Order Estimation for a Special Class of Hidden Markov Sources and Binary Renewal Processes

Sanjeev Khudanpur, *Member, IEEE*, and
Prakash Narayan, *Fellow, IEEE*

Invited

***Abstract*—We consider the estimation of the order, i.e., the number of hidden states, of a special class of discrete-time finite-alphabet hidden Markov sources. This class can be characterized in terms of equivalent renewal processes. No *a priori* bound is assumed on the maximum permissible order. An order estimator based on renewal types is constructed, and is shown to be strongly consistent by computing the precise asymptotics of the probability of estimation error. The probability of underestimation of the true order decays exponentially in the number of observations while the probability of overestimation goes to zero sufficiently fast. It is further shown that this estimator has the best possible error exponent in a large class of estimators. Our results are also valid for the general class of binary independent-renewal processes with finite mean renewal times.**

***Index Terms*—Error exponent, hidden Markov sources, order estimation, renewal processes, renewal types.**

I. Introduction

The problem of order estimation for Markov processes, and more generally, for the class of hidden Markov processes (also known, and hereafter referred to, as hidden Markov sources (HMSs)), has received wide attention. Several relevant results can be found in the survey paper [2].

Most existing results on the order estimation for a Markov process assume a prior bound on the permissible order; here, the notion of order corresponds to the depth of memory of the Markov process. Exceptions include the estimator based on Lempel–Ziv parsing in [17] and that based on the Bayesian Information Criterion in [6] (cf. also [3]). Another set of results of interest to us appear in [10], where a strongly consistent order estimator is presented, which additionally possesses the best "error exponent," i.e., the best attainable rate of decay—with respect to sample size—of the probability of order underestimation in a suitably large class of estimators. Such a feature is desirable in applications in which the penalty for choosing an incorrectly simpler model is unacceptably high. This result, however, is based on the nettlesome assumption of a known upper bound on the model order.

A number of strongly consistent estimators have also been proposed for the order estimation for a hidden Markov source (cf. e.g., several relevant results cited in [2], [15]), where order now refers to the number of states of the underlying Markov chain. Most of these estimators also assume a prior bound on model order. Furthermore, while order estimators with exponentially decaying probabilities of underestimation error are known, the best error exponent is yet to be characterized. (See, however, the comment in Section V concerning a recent manuscript [11] which tackles this problem.)

In this correspondence, we focus attention on a restricted class of hidden Markov sources, namely, those endowed with a special renewal structure, which enables an equivalent representation in terms of binary independent-renewal processes. Such processes arise in several applications, e.g., teletraffic modeling in networks [18]. This renewal structure leads to an interesting instance of a class of HMSs for which a strongly consistent order estimator can be constructed which possesses the best error exponent; furthermore, such an estimator does not require a prior bound on the model order. This class of processes lends itself to an analysis based on the notion of renewal types introduced in [7], which is then exploited in constructing an estimator along the lines of that proposed in [10].

In fact, our results are valid for the general class of binary independent-renewal processes with finite mean renewal times. While there is no established notion of order for this class of processes unlike in the case of the HMSs above, a suitable notion of order can be defined in terms of the indexes of any countable partition of the set of independent-renewal measures on $\{0, 1\}^\infty$. The estimator above (under some fairly mild conditions) is strongly consistent in identifying the subset of such measures which contains the measure generating the data. From a practical standpoint, a specific partition of interest is one which classifies a binary independent-renewal measure according to the number of "free parameters" needed to describe the associated probability mass function (pmf) of the independent and identically distributed (i.i.d.) renewals. An increase in order then corresponds to an increasing number of free parameters, a feature which is compatible with the standard notion of model complexity (cf. e.g., [19]).

The remainder of this correspondence is organized as follows. The special class of HMSs under consideration is defined in Section II. The notion of renewal types from [7] and relevant properties are developed in Section III, including bounds on probabilities using renewal type-counting arguments along the lines of well-known (i.i.d.) types, and attributes of differentiators between renewal processes. The order estimator is described and its performance examined in Section IV. We conclude in Section V with a discussion of our results, including order estimation for general binary processes with independent renewals.

II. Preliminaries

Let $\mathcal{S} = \{1, 2, \ldots, k\}$, $k \geq 1$, be a finite set of integers. Let $\{S_n\}_{n=0}^\infty$ be an $\mathcal{S}$-valued first-order Markov process, generated by a $k \times k$-stochastic matrix $A = \{a_{uv}\}$ with *strictly positive elements*, and with (initial state) $S_0 = 1$.[1] Here

$$a_{uv} \triangleq P(S_n = v | S_{n-1} = u), \qquad u, v \in \mathcal{S}$$

denote the (time-homogeneous) transition probabilities of the process $\{S_n\}_{n=0}^\infty$. Throughout, we shall use the notation s_m^n to refer to the subsequence $(s_m, \ldots, s_n)$, $1 < m < n$, of symbols from $\mathcal{S}$; the sequence $(s_1, \ldots, s_n)$ will simply be denoted by s^n.

Let $\mathcal{X} = \{0, \ldots, q-1\}$, $q \geq 2$, be a finite set of integers. Let $\{X_n\}_{n=1}^\infty$ be a $\mathcal{X}$-valued stochastic process which is generated by the process $\{S_n\}_{n=0}^\infty$ according to the following conditional pmf:

$$\begin{aligned} b_{il} &\triangleq P(X_n = l | S_n = i, S_{n-1}, \ldots, S_0, X_{n-1}, \ldots, X_1) \\ &= P(X_n = l | S_n = i) \end{aligned} \tag{1}$$

where $B = \{b_{il}\}$ is a $k \times q$-stochastic matrix. The process $\{X_n\}_{n=1}^\infty$ so generated, which is a function of the Markov chain $\{S_n\}_{n=0}^\infty$, is called

[1]The assumption $S_0 = 1$ facilitates subsequent exposition and does not entail any loss of essential generality.

Manuscript received June 24, 2001; revised December 12, 2001. This work was supported by the Institute for Systems Research at the University of Maryland, College Park, under NSF Grant OIR-85-00108; by the Center for Satellite and Hybrid Communication Networks, a NASA Commercial Space Center, under NASA Cooperative Agreement NCC3-528; and by DARPA under Grant N66001-00-C-8063.

S. Khudanpur is with the Department of Electrical and Computer Engineering and the Center for Language and Speech Processing, The Johns Hopkins University, Baltimore, MD 21218 USA (e-mail: khudanpur@jhu.edu).

P. Narayan is with the Department of Electrical and Computer Engineering and the Institute for Systems Research, University of Maryland, College Park, MD 20742 USA (e-mail: prakash@eng.umd.edu).

Communicated by S. Shamai, Guest Editor.

Publisher Item Identifier S 0018-9448(02)04032-4.

a *hidden Markov source* (HMS) or a hidden Markov process. The n-dimensional joint pmf of the HMS $\{X_n\}_{n=1}^\infty$ is completely determined by the stochastic matrices A and B. In particular

$$\begin{aligned} P(X^n = x^n | S_0 = 1) &= \sum_{s^n \in \mathcal{S}^n} P(X^n = x^n | S^n = s^n, S_0 = 1) P(S^n = s^n | S_0 = 1) \\ &= \sum_{s^n \in \mathcal{S}^n} \left(\prod_{t=1}^n P(X_t = x_t | S_t = s_t) P(S_t = s_t | S_{t-1} = s_{t-1}) \right) \\ &= \sum_{s^n \in \mathcal{S}^n} \left(\prod_{t=1}^n b_{s_t x_t} a_{s_{t-1} s_t} \right) \end{aligned} \tag{2}$$

where, in keeping with the previous notation, $X^n \triangleq (X_1, \ldots, X_n)$, $x^n \triangleq (x_1, \ldots, x_n)$, $n > 1$, and we interpret $P(S_1 = s_1 | S_0 = s_0) = a_{s_0 s_1}$ as being $a_{1 s_1}$. Loosely speaking, the *order of the HMS* $\{X_n\}_{n=1}^\infty$ refers to the cardinality k of the state space $\mathcal{S}$ of the underlying Markov chain $\{S_n\}_{n=0}^\infty$.

The order estimation problem for the HMS entails obtaining a consistent estimate of the (unknown) order k based on observations of the process $\{X_n\}_{n=1}^\infty$; the stochastic matrices A and B are assumed to be unknown. For each $k \geq 1$, let $\tilde{\Theta}_k$ denote the set of all pairs of stochastic matrices (A, B), where A is a $k \times k$-stochastic matrix with strictly positive elements and B is a $k \times q$-stochastic matrix. For each $\theta \in \tilde{\Theta}_k$, $k \geq 1$, let P_θ denote a probability measure on $\mathcal{X}^\infty$ generated in accordance with (1) and (2). Note that this formulation yields an increasing sequence of sets of measures $\tilde{\mathcal{P}}_k$, $k \geq 1$, on $\mathcal{X}^\infty$, corresponding, respectively, to the sets $\tilde{\Theta}_k$, $k \geq 1$. Considering the mutually disjoint sequence of measures $\mathcal{P}_k = \tilde{\mathcal{P}}_k \setminus \tilde{\mathcal{P}}_{k-1}$, $k \geq 1$, on $\mathcal{X}^\infty$ (with $\mathcal{P}_1 = \tilde{\mathcal{P}}_1$), let Θ_k denote the subset of $\tilde{\Theta}_k$ which is in one-to-one correspondence with $\mathcal{P}_k$, $k \geq 1$.

