

# Heavy Tailed Hidden Semi-Markov Models

Sidney Resnick\* and Ajay Subramanian  
School of ORIE and Center for Applied Mathematics  
Rhodes Hall  
Cornell University, Ithaca, NY 14853

March 4, 1997

## Abstract

Hidden semi-Markov models have been proposed in Meier-Hellstern et al (1991) to model the times between transmission of packets at a source. In this paper, we specifically investigate such models where the distributions associated with one or more of the states are heavy-tailed. We study the asymptotic properties of the sample covariance and correlation functions and show that the correlation function at nonzero lags is asymptotically equal to zero. We simulate a five state hidden semi-Markov model to illustrate our results.

## 1 Introduction.

Hidden semi-Markov models have often been used in speech recognition and computational biology; see Krogh et al (1994) and Juang and Rabiner (1991) for reviews. Such models have also often been proposed in the teletraffic literature. In the heavy tailed context, they were recently used by Meier-Hellstern et al (1991) to model the times between transmission of packets at a source. In the process of fitting these models to the data, it was observed that the distributions associated with one or more of the states of the model were heavy-tailed.

In this paper, we investigate hidden semi-Markov models where the distributions associated with one or more of the states are heavy-tailed. Our goal is to study the asymptotic properties of the sample covariance and autocorrelation function (acf) which are important in the processes of parameter estimation and fitting of actual data to such models. We use point process techniques to show that the sample correlation function at nonzero lags is asymptotically equal to zero.

The plan of the paper is as follows: In Section 2, we give a detailed description of the model and construct a stationary version. In Section 3, we prove several convergence results for point processes associated with the model with the aim of studying the asymptotic properties of the sample covariance and correlation functions. In Section 4, we prove the major result of the paper, that the sample correlation function at nonzero lags is asymptotically equal to zero. In Section 5, we offer some numerical illustrations and some concluding remarks.

---

\*Partially supported by NSF Grant DMS-9400535 at Cornell University.

The reason for the interest in the sample autocorrelation function is the following. Examination of heavy tailed data frequently reveals the characteristic that the acf plots for different subsets of the data do not look similar. Some examples are given in Resnick (1996). Various explanations for this phenomenon are possible: maybe the data exhibits lack of stationarity or maybe the underlying model is non-linear. The following plot, from Davis and Resnick (1996) shows an extreme case: Plots of the acf from independent samples of the same stationary nonlinear process look completely different.

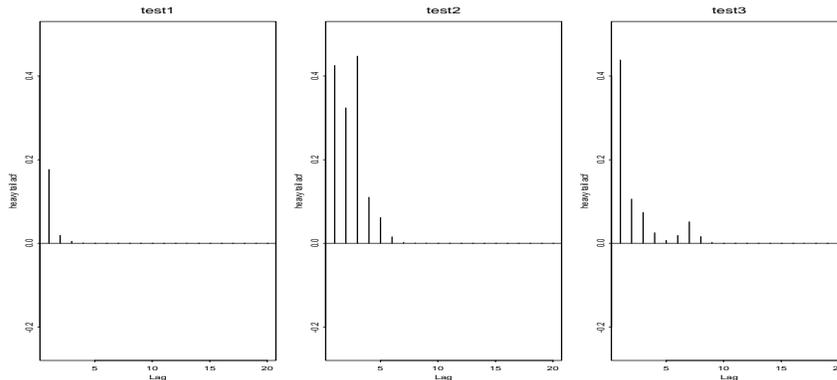


Figure 1: ACF of three independent samples from a bilinear process.

Subset analysis of the acf of a data set is often sufficient to rule out the iid or linear time series models as candidate models for fitting to the data. We were curious to see if the hidden Markov model was a possibility and hence set out to investigate the asymptotic behavior of the sample acf. Unfortunately our results rule out hidden Markov models as well since the behavior of the acf for such models is asymptotically the same as for the iid model. Unless one gives up finite mean sojourn times (and thus it is likely one must also give up stationarity) such models will be difficult to fit to data whose sample acf fails to exhibit subset stationarity.

## 2 The model.

Let  $\{J_n, n \geq 0\}$  be an ergodic Markov chain with state space  $\{1, \dots, p\}$ . Let the transition probability matrix of this chain be  $P = \{p_{ij}, 1 \leq i, j \leq p\}$  and the stationary distribution be  $\pi' = (\pi_1, \dots, \pi_p)$ . Suppose for  $i = 1, \dots, p$  we are given holding time distributions  $\{q_n^{(i)}\}$  concentrating on  $\{1, 2, \dots\}$  and that for  $i = 1, \dots, p$ ,  $\{D_n^{(i)}, n \geq 0\}$  are iid with common distribution  $\{q_n^{(i)}\}$  and further,  $E[q_n^{(i)}] = \tau_i < \infty$ . The sequences

$$\left\{ \{D_n^{(i)}\}, i = 1, \dots, p \right\}$$

are independent. Define the semi-Markov chain  $\{V_n, n \geq 0\}$  by

$$\begin{aligned} V_j &= J_0, \text{ if } 0 \leq j < D_0^{(J_0)}, \\ &= J_1, \text{ if } D_0^{(J_0)} \leq j < D_0^{(J_0)} + D_1^{(J_1)}, \\ &= J_2, \text{ if } D_0^{(J_0)} + D_1^{(J_1)} \leq j < D_0^{(J_0)} + D_1^{(J_1)} + D_2^{(J_2)}, \\ &\vdots \quad \vdots \end{aligned}$$

