https://doi.org/10.1109/CDC.1998.758231

Faster parameter estimation using risk-sensitive filters

Sanjeewa Athuraliya†, Jason Ford‡ and John Moore*

† Department of Electrical and Electronic Engineering, University of Melbourne, Australia.

‡ Weapons Systems Division, Defence Science and Technology Organisation, Australia;

Email: Jason.Ford@dsto.defence.gov.au

*Department of Systems Engineering Australian National University, Australia.

Abstract

In this paper, we propose a risk-sensitive approach to parameter estimation for hidden Markov models (HMMs). The parameter estimation approach considered exploits estimation of various functions of the state, based on model estimates. We propose certain practical suboptimal risk-sensitive filters to estimate the various functions of the state during transients, rather than optimal risk-neutral filters as in earlier studies. The estimates are asymptotically optimal, if asymptotically risk neutral, and can give significantly improved transient performance, which is a very desirable objective for certain engineering applications.

To demonstrate the improvement in estimation simulation studies are presented that compare parameter estimation based on risk-sensitive filters with estimation based on risk-neutral filters.

1 Introduction

Hidden Markov models (HMMs) are a powerful tool in the field of signal processing [2] with application to speech processing[8], digital communication systems[3] and biological signal processing[5].

Hidden Markov models in discrete time can be viewed as having a state X_k at time k belonging to a discrete set, without loss of generality denoted as $S = \{e_1, e_2, ..., e_N\}$, where e_i is a vector that is zero everywhere excepting the ith element which is 1. There are transitions between states described by fixed probabilities which form a matrix $A = (a_{ij})$ where a_{ij} is the probability of transferring from state e_j to state e_i. Observations of the Markov state are made. We consider a Poisson observation process, where the Markov chain modulates the rate of a observed Poisson process.

Several schemes for estimating the parameters of a HMM have been proposed including an EM approach [1] and RPE approaches [4, 6]. All these approaches require estimation of various functions of the state. For example, the number of transitions between the possible states of the Markov chain is used to estimate the transition probability matrix. Previous approaches use risk-neutral filters conditioned on model estimates to generate estimates of various state quantities. However, risk-neutral filters are only optimal in the trivial case when the model estimates are equal to the true model.

In [7] optimal risk-sensitive filters and smoothers for known hidden Markov models, are proposed. The risk-sensitive filter is finite dimensional and evanesces to the risk-neutral filter as the risk parameter θ approaches zero. Simulation studies illustrate that the risk-sensitive filter can perform better than the risk-neutral filter on limited finite data when the HMM is not known accurately.

The key proposal of this paper is that *risk-sensitive* filters should be used to improve performance of certain HMM parameter estimators. Unfortunately, optimal risk-sensitive filters for the various functions of the states are computational prohibitive and we propose suboptimal versions. In this paper we consider off-line estimation of HMM parameters.

This paper is organized as follows: In Section 2 we introduce the notation used for HMMs in this paper. In Section 3 the parameter estimation problem is introduced. In Section 4 the risk-sensitive filtering problem is introduced and our suboptimal risk-sensitive approach to parameter estimation is proposed. In Section 5 simulation studies are presented. Finally, in Section 6 some conclusions are presented.

2 State Dynamics, and Observation Process

2.1 The State Process

Let X_k be a discrete-time homogeneous, first order Markov process, belonging to a finite set. The state space, X, *without loss of generality*, can be identified with a set of unit vectors, $S = \{e_1, e_2, .., e_N\}, e_i = (0, \dots, 0, 1, 0, \dots, 0)' \in A^N$ with 1 in the ith position. The transition probability matrix is

$$A = (a_{ij}) \text{ for } 1 \leq i, j \leq N$$

where $a_{ij} = P(X_{k+1} = e_i | X_k = e_j)$, so that

$$E[X_{k+1}|X_k] = AX_k \tag{1}$$

where $E[.]$ denotes the expectation operator. We also denote $\{\mathcal{F}_\ell, \ell \in \mathcal{Z}^+\}$ the complete filtration generated by X, that is, for any $k \in \mathcal{Z}^+, \mathcal{F}_k$ is the complete filtration generated by $X_\ell, \ell \leq k$. For a brief introduction of the concept of filtration in this context see [2].