Definition: The order[2] of an HMS $\{X_n\}_{n=1}^\infty$ is defined as the (smallest) value of $k \geq 1$ for which there exists a parameter $\theta \in \Theta_k$, such that $P_\theta \in \mathcal{P}_k$ and the measure on $\mathcal{X}^\infty$ corresponding to $\{X_n\}_{n=1}^\infty$ are equal.

We shall limit ourselves to a study of the order estimation problem for a special class of HMSs in (1) and (2), namely, the class of *renewal processes*. Specifically, we consider those $\{0, 1\}$-valued HMSs (i.e., with $q = 2$) for which the $k \times 2$-stochastic matrix B in (1) is of the form

$$b_{il} = \begin{cases} 1, & \text{if } (i = 1 \ \& \ l = 1) \text{ or } (i \neq 1 \ \& \ l = 0) \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

Thus, the observed symbol is 1 iff the underlying Markov process is in the state 1, so that the process $\{S_n\}_{n=0}^\infty$ remains by and large "hidden," and an occurrence of $S_n = 1$ is revealed when, and only when, we observe $X_n = 1$. It can be shown that the HMS $\{X_n\}_{n=1}^\infty$ is now a $\{0, 1\}$-valued renewal process. The n's for which $X_n = 1$ constitute renewal epochs, and one plus the number of intervening 0's between successive 1's corresponds to a renewal time (cf., e.g., [4, Example 9.1.20]). Formally, let Γ_j, $j \geq 1$, be the *epoch* of the jth renewal, i.e., the instant of the jth occurrence of the symbol 1 in $\{X_n\}_{n=1}^\infty$. Setting

$$Y_j = \begin{cases} \Gamma_1, & j = 1 \\ \Gamma_j - \Gamma_{j-1}, & j \geq 2 \end{cases} \tag{4}$$

it can be shown from (1)–(3) that the sequence $\{Y_j\}_{j=1}^\infty$ of *renewal times*, which depicts the interarrival times of the symbols 1 in $\{X_n\}_{n=1}^\infty$ is an i.i.d. sequence under P_θ, $\theta \in \Theta_k$, $k \geq 1$ [4, Proposition 5.2.1]. Formally, for each $k \geq 1$ and $\theta \in \Theta_k$, we have from [4, Proposition 5.2.1] that for all $j \geq 1$, $y \geq 2$

$$\begin{aligned} P_\theta(Y_{j+1} = y | Y_1, \ldots, Y_j, S_0 = 1) &= P_\theta(\Gamma_{j+1} - \Gamma_j = y | \Gamma_1, \ldots, \Gamma_j, S_0 = 1) \\ &= P_\theta(\Gamma_1 = y | S_0 = 1) \quad \text{a.s.} \\ &= P_\theta(Y_1 = y | S_0 = 1) \quad \text{a.s.} \end{aligned} \tag{5}$$

where,[3] of course,

$$P_\theta(Y_1 = y | S_0 = 1) = \begin{cases} P_\theta(X_1 = 1 | S_0 = 1), & y = 1 \\ P_\theta(X_1 = 0, \ldots, X_{y-1} = 0, \\ \qquad X_y = 1 | S_0 = 1), & y \geq 2. \end{cases} \tag{6}$$

Hereafter, our standing assumption throughout that $S_0 = 1$ will not be displayed explicitly. All statements regarding probability values and expectations under the measures P_θ, $\theta \in \Theta_k$, $k \geq 1$, will be tacitly understood as being conditioned on the event $\{S_0 = 1\}$. For instance, $P_\theta(\cdot | S_0 = 1)$ will simply be denoted by $P_\theta(\cdot)$, etc.

Note that for each $k \geq 1$, the Markov process $\{S_n\}_{n=0}^\infty$ has no transient states, by the assumed positivity of the elements of the stochastic matrix A; also, no state is null recurrent (cf. e.g., [20, Problem 4.11, p. 135]). Thus, all the states are positive recurrent. It then follows for each $\theta \in \Theta_k$, $k \geq 1$, that

$$\mu_\theta \triangleq E_\theta[Y_1] < \infty. \tag{7}$$

III. Renewal Types and Properties

We now describe the notion of *renewal types*, introduced in [7], which constitute a prime technical tool, and derive several useful properties. Given a sequence $x^n \in \{0, 1\}^n$, $n \geq 1$, let

$$N = N(x^n) = \sum_{t=1}^n x_t$$

denote the number of renewals in x^n. Let $\gamma_j = \gamma_j(x^n)$, $0 \leq j \leq N$, be the renewal epochs in x^n, where we take $\gamma_0 = 0$ corresponding to $N = 0$. For $N \geq 1$, let $y_j = y_j(x^n)$, $1 \leq j \leq N$, be the renewal times of the realization x^n as defined in (4).

The *renewal type* of a sequence $x^n \in \{0, 1\}^n$, $n \geq 1$, with $N \geq 1$, denoted $K(x^n) = (k_1, \ldots, k_n)$, is an n-tuple of nonnegative integers defined by

$$k_m = k_m(x^n) \triangleq \sum_{j=1}^N \mathbb{1}(y_j = m), \qquad 1 \leq m \leq n \tag{8}$$

where $\mathbb{1}(\cdot)$ is the indicator function. Note that by construction, $\sum_{m=1}^n k_m = N$.

The following observations concerning renewal types are straightforward counterparts of well-known results on i.i.d. types (cf. e.g., [5]).

[2]This notion of order is distinct from that which arises in the context of order estimation for Markov chains, where it generally refers to the depth of memory. An HMS may not have a finite depth of memory, and the present (standard) definition of order, which relates directly to model complexity in terms of the number of "free parameters" (cf. e.g., [19]), is appropriate.

[3]In fact, the right-hand side of (5) remains unchanged when, in the left-hand side $S_0 = 1$ is replaced by $S_0 = u$ for *any* $1 \leq u \leq k$. Thus, our standing assumption throughout that $S_0 = 1$ entails no loss of essential generality.

The renewal type of $x^n \in \{0, 1\}^n$ suffices for computing $P_\theta(X^n = x^n)$, $\theta \in \Theta_k$, $k \geq 1$. Specifically, for each $x^n \in \{0, 1\}^n$ with $N \geq 1$

$$\begin{aligned} P_\theta(X^n = x^n) &= \left[\prod_{j=1}^{N} P_\theta(Y_j = y_j)\right] \times P_\theta(\Gamma_{N+1} > n | \Gamma_N = \gamma_N) \\ &= \left[\prod_{j=1}^{N} P_\theta(Y_1 = y_j)\right] \times P_\theta(Y_1 > n - \gamma_N) \\ &= \prod_{m=1}^{n} p_\theta(m)^{k_m} \times \overline{p}_\theta(n - \gamma_N) \end{aligned} \tag{9}$$

by (4), (5), and (8), where

$$p_\theta(m) \triangleq P_\theta(Y_1 = m), \qquad m \geq 1 \tag{10}$$

is a pmf on $\mathbb{N}$, the set of positive integers, and

$$\overline{p}_\theta(n - \gamma_N) \triangleq P_\theta(Y_1 > n - \gamma_N). \tag{11}$$

Note that $\gamma_N = \sum_{m=1}^{n} m k_m$. Upon defining a pmf $q = q_{K(x^n)}$ on $\{1, \ldots, n\}$ by

$$q(m) = q_{K(x^n)}(m) \triangleq \frac{k_m}{N}, \qquad 1 \leq m \leq n \tag{12}$$

we obtain from (9) in a straightforward manner, for each $x^n \in \{0, 1\}^n$ with $N \geq 1$, that

$$P_\theta(X^n = x^n) = \exp\{-N[H(q) + D(q \| p_\theta)]\} \times \overline{p}_\theta(n - \gamma_N) \tag{13}$$

where $H(q)$ denotes the entropy[4] of the pmf q and $D(q\|p_\theta)$ is the (Kullback–Leibler) divergence between q, viewed as a pmf on $\mathbb{N}$, and p_θ. For $x^n \in \{0, 1\}^n$ with $N = 0$, the expression in (13) remains valid with $\gamma_0 = 0$ and upon setting $\exp\{-N[H(q) + D(q\|p_\theta)]\} = 1$.

Next, let $\mathcal{K}^{(n)}$ represent the set of all *bona fide* renewal types for sequences in $\{0, 1\}^n$, i.e., the set of all $(k_1, \ldots, k_n) \in \{0, 1, \ldots, n\}^n$ satisfying $\sum_{m=1}^{n} m k_m \leq n$. Given a renewal type $K = (k_1, \ldots, k_n) \in \mathcal{K}^{(n)}$, set

$$N_K \triangleq \sum_{m=1}^{n} k_m \tag{14}$$

and, whenever $N_K > 0$, define a pmf q_K on $\{1, \ldots, n\}$ by

$$q_K(m) \triangleq \frac{k_m}{N_K}, \qquad 1 \leq m \leq n. \tag{15}$$

For $K \in \mathcal{K}^{(n)}$ with $N_K > 0$, let[5]

$$T_K \triangleq \{x^n \in \{0, 1\}^n \colon K(x^n) = K\}. \tag{16}$$

Note that all the sequences in T_K have the same last renewal epoch γ_{N_K}, and can differ from each other only by a permutation of the "completed" renewal times Y_j, $1 \leq j \leq N_K$. Hence, in analogy with [5, Lemma 2.3], the cardinality of the set T_K can be bounded according to

$$(N_K + 1)^{-\|K\|} \exp\{N_K H(q_K)\} \leq |T_K| \leq \exp\{N_K H(q_K)\} \tag{17}$$

where $\|K\| \triangleq \sum_{m=1}^{n} \mathbb{1}(k_m > 0)$. Since

$$\begin{aligned} \max_{K \in \mathcal{K}^{(n)}} \|K\| &= \max\{r \colon 1 + \cdots + r \leq n\} \\ &= \max\left\{r \colon \frac{r(r+1)}{2} \leq n\right\} \\ &= \max\{r \colon r^2 + r - 2n \leq 0\} \\ &\leq \sqrt{2n} \end{aligned} \tag{18}$$

we get from (17) and (18) that

$$(n+1)^{-\sqrt{2n}} \exp\left\{n \frac{H(q_K)}{\mu_K}\right\} \leq |T_K| \leq \exp\left\{n \frac{H(q_K)}{\mu_K}\right\} \tag{19}$$

where $\mu_K \triangleq \frac{n}{N_K} \geq 1$ is the mean of the pmf q_K associated with the renewal type K (cf. (14) and (15)).

It will also be convenient to consider the *circular version* of a renewal type. Specifically, the *circular version* of a renewal type

$$K = (k_1, \ldots, k_n) \in \mathcal{K}^{(n)}, \qquad \text{with } N_K > 0$$

is the renewal type $\mathring{K} \in \mathcal{K}^{(n+1)}$ defined by[6]

$$\mathring{k}_m \triangleq \begin{cases} k_m + \mathbb{1}(m = z_K + 1), & 1 \leq m \leq n \\ 0, & m = n + 1 \end{cases} \tag{20}$$

where

$$z_K \triangleq n - \sum_{m=1}^{n} m k_m = n - \gamma_N \tag{21}$$

is the length of the sequence of trailing 0's in any $x^n \in T_K$.

The circular version $\mathring{K}$ of the renewal type K can be interpreted as the renewal type of the sequence $x^{n+1} \in \{0, 1\}^{n+1}$ with $x^n \in T_K$ and $x_{n+1} = 1$; note that $\mathring{K}$ differs from K only in the count of one renewal length, namely, $z_K + 1$. Defining a pmf $q_{\mathring{K}}$ on $\{1, \ldots, n\}$ by

$$q_{\mathring{K}}(m) \triangleq \frac{\mathring{k}_m}{N_K + 1}, \qquad 1 \leq m \leq n \tag{22}$$

we see that the variational distance between $q_{\mathring{K}}$ and q_K can be bounded according to

$$\begin{aligned} \left\| q_{\mathring{K}} - q_K \right\| &\triangleq \sum_{m=1}^{n} \left| q_{\mathring{K}}(m) - q_K(m) \right| \\ &\leq \frac{2}{N_K + 1} = \frac{2\mu_K}{n + \mu_K}. \end{aligned} \tag{23}$$

For a given renewal type $K \in \mathcal{K}^{(n)}$ with $N_K > 0$, the cardinality of T_K as well as the probabilities of sequences in T_K are related to the circular version $\mathring{K}$ of K. By definition, every $(n+1)$-length sequence in $T_{\mathring{K}}$ satisfies $z_{\mathring{K}} = 0$ and contains at least one renewal of length $z_K + 1$. Now, fix $x^n \in T_K$ and note that its prefix x^{n-z_K} contains exactly N_K 1's. Upon inserting into x^{n-z_K} a substring of length $z_K + 1$ (of z_K 0's followed by a 1) immediately after any of these N_K 1's or as a prefix, we obtain an $(n+1)$-length sequence in $T_{\mathring{K}}$. Conversely, by stripping off any one of the $(z_K + 1)$-length substrings of the same kind as above from a sequence $\tilde{x}^{n+1} \in T_{\mathring{K}}$, and appending a substring of z_K 0's, we obtain an n-length sequence in T_K. Therefore,

$$|T_K| \leq \left| T_{\mathring{K}} \right| \leq (N_K + 1) |T_K|. \tag{24}$$

Combining this with (17), we get that the cardinality of T_K can be bounded in terms of quantities involving $\mathring{K}$ according to

$$\begin{aligned} \left(N_{\mathring{K}} + 1\right)^{-\|\mathring{K}\|+1} &\exp\left\{N_{\mathring{K}} H\left(q_{\mathring{K}}\right)\right\} \\ &\leq |T_K| \leq \exp\left\{N_{\mathring{K}} H\left(q_{\mathring{K}}\right)\right\}. \end{aligned} \tag{25}$$

[4]All logarithms and exponentials are with respect to the base 2.

[5]There is only one renewal type $K \in \mathcal{K}^{(n)}$ with $N_K = 0$, containing only the all-zero sequence, which can be dealt with separately.

[6]For the lone renewal type K with $N_K = 0$, the circular version is given by $\mathring{k}_m = 0$ for $1 \leq m \leq n$ and $\mathring{k}_{n+1} = 1$. Since this type contains only the all-zero sequence, it again can be dealt with separately.

Next, for every $x^n \in T_K$, where $K \in \mathcal{K}^{(n)}$ with $N_K > 0$, we see for each $\theta \in \Theta_k$, $k \geq 1$, that

$$\begin{aligned} P_\theta(X^n = x^n) &\geq P_\theta(X^n = x^n, X_{n+1} = 1) \\ &= \prod_{m=1}^{n+1} \{p_\theta(m)\}^{\overset{\circ}{k}_m} \\ &= \exp\left\{-N_{\overset{\circ}{K}}\left[H\left(q_{\overset{\circ}{K}}\right) + D\left(q_{\overset{\circ}{K}}\middle\| p_\theta\right)\right]\right\}. \end{aligned} \quad (26)$$

We are thus led to the following bounds on the total probability of all sequences of a particular renewal type $K \in \mathcal{K}^{(n)}$.

Lemma 1: Let $K \in \mathcal{K}^{(n)}$ be a renewal type with $N_K > 0$. Then, for each $\theta \in \Theta_k$, $k \geq 1$, it holds that

$$\begin{aligned} &(n+2)^{-(\sqrt{2n}+1)} \exp\left\{-(n+1)\frac{D\left(q_{\overset{\circ}{K}}\middle\| p_\theta\right)}{\mu_{\overset{\circ}{K}}}\right\} \\ &\qquad \leq P_\theta(X^n \in T_K) \leq \exp\left\{-n\frac{D(q_K\|p_\theta)}{\mu_K}\right\}. \end{aligned} \quad (27)$$

Proof: The upper bound follows from (13) and (19), and the lower bound from (25) and (26). □

Remark: For $K \in \mathcal{K}^{(n)}$ with $N_K = 0$, the upper bound in (27) holds trivially with the convention $\mu_K = \infty$. The lower bound also holds since it can be verified that it does not exceed $P_\theta(X^n \in T_K) = P_\theta(Y_1 > n)$.

Given two (renewal) pmfs p', p on $\mathbb{N}$, the set of positive integers, the quantity $\frac{D(p'\|p)}{\mu_{p'}}$, with $\mu_{p'}$ denoting the mean of p', will play a crucial role in discriminating between renewal processes with renewal times distributed according to p' and p. We close this section by assembling below certain relevant technical properties. The notation $p' \ll p$ in what follows indicates that the support of p' is contained in that of p.

Lemma 2: Let p be a pmf on $\mathbb{N}$ with $\mu_p < \infty$. Then, for all n sufficiently large, it holds that

i) $$\min_{K\in\mathcal{K}^{(n)}} \frac{D(q_K\|p)}{\mu_K} = o\left(\frac{1}{\sqrt{n}}\right) \quad (28)$$

ii) $$\min_{K\in\mathcal{K}^{(n)}} D(q_K\|p) = o\left(\frac{1}{\sqrt{n}}\right) \quad (29)$$

where μ_K is the mean of the pmf q_K associated with the renewal type K (cf. (15)).

Proof: The proof is relegated to Appendix I. □

Lemma 3: Let p be a pmf on $\mathbb{N}$. Then

i) $\frac{D(p'\|p)}{\mu_{p'}}$ is continuous in p' on the set of pmfs $\{p' : p' \ll p\}$.

ii) Assume that the moment generating function of p exists in a neighborhood of zero. For $1 \leq \mu^* < \infty$, let

$$\Lambda(\mu^*) = \{p' : \mu_{p'} = \mu^*\}.$$

Then

$$D(\Lambda(\mu^*)\|p) \triangleq \min_{p'\in\Lambda(\mu^*)} \frac{D(p'\|p)}{\mu_{p'}} \quad (30)$$

is an increasing function of μ^* on $[\mu_p, \infty)$. In particular

$$\lim_{\mu^*\to\infty} D(\Lambda(\mu^*)\|p) > 0. \quad (31)$$

Proof: See Appendix II. □

IV. An Optimal Order Estimator Based on Renewal Types

We now present our estimator of the HMS order, which is based on the renewal type of the observed sequence $x^n \in \{0, 1\}^n$, $n > 1$, and examine its consistency and optimality properties.

It will be convenient to introduce the following sets of renewal types:

$$\mathcal{A}^{(n)}(\theta) = \left\{K \in \mathcal{K}^{(n)} : \mu_K < \rho_n \text{ and } \frac{D(q_K\|p_\theta)}{\mu_K} > \epsilon_n\right\} \quad (32)$$

$$\mathcal{B}^{(n)}(\theta) = \left\{K \in \mathcal{K}^{(n)} : \mu_K < \rho_n \text{ and } \frac{D(q_K\|p_\theta)}{\mu_K} \leq \epsilon_n\right\} \quad (33)$$

$$\mathcal{C}^{(n)} = \left\{K \in \mathcal{K}^{(n)} : \mu_K \geq \rho_n\right\} \quad (34)$$

where ρ_n and ϵ_n, $n \geq 1$, are as described next.

For some constant $c > 0$ (prescribed later in the proof of Proposition 1) and $\delta > 0$, let

$$\epsilon_n \triangleq \left(c\frac{\sqrt{n}}{\log n} + \delta\right)\frac{\log n}{n} \quad (35)$$

and let ρ_n, $n \geq 1$, be a sequence such that

$$\rho_n > 1, \qquad n \geq 1 \quad (36)$$

$$\lim_{n\to\infty} \rho_n = +\infty \quad (37)$$

and

$$\lim_{n\to\infty} \rho_n \epsilon_n = 0. \quad (38)$$

(For instance, $\rho_n = n^{\frac{1}{4}}$ will suffice.)

Definition: Given $x^n \in \{0, 1\}^n$, $n > 1$, define the order estimator $\hat{k}_n$ according to (39) shown at the bottom of the page.

Remark: Note that $\hat{k}_n$ does not require any prior bound on the HMS order.[7]

The choice of the estimator above is motivated by the fact that the renewal type $K(x^n)$ of an observed sample x^n will eventually be trapped in a normalized "divergence neighborhood" $\mathcal{B}^{(n)}(\theta)$ of the renewal measure p_θ, $\theta \in \Theta_k$, which generated the sample. The "radius" ϵ_n of this neighborhood is carefully chosen (cf. (35)) so as to appropriately control the probabilities of underestimation and overestimation. Specifically, the neighborhood shrinks rapidly enough so as to eventually exclude $K(x^n)$ from the proximity of any lower order measure, thereby diminishing the likelihood of underestimation; at the same time, it shrinks slowly enough for $K(x^n)$ to be eventually

[7]For completeness, the order estimate corresponding to an observed all-zero sequence is set to $+\infty$. This does not materially change any of the subsequent calculations, as the P_θ-probability of the all-zero sequence vanishes exponentially fast in n.

$$\hat{k}_n(x^n) \triangleq \begin{cases} k, & \text{if } K(x^n) \in \left[\bigcap_{k'<k}\bigcap_{\theta'\in\Theta_{k'}} \mathcal{A}^{(n)}(\theta')\right] \bigcap \left[\bigcup_{\theta\in\Theta_k} \mathcal{B}^{(n)}(\theta)\right] \\ +\infty, & \text{if } K(x^n) \in \mathcal{C}^{(n)} \bigcup \left[\bigcap_{k'\geq 1}\bigcap_{\theta'\in\Theta_{k'}} \mathcal{A}^{(n)}(\theta')\right] \end{cases}. \quad (39)$$

included in it with high probability, thereby diminishing the likelihood of overestimation.

A. Probability of Overestimation

We first bound the probability of overestimation of this estimator as follows.

Proposition 1: Consider the order estimator in (39). For every $\theta \in \Theta_k$, $k \geq 1$, it holds that

$$\lim_{n\to\infty} P_\theta(\hat{k}_n(X^n) > k) = 0 \tag{40}$$

for a suitable choice of the constant c and any $\delta > 0$ in (35).

Proof: Fix k and $\theta \in \Theta_k$. Then

$$\begin{aligned} P_\theta(\hat{k}_n(X^n) > k) &\leq P_\theta\left(X^n \in \bigcup_{K\in\mathcal{A}^{(n)}(\theta)} T_K\right) + P_\theta\left(X^n \in \bigcup_{K\in\mathcal{C}^{(n)}} T_K\right) \\ &\leq \left|\mathcal{A}^{(n)}(\theta)\right| \max_{K\in\mathcal{A}^{(n)}(\theta)} P_\theta\left(X^n \in T_K\right) \\ &\quad + \left|\mathcal{C}^{(n)}\right| \max_{K\in\mathcal{C}^{(n)}} P_\theta\left(X^n \in T_K\right) \\ &= \left|\mathcal{A}^{(n)}(\theta)\right| \max_{K\in\mathcal{A}^{(n)}(\theta)} P_\theta\left(X^n \in T_K\right) \\ &\quad + \left|\mathcal{C}^{(n)}\right| \max_{\mu\geq\rho_n} \max_{K\in\mathcal{C}^{(n)}:\, \mu_K=\mu} P_\theta\left(X^n \in T_K\right) \\ &\leq \left|\mathcal{A}^{(n)}(\theta)\right| \max_{K\in\mathcal{A}^{(n)}(\theta)} P_\theta\left(X^n \in T_K\right) \\ &\quad + \left|\mathcal{C}^{(n)}\right| \max_{\mu\geq\rho_n} \exp(-nD(\Lambda(\mu)\|p_\theta)) \\ &\leq \left|\mathcal{A}^{(n)}(\theta)\right| \max_{K\in\mathcal{A}^{(n)}(\theta)} P_\theta\left(X^n \in T_K\right) \\ &\quad + \left|\mathcal{C}^{(n)}\right| \exp\left(-n\min_{\mu\geq\rho_n} D(\Lambda(\mu)\|p_\theta)\right) \\ &\leq \left|\mathcal{K}^{(n)}\right| \max\{\exp(-n\epsilon_n),\, \exp(-nD(\Lambda(\rho_n)\|p_\theta))\} \end{aligned} \tag{41}$$

for all sufficiently large n, where the bounds follow from Lemma 1 and Lemma 3 ii). Next, note that $|\mathcal{K}^{(n)}| \leq \exp(c\sqrt{n})$ by the Hardy–Ramanujan theorem [12] (cf. e.g., [7, p. 2068]). Furthermore, by (31) and (36)–(38), $D(\Lambda(\rho_n)\|p_\theta)$ is bounded strictly away from zero for all n sufficiently large (depending on θ). We can then complete the bounding in (41) as

$$\begin{aligned} P_\theta(\hat{k}_n(X^n) > k) &\leq \exp\left(c\sqrt{n}\right)\exp(-n\epsilon_n) \\ &= \exp\left\{-n\left(\epsilon_n - \frac{c}{\sqrt{n}}\right)\right\} \end{aligned} \tag{42}$$

for all n sufficiently large. Finally, by (35) and (42), where c is chosen according to [12], we see that for every $\delta > 0$, it holds that

$$P_\theta(\hat{k}_n(X^n) > k) \leq n^{-\delta} \tag{43}$$

for all n sufficiently large, thereby establishing the proposition. □

B. Probability of Underestimation

Before examining the probability of underestimation of $\hat{k}_n$, we present a relevant technical lemma. To this end, fix $k > 1$, $\theta \in \Theta_k$, and $k' < k$. For ϵ_n, ρ_n, $n \geq 1$, as in (35) and (36)–(38), set

$$\mathcal{B}^{(n)}(\Theta_{k'}) \triangleq \left\{K \in \mathcal{K}^{(n)}\colon \mu_K < \rho_n \text{ and } \frac{D(q_K\|p_{\theta'})}{\mu_K} \leq \epsilon_n \text{ for some } \theta' \in \Theta_{k'}\right\}. \tag{44}$$

Lemma 2 ii) shows, together with (36)–(38), that $\mathcal{B}^{(n)}(\Theta_{k'})$ is nonempty for all sufficiently large n. The set $\mathcal{B}^{(n)}(\Theta_{k'})$ does indeed contain—for each $p_{\theta'}$, $\theta' \in \Theta_{k'}$—the renewal types $\tilde{K}_N$ constructed in the proof of Lemma 2 ii) (cf. Appendix I) and neighboring renewal types.

Lemma 4: Fix $1 \leq k' < k$ and $\theta \in \Theta_k$. Then

$$\lim_{n\to\infty} \min_{K\in\mathcal{B}^{(n)}(\Theta_{k'})} \frac{D(q_K\|p_\theta)}{\mu_K} = \inf_{\theta'\in\Theta_{k'}} \frac{D(p_{\theta'}\|p_\theta)}{\mu_{\theta'}} \triangleq D(\Theta_{k'}\|P_\theta). \tag{45}$$

Remark: The infimum on the right-hand side of (45), henceforth denoted as $D(\Theta_{k'}\|P_\theta)$, is positive by [15, Lemma 3.2] and Lemma 3 ii). First, for any sequence in $\Theta_{k'}$, $k' < k$, for which the sequence of the corresponding $\mu_{\theta'}$'s has a finite upper bound, the infimum in (45) is positive by virtue of the positivity of the infimum of the numerator alone [15, Lemma 3.2]. Alternatively, if the sequence in $\Theta_{k'}$, $k' < k$, is such that the sequence of the corresponding $\mu_{\theta'}$'s is unbounded, the infimum in (45) is positive by Lemma 3 ii).

Proof: Let $\theta'_* \in \Theta_{k'}$ be such that for a given $\eta > 0$

$$\frac{D\left(p_{\theta'_*}\middle\| p_\theta\right)}{\mu_{\theta'_*}} \leq \inf_{\theta'\in\Theta_{k'}} \frac{D(p_{\theta'}\|p_\theta)}{\mu_{\theta'}} + \frac{\eta}{2}. \tag{46}$$

Let $K^n = K^n(\theta'_*)$ be a renewal type for which the variational distance from $p_{\theta'_*}$ satisfies

$$\|q_{K^n} - p_{\theta'_*}\| \leq \sqrt{2\ln 2 D\left(q_{K^n}\middle\|p_{\theta'_*}\right)} \leq \sqrt{2\ln 2\epsilon_n}. \tag{47}$$

Such types exist for all sufficiently large n, by Lemma 2 ii). Clearly, $K^n \in \mathcal{B}^{(n)}(\Theta_{k'})$. By the continuity property of $\frac{D(p'\|p_\theta)}{\mu_{p'}}$ in p', it holds for all sufficiently large n that

$$\frac{D(q_{K^n}\|p_\theta)}{\mu_{K^n}} \leq \frac{D\left(p_{\theta'_*}\middle\| p_\theta\right)}{\mu_{\theta'_*}} + \frac{\eta}{2} \tag{48}$$

and it follows, therefore, that for all sufficiently large n

$$\min_{K\in\mathcal{B}^{(n)}(\Theta_{k'})} \frac{D(q_K\|p_\theta)}{\mu_K} \leq \inf_{\theta'\in\Theta_{k'}} \frac{D(p_{\theta'}\|p_\theta)}{\mu_{\theta'}} + \eta. \tag{49}$$

Next, for each n, let

$$K_*^n = \arg\min_{K\in\mathcal{B}^{(n)}(\Theta_{k'})} \frac{D(q_K\|p_\theta)}{\mu_K}. \tag{50}$$

Since $K_n^* \in \mathcal{B}^{(n)}(\Theta_{k'})$, by (44), there exists $\theta'_n \in \Theta_{k'}$ such that

$$\begin{aligned} \left\|q_{K_*^n} - p_{\theta'_n}\right\| &\leq \sqrt{2\ln 2D\left(q_{K_*^n}\middle\| p_{\theta'_n}\right)} \\ &\leq \sqrt{2\ln 2\mu_{K_*^n}\epsilon_n} \leq \sqrt{2\ln 2\rho_n\epsilon_n}. \end{aligned} \tag{51}$$

Therefore, for sufficiently large n, we have

$$\frac{D\left(q_{K_*^n}\middle\| p_\theta\right)}{\mu_{K_*^n}} \geq \frac{D\left(q_{\theta'_n}\middle\| p_\theta\right)}{\mu_{\theta'_n}} - \eta \geq \inf_{\theta'\in\Theta_{k'}} \frac{D(p_{\theta'}\|p_\theta)}{\mu_{\theta'}} - \eta. \tag{52}$$

The assertion of the lemma follows. □

Proposition 2: Consider the order estimator in (39). For each $\theta \in \Theta_k$, $k \geq 1$, given any $\eta > 0$, it holds that

$$P_\theta(\hat{k}_n(X^n) < k) \leq \exp\left\{-n\left[D\left(\Theta^{(k-1)}\middle\| P_\theta\right) - \eta\right]\right\} \tag{53}$$

for all sufficiently large n, where

$$D\left(\Theta^{(k-1)} \middle\| P_\theta\right) \triangleq \min_{k'<k} \inf_{\theta' \in \Theta_{k'}} \frac{D(p_{\theta'} \| p_\theta)}{\mu_{\theta'}} = \min_{k'<k} D(\Theta_{k'} \| P_\theta) > 0. \tag{54}$$

Remark: The estimator $\hat{k}_n$ in (39) is strongly consistent, as can be seen from (43) (with $\delta > 1$) and (53) by a standard application of the Borel–Cantelli lemma.

Proof: Fix $k' < k$ and $\theta \in \Theta_k$. Then

$$\begin{aligned} P_\theta(\hat{k}_n(X^n) = k') &\le P_\theta\left(\bigcup_{K \in \mathcal{B}^{(n)}(\Theta_{k'})} T_K\right) \\ &= \sum_{K \in \mathcal{B}^{(n)}(\Theta_{k'})} P_\theta(T_K) \\ &\le \exp\left\{c\sqrt{n}\right\} \exp\left\{-n \min_{K \in \mathcal{B}^{(n)}(\Theta_{k'})} \frac{D(q_K \| p_\theta)}{\mu_K}\right\} \\ &\le \exp\left\{-n\left[D(\Theta_{k'} \| P_\theta) - \frac{\eta}{2}\right]\right\} \end{aligned} \tag{55}$$

for all $n > N(\eta, \theta, k')$, by Lemma 4. Hence,

$$\begin{aligned} P_\theta(\hat{k}_n(X^n) < k) &= \sum_{k'<k} P_\theta(\hat{k}_n(X^n) = k') \\ &\le (k-1)\exp\left\{-n\left[\min_{k'<k} D(\Theta_{k'} \| P_\theta) - \frac{\eta}{2}\right]\right\} \\ &\le \exp\left\{-n\left[D\left(\Theta^{(k-1)} \middle\| P_\theta\right) - \eta\right]\right\} \end{aligned} \tag{56}$$

for all sufficiently large n. □

C. Optimality of the Estimator

In the spirit of the results in [10], and mimicking the same approach, it can be shown that the estimator in (39) is optimal in the sense that among the class of estimators whose probability of overestimation error is *uniformly* bounded away from unity, none can be found with a smaller probability of underestimation (i.e., with a larger error exponent) than in Proposition 2.

Proposition 3: Let k_n be an estimator which, for each $\theta \in \Theta_k$, $k \ge 1$, satisfies

$$\limsup_{n \to \infty} P_\theta(k_n(X^n) > k) < \alpha \tag{57}$$

for some $0 < \alpha < 1$. Then, for every $\epsilon > 0$, for all $\theta \in \Theta_k$, $k \ge 1$, and for all sufficiently large n (depending on α, ϵ, k and θ), it holds that

$$P_\theta(k_n(X^n) < k) \ge \exp\left\{-n\left[D\left(\Theta^{(k-1)} \middle\| P_\theta\right) + \epsilon\right]\right\}. \tag{58}$$

Proof: The proof, which is along the lines of [10, Theorem 2], is relegated to Appendix III. □

V. Discussion

Given an order estimation problem, the optimal error exponent is, loosely speaking, the best rate of (exponential) decay—as a function of sample size—of the probability of underestimation among all estimators whose probability of overestimation is uniformly bounded away from unity over the entire range of (unknown) measures that could have generated the observations. For the problem of estimating the order of dependence, i.e., the depth of memory, of a Markov process, the optimal error exponent has been characterized in [10]. Furthermore, a strongly consistent estimator of Markov order was constructed, which achieved this error exponent.

General HMSs: For the more general class of HMSs, the determination of the optimal error exponent had remained an open problem.[8] In [15], a strongly consistent order estimator based on "mixture" probabilities, with an exponentially decaying underestimation probability, was presented: however, the corresponding exponent was not explicitly characterized. In [13], another strongly consistent order estimator, based on maximum-likelihood parameter estimates, was presented, also with an exponentially decaying underestimation probability. It had been unclear if this exponent were the best possible.

The results presented here constitute a solution for a special class of HMSs which possess a renewal structure. This special structure also enables us to construct an optimal estimator without assuming a prior bound on the order of the HMS.

A natural candidate for the optimal error exponent in the order estimation for a general HMS is developed below by establishing a lower bound for the probability of underestimation of any order estimator whose overestimation probability is uniformly bounded away from one. For the special cases of Markov processes and the class of renewal processes considered earlier, this candidate does indeed coincide with the known optimal error exponents.

To this end, the divergence between two stationary and ergodic HMSs P_θ and $P_{\theta'}$ on $\mathcal{X}^\infty$ is defined [9], [14] as

$$D(P_{\theta'} \| P_\theta) \triangleq \lim_{n \to \infty} E_{P_{\theta'}}\left[\log \frac{P_{\theta'}(X_{n+1} | X^n)}{P_\theta(X_{n+1} | X^n)}\right] \tag{59}$$

where $E_{P_{\theta'}}$ denotes expectation with respect to the measure $P_{\theta'}$. An alternative equivalent characterization of $D(P_{\theta'} \| P_\theta)$ is as the $P_{\theta'}$-almost sure limit (in n) of $[\frac{1}{n}\log\frac{P_{\theta'}(X^n)}{P_\theta(X^n)}]$ [9], [14].

The following large deviations result, due to [1] (cf. also [8]), is relevant in characterizing the above-mentioned candidate for the optimal error exponent. Let $A_n \subset \mathcal{X}^n$, $n = 1, 2, \ldots$, be a sequence of sets for which

$$\liminf_{n \to \infty} P_{\theta'}(A_n) > 0. \tag{60}$$

Then it must hold that

$$\liminf_{n \to \infty} \frac{1}{n} \log P_\theta(A_n) \ge -D(P_{\theta'} \| P_\theta). \tag{61}$$

The following lower bound on the probability of underestimation for order estimators of HMSs is an immediate consequence.

Proposition 4: Let k_n be an estimator of the order of an HMS such that for each $\theta \in \Theta_k$, $k \ge 1$

$$\limsup_{n \to \infty} P_\theta(k_n(X^n) > k) < \alpha \tag{62}$$

for some $0 < \alpha < 1$. Then for every $\theta \in \Theta_k$, $k \ge 1$, it holds that

$$\liminf_{n \to \infty} \frac{1}{n} \log P_\theta(k_n(X^n) < k) \ge -D\left(\Theta^{(k-1)} \middle\| P_\theta\right) \tag{63}$$

where

$$D\left(\Theta^{(k-1)} \middle\| P_\theta\right) \triangleq \min_{k'<k} \inf_{\theta' \in \Theta_{k'}} D(P_{\theta'} \| P_\theta). \tag{64}$$

[8]See, however, the comment in the paragraph following Proposition 4 concerning a recent manuscript [11] which addresses a resolution of this problem.

Proof: Fix $\theta \in \Theta_k$, $k \geq 1$, $k' < k$, and choose any $\theta' \in \Theta_{k'}$. From (62), it holds that

$$\limsup_{n\to\infty} P_{\theta'}(k_n(X^n) > k') < \alpha \tag{65}$$

and, therefore,

$$\liminf_{n\to\infty} P_{\theta'}(k_n(X^n) \leq k') > 1-\alpha \tag{66}$$

so that

$$\liminf_{n\to\infty} P_{\theta'}(k_n(X^n) < k) > 1-\alpha. \tag{67}$$

A simple application of (60) and (61) then yields

$$\liminf_{n\to\infty} \frac{1}{n}\log P_\theta(k_n(X^n) < k) \geq -D(P_{\theta'}\|P_\theta). \tag{68}$$

Since this holds for any choice of $k' < k$ and $\theta' \in \Theta_{k'}$, and since the choice only affects the right-hand side of (68), the claim of the proposition follows.[9] □

We had conjectured, in both [13] and the draft version of this correspondence presented for review, that the exponent defined in (64) is the optimal error exponent for order estimation for the general HMS. For the special class of HMSs, considered above, viz. those endowed with a renewal structure, the notion of renewal types is exploited here in analyzing the precise behavior of the "polynomial" terms in the probability of underestimation of our estimator. This enables us to establish its optimality. While several estimators have been proposed for the order estimation for general HMSs, a similar precise analysis is as yet unavailable. However, while preparing the final version of this correspondence for publication, we have just become aware of a recent manuscript [11], where the issue of the achievability of the error exponent of (64) is tackled, using techniques from the theory of large deviations; the optimality of suitable modifications of the estimators in [15] and [6] is addressed, without any assumed prior bound on order.

Binary Independent-Renewal Processes: We conclude by addressing the validity of our results for the order estimation of general binary independent-renewal processes with finite mean renewal times. The special structure (3) of the HMSs considered above, as mentioned earlier, lends itself to an alternative characterization in terms of $\{0, 1\}$-valued independent-renewal processes. This enables the subsequent use of renewal type-counting arguments to bound the number of realizations of a particular renewal type and the total probability of all such realizations under an HMS-derived renewal pmf p_θ. Our arguments, however, do not require the specific HMS nature of these measures; indeed, they hold as well for any binary independent-renewal measure. Consequently, for any countable partition of the set of independent-renewal measures on $\{0, 1\}^\infty$, an estimator constructed as in (39) with the objective of identifying that subset of such measures which contains the generating measure, is indeed well-defined and retains all of its attributes provided the minimum in (54) is strictly positive.

What is not obvious for the class of binary independent-renewal processes, on the other hand, is the choice of a natural, useful, or standard notion of order. The previous partition should be chosen to suit the identification task at hand. The resulting estimator, as well as the optimal error exponent for the probability of underestimation, will then be a function of the partition selected.

[9]Clearly, Proposition 3 is a special case of Proposition 4. However, Proposition 3 has been proved based on renewal type-counting arguments, without recourse to the more sophisticated theory of large deviations.

Of potential practical relevance is a partition as above which divides the set of independent-renewal measures on $\{0, 1\}^\infty$ according to the number of "free parameters" needed to describe p_θ, the underlying pmf of the i.i.d. renewals. Specifically, let $\tilde{\Theta}_k$, $k \geq 1$, contain all the pmfs on the set of positive integers $\mathbb{N}$, which can be parameterized by exactly k parameters.[10] For instance, all pmfs which are geometric (θ) with $\theta \in (0, 1)$, or Poisson (θ) with $\theta > 0$, or uniform $[1, \theta]$ with $\theta \in \mathbb{N}$, are included in $\tilde{\Theta}_1$; all pmfs which are binomial (θ_1, θ_2) with $\theta_1 \in \mathbb{N}$, $\theta_2 \in (0, 1)$, or uniform $[\theta_1, \theta_2]$ with $\theta_1, \theta_2 \in \mathbb{N}$, $\theta_1 < \theta_2$, belong to $\tilde{\Theta}_2$, etc. The resulting interpretation of the order k is compatible with the standard notion of model complexity (cf. e.g., [19]). Note that this formulation yields an increasing sequence of sets of independent-renewal measures $\tilde{\mathcal{P}}_k$, $k \geq 1$, on $\{0, 1\}^\infty$, corresponding, respectively, to the sets $\tilde{\Theta}_k$, $k \geq 1$. Then as in Section II, we can consider the mutually disjoint sequence of measures $\mathcal{P}_k = \tilde{\mathcal{P}}_k \setminus \tilde{\mathcal{P}}_{k-1}$, $k \geq 1$, on $\{0, 1\}^\infty$ (with $\mathcal{P}_1 = \tilde{\mathcal{P}}_1$). It can also be seen in this case that the minimum defined in (54) is strictly positive. This then implies that the optimal error exponent for this problem is strictly positive, and that our (consistent) estimator is optimal in that it achieves the least possible probability of underestimation error up to the first order in the exponent.

APPENDIX I
PROOF OF LEMMA 2

i) Fix $\epsilon \in (0, \frac{1}{4})$. Given a renewal pmf p, set

$$M_n = M(p, \epsilon, n) = \sqrt{\frac{n\epsilon}{\mu_p}}. \tag{69}$$

Since $\mu_p < \infty$ and M_n is increasing in n, there exists $n_0 = n_0(p, \epsilon)$ such that for every $n \geq n_0$

$$\sum_{m=\lfloor M_n\rfloor+1}^{\infty} mp(m) < \epsilon \tag{70}$$

and, therefore,

$$\sum_{m=\lfloor M_n\rfloor+1}^{\infty} p(m) < \frac{\epsilon}{\lfloor M_n\rfloor+1} \leq \frac{\epsilon}{\lceil M_n\rceil}. \tag{71}$$

For every $n \geq n_0$, pick $L_n = L(p, \epsilon, n) = \frac{n}{\mu_p}$ and define $\tilde{K}_n = (\tilde{k}_1, \ldots, \tilde{k}_n) \in \mathcal{K}^{(n)}$ as

$$\tilde{k}_m = \lfloor L_n p(m)\rfloor, \qquad 1 \leq m \leq n. \tag{72}$$

Note that $\sum_{m=1}^{n} m\tilde{k}_m \leq L_n\mu_p = n$. Furthermore

$$\begin{aligned} D\left(q_{\tilde{K}_n}\middle\| p\right) &= \sum_{m=1}^{n} q_{\tilde{K}_n}(m)\log\frac{q_{\tilde{K}_n}(m)}{p(m)} \\ &= \sum_{m=1}^{n} \frac{\tilde{k}_m}{N_{\tilde{K}_n}}\log\frac{\left(\frac{\tilde{k}_m}{N_{\tilde{K}_n}}\right)}{p(m)} \\ &\leq \sum_{m=1}^{n} \frac{\tilde{k}_m}{N_{\tilde{K}_n}}\log\frac{L_n p(m)}{N_{\tilde{K}_n}p(m)} = \log\frac{L_n}{N_{\tilde{K}_n}} \\ &= \log\frac{L_n}{\sum_{m=1}^{n}\tilde{k}_m} \leq \log\frac{L_n}{\sum_{m=1}^{\lfloor M_n\rfloor}\tilde{k}_m} \end{aligned} \tag{73}$$

[10]All nonparametric pmfs with infinite support may be assigned a "true" order of $+\infty$.

since $n \geq \lfloor M_n \rfloor$ by virtue of $\epsilon < \frac{1}{4}$ and $\mu_p \geq 1$. Next, using (72)

$$\begin{aligned} D\left(q_{\tilde{K}_n} \| p\right) &\leq \log \frac{L_n}{\sum_{m=1}^{\lfloor M_n \rfloor} (L_n p(m) - 1)} \\ &= \log \frac{L_n}{\left(L_n \sum_{m=1}^{\lfloor M_n \rfloor} p(m)\right) - \lfloor M_n \rfloor} \\ &\leq \log \frac{L_n}{\left(1 - \frac{\epsilon}{\lceil M_n \rceil}\right) L_n - \lfloor M_n \rfloor} \\ &\leq \log \frac{1}{\left(1 - \frac{\epsilon}{\lceil M_n \rceil}\right) - \frac{M_n}{L_n}} \\ &\leq \log \frac{1}{1 - \frac{2\epsilon}{M_n}} \leq \frac{4\epsilon}{M_n} = \sqrt{\frac{16\epsilon\mu_p}{n}} \end{aligned} \tag{74}$$

provided $n \geq n_1 = n_1(p, \epsilon) = 16\mu_p\epsilon$. Next, observe that for every renewal type $\tilde{K}_n$ defined in (72), it holds that

$$\begin{aligned} \mu_{\tilde{K}_n} &= \sum_{m=1}^{n} m \frac{\tilde{k}_m}{N_{\tilde{K}_n}} \\ &\geq \frac{1}{N_{\tilde{K}_n}} \sum_{m=1}^{\lfloor M_n \rfloor} m(L_n p(m) - 1) \\ &\geq \frac{L_n}{N_{\tilde{K}_n}} \sum_{m=1}^{\lfloor M_n \rfloor} \left(mp(m) - \frac{m}{L_n}\right) \\ &\geq \left(\sum_{m=1}^{\lfloor M_n \rfloor} mp(m)\right) - \frac{\lfloor M_n \rfloor(\lfloor M_n \rfloor + 1)}{2L_n} && (75) \\ &\geq (\mu_p - \epsilon) - \frac{M_n(M_n+1)}{2L_n} \geq \mu_p \left(1 - \frac{2\epsilon}{\mu_p}\right) && (76) \end{aligned}$$

for all $n \geq n_2 = n_2(p, \epsilon) = \frac{6\mu_p}{\epsilon}$, where we have used the fact that

$$N_{\tilde{K}_n} = \sum_{m=1}^{n} \tilde{k}_m \leq L_n$$

by (72). Therefore, by (74) and (76), it holds for all $n \geq \max\{n_0, n_1, n_2\}$ that

$$\frac{D\left(q_{\tilde{K}_n} \| p\right)}{\mu_{\tilde{K}_n}} \leq \frac{1}{\mu_p}\left(1 + \frac{4\epsilon}{\mu_p}\right)\frac{4\epsilon}{M_n} \leq \frac{8}{\mu_p}\frac{\epsilon}{M_n} = \sqrt{\frac{64\epsilon}{n\mu_p}} \tag{77}$$

for $0 < \epsilon < \frac{1}{4}$. The first claim of the lemma follows. □

ii) This follows from (74). □

APPENDIX II
PROOF OF LEMMA 3

i) This is an immediate consequence of the continuity of $D(p' \| p)$ in p' on the set $\{p' \colon p' \ll p\}$ and of the continuity of $\mu_{p'} > 1$ in p'. □

ii) By standard calculations, the minimum in the definition of $D(\Lambda(\mu^*) \| p)$ is achieved by $p'_\lambda \in \Lambda(\mu^*)$ given by

$$p'_\lambda(m) = \frac{p(m)\exp\{m\lambda\}}{\sum_{m'=1}^{\infty} p(m')\exp\{m'\lambda\}}, \qquad m \geq 1 \tag{78}$$

where λ is chosen to satisfy the constraint

$$\mu_{p'_\lambda} \triangleq \sum_{m=1}^{\infty} m \frac{p(m)\exp\{m\lambda\}}{\sum_{m'=1}^{\infty} p(m')\exp\{m'\lambda\}} = \mu^*. \tag{79}$$

Note that $\mu_{p'_\lambda} = \mu_p$ at $\lambda = 0$. Therefore, by the assumed existence of the moment generating function of p in a neighborhood of zero, (79) has a solution for every $\mu^* \in [\mu_p, \infty)$. Furthermore

$$\frac{d}{d\lambda}\mu_{p'_\lambda} = \sum_{m=1}^{\infty} m^2 p'_\lambda(m) - \left(\mu_{p'_\lambda}\right)^2. \tag{80}$$

It can also be shown, after some algebraic manipulations, that

$$\begin{aligned} &\frac{d}{d\lambda}\frac{D(p'_\lambda \| p)}{\mu_{p'_\lambda}} \\ &\quad = \frac{d}{d\lambda}\left[\lambda - \frac{1}{\mu_{p'_\lambda}} \log \sum_{m=1}^{\infty} p(m)\exp\{m\lambda\}\right] \\ &\quad = 1 + \frac{1}{\left(\mu_{p'_\lambda}\right)^2}\left(\frac{d}{d\lambda}\mu_{p'_\lambda}\right)\log \sum_{m=1}^{\infty} p(m)\exp\{m\lambda\} \\ &\qquad - \frac{1}{\mu_{p'_\lambda}} \frac{1}{\sum_{m=1}^{\infty} p(m)\exp\{m\lambda\}} \frac{d}{d\lambda} \sum_{m=1}^{\infty} p(m)\exp\{m\lambda\} \\ &\quad = \frac{\sum_{m=1}^{\infty} m^2 p'_\lambda(m) - \left(\mu_{p'_\lambda}\right)^2}{\left(\mu_{p'_\lambda}\right)^2} \log \sum_{m=1}^{\infty} p(m)\exp\{m\lambda\}. \end{aligned} \tag{81}$$

From (80) and (81), we get

$$\frac{d}{d\mu^*}\frac{D(p'_\lambda \| p)}{\mu_{p'_\lambda}} = \frac{\log \sum_{m=1}^{\infty} p(m)\exp\{m\lambda\}}{\mu^{*2}} \tag{82}$$

where λ satisfies (79). It is also clear that for $\mu^* \in (\mu_p, \infty)$, we have from (79) that $\lambda > 0$, and, as a consequence, $\frac{d}{d\mu^*}\frac{D(p'_\lambda \| p)}{\mu_{p'_\lambda}} > 0$. The claims of the lemma follow. □

APPENDIX III
PROOF OF PROPOSITION 3

We first prove the following technical lemma.

Lemma 5: Let k_n be an estimator which, for some $0 < \alpha < 1$, and for all $\theta \in \Theta_k$, $k \geq 1$, satisfies

$$\limsup_{n \to \infty} P_\theta(k_n(X^n) > k) < \alpha. \tag{83}$$

Then, for fixed k, $k' < k$, and each $\tilde{\mu} > 1$, there exist $\eta = \eta(\alpha, k')$ and $N = N(\alpha, k', \tilde{\mu})$ such that for any $\theta' \in \Theta_{k'}$ satisfying $\mu_{\theta'} < \tilde{\mu}$, there exists a renewal type $K = K(\theta')$ which satisfies

$$\mu_K < \rho_n \tag{84}$$

$$\frac{D(q_K \| p_\theta)}{\mu_K} \leq \epsilon_n \tag{85}$$

and

$$\left| T_K \bigcap \{x^n \colon k_n(x^n) < k\} \right| \geq \eta |T_K| \tag{86}$$

for all $n > N$.

Proof: Assume to the contrary that for every $\eta > 0$ (however small) and every $N < \infty$ (however large), there exist $\theta' = \theta'(\eta) \in \Theta_{k'}$ with $\mu_{\theta'} < \tilde{\mu}$ and $n > N$ such that every renewal type $K \in \mathcal{K}^{(n)}$ which satisfies

$$\mu_K < \rho_n \tag{87}$$

and

$$\frac{D(q_k \| p_\theta)}{\mu_K} \le \epsilon_n \tag{88}$$

also satisfies

$$\left| T_K \bigcap \{x^n : k_n(x^n) < k\} \right| < \eta |T_K|. \tag{89}$$

Then, infinitely often in n, it must hold that

$$\begin{aligned} P_{\theta'} & \left(\{x^n : k_n(x^n) < k\} \bigcap \bigcup_{K \in \mathcal{B}^{(n)}(\theta')} T_K \right) \\ &= \sum_{K \in \mathcal{B}^{(n)}(\theta')} P_{\theta'} \left(T_K \bigcap \{x^n : k_n(x^n) < k\} \right) \\ &< \eta \sum_{K \in \mathcal{B}^{(n)}(\theta')} P_{\theta'}(T_K) \le \eta. \end{aligned} \tag{90}$$

Next

$$\begin{aligned} & P_{\theta'}(k_n(X^n) > k') \\ &\quad \ge P_{\theta'}(k_n(X^n) \ge k) = 1 - P_{\theta'}(k_n(X^n) < k) \\ &\quad = 1 - P_{\theta'} \left(\{x^n : k_n(x^n) < k\} \bigcap \bigcup_{K \in \mathcal{B}^{(n)}(\theta')} T_K \right) \\ &\qquad - P_{\theta'} \left(\{x^n : k_n(x^n) < k\} \bigcap \bigcup_{K \in \mathcal{A}^{(n)}(\theta')} T_K \right) \\ &\qquad - P_{\theta'} \left(\{x^n : k_n(x^n) < k\} \bigcap \bigcup_{K \in \mathcal{C}^{(n)}(\theta')} T_K \right) \\ &\quad \ge 1 - \eta - P_{\theta'} \left(\bigcup_{K \in \mathcal{A}^{(n)}(\theta')} T_K \right) - P_{\theta'} \left(\bigcup_{K \in \mathcal{C}^{(n)}(\theta')} T_K \right) \\ &\quad \ge 1 - \eta - \exp \left[-n \min \left\{ \epsilon_n - \frac{c}{\sqrt{n}},\, D(\Lambda(\rho_n) \| p_{\theta'}) \right\} \right] \\ &\quad \ge 1 - \eta - \exp \left[-n \min \left\{ \epsilon_n - \frac{c}{\sqrt{n}},\, D(\Lambda(2\tilde{\mu}) \| p_{\theta'}) \right\} \right] \\ &\quad \ge 1 - 2\eta \end{aligned} \tag{91}$$

for all $n > N(\eta, \tilde{\mu})$ for which (87)–(89) also hold, where we have used steps similar to the derivation in (41) and (42). Note that $N(\eta, \tilde{\mu})$ does not depend on θ', because $\mu_{\theta'} < \tilde{\mu}$ by assumption. Finally, since the choice of η is also arbitrary, we choose $\eta = \frac{1-\alpha}{2}$, which leads to

$$\limsup_{n \to \infty} P_{\theta'}(k_n(X^n) > k') \ge \alpha \tag{92}$$

contradicting the bound on the probability of overestimation. □

Next, turning to the proof of Proposition 3, for a given $\theta \in \Theta_k$ and $\epsilon > 0$, let $\theta'_* \in \bigcup_{k' < k} \Theta_{k'}$, be such that

$$\frac{D\left(p_{\theta'_*} \| p_\theta\right)}{\mu_{\theta'_*}} \le D\left(\Theta^{(k-1)} \middle\| P_\theta\right) + \frac{\epsilon}{3} \tag{93}$$

and set $\tilde{\mu} = 2\mu_{\theta'_*}$. For each $k' < k$, and each $\theta' \in \Theta_{k'}$, with $\mu_{\theta'} < \tilde{\mu}$, there is a special renewal type $K^n(\theta')$ described in Lemma 5 which has a significant intersection with the region of underestimation for k. Specifically, for some $\eta = \eta(\alpha, k')$ and for all $n > N(\alpha, k', \tilde{\mu})$

$$\mu_{K^n(\theta')} < \rho_n \tag{94}$$

$$\frac{D(q_k \| p_\theta)}{\mu_{K^n(\theta')}} \le \epsilon_n \tag{95}$$

and

$$\left| T_{K^n(\theta')} \bigcap \{x^n : k_n(x^n) < k\} \right| \ge \eta \left| T_{K^n(\theta')} \right|. \tag{96}$$

Applying the lower bound (27) on the probability of the renewal type from Lemma 1, we get

$$\begin{aligned} & P_\theta(k_n(X^n) < k) \\ &\qquad \ge \eta (n+2)^{-(\sqrt{2n}+1)} \exp \left\{ -(n+1) \frac{D\left(q_{\mathring{K}(\theta')} \middle\| p_\theta\right)}{\mu_{\mathring{K}(\theta')}} \right\} \end{aligned} \tag{97}$$

where $\mathring{K}(\theta')$ is the circular version of $K^n(\theta')$. Since this holds for all $k' < k$, and for all $\theta' \in \Theta_{k'}$ with $\mu_{\theta'} < \tilde{\mu}$, the equation at the bottom of the page must hold for all $n > \max_{k'<k} N(\alpha, k', \tilde{\mu})$. Thus, there exists $N = N(\alpha, k, \tilde{\mu}, \epsilon) \ge \max_{k'<k} N(\alpha, k', \tilde{\mu})$, such that for all $n > N$

$$\begin{aligned} & P_\theta(k_n(X^n) < k) \\ &\qquad \ge \exp \left\{ -n \left[\min_{k'<k} \inf_{\theta' \in \Theta_{k'}} \frac{D\left(q_{\mathring{K}(\theta')} \middle\| p_\theta\right)}{\mu_{\mathring{K}(\theta')}} + \frac{\epsilon}{3} \right] \right\}. \end{aligned} \tag{98}$$

$$\begin{aligned} P_\theta(k_n(X^n) < k) &\ge (n+2)^{-(\sqrt{2n}+1)} \max_{k'<k} \left[\eta \sup_{\theta' \in \Theta_{k'} :\, \mu_{\theta'} < \tilde{\mu}} \exp \left\{ -(n+1) \frac{D\left(q_{\mathring{K}(\theta')} \middle\| p_\theta\right)}{\mu_{\mathring{K}(\theta')}} \right\} \right] \\ &\ge (n+2)^{-(\sqrt{2n}+1)} \max_{k'<k} \left[\eta \exp \left\{ -(n+1) \inf_{\theta' \in \Theta_{k'} :\, \mu_{\theta'} < \tilde{\mu}} \frac{D\left(q_{\mathring{K}(\theta')} \middle\| p_\theta\right)}{\mu_{\mathring{K}(\theta')}} \right\} \right] \\ &\ge (n+2)^{-(\sqrt{2n}+1)} \eta \exp \left\{ -(n+1) \min_{k'<k} \left[\inf_{\theta' \in \Theta_{k'} :\, \mu_{\theta'} < \tilde{\mu}} \frac{D\left(q_{\mathring{K}(\theta')} \middle\| p_\theta\right)}{\mu_{\mathring{K}(\theta')}} \right] \right\} \\ &\ge (n+2)^{-(\sqrt{2n}+1)} \exp \left\{ -(n+1) \left(\min_{k'<k} \left[\inf_{\theta' \in \Theta_{k'} :\, \mu_{\theta'} < \tilde{\mu}} \frac{D\left(q_{\mathring{K}(\theta')} \middle\| p_\theta\right)}{\mu_{\mathring{K}(\theta')}} \right] - \frac{\log \eta}{n+1} \right) \right\}. \end{aligned}$$

Now, for $n > N$, the special type of Lemma 5 for θ'_*, denoted $K^n(\theta'_*)$, is close in variational distance to $p_{\theta'_*}$ in the sense that

$$\left\|q_{K^n(\theta'_*)} - p_{\theta'_*}\right\| \le \sqrt{2\ln 2\rho_n\epsilon_n}. \tag{99}$$

Next, the circular version $\overset{\circ}{K}(\theta'_*)$ of $K^n(\theta'_*)$ is also close to $p_{\theta'_*}$ in variational distance, because

$$\left\|q_{\overset{\circ}{K}(\theta'_*)} - q_{K^n(\theta'_*)}\right\| \le \frac{2\mu_{K^n(\theta'_*)}}{n+\mu_{K^n(\theta'_*)}} \le \frac{2\rho_n}{n} \le \frac{2\rho_n\epsilon_n}{c\sqrt{n}}. \tag{100}$$

Therefore, by continuity

$$\frac{D\left(q_{\overset{\circ}{K}(\theta')}\middle\| p_\theta\right)}{\mu_{\overset{\circ}{K}(\theta')}} \le \frac{D\left(p_{\theta'_*}\middle\| p_\theta\right)}{\mu_{\theta'_*}} + \frac{\epsilon}{3} \tag{101}$$

for all sufficiently large n, so that

$$\min_{k'<k}\inf_{\theta'\in\Theta_{k'}} \frac{D\left(q_{\overset{\circ}{K}(\theta')}\middle\| p_\theta\right)}{\mu_{\overset{\circ}{K}(\theta')}} + \frac{\epsilon}{3} \le \frac{D\left(p_{\theta'_*}\middle\| p_\theta\right)}{\mu_{\theta'_*}} + \frac{2\epsilon}{3} \le D\left(\Theta^{(k-1)}\middle\| P_\theta\right) + \epsilon \tag{102}$$

and the result of the proposition follows. □

ACKNOWLEDGMENT

The authors gratefully acknowledge that this correspondence has benefited significantly from an excruciatingly thorough review by Lorenzo Finesso who also provided numerous constructive suggestions. The authors also thank the anonymous referees for their comments.

REFERENCES

[1] R. R. Bahadur, S. L. Zabell, and J. C. Gupta, "Large deviations, tests and estimates," in *Asymptotic Theory of Statistical Tests and Estimation*, I. M. Chakravarty, Ed. Boston, MA: Academic, 1980.

[2] A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2743–2760, Oct. 1998.

[3] P. Bühlman and A. Wyner, "Variable length Markov chains," *Ann. Statisti.*, vol. 27, pp. 480–513, 1999.

[4] E. Çinlar, *Introduction to Stochastic Processes*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[5] I. Csiszár and J. Körner, *Information Theory: Coding Theory for Discrete Memoryless Systems*. New York, NY: Academic, 1981.

[6] I. Csiszár and P. C. Shields, "Consistency of the BIC Markov order estimator," in *Proc. IEEE Int. Symp. Information Theory*, Sorrento, Italy, 2000.

[7] ——, "Redundancy rates for renewal and other processes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 2065–2072, Nov. 1996.

[8] A. Dembo and O. Zeitouni, *Large Deviation Techniques and Applications*. Boston, MA: Jones and Bartlett, 1993.

[9] L. Finesso, "Estimation of the order of a finite Markov chain," in *Recent Advances in the Mathematical Theory of Systems, Control, Networks, Signals*, H. Kimura and S. Kodama, Eds. Tokyo, Japan: Mita Press, 1992, pp. 643–645.

[10] L. Finesso, C.-C. Liu, and P. Narayan, "The optimal error exponent for Markov order estimation," *IEEE Trans. Inform. Theory*, vol. 42, pp. 1488–1497, Sept. 1996.

[11] E. Gassiat and S. Boucheron. Optimal error exponents in hidden Markov models order estimation. [Online]. Available: http://www.math.u-psud.fr/~gassiat/liste-publi.html

[12] G. H. Hardy and S. Ramanujan, "Asymptotic formulae in combinatorial analysis," in *Proc. London Math. Soc.*, ser. 2, 1918, vol. XVII, pp. 75–115.

[13] S. Khudanpur, "Model selection and universal data compression," PH.D. dissertation, Dept. Elec. Comput. Eng., Univ. Maryