given P^{θ_0} and P^{θ_1}, the Kullback-Leibler information numbers are defined as

$$K(P^{\theta_0},P^{\theta_1})=E_{P^{\theta_0}}\left(\log\frac{\| M_1(\theta_0)M_0(\theta_0)\pi^{\theta_0}\|}{\| M_1(\theta_1)M_0(\theta_1)\pi^{\theta_1}\|}\right),\quad K(P^{\theta_1},P^{\theta_0})=E_{P^{\theta_1}}\left(\log\frac{\| M_1(\theta_1)M_0(\theta_1)\pi^{\theta_1}\|}{\| M_1(\theta_0)M_0(\theta_0)\pi^{\theta_0}\|}\right).$$

To derive a second-order approximation for the average run length of the SRP rule, we will apply relevant results from nonlinear Markov renewal theory developed in Section 3 of Fuh (2004a). For this purpose, we rewrite the stopping time $N_b:=N_b^{\Psi}$ in the form of a Markov random walk crossing a constant threshold plus a nonlinear term that is slowly changing. Note that the stopping time can be written in the following form

$$N_b=\inf\{n\geq 1: S_n+\eta_n\geq b\},\quad b=\log B, \tag{3.1}$$

where S_n is defined in (2.15) of Fuh(2004a)with $E_1S_1=K(P^{\theta_1},P^{\theta_0})$,and $\eta_n=\log\{1+\sum_{k=1}^{n-1}e^{-S_k}\}$.

For $b>0$, define $N_b^*=\inf\{n\geq 1: S_n\geq b\}$, and let $R_b=S_{N_b^*}-b$ (on $\{N_b^*<\infty\}$).When $b=0$, we denote N_b^* as N_+^*. For given $\tilde{w}\in X$, it is known [Theorem 1 of Fuh (2004a)] that $\lim_{b\to\infty}E_1(R_b|W_0=\tilde{w})=E_{m_+}S_{N_+^*}^2/2E_{m_+}S_{N_+^*}$,where m_+ is defined in Fuh (2004a).

Note that $S_{N_b}=b-\eta_{N_b}+\chi_b$ on $\{N_b<\infty\}$ where $\chi_b=S_{N_b}+\eta_{N_b}-b$. Taking the expectations on both sides, we obtain

$$\begin{aligned}&K(P^{\theta_1},P^{\theta_0})\,E_1(N_b|W_0=\tilde{w})\textstyle\int_X\Delta(w')\,dm_+(w')+\Delta(\tilde{w})\\&=E_1(S_{N_b}|W_0=\tilde{w})=b-E_1(\eta_{N_b}|W_0=\tilde{w})+E_1(\chi_b|W_0=\tilde{w}),\end{aligned}\tag{3.2}$$

where $\Delta: X\to R$ solves the Poisson equation $E_w\Delta(W_1)-\Delta(w)=E_wS_1-E_mS_1$ for almost every $w\in X$ with $E_m\Delta(W_1)=0$.

Theorem 1. Assume C1-C2 in Fuh (2004a) hold.

i) Assume that S_1 is nonarithmetic with respect to P_∞ and P_1. If $0<K(P^{\theta_1},P^{\theta_0})<\infty$, $0<K(P^{\theta_0},P^{\theta_1})<\infty$, and $E_1|S_1|^2<\infty$, then for $w\in X$, as $b\to\infty$

$$E_1(N_b|W_0=\tilde{w})=\frac{1}{K(P^{\theta_1},P^{\theta_0})}\left(b-E_{m_+}\eta+\frac{E_{m_+}S_{N_+^*}^2}{2(E_{m_+}S_{N_+^*})}-\textstyle\int_x\Delta(w)dm_+(w)+\Delta(\tilde{w})\right)+o(1). \tag{3.3}$$

ii) Suppose that the P_∞-distribution of LR_1 is nonarithmetic. Then for any $1<\gamma<\infty$, there exists a constant $\delta<b=\log B<\infty$ and a probability measure Ψ such that $\gamma=E_\infty N_b$ and such that if N is any stopping time which satisfies $E_\infty N\geq\gamma$, then

$$\sup_{1\leq\omega<\infty}E_\omega(N-\omega|N\geq\omega)\geq\sup_{1\leq\omega<\infty}E_\omega(N_b^{\Psi}-\omega|N_b^{\Psi}\geq\omega)+o(1), \tag{3.4}$$

where $o(1)\to 0$ as $\gamma\to\infty$, $E_\omega(N_b^{\Psi}-\omega|N_b^{\Psi}\geq\omega)$ is a constant for $1<\omega<\infty$.

Conclusion

It is worth noting that while the SRP rule (2.4) is asymptotically minimax in independent cases [Pollak (1985)], it is nontrivial whether this is still true in hidden Markov models? By making use of a Markov chain representation of the likelihood ratio LR_n^k, and a nonlinear Markov renewal theory, this piece of work provided a definite answer to this longstanding problem. These methods were developed in a series of papers by Fuh, and have been shown to be a useful tool for theoretical investigation, including likelihood estimation and change point detection, in hidden Markov models. Since hidden Markov model plays an important role in the study of problems raised in many applied fields, a deeper understanding of its theoretical properties, finding a suitable way for model fitting, and providing efficient computational methods for the models may lead to both interesting and important.

Cheng-Der Fuh

Institute of Statistical Science, Academia Sinica

Annals of Statistics **32**, 2305-2339 (**2004**)

References

1. Fuh, C. D. (2004a). Asymptotic operating characteristics of an optimal change point detection in hidden Markov models. *Ann. Statist*. 2305-2339.
2. Fuh, C. D. (2004b). Uniform Markov renewal theory and ruin probabilities in Markov random walks. *Ann. Appl. Probab*. 1202-1241.
3. Pollak, M. (1985). Optimal detection of a change in distribution. *Ann. Statist*. 206-227.
4. Roberts, S. W. (1966). A comparison of some control chart procedures. *Technometrics* 411-430.
5. Shiryayev, A. N. (1963). On optimum methods in quickest detection problems. *Theory Probab. Appl*. 22-46.