Lemma 2.1 *The dynamics of X_k are given by the state equation*

$$X_{k+1} = AX_k + V_{k+1} \tag{2}$$

where V_{k+1} is a $(A, \mathcal{F}_k)$ martingale increment, in that $E[V_{k+1}|\mathcal{F}_k] = 0$.

Proof: See [2].

We can also write the initial state probability vector for the Markov chain as $\pi = (\pi_i)$ where $\pi_i = P(X_1 = e_i)$.

2.2 The Observation Process

We assume X_k is hidden, that is, indirectly observed by measurements y_k (a Poisson process modulated by X_k). That is, the state modulates the rate (ie. the rate z_k is given by $z_k = \lambda X_k$, where λ is a vector of Poisson rates) of an observed Poisson process. Hence we have observations $\{y_0, y_1, \dots\}$ that obey the following Poisson density:

$$P\left((y_k - y_{k-1}) = n | x_k\right) = \frac{z_k^n}{n!} e^{-z_k}, n = 0, 1, \dots \tag{3}$$

The HMM described by (2),(3) is denoted by $M^P = (A, \lambda, \pi)$.

3 Parameter Estimation

In this section we consider the problem of estimating the parameters of a hidden Markov model (2),(3) from observations y_k.

From manipulation of (2), by multiplication by X_k' and summing over k we obtain

$$\sum_{\ell=1}^{k} X_{\ell+1} X_\ell' = A \sum_{\ell=1}^{k} X_\ell X_\ell' + \sum_{\ell=1}^{k} V_k X_\ell', \tag{4}$$

or

$$\mathcal{J}_k = A\mathcal{O}_k + \sum_{\ell=1}^{k} V_k X'_\ell \quad (5)$$

where

$$\mathcal{J}_k = \sum_{\ell=1}^{k} X_{\ell+1}X'_\ell \quad \text{and} \quad \mathcal{O}_k = \sum_{\ell=1}^{k} X_\ell X'_\ell, \quad (6)$$

A reasonable estimate of A is $\hat{A}^*_k = \mathcal{J}_k\left(\mathcal{O}_k\right)^{-1}$.

To estimate λ we note that (3) can be rewritten as

$$y_k - y_{k-1} = \lambda X_k + \omega_k \quad (7)$$

where ω_k is zero mean $\omega_k \geq -\lambda X_k$ and $(\lambda X_k + \omega_k) \in Z^+$ where Z^+ is the set of non-negative integers. Hence, a reasonable estimate of λ is $\hat{\lambda}^*_k = \bar{\mathcal{T}}_k\left(\mathcal{O}_k\right)^{-1}$ where

$$\bar{\mathcal{T}}_k = \sum_{\ell=1}^{k} \bar{y}_\ell X'_\ell \quad (8)$$

and where $\bar{y}_\ell = y_\ell - y_{\ell-1}$.

3.1 Conditional mean estimates

When $\mathcal{J}_k, \mathcal{O}_k$ and $\bar{\mathcal{T}}_k$ are not available directly we work with estimates of these quantities. Conditional mean estimates can be obtained using the filters given in [2] if the true model is known. Thus define,

$$\begin{aligned} \mathcal{J}^X_k &:= X_k \text{row vec } \mathcal{J}_k \in \Re^{N\times N^2} & (9)\\ \text{row vec } \mathcal{J}_k &= \underline{1}'\mathcal{J}^X_k \quad \in \Re^{1\times N^2}. & (10)\end{aligned}$$

