LEVEL ESTIMATION IN NONLINEARLY DISTORTED HIDDEN MARKOV MODELS USING STATISTICAL EXTREMES

Kutluyıl Doğançay

Inst. of Info. Sciences & Technology
Massey University
Palmerston North, New Zealand
Email: K.Dogancay@massey.ac.nz

Vikram Krishnamurthy

Dept. of Electrical & Electronic Engineering
The University of Melbourne
Parkville, VIC 3052, Australia
Email: v.krishnamurthy@ee.mu.oz.au

ABSTRACT

Estimation of the state levels of a discrete-time, finite-state Markov chain hidden in coloured Gaussian noise and subjected to unknown nonlinear distortion is considered. If the nonlinear distortion has almost linear behaviour for small values near zero or for large values, extreme value theory can be applied to the level estimation problem, resulting in simple estimation algorithms. The extreme value-based level estimator is computationally inexpensive and has potential applications in data measurement systems where inaccuracies are introduced by dead zones or saturation in sensor characteristics. The effectiveness of the new level estimator is demonstrated by way of computer simulations.

1. INTRODUCTION

A significant amount of research has been directed at the estimation of hidden Markov models (HMM) using maximum likelihood (ML) estimation methods. In this paper we explore the use of statistical extremes in estimating the levels of an HMM subjected to unknown nonlinear distortion, which makes the application of ML methods computationally infeasible. The nonlinear distortion is assumed to represent an *unknown nonignorable missing data mechanism*. If the missing data mechanism produces left or right censored observations, which implies that either the right or left tail of the data distribution can be observed without distortion, respectively, extreme value theory can be applied to the level estimation problem. Likewise, if the observed data are subject to double censoring, such as truncation, the minima of the absolute values of observations can be used to estimate the levels of an HMM by invoking extreme value theory.

Suppose that the observations available to us are generated by the functional

$$y(k) = \mathcal{F}\big(\langle \boldsymbol{x}(k), \boldsymbol{g}_\lambda \rangle + n(k)\big) \qquad (1)$$

where $\mathcal{F} : \mathbb{R} \to \mathbb{R}$ is an unknown nonlinear function with linear behaviour either for large positive and/or negative values or for small values of its argument near zero, $\langle \cdot, \cdot \rangle$ denotes inner product, $\boldsymbol{x}(k)$ is a discrete-time, finite-state Markov chain with level vector $\boldsymbol{g}_\lambda$, and $n(k)$ is a possibly coloured noise process with known marginal distribution. Assume that the levels of the Markov chain are parametrised by an *unknown* parameter λ. For example in digital modulation the actual signal levels are determined by a single parameter related to signal amplification. We wish to estimate λ from the observations $\{y(k)\}$ in (1). The main idea in this paper is to show that extreme value theory for HMMs yields three possible invariant distributions for the extremes of a sequence of HMM observations. The unknown level parameter λ can be estimated from these distributions by simple curve fitting with a unique global solution unlike the ML estimation of HMMs, which can be plagued by local stationary solutions.

The hidden Markov model in (1) is encountered in signal processing and control applications where sensors used for data measurement may have dead zones or nonlinear response for small input signals, and exhibit saturation for large input signals, resulting in missing or nonlinearly distorted measurements. Extreme value theory can be applied to all these cases to glean useful information about the underlying signal from statistical extremes of nonlinearly distorted observations.

The paper is organised as follows. Section 2 provides an overview of extreme value theory. In Section 3 the level estimation problem is defined and extreme value-based algorithms for level estimation are developed. Section 4 presents simulation examples to illustrate the application of the algorithms.

This work was supported by the Australian Research Council (ARC) and the Cooperative Research Centre for Sensor, Signal and Information Processing (CSSIP).