Thus

$$V_j = \sum_{k=0}^{\infty} J_k 1_{[\sum_{l=0}^{k-1} D_l^{(J_l)} \leq j < \sum_{l=0}^k D_l^{(J_l)}]}$$

The observable process  $\{X_n\}$  consists of random elements of  $\mathbb{R}_+$  defined on  $\{V_n\}$  such that the conditional distribution of  $X_n$  given the state of  $V_n$  is  $F_i$ . More precisely, given distributions  $F_1, \dots, F_p$  on  $\mathbb{R}_+$  and iid uniform random variables with support  $[0, 1]$  which we call  $\{U_n, n \geq 0\}$ , we then define for  $n \geq 0$

$$X_n = F_{V_n}^{\leftarrow}(U_n)$$

and assume  $\{U_n\}, \{J_n\}, \{D_n^{(i)}, n \geq 0, 1 \leq i \leq p\}$  are all independent.

Thus, changes of state follow the Markov chain  $\{J_n\}$  and a transition from  $i$  to  $j$  occurs with probability  $p_{ij}$ . Having entered state  $i$ , the system stays in state  $i$  for  $k$  time units with probability  $q_k^{(i)}$ . While the  $V$ -process is in state  $i$ , random variables are generated from a distribution  $F_i$ .

One or more of the distributions  $F_i$  may be heavy-tailed. We assume that  $F_1$  has the heaviest tail with index of regular variation  $\alpha$ . Thus

$$\lim_{t \rightarrow \infty} (1 - F_1(tx)) / (1 - F_1(t)) = x^{-\alpha} \quad (2.1)$$

and

$$\lim_{t \rightarrow \infty} (1 - F_j(t)) / (1 - F_1(t)) = 0, \quad j = 2, \dots, p. \quad (2.2)$$

For a distribution function  $G$ , we will abbreviate the tail function by  $\bar{G} := 1 - G$ .

The above is the definition of a basic hidden semi-Markov model. However, the model, as we have defined it, is not stationary. Note that if  $\{V_n\}$  is stationary, then so is  $\{X_n\}$  since

$$P[X_j \leq x_j, j = 0, \dots, k] = E \left( \prod_{j=0}^k F_{V_j}(x) \right).$$

Since the holding time distributions  $\{q^{(i)}\}$  have finite expectations, it follows that the semi-Markov process  $\{V_n\}$  has a stationary distribution  $\{\nu_i\}$  with

$$\nu_i = \pi_i \tau_i / \sum_{j=1}^p \pi_j \tau_j$$

and where  $\{\nu_i\}$  is also the *asymptotic* distribution of  $\{V_n\}$ . Since time series modeling is usually done with stationary models we shall modify the definition of the model so that the process  $\{V_n\}$  is stationary and therefore,  $\{X_n\}$  is stationary. This construction can be carried out following the procedure described by Cinlar (1969).

Choose  $V_0$  such that

$$P[V_0 = j] = \nu_j$$

If  $A_{jk}, A'_{jk}$  are the semi-Markov matrices associated with the *delayed* semi-Markov process  $\{V_n\}$ , following Cinlar, choose

$$A'_{jk}(n) = (1/\tau_j) \sum_{i=0}^n [\tilde{A}_{jk} - A_{jk}(i)]$$

where

$$\tilde{A}_{jk} = A_{jk}(+\infty)$$

In this paper, we shall assume that the underlying semi-Markov process  $\{V_n\}$  is stationary and therefore the hidden semi-Markov model  $\{X_n\}$  is stationary.

### 3 Point processes associated with the model.

In this section, we use point process techniques to prove a theorem that enables us to study the asymptotic properties of the sample correlation function in the next section.

Let  $\{X_n\}$  be the stationary process defined in Section 2 and define

$$a_n = \inf\{x : \sum_{i=1}^p \nu_i \bar{F}_i(x) \geq n^{-1}\}$$

so that

$$\sum_{i=1}^p \nu_i \bar{F}_i(a_n) \sim \frac{1}{n}.$$

Furthermore we have

$$nP[a_n^{-1} X_1 \in \cdot] \xrightarrow{v} \lambda(\cdot) \tag{3.1}$$

on  $(0, \infty]$  where  $\lambda$  is a measure on  $(0, \infty]$  such that  $\lambda(x, \infty] = x^{-\alpha}$  and the convergence above is in the vague topology on the space of Radon measures on  $(0, \infty]$ . To verify (3.1), use the definition of  $\{X_n\}$  to get

$$\begin{aligned} nP[X_1 > a_n x] &= \sum_{i=1}^p nP[X_1 > a_n x | V_1 = i] P[V_1 = i] \\ &\sim \sum_{i=1}^p \nu_i \bar{F}_i(a_n x) / \left( \sum_{i=1}^p \nu_i \bar{F}_i(a_n) \right). \end{aligned}$$

Dividing numerator and denominator by  $\bar{F}_1(a_n)$  and using equations (2.1) and (2.2) we easily see that

$$\lim_{n \rightarrow \infty} nP[X_1 > a_n x] = x^{-\alpha}$$

which suffices for (3.1).  $\square$

We need a result about the frequency of visits to each state.