Conditional mean estimates are obtained as follows[2],

$$\begin{aligned} \sigma(\mathcal{J}^X_{k|k}) &= AB(y_k)\sigma(\mathcal{J}^X_{k-1|k-1})\\ &\quad + ((Ae_1)_{\text{diag}},\ (Ae_2)_{\text{diag}}, \ldots, (Ae_N)_{\text{diag}})\\ &\quad \times (B(y_k)(\alpha_{k-1})_{\text{diag}} \otimes I)\\ \text{where} &\quad \text{row vec } \sigma(\mathcal{J}_{k|k}) = \underline{1}'\sigma(\mathcal{J}^X_{k|k}) \quad (11)\\ \sigma(\bar{\mathcal{T}}^X_{k|k}) &= AB(\bar{y}_k)\sigma(\bar{\mathcal{T}}^X_{k-1|k-1})\\ &\quad + AB(\bar{y}_k)(\alpha_{k-1})_{\text{diag}}\bar{y}_k\\ \text{where} &\quad \text{row vec } \sigma(\bar{\mathcal{T}}_{k|k}) = \underline{1}'\sigma(\bar{\mathcal{T}}^X_{k|k}). \quad (12)\end{aligned}$$

where $\sigma(\mathcal{J}^X_{k|k})$ is the unnormalised conditional mean estimate for J^X_k. These recursions assume knowledge of M and hence these filters are termed risk-neutral (RN) filters.

Off-line Parameter Estimation

In off-line estimation, after each pass through the data set, parameter estimates are obtained and the model estimate is updated. That is, given the data set $\bar{y}_0, \ldots, \bar{y}_T$ we calculate estimates for $\mathcal{J}_T, \mathcal{O}_T$ and $\bar{\mathcal{T}}_T$ based on the last off-line model, $\hat{M}^P_{\ell-1}$, obtained after the $\ell - 1$th pass through the data as follows

$$\begin{aligned} \hat{\mathcal{J}}_{T|T,\hat{M}^P_{\ell-1}} &= E\left[\mathcal{J}_T \middle| \hat{M}^P_{\ell-1}\right], & (13)\\ \hat{\mathcal{O}}_{T|T,\hat{M}^P_{\ell-1}} &= E\left[\mathcal{O}_T \middle| \hat{M}^P_{\ell-1}\right], & (14)\\ \hat{\bar{\mathcal{T}}}_{T|T,\hat{M}^P_{\ell-1}} &= E\left[\bar{\mathcal{T}}_T \middle| \hat{M}^P_{\ell-1}\right]. & (15)\end{aligned}$$

Then, after passing through the data our new parameter estimates are:

$$\begin{aligned} \bar{A}_\ell &= \hat{\mathcal{J}}_{T|T,\hat{M}^P_{\ell-1}}\left(\hat{\mathcal{O}}_{T|T,\hat{M}^P_{\ell-1}}\right)^{-1},\\ \bar{\lambda}_\ell &= \hat{\mathcal{J}}_{T|T,\hat{M}^P_{\ell-1}}\left(\hat{\mathcal{O}}_{T|T,\hat{M}^P_{\ell-1}}\right)^{-1} & (16)\end{aligned}$$

The model estimate, $\hat{M}^P_\ell$, is then created for use in the next pass and $\mathcal{J}_k$, $\mathcal{O}_k$ and $\bar{\mathcal{T}}_k$ are again estimated.

Local convergence of $\bar{A}_\ell$ to A and $\bar{\lambda}_\ell$ to λ follows by applying the EM algorithm [8].

These conditional mean filters (11) – (12) are optimal for pass ℓ if $\hat{M}^P_\ell = M^P$. However, it is the nature of the estimation problem that in general $\hat{M}^P_\ell \neq M^P$. When $\hat{M}^P_\ell \neq M^P$ the RN filter may have poor performance even when the error in $\hat{M}^P_\ell$ is small (ie. $\hat{M}^P_\ell$ is close M^P). The question explored in this paper is whether or not RS filters can be shown to perform better than RN filters when used for the purpose of parameter estimation. RS filters do not assume an average noise situation, but rather move towards a worse case noise situation, which may represent in an useful way this sort of model uncertainty. In the following section we investigate RS filters and parameter estimation.

4 Risk-Sensitive Filters

In this section we define the RS filtering task and propose practical RS filters for the $\mathcal{J}_k$, $\mathcal{O}_k$, and $\bar{\mathcal{T}}_k$ quantities. Estimators for the model parameters A and λ are then proposed.

Risk-Sensitive Performance Index

The RS estimation problem is to determine the estimate $\hat{\mathcal{X}}_k$ such that

$$\hat{X}_k = \arg\min_{\zeta};\, J_k(\sigma) \text{ for all } k = 0, 1, \ldots, \tag{17}$$

where

$$J_k(\zeta) = E\left[\theta \exp\left(\theta \Psi_{0,k}(\zeta)\right) \middle| \mathcal{Y}_k\right] \tag{18}$$

is the RS cost function. Here,

$$\Psi_{0,k}(\zeta) = \hat{\zeta}_{0,k-1} + \frac{1}{2}(X_k - \zeta)' Q_k (X_k - \zeta), \tag{19}$$

where $Q_k \geq 0$ for all k and

$$\hat{\Psi}_{m,n} = \frac{1}{2}\sum_{i=m}^{n}\left(X_i - \hat{X}_i\right)' Q_i \left(X_i - \hat{X}_i\right) \tag{20}$$

This RS filtering problem for HMMs was solved by Dey and Moore [7].

Risk-Sensitive State Estimation

In [7] it is shown that a information state, α_k, for the state X_k is given by

$$\alpha_{k+1} = A D_k B_k \alpha_k \tag{21}$$

where

$$D_k = diag\left\{\exp\left(\frac{\theta}{2}\left(e_1 - \hat{X}_k\right)' Q_k \left(e_1 - \hat{X}_k\right)\right), \ldots, \exp\left(\frac{\theta}{2}\left(e_N - \hat{X}_k\right)' Q_k \left(e_N - \hat{X}_k\right)\right)\right\} \tag{22}$$

and

$$B_k = diag\left\{\frac{(\lambda e_1)^n}{n!} e^{-(\lambda e_1)}, \ldots, \frac{(\lambda e_N)^n}{n!} e^{-(\lambda e_N)}\right\}. \tag{23}$$

Note that for the Poisson process the differences, $y_k - y_{k-1}$, can be considered the observations. By comparing with the standard HMM filter (ie. $\alpha_{k+1} = A B_k \alpha_k$) it is clear that the RS filter is different only in that the diagonal matrix B_k has been modified to the diagonal matrix $D_k B_k$. The D_k matrix has the effect of increasing the tails of the noise probability density in an appropriate manner. This can be interpreted as allowing for parameter uncertainty by allowing for more observation noise.

Here we are interested in RS filters for the quantities $\mathcal{J}_k$, $\mathcal{O}_k$ and $\bar{\mathcal{T}}_k$. The RS filtering problem is computationally intensive for $\mathcal{J}_k, \mathcal{O}_k$ and $\bar{\mathcal{T}}_k$ because it requires maximization over N^k elements, at each time instant k so in the next section we consider suboptimal risk-sensitive filters for these quantities.

Proposed Suboptimal Parameter Estimators

A key result of this paper is to propose suboptimal RS filters in which the noise model is modified to allow for parameter uncertainty. Our pseudo RS filter for quantities $\mathcal{J}_k, \mathcal{O}_k$ and $\bar{\mathcal{T}}_k$ are constructed from the optimal RN filter by replacing the B_k terms by $D_k B_k$ terms, in the same way as occurs for the RS state estimation filtering problem. This is partially justifiable from the interpretation given above of D_k as modifying the noise model.

That is, our suboptimal RS estimates for $\mathcal{J}_k, \mathcal{O}_k$ and $\bar{\mathcal{T}}_k$ are as follows:

$$\begin{aligned}
\sigma(\mathcal{J}^X_{k|k}) &= ADB(y_k)\sigma(\mathcal{J}^X_{k-1|k-1}) \\
&\quad + ((Ae_1)_{\text{diag}}, (Ae_2)_{\text{diag}}, \ldots, (Ae_N)_{\text{diag}}) \\
&\quad \times (DB(y_k)(\alpha_{k-1})_{\text{diag}} \otimes I) \\
\text{where} &\quad \text{row vec } \sigma(\mathcal{J}_{k|k}) = \underline{1}'\sigma(\mathcal{J}^X_{k|k}) \qquad (24) \\
\sigma(\bar{\mathcal{T}}^X_{k|k}) &= ADB(\bar{y}_k)\sigma(\bar{\mathcal{T}}^X_{k-1|k-1})
\end{aligned}$$

$$+ADB(\bar{y}_k)(\alpha_{k-1})_{\text{diag}}\bar{y}_k$$

$$\text{where} \qquad \text{row vec } \sigma(\bar{\mathcal{T}}_{k|k}) = \underline{1}'\sigma(\bar{\mathcal{T}}^X_{k|k}). \qquad (25)$$

Here $\sigma(\mathcal{J}^X_{k|k})$ is the unnormalised conditional mean estimate for J^X_k. We obtain unnormalised conditional mean estimates for J_k from $\sigma(\mathcal{J}^X_{k|k})$ using (24).

This suboptimal RS filter is not optimal in a RS sense but hopefully will be able to handle parameter uncertainty better than the RN filter.

Using the suboptimal RS filters we estimate parameters in an off-line manner as follows.

$$\begin{aligned} \bar{A}^{RS}_{\ell+1} &= \hat{\mathcal{J}}^{RS}_{T|T,\hat{M}^P_\ell}\left(\hat{\mathcal{O}}^{RS}_{T|T,\hat{M}^P_\ell}\right)^{-1}, \\ \bar{\lambda}^{RS}_{\ell+1} &= \hat{\hat{\mathcal{T}}}^{RS}_{T|T,\hat{M}^P_\ell}\left(\hat{\mathcal{O}}^{RS}_{T|T,\hat{M}^P_\ell}\right)^{-1} \end{aligned} \qquad (26)$$

where $\bar{A}^{RS}_\ell$ etc. are the RS estimates.

5 Simulations Studies

In this section, simulation studies are presented to illustrate the performance improvement of RS estimation of HMM parameters. The example presented is representative of the sort of performance improvement possible. Of course, it would be more convincing to have theorems giving relative rates of convergence during transients on finite data, but because estimation here is inherently a nonlinear exercise for which there is no precedent for such results, we settle here for simulation studies.

5.1 Off Line estimation HMM parameters

A 3000 point, 2-state HMM with parameters: $A = [0.8, 0.3; 0.2, 0.7]$ and $\lambda = [10, 20]'$ is generated.

Initial estimates of the HMM parameter are: $a_{ij} = 0.5$ for all i, j. λ is assumed known. Adaptive estimation of A is performed using both (16) and (26) ($\theta = 0.1$ and $Q = I_2$) with A updated after each pass through the data.

Figure 1 shows the error in a_{22} estimates plotted against pass number. Convergence in the other parameters is similar. The RS parameter estimator appears to converge better from the initialization. A reasonable approach may be to estimated parameters by using RS filters for the initial passes through the data and then use the RN estimator once convergence close to the true values has occurred.

5.2 Effect of θ choices on RS estimation

A HMM was generated and estimation of A and λ was performed using a online RS approach (not shown here) using three different choices of θ. Figure 2 shows the evolution of parameter estimates.

For small values of θ the performance of the risk sensitive filter is, not surprisingly, very similar to the optimal filter. As the value of θ is increased the performance improves until some point. For larger values of θ the RS estimator does not perform as well. It appears that moderate choices of θ result in performance gains.

6 Conclusion

In this paper we have proposed a practical risk-sensitive approach to estimation of hidden Markov model (HMM) parameters. It is not surprising that a risk-sensitive approach, being inherently risk averse or robust to modelling errors, can give better transient performance than working with risk-neutral filters, and yet achieve asymptotic optimality if asymptotically risk-neutral. The simulation examples show that in an off-line situations that using risk-sensitive filtering results in a improvement in transient performance of the parameter estimation from poor initializations. We conclude that the risk sensitive approach to HMM parameter estimation should be seen an alternative approach to the HMM parameter estimation problems.

References

[1] L.E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Ann. Math. Stat.*, Vol. 37, pp. 1554-1563, 1966.

[2] R.J. Elliott, L. Aggoun and J.B. Moore, *Hidden Markov Models, Estimation and Control*, Springer-Verlag, 1995.

[3] I.B. Collings and J.B. Moore, "An Adaptive Hidden Markov Model Approach to FM and M-ary DPSK Demodulation in Noisy Fading Channels", *Signal Processing*, Vol. 47, No. 1, pp. 71-84, Nov. 1995.

[4] I.B. Collings, V. Krishnamurthy and J.B. Moore, "Online Identification of Hidden Markov Models via Recursive Prediction Error Techniques," *IEEE Trans. on Signal Processing*, Vol. SP-42, No. 12, pp. 3535-3539, Dec. 1994.

[5] S.H. Chung, V. Krishnamurthy and J.B. Moore, "Adaptive processing techniques based on Hidden Markov Models for characterising very small channel currents buried in noise and deterministic interferences," *Philosophical Transactions of the Royal Society of London*, Series B, Vol. 334, pp. 357-384, 1991.

[6] J.J. Ford and J.B. Moore, "Adaptive Estimation of HMM Transition Probabilities," *IEEE Transactions on Signal Processing*, Vol. 46, No. 5, pp. 1374-1385, May, 1998.

[7] S. Dey and J.B. Moore, *Risk Sensitive Filtering and Smoothing for Hidden Markov Models*, June 27 1994.

[8] L.R. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition," *Proc. of IEEE*, Vol. 77, No.2, pp. 257-285, 1989.

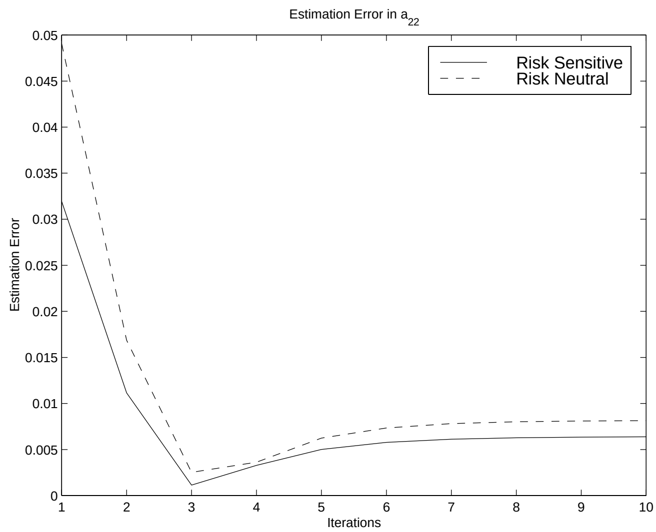

Figure 1: Off-line Estimation of A in Poisson Model.

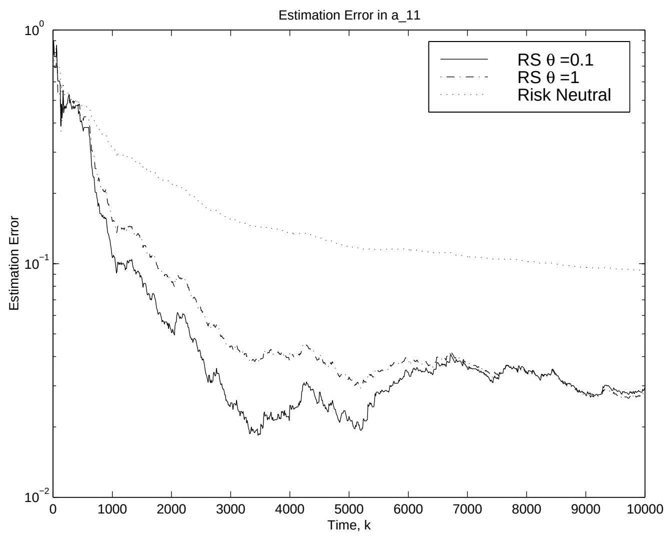

Figure 2: Variation of Risk Sensitive Parameter θ on Estimation.