2. EXTREME VALUE THEORY

2.1. Maxima of i.i.d. Sequences

Let M_N denote the *maximum* in a sequence of N independent identically distributed (i.i.d.) random variables $\{y(1), y(2), \ldots, y(N)\}$ with common cumulative distribution function (c.d.f.) $F(x) = \Pr\{y(i) \le x\}, i = 1, 2, \ldots, N$, i.e.,

$$M_N = \max\big(y(1), y(2), \ldots, y(N)\big). \qquad (2)$$

Classical extreme value theory is concerned with the asymptotic distribution of M_N as $N \to \infty$. If $F(x)$ is known, the c.d.f. of M_N is given by

$$\Pr\{M_N \le x\} = \Pr\{y(1) \le x, y(2) \le x, \ldots, y(N) \le x\} = F^N(x). \qquad (3)$$

According to extreme value theory, the asymptotic nondegenerate c.d.f. of M_N must belong to one of three possible distributions if it exists. The particular asymptotic distribution can be determined from only a limited knowledge of the tail distribution of $y(i)$, thereby making knowledge of $F(x)$ for all x unnecessary.

We are interested in determining the limiting c.d.f. $G(x)$ that $F(x)$ converges to as $N \to \infty$ under appropriate choices of normalising constants $a_N > 0$ and b_N

$$\Pr\{a_N(M_N - b_N) \le x\} \xrightarrow{w} G(x), \quad N \to \infty \tag{4}$$

where $\xrightarrow{w}$ denotes convergence at continuity points of $G(x)$. If such a nondegenerate distribution $G(x)$ exists, then it must belong to the class of *max-stable* distributions that satisfy the relationship $G^N(a_N x + b_N) = G(x)$ for a given N and some constants a_N and b_N. All possible max-stable distributions have one of the following three parametric types, which are also known as the *extreme value distributions*:

Type I (Gumbel distribution):

$$G(x) = \exp(-e^{-x}), \quad -\infty < x < \infty.$$

Type II (Fréchet distribution):

$$G(x) = \begin{cases} 0 & x \le 0 \\ \exp(-x^{-\alpha}) & \alpha > 0,\ x > 0. \end{cases}$$

Type III (Weibull distribution):

$$G(x) = \begin{cases} \exp(-(-x)^{\alpha}) & \alpha > 0,\ x \le 0 \\ 1 & x > 0. \end{cases}$$

Equation (4) can be rewritten as

$$F^N(u_N) \xrightarrow{w} G(x), \quad u_N \triangleq x/a_N + b_N. \tag{5}$$

If the convergence in (5) holds, $F(x)$ is said to be in the *domain of attraction* of $G(x)$. The necessary and sufficient conditions for convergence of $F(x)$ to one of the extreme value distributions have been documented in [1]. In general, if $F(x)$ has a finite right endpoint, i.e., $\sup\{x : F(x) < 1\} < \infty$, and a jump discontinuity at it, then $G(x)$ is degenerate. For continuous $F(x)$, $G(x)$ is usually of Type I. If the right endpoint of $F(x)$ is finite with no jump at it, $G(x)$ will have the Type III limit.

2.2. Maxima of HMM Observations

We are interested in extending the classical extreme value theory results for i.i.d. sequences to dependent sequences such as HMMs. Let us assume that $\{y(1), y(2), \dots, y(N)\}$ is a stationary and dependent sequence, e.g., an HMM. In this case (3) takes the form

$$\begin{aligned}\Pr\{M_N \le x\} &= \Pr\{y(1) \le x, y(2) \le x, \dots, y(N) \le x\} \\ &\triangleq F_{1,2,\dots,N}(x, x, \dots, x)\end{aligned}$$

where $F_{1,2,\dots,N}(x, x, \dots, x)$ is the joint distribution of the dependent sequence $\{y(i)\}$, which may be unknown or hard to compute. The advantage of extreme value theory in this case is that it provides an asymptotic result for extreme order statistics that depends only on the marginal distribution of $\{y(i)\}$ under some weak conditions on the dependent sequence.

The existence of a nondegenerate distribution $G(x)$ such that for a suitably chosen linear normalisation $a_N > 0$ and b_N

$$F_{1,2,\dots,N}(u_N, u_N, \dots, u_N) \xrightarrow{w} G(x) \tag{6}$$

hinges on the rapid decay of dependence between the successive elements of $\{y(i)\}$ [1], which is assured by the satisfaction of the condition $D(u_N)$.

Definition (The Condition $D(u_N)$). For a given sequence $\{u_N\}$ and any positive integers $1 \le i_1 < i_2 < \cdots < i_p < j_1 < j_2 < \cdots < j_q \le N$ with $j_1 - i_p \ge l$, if we have

$$\begin{aligned}&\big|F_{i_1,\dots,i_p,j_1,\dots,j_q}(u_N, \dots, u_N) \\ &\quad - F_{i_1,\dots,i_p}(u_N, \dots, u_N) F_{j_1,\dots,j_q}(u_N, \dots, u_N)\big| < \alpha_{N,l}\end{aligned}$$

where $\alpha_{N,l_N} \to 0$ as $N \to \infty$ for some sequence $l_N = o(N)$, the condition $D(u_N)$ is said to hold.

Let $\{\hat{y}(i)\}$ denote the *associated independent sequence* which has the same marginal c.d.f. as $\{y(i)\}$, but with i.i.d. elements. Defining $\hat{M}_N \triangleq \max(\hat{y}(1), \hat{y}(2), \dots, \hat{y}(N))$, the relationship between the dependent sequence $\{y(i)\}$ and the associated independent sequence $\{\hat{y}(i)\}$ is given by the *extremal index* θ:

$$\Pr\{\hat{M}_N \le v_N\} \xrightarrow{w} G(x) \Leftrightarrow \Pr\{M_N \le v_N\} \xrightarrow{w} G^{\theta}(x) \tag{7}$$

where $v_N = x/a_N + b_N$ and $0 < \theta \le 1$. The c.d.f.s $G(x)$ and $G^{\theta}(x)$ are of the same type since $G(x)$ is a max-stable distribution.

2.3. Distribution of Minima

Let m_N denote the *minimum* term in the sequence $\{y(1), y(2), \dots, y(N)\}$, i.e.,

$$m_N = \min(y(1), y(2), \dots, y(N)). \tag{8}$$

The distribution of m_N is given by

$$\begin{aligned}\Pr\{m_N \le x\} &= \\ &1 - \Pr\{y(1) > x, y(2) > x, \dots, y(N) > x\} \\ &= 1 - (1 - F(x))^N\end{aligned}$$

where $F(x)$ is the marginal c.d.f. of $\{y(i)\}$. The possible nondegenerate limiting distributions $H(x)$ satisfying

$$\Pr\{c_N(m_N - d_N) \le x\} \xrightarrow{w} H(x) \tag{9}$$

for suitable linear normalisation constants $c_N > 0$ and d_N are

Type I (Gumbel distribution):

$$H(x) = \exp(-e^{-x}), \quad -\infty < x < \infty.$$

Type II (Fréchet distribution):

$$H(x) = \begin{cases} 1 - \exp(-(-x)^{-\beta}) & \beta > 0,\ x < 0 \\ 1 & x \ge 0. \end{cases}$$

Type III (Weibull distribution):

$$H(x) = \begin{cases} 0 & x < 0 \\ 1 - \exp(-x^{\beta}) & \beta > 0,\ x \ge 0. \end{cases}$$

The three distributions listed above are *min-stable* distributions defined by $(1 - H(c_N x + d_N))^N = 1 - H(x)$ for appropriate constants c_N and d_N.

3. HMM LEVEL ESTIMATION

We will first state the model assumptions for the estimation problem and then propose estimation procedures for two general forms of nonlinearity.

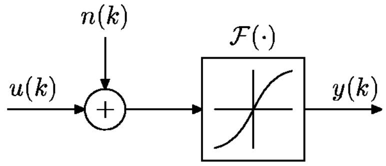

Figure 1: Observation model.

3.1. Observation Model Assumptions

The generic observation model of (1) is schematically shown in Fig. 1. Let $\boldsymbol{x}(i)$, $i \in \mathbb{Z}^+$, define an M-state ($M < \infty$) homogeneous Markov chain with state space $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_M\}$ where $\boldsymbol{e}_i$ is the ith unit vector in $\mathbb{R}^M$ with zero entries except for its ith entry which is one. Let $\boldsymbol{P}$ denote the $M \times M$ transition probability matrix with its (i,j)th entry given by $p_{ij} = \Pr\{\boldsymbol{x}(k+1) = \boldsymbol{e}_j \mid \boldsymbol{x}(k) = \boldsymbol{e}_i\}$ and $\sum_{j=1}^M p_{ij} = 1 \; \forall i \in \{1, 2, \ldots, M\}$ (i.e., $\boldsymbol{P}$ is a stochastic matrix).

The stationary Markov chain distribution $\boldsymbol{s} = [s_1, \ldots, s_M]^T$ is given by the *Perron-Frobenius* (PF) eigenvector of $\boldsymbol{P}$, where $s_i = \Pr\{\boldsymbol{x}(k) = \boldsymbol{e}_i\}$ as $k \to \infty$. The PF eigenvector is defined by

$$\boldsymbol{s}^T \boldsymbol{P} = \boldsymbol{s}^T, \quad s_i \geq 0, \quad \sum_{i=1}^M s_i = 1.$$

We assume that $\boldsymbol{s}$ is known. In most cases of interest $\boldsymbol{s}$ will have the form $\boldsymbol{s} = [1/M, \ldots, 1/M]^T$, i.e., equally likely Markov states as $k \to \infty$.

The output of the Markov chain is given by the inner product $u(k) = \langle \boldsymbol{x}(k), \boldsymbol{g}_\lambda \rangle$ where $\boldsymbol{g}_\lambda = [g_1(\lambda), g_2(\lambda), \ldots, g_M(\lambda)]^T$ is the vector of Markov chain levels parametrised by an *unknown* parameter λ. The functions $g_i(\lambda)$ are assumed to have a known explicit form. For example, in the case of a binary Markov chain, we will have $\boldsymbol{g}_\lambda = [-\lambda, \lambda]^T$.

The noise process $n(k)$ in which the Markov chain is embedded is assumed to be stationary and possibly coloured, to have a known marginal distribution, to be in the domain of attraction of a nondegenerate extreme value distribution, and to satisfy the distributional mixing condition $D(u_N)$. Hidden in additive stationary noise, the Markov chain outputs can be regarded as a chain-dependent process [2, 3].

Below we outline two algorithms for estimating λ based on the invariant distribution of the extremes of an HMM.

3.2. Algorithm 1: Estimation of Levels from Maxima of Mixtures

In most practical cases the additive noise $n(k)$ will have a Gaussian marginal distribution with zero mean and variance σ^2, i.e.,

$$F(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left(-\frac{u^2}{2\sigma^2}\right) du.$$

If the condition $D(u_N)$ is satisfied by $n(k)$, then the maxima of $\{y(k)\}$ will be in the domain of attraction of the Type I distribution [3]. The estimation procedure consists of segmenting the observation sequence $\{y(k)\}$ into L subsequences of length N and then fitting the maxima of individual subsequences to the Gumbel distribution in accordance with (7)

$$\Pr\{M_{N,i} \leq v_N\} \xrightarrow{w} G^\theta(x), \quad i = 1, \ldots, L$$

where $M_{N,i}$ is the maximum of the ith subsequence. Since N is finite, an exact fit is not possible. The parameters a and b of the Gumbel distribution $G(a(x+b)) = \exp(-e^{-a(x+b)})$ are given by $a = 1/a_N$ and $b = a_N(b_N + \ln\theta)$. Maxima of the Markov chain hidden in coloured noise will have the following marginal mixture distribution

$$\left(\frac{1}{M}\sum_{i=1}^M F(x - g_i(\lambda))\right)^N \xrightarrow{w} G(a(x+b)). \tag{10}$$

Once the estimates of a and b have been obtained, the parameter λ can be estimated by minimising a distance measure, such as the squared ℓ_2 norm, between the left and right sides of (10).