**Lemma 1** *Let  $N_n(j)$  be the number of “visits” to state  $j$  by the semi-Markov chain up to time  $n$  so that*

$$N_n(j) = \sum_{l=1}^n 1_{[V_l=j]}.$$

*Then we have*

$$\lim_{n \rightarrow \infty} N_n^{(j)}/n = \nu_j.$$

*Proof.* The proof is almost exactly as in Resnick and Starica (1996). The semi-Markov process  $\{V_j\}$  changes states at times  $\{S_n\}$  where

$$S_n = \sum_{q=0}^n D_q^{(J_q)}$$

Let  $\mu_n(j)$  be the number of “visits” to state  $j$  by the underlying Markov chain  $\{J_n\}$ . As  $n \rightarrow \infty$ , we have

$$\begin{aligned} S_n/n &\stackrel{d}{=} (1/n) \sum_{k=1}^p \sum_{i=1}^{\mu_n(k)} D_i^{(k)} \\ &= \sum_{k=1}^p \left[ \sum_{i=1}^{\mu_n(k)} D_i^{(k)} / \mu_n(k) \right] [\mu_n(k)/n] \rightarrow \sum_{k=1}^p E D_1^{(k)} \pi_k \\ &= \sum_{k=1}^p \tau_k \pi_k \end{aligned}$$

Now define the process inverse to  $\{S_n\}$  as

$$M(t) = \sup\{n : S_n \leq t\}$$

so that as  $t \rightarrow \infty$

$$M(t)/t \rightarrow 1 / \left( \sum_{k=1}^p \tau_k \pi_k \right).$$

Note that  $M(t)$  represents the number of changes of state up to time  $t$ . Now

$$\begin{aligned}
N_n(k)/n &\leq (1/n) \sum_{q=1}^{M(n)+1} (S_q - S_{q-1}) 1_{[J_{q-1}=k]} \\
&\stackrel{d}{=} (1/n) \sum_{q=0}^{\mu_{(M(n)+1)}(k)} D_q^{(k)} \\
&= \left( \sum_{q=0}^{\mu_{(M(n)+1)}(k)} D_q^{(k)} / \mu_{(M(n)+1)}^{(k)} \right) (\mu_{(M(n)+1)}^{(k)} / (M(n)+1)) ((M(n)+1)/n) \\
&\rightarrow \tau_k \pi_k / \sum_{j=1}^p \tau_j \pi_j.
\end{aligned}$$

A lower bound is obtained similarly and this completes the proof.  $\square$

For what follows, for a nice metric space  $\mathbb{E}$ , we define  $M_p(\mathbb{E})$  to be the Radon point measures on  $\mathbb{E}$ , topologized by the vague topology. A point measure is represented by

$$\sum_i \epsilon_{x_i}$$

where  $x_i \in \mathbb{E}$ . A Poisson process with mean measure  $\lambda$  on  $\mathbb{E}$  will be denoted  $\text{PRM}(\lambda)$ .

We now consider point processes based on the observed process  $\{X_n\}$ .

**Proposition 1** *For the stationary process  $\{X_n\}$  defined in Section 2,*

$$\sum_{k=1}^n \epsilon_{a_n^{-1} X_k} \Rightarrow \text{PRM}(\lambda)$$

in  $M_p((0, \infty])$ .

*Proof.* We show the convergence in the space of point measures by proving that the corresponding Laplace functionals converge. Therefore, we need to show that for  $f \in C_K^+((0, \infty])$

$$\Psi_n(f) = E \exp\left\{-\sum_{j=1}^n f(X_j/a_n)\right\} \rightarrow \exp\left\{-\int_{(0, \infty]} (1 - \exp(-f(x))) \lambda(dx)\right\} \quad (3.2)$$

Let  $N_n(j)$  be the number of ‘‘visits’’ to state  $j$  by the semi-Markov chain up to time  $n$ . Then

$$\begin{aligned}
\Psi_n(f) &= E(E(\exp\{-\sum_{j=1}^n f(X_j/a_n)\} | V_0, \dots, V_n)) \\
&= E \prod_{j=1}^p \left( \int_{(0, \infty]} \exp(-f(x)) F_j(a_n dx) \right)^{N_n(j)} \\
&= E \prod_{j=1}^p \left( 1 - \int_{(0, \infty]} \frac{(1 - e^{-f(x)}) n F_j(a_n dx)}{n} \right)^{n \cdot N_n(j)/n}
\end{aligned}$$

In the previous Lemma 1, we have shown that

$$\lim_{n \rightarrow \infty} N_n(j)/n = \nu_j$$

where  $\{\nu_j\}$  is the stationary distribution for the semi-Markov chain. For  $j = 2, \dots, p$

$$\lim_{n \rightarrow \infty} nF_j(a_n dx) = 0$$

and

$$\lim_{n \rightarrow \infty} nF_1(a_n dx) = (1/\nu_1)\lambda(dx).$$

Therefore,

$$\begin{aligned} & \prod_{j=1}^p \left( 1 - \int_{(0, \infty]} \frac{(1 - e^{-f(x)})nF_j(a_n dx)}{n} \right)^{n \cdot N_n(j)/n} \\ & \rightarrow \prod_{j=1}^p \exp\left\{- \int_{(0, \infty]} (1 - e^{-f(x)}) \lim_{n \rightarrow \infty} nF_j(a_n dx) \nu_j\right\} \\ & = \exp\left\{- \int_{(0, \infty]} (1 - \exp(-f(x)))\lambda(dx)\right\}. \end{aligned}$$

We can now apply the dominated convergence theorem to conclude

$$\Psi(f) \rightarrow \exp\left\{- \int_{0, \infty]} (1 - e^{-f(x)})\lambda(dx)\right\}$$

This completes the proof of the proposition.  $\square$

Our methods require the following definition of a new sequence  $\{\mathbf{Y}_k\}$ :

$$\mathbf{Y}_k = (X_k, \dots, X_{k-m+1})$$

for  $k \geq 1$  and fixed  $m \geq 1$ . The sequence  $\{\mathbf{Y}_k\}$  is conditionally  $m$ -dependent given the sequence  $\{V_n\}$ .

We now define two new processes:

$$\begin{aligned} I_n &= \sum_{k=1}^n \epsilon_{a_n^{-1}} \mathbf{Y}_k \\ I &= \sum_{k=1}^{\infty} \sum_{i=1}^m \epsilon_{j_k} \hat{e}_i \end{aligned}$$

where  $\hat{e}_i \in \mathbb{R}^m$  is the basis element with  $i$ th component equal to 1 and the rest equal to zero. Let  $\mathbb{E} = ([0, \infty]^m \setminus \{(0, \dots, 0)\})$ . Let  $\mathcal{S}$  be the collection of all sets  $B$  of the form

$$B = [b_1, c_1] \times \dots \times [b_m, c_m]$$

where  $[b_1, c_1] \times \dots \times [b_m, c_m]$  is bounded away from  $(0, \dots, 0)$ ,  $b_i < c_i$ ,  $b_0 \geq 0$ . It is easy to see that either  $\prod_{i=1}^m [b_i, c_i]$  does not intersect any axis; that is,

$$(C1) \quad [b_1, c_1] \times \dots \times [b_m, c_m] \cap \{y \hat{e}_i : y \in \mathbb{R}_+\} = \emptyset, \text{ for } i = 1, \dots, m$$

or  $\prod_{i=1}^m [b_i, c_i]$  intersects exactly one, say the  $i^{\text{th}}$ ; that is

$$(C2) \quad [b_1, c_1] \times \cdots \times [b_m, c_m] \cap \{y\hat{e}_j : y \in \mathbb{R}_+\} = \begin{cases} [b_i, c_i], & \text{for } j = i, \\ \emptyset, & \text{for } j \neq i. \end{cases}$$

In (C2),  $b_j = 0 < c_i$  and for  $j \neq i$  we have  $b_i > 0$ .

We now give some properties of  $I_n$  and  $I$ .

**Lemma 2** *We have*

$$P[I(\partial B) = 0] = 1$$

for all  $B \in S$ .

*Proof.* If  $B$  satisfies (C1), the assertion is trivial since  $I$  has all its points on the axes. If  $B$  satisfies (C2),

$$I(\partial B) \leq \sum_{k=1}^{\infty} \epsilon_{j_k}(\{b_i, c_i\}) = 0$$

a.s. since the mean measure of  $\sum_{k=1}^{\infty} \epsilon_{j_k}$  is atomless.  $\square$

**Lemma 3** (a) *If  $B \in S$  satisfies (C1), then*

$$P[I(B) = 0] = 1 \text{ and } EI_n(B) \rightarrow 0.$$

(b) *If  $B \in S$  satisfies (C2),*

$$P[I(B) = 0] = \exp\{-\lambda([b_i, c_i])\}$$

and

$$E(I_n(B)) \rightarrow \lambda([b_i, c_i]).$$

*Proof.* (a) Let

$$\nu_{i_m, \dots, i_1} = P[V_m = i_m, \dots, V_1 = i_1].$$

We have

$$\begin{aligned} EI_n(B) &= nP(a_n^{-1}X_m \in [b_1, c_1], \dots, a_n^{-1}X_1 \in [b_m, c_m]) \\ &= n \sum_{i_m, \dots, i_1} P\left(\bigcap_{j=1}^n [a_n^{-1}X_{m-j+1} \in [b_j, c_j]] \mid V_m = i_m, \dots, V_1 = i_1\right) \nu_{i_m, \dots, i_1} \\ &= \sum_{i_m, \dots, i_1} F_{i_m}(a_n[b_1, c_1]) \cdots F_{i_1}(a_n[b_m, c_m]) \cdot \nu_{i_m, \dots, i_1}. \end{aligned}$$

Since  $B$  has empty intersection with the coordinate axes and the inner summation is finite, the above goes to zero as  $n \rightarrow \infty$ .

(b) We have

$$\begin{aligned}
EI_n(B) &= nP[a_n^{-1}X_k \in [b_1, c_1], \dots, a_n^{-1}X_{k-m+1} \in [b_m, c_m]] \\
&= nP\left(\bigcap_{l=1}^n [a_n^{-1}X_{m-l+1} \in [b_l, c_l] | V_m = i_m, \dots, V_1 = i_1\right) \nu_{i_m, \dots, i_1} \\
&= \sum_{i_m, \dots, i_1} F_{i_m}(a_n[b_1, c_1]) \dots F_{i_1}(a_n[b_m, c_m]) \cdot \nu_{i_m, \dots, i_1}.
\end{aligned}$$

Since  $b_j = 0 < c_j$  for  $j \neq i$ ,  $F_{i_l}(a_n[b_j, c_j]) \rightarrow 1$  as  $n \rightarrow \infty$  for  $j \neq i$ . Therefore,

$$E(I_n(B)) \sim P[V_1 = 1]F_1(a_n[b_i, c_i]) / (\nu_1 F_1(a_n))$$

and so

$$E(I_n(B)) \rightarrow \lambda[b_i, c_i]$$

This completes the proof of the lemma.  $\square$

**Proposition 2** Let  $\tilde{I}_n = \sum_{k=1}^n \sum_{i=1}^m \epsilon_{(a_n^{-1}\mathbf{Y}_k, \hat{c}_i)}$ . Then

$$I_n(B) - \tilde{I}_n(B) \xrightarrow{P} 0$$

for all  $B \in \mathcal{S}$ .

*Proof.* We can assume, without loss of generality, that  $B = [b_1, c_1] \times \dots \times [b_m, c_m]$ . If  $B$  satisfies (C1),  $EI_n(B) \rightarrow 0$  and  $\tilde{I}_n(B) = 0$  which proves the result in this case.

Suppose  $B$  satisfies (C2). Then  $b_j = 0$  for  $j \neq i$  and  $b_i > 0$ . Write

$$I_n(B) = \sum_{k=1}^{i-1} \epsilon_{(a_n^{-1}\mathbf{Y}_k)}(B) + \sum_{k=i}^n \epsilon_{(a_n^{-1}\mathbf{Y}_k)}(B). \quad (3.3)$$

By stationarity of the process  $\{X_n\}$ , the expectation of the first term is bounded above by  $(i-1) \cdot P[a_n^{-1}X_1 \in [b_i, c_i]]$  which tends to zero as  $n \rightarrow \infty$  since  $b_i > 0$ . This expresses

$$I_n(B) = o_p(1) + I_n^{(1)}(B)$$

and it suffices to show

$$I_n^{(1)}(B) - \tilde{I}_n(B) \xrightarrow{P} 0. \quad (3.4)$$

The second term,  $I_n^{(1)}(B)$  in (3.3) can be bounded above by

$$\sum_{j=1}^{n-i+1} \epsilon_{(a_n^{-1}X_j)}([b_i, c_i]) \leq \sum_{j=1}^n \epsilon_{a_n^{-1}X_j}([b_i, c_i]) = \tilde{I}_n(B).$$

Also

$$\begin{aligned} E\tilde{I}_n(B) &= \sum_{j=1}^n P[a_n^{-1}X_j \in [b_i, c_i]] \\ &= nP(a_n^{-1}X_1 \in [b_i, c_i]) \\ &\rightarrow \lambda([b_i, c_i]) \end{aligned}$$

from (3.1). We thus conclude that

$$0 \leq \tilde{I}_n(B) - I_n(B) \xrightarrow{L_1} 0$$

and therefore,

$$\tilde{I}_n(B) - I_n^{(1)}(B) \xrightarrow{P} 0$$

as required. This completes the proof of the proposition.  $\square$

**Proposition 3** *We have*

$$\sum_{k=1}^n \epsilon_{(a_n^{-1}\mathbf{Y}_k)} \implies \sum_{k=1}^{\infty} \sum_{i=1}^m \epsilon_{(j_k \hat{e}_i)}$$

in  $M_p(\mathbb{E})$ , where  $\mathbb{E} := [0, \infty)^m \setminus \{\mathbf{0}\}$  and  $\{(j_k) : k \geq 1\}$  are the points of a PRM( $\lambda$ ) on  $(0, \infty]$ .

*Proof.* By the result of the previous Proposition 2, it clearly suffices to show that

$$\tilde{I}_n = \sum_{k=1}^n \sum_{i=1}^m \epsilon_{a_n^{-1}\mathbf{Y}_{k \cdot \hat{e}_i}} \implies \sum_{k=1}^{\infty} \sum_{i=1}^m \epsilon_{j_k \hat{e}_i}.$$

By the result of Lemma 2

$$\sum_{k=1}^n \epsilon_{a_n^{-1}X_k} \implies \sum_{k=1}^{\infty} \epsilon_{j_k}$$

in  $M_p((0, \infty])$ . Further, the mapping

$$\epsilon_{v_k} \mapsto \left( \sum_{k=1}^{\infty} \epsilon_{v_k \cdot \hat{e}_1}, \dots, \sum_{k=1}^{\infty} \epsilon_{v_k \cdot \hat{e}_m} \right) \mapsto \sum_{k=1}^{\infty} \sum_{i=1}^m \epsilon_{v_k \cdot \hat{e}_i}$$

is continuous from  $M_p((0, \infty])$  into  $M_p(\mathbb{E})$ . By the continuous mapping theorem, it follows that

$$\tilde{I}_n \implies \sum_{k=1}^{\infty} \sum_{i=1}^m \epsilon_{(j_k \cdot \hat{e}_i)}$$

This completes the proof of the proposition.  $\square$

## 4 Asymptotic properties of sample correlation function.

This section concentrates on describing the asymptotic behavior of the sample autocorrelation function (acf) for the heavy tailed hidden Markov model. The classical sample acf is defined in time series analysis as

$$\hat{\rho}(h) = \frac{\sum_{t=1}^{n-h} (X_t - \bar{X})(X_{t+h} - \bar{X})}{\sum_{t=1}^n X_t^2}, \quad h = 0, 1, \dots$$

In heavy tail analysis where neither variances nor means are guaranteed to exist, there is little point to centering by means of  $\bar{X}$  and so we define

$$\hat{\rho}_H(h) = \frac{\sum_{t=1}^{n-h} X_t X_{t+h}}{\sum_{t=1}^n X_t^2}, \quad h = 0, 1, \dots$$

See Davis and Resnick (1985, 1986)

From the Proposition 3, we have

$$\sum_{k=1}^n \epsilon_{a_n^{-1} \mathbf{Y}_k} \Rightarrow \sum_{k=1}^{\infty} \sum_{i=1}^m \epsilon_{j_k \cdot \hat{e}_i} \quad (4.1)$$

Let  $\gamma > 0$ , and for each integer  $h$ ,  $0 \leq h \leq m-1$ , define the mapping

$$\Phi_{h,\gamma} : M_p(\mathbb{E}) \mapsto \mathbb{R}_+$$

by

$$\Phi_{h,\gamma} \left( \sum_{k=1}^{\infty} \epsilon_{(u_{k,0}, \dots, u_{k,m-1})} \right) = \sum_{k=1}^{\infty} u_{k,0} u_{k,h} 1_{[|u_{k,0}| > \gamma \text{ or } |u_{k,h}| > \gamma]}. \quad (4.2)$$

The map  $\Phi_{h,\gamma}$  is a.s. continuous relative to the limit point process in (4.1). Therefore, by the continuous mapping theorem,

$$\begin{aligned} & \left( \Phi_{h,\gamma} \left( \sum_{k=1}^n \epsilon_{a_n^{-1} (X_k, \dots, X_{k-m+1})} \right), h = 0, \dots, m-1 \right) \\ &= \left( a_n^{-2} \sum_{k=1}^n X_k X_{k-h} 1_{[|X_k| > a_n \gamma \text{ or } |X_{k-h}| > a_n \gamma]}, h = 0, 1, \dots, m-1 \right) \end{aligned}$$

converges in distribution in  $[0, \infty]^m$  to the random variable whose components are

$$\Phi_{h,\gamma} \left( \sum_{k=1}^{\infty} \sum_{i=1}^m \epsilon_{j_k \hat{e}_i} \right) = \begin{cases} 0, & \text{for } h > 0, \\ \sum_{k=1}^{\infty} j_k^2 1_{[|j_k| > \gamma]}, & \text{for } h = 0, \end{cases}$$

where  $(j_k^2, k \geq 1)$  are the points of a PRM( $\tilde{\lambda}$ ) on  $(0, \infty]$  with  $\tilde{\lambda}(dx) = x^{-\alpha/2-1}(\alpha/2)dx$ .

As  $\gamma \rightarrow 0$ , the limit random variable approaches  $\sum_{k=1}^{\infty} j_k^2$  for  $h = 0$  and equals zero for  $h > 0$ .

**Lemma 4** For any  $\eta > 0$ , if  $0 < \alpha < 2$ , then

$$\lim_{\gamma \rightarrow 0} \limsup_{n \rightarrow \infty} P[a_n^{-2} \left| \sum_{k=1}^n X_k X_{k-h} 1_{[|X_k| \leq a_n \gamma, |X_{k-h}| \leq a_n \gamma]} \right| > \eta] = 0$$

*Proof.* The probability in the hypothesis of the lemma is bounded above by

$$(n/(a_n^2 \eta)) E[|X_k| 1_{[|X_k| \leq a_n \gamma]} |X_{k-h}| 1_{[|X_{k-h}| \leq a_n \gamma]}]$$

By the Cauchy-Schwarz inequality, the above has an upper bound

$$(n/(a_n^2 \eta)) [E X_k^2 1_{[|X_k| \leq a_n \gamma]}]^{1/2} [E X_{k-h}^2 1_{[|X_{k-h}| \leq a_n \gamma]}]^{1/2}$$

By stationarity, the above equals  $(n/(a_n^2 \eta)) E X_1^2 1_{[|X_1| \leq a_n \gamma]}$ . From (3.1),  $X_1$  has regularly varying tail probabilities. If  $0 < \alpha < 2$ , the result follows from Karamata's Theorem.  $\square$

Now for the result about the sample acf.