The maximum likelihood estimates $\hat{a}$ and $\hat{b}$ are given by the solution of the following system of equations

$$e^{ab} \sum_{i=1}^L e^{-aM_{N,i}} = L$$

$$\frac{1}{L} + \frac{\sum_{i=1}^L M_{N,i} e^{-aM_{N,i}}}{\sum_{i=1}^L e^{-aM_{N,i}}} = \frac{1}{L}\sum_{i=1}^L M_{N,i}.$$

3.3. Algorithm 2: Estimation of Levels from Truncated Data

The procedure for level estimation from minima is similar to the procedure in the previous subsection. Assuming that the values of $y(k)$ near zero are not distorted nonlinearly, we will divide $\{|y(k)|\}$ into L subsequences of length N and then fit the minima of subsequences to the Type III distribution (i.e., the Weibull distribution):

$$\Pr\{c_N(m_{N,i} - d_N) \leq x\} \xrightarrow{w} H(x), \quad i = 1, \ldots, L$$

where $m_{N,i}$ is the minimum of the ith subsequence. The parameters a and β of the Weibull distribution $H(x) = 1 - \exp(-x^\beta/a)$ are estimated from the $m_{N,i}$ with marginal mixture distribution

$$1 - \left(1 - \frac{2}{M}\sum_{i=1}^{M/2}\big(F(x - g_i(\lambda)) - F(-x - g_i(\lambda))\big)\right)^N \xrightarrow{w} H(x) \tag{11}$$

where the left-hand side reflects the folding effect of absolute value operation. The level parameter λ is estimated by minimising a distance measure between the left and right sides of (11).

The maximum likelihood estimates $\hat{a}$ and $\hat{\beta}$ are obtained by solving the following system of equations for a and β

$$a = \frac{1}{L}\sum_{i=1}^L m_{N,i}^\beta$$

$$\frac{L}{\beta} = \frac{L\sum_{i=1}^L m_{N,i}^\beta \ln m_{N_i}}{\sum_{i=1}^L m_{N,i}^\beta} - \sum_{i=1}^L \ln m_{N,i}.$$

4. SIMULATION EXAMPLES

Consider an HMM with transition probability matrix

$$\boldsymbol{P} = \begin{bmatrix} 0.3 & 0.2 & 0.2 & 0.3 \\ 0.2 & 0.4 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.4 & 0.2 \\ 0.3 & 0.2 & 0.2 & 0.3 \end{bmatrix}$$

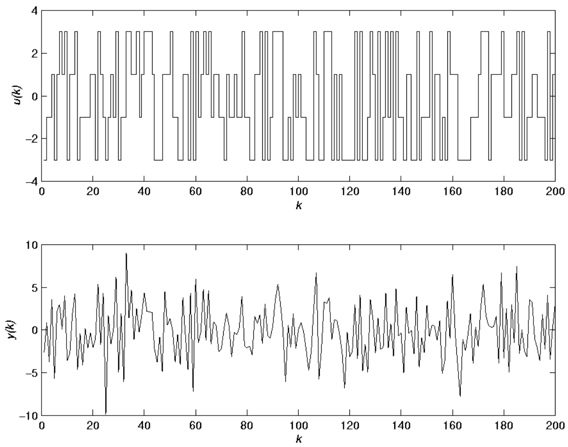

Figure 2: Markov chain output $\{u(k)\}$ and its noisy and distorted version $\{y(k)\}$.

and levels $g_\lambda = [-3\lambda, -\lambda, \lambda, 3\lambda]^T$. The additive noise $n(k)$ is Gaussian with zero mean and autocorrelation $R_n(0) = 5.00$, $R_n(1) = -2.78$, $R_n(2) = 1.40$, $R_n(3) = -1.28$, $R_n(4) = -0.20$, $R_n(5) = 1.00$.

Suppose $\{y(k)\}$ is nonlinearly distorted for small HMM outputs. We can then use the maxima of subsequences of $\{y(k)\}$ to estimate λ. For $\lambda = 1$, the first 200 points of $u(k)$ and $y(k)$ are shown in Fig. 2. For $N = 50$ and $L = 100$, the Gumbel distribution and the distribution of the associated independent maxima fitted to observed maxima are shown in Fig. 3. The Gumbel parameter estimates are $\hat{a} = 0.8602$ and $\hat{b} = 6.2365$, and the level estimate is $\hat{\lambda} = 0.9943$. Fig. 3 shows that the level estimation will not be affected by nonlinear distortion for $y(k) < 4.5$.

If the HMM outputs go through an unknown nonlinearity causing saturation, then the minima of $\{|y(k)|\}$ can be used. The first 200 points of $u(k)$ and $|y(k)|$ are shown in Fig. 4. The Weibull distribution and the c.d.f. of associated minima fitted to the minima of $\{|y(k)|\}$ are shown in Fig. 5. The fit results in $\hat{\alpha} = 0.0748$, $\hat{\beta} = 1.0380$ and $\hat{\lambda} = 0.9679$. According to Fig. 5, any nonlinear distortion for $|y(k)| > 0.3$ will not affect the level estimate.

5. CONCLUSION

We have developed level estimation algorithms for HMMs in coloured noise observed through nonlinear distortion. Subject to the nature of the nonlinearity, maxima or minima of distorted HMM outputs are shown to be usable for the purpose of level estimation in very small signal-to-noise ratios.

6. REFERENCES

[1] M. R. Leadbetter, G. Lindgren, and H.Rootzen, *Extremes and Related Properties of Random Sequences and Processes*, New York: Springer Verlag, 1983.

[2] S. I. Resnick and M. F. Neuts, "Limit laws for maxima of a sequence of random variables defined on a Markov chain," *Adv. Appl. Prob.*, no. 2, pp. 323–343, 1970.

[3] G. E. Denzel and G. L. O'Brien, "Limit theorems for extreme values of chain-dependent processes," *Annals of Probability*, vol. 3, no. 5, pp. 773–779.

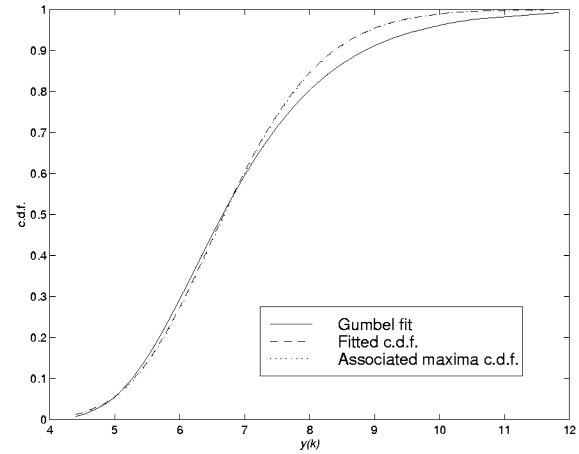

Figure 3: Extreme value distributions fitted to maxima.

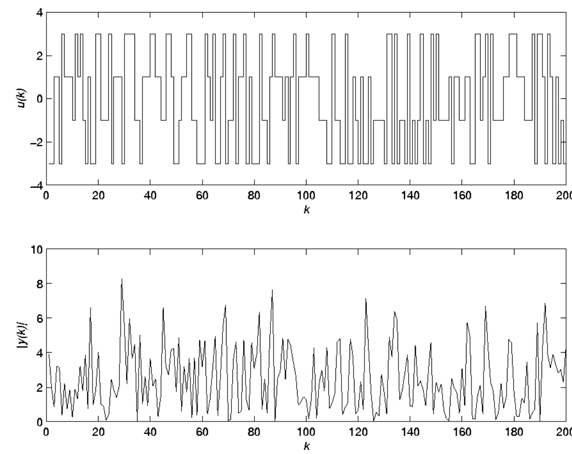

Figure 4: Markov chain output $\{u(k)\}$ and distorted HMM observations $\{|y(k)|\}$.

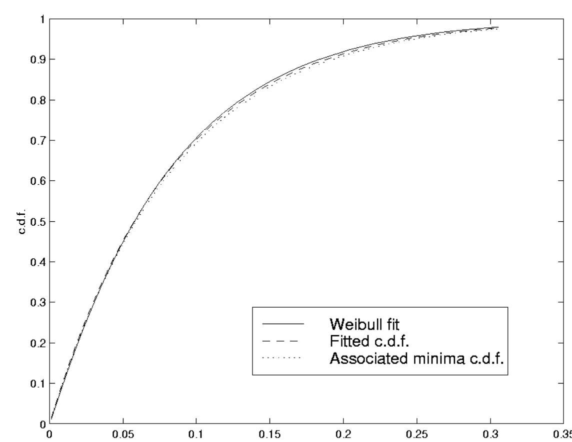

Figure 5: Extreme value distributions fitted to minima.