**Proposition 4** Assume  $0 < \alpha < 2$ . Then we have

$$(\hat{\rho}_H(l), l = 1, \dots, m) = \left( \sum_{t=1}^{n-l} \frac{X_t X_{t+1}}{\sum_{t=1}^n X_t^2} l = 1, \dots, m \right) \Rightarrow (0, \dots, 0).$$

Thus the sample correlation function at nonzero lags is asymptotically equal to zero. This is the same behavior exhibited by the sample acf of heavy tailed iid sequences.

*Proof.* Using the result of the previous Lemma 4, we can apply Theorem 4.2 in Billingsley (1968) to conclude that

$$a_n^{-2} \left( \sum_{t=1}^n X_t^2, \sum_{t=1}^n X_t X_{t-1}, \dots, \sum_{t=1}^n X_t X_{t-m+1} \right) \Rightarrow \left( \sum_{k=1}^{\infty} j_k^2, 0, \dots, 0 \right).$$

Again, by the continuous mapping theorem, it follows that

$$\left( \sum_{t=1}^n X_t X_{t-1} / \sum_{t=1}^n X_t^2, \dots, \sum_{t=1}^n X_t X_{t-m+1} / \sum_{t=1}^n X_t X_{t-m+1} \right) \Rightarrow (0, \dots, 0).$$

This completes the proof of the proposition.  $\square$

## 5 Concluding remarks.

To illustrate the results, we simulated a five state semi-Markov process with transition matrix

$$P = \begin{pmatrix} 0.0 & 0.2 & 0.3 & 0.2 & 0.3 \\ 0.4 & 0.0 & 0.2 & 0.3 & 0.1 \\ 0.3 & 0.1 & 0.0 & 0.4 & 0.2 \\ 0.1 & 0.2 & 0.2 & 0.0 & 0.5 \\ 0.4 & 0.1 & 0.2 & 0.1 & 0.2 \end{pmatrix}.$$

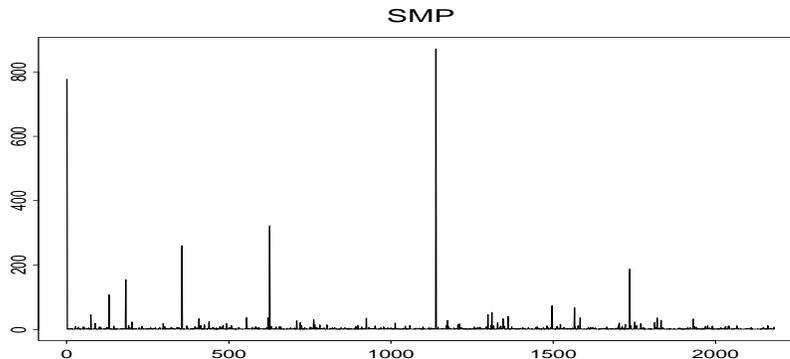


Figure 2: Time series plot.

The holding times in each state have discrete Pareto distributions with exponents 1.3, 1.4, 1.5, 1.6, 1.7 and the  $F_i$ 's were Pareto distributions with parameters 0.7, 0.9, 1.3, 1.5, 1.8. The realization of the resulting time series  $\{X_n, n \geq 1\}$  is of length 2182 and is labelled SMP. The time series plot of SMP is given next.

A popular estimator of tail heaviness is the Hill estimator and we were curious to see how this estimator performed on our data. Usually the Hill estimator is quite accurate for Pareto data and the question arose: How accurate would it be for data produced by the mixture method induced by the semi-Markov process. The results are satisfactory although one must push the Hill estimator using smoothing and alternate plotting. The Hill estimator is obtained by first computing the order statistics of the data

$$X_{(1)} \geq X_{(2)} > \cdots > X_{(n)}$$

and defining

$$H_{k,n} = \frac{1}{k} \sum_{i=1}^k \log \left( \frac{X_{(i)}}{X_{(k+1)}} \right)$$

where  $k$  is the number of order statistics used in estimation. The estimate of  $\alpha$  is  $H_{k,n}^{-1}$ . The way estimation is done in practice is to make a *Hill plot* of  $\{(k, H_{k,n}^{-1}), 1 \leq k \leq n\}$ . Resnick and Starica (1997) recommend using a supplementary plot, called the *alt-plot*, where one plots

$$\{(\theta, H_{[n^\theta],n}^{-1}), 0 \leq \theta \leq 1\}$$

and also *smoothing*. The Hill plot is given in four views in Figure 3. Considering the correct answer is 0.7, the ordinary Hill plot is not overly revealing. However, the alt plot and the smoothed alt plot come quite close to the known value of 0.7.

The final plot gives the sample acf. The values for 20 lags are quite small, with the biggest being 0.0204. Note the vertical scale extends from 0 to 0.02.

## References

- [1] P. Billingsley (1968) *Convergence of Probability Measures*. John Wiley and Sons, Inc.

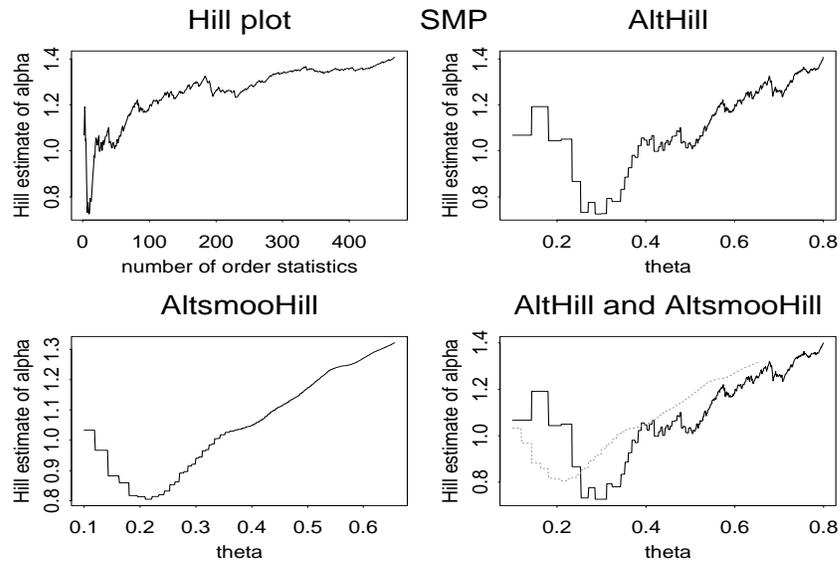


Figure 3: Hill plot.

- [2] E. Cinlar (1969) "Markov Renewal Theory," *Adv. in Appl. Probability*, 1, 123-187.
- [3] R. Davis and S. Resnick (1985) "Limit theory for moving averages of random variables with regularly varying tail probabilities," *Ann. Probability*, 13, 179-195.
- [4] R. Davis and S. Resnick (1986) "Limit theory for sample covariance and correlation functions," *Ann. Statist.*, 14, 533-558.
- [5] R. Davis and S. Resnick, (1996) "Limit theory for bilinear processes with heavy tailed noise," *Ann. Applied Probability*, 6, 1191-1210.

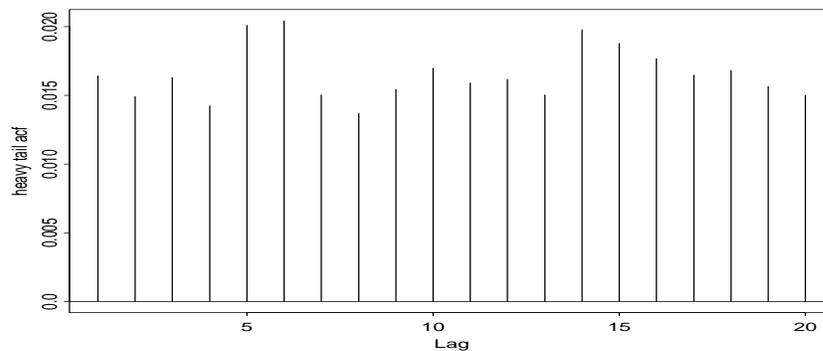


Figure 4: ACF plot of SMP

- [6] B.H. Juang and L.R. Rabiner (1991) “Hidden Markov models for speech recognition,” *Technometrics*, 33, 251–272.
- [7] A. Krogh, M. Brown, I. S. Mian, K. Sjolander and D. Haussler (1994) “Hidden Markov models in Computational Biology; Applications to protein modeling,” *J. Mol. Biol.*, 235, 1501–1531.
- [8] K.S. Meier-Hellstern, P.E. Wirth, Y.L. Yan and D.A. Hoeflin (1991) “Traffic models for ISDN data users : office automation application. In: Teletraffic and Datatraffic in a Period of Change”, *Proceedings of the 13th ITC*, editors: A.Jensen and V.B.Iversen, 167-192, North Holland, Amsterdam, The Netherlands.
- [9] S. Resnick (1987) *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York.
- [10] S. Resnick, (1996) “Why non-linearities can ruin the heavy tailed modeler’s day.” Preprint: Available as TR1157.ps.Z at <http://www.orie.cornell.edu/trlist/trlist.html>.
- [11] S. Resnick and C. Starica (1996) “ Tail Index Estimation for Dependent Data,” *School of ORIE, Cornell University, Technical Report no. 1174*.
- [12] S. Resnick and C. Stărică (1997) “ Smoothing the Hill estimator,” *J. Applied Probability*.

Sidney Resnick, School of ORIE, Cornell University, Ithaca NY 14853 USA  
sid@orie.cornell.edu

Ajay Subramanian, Center for Applied Mathematics, Cornell University, Ithaca NY 14853 USA  
ajay@cam.cornell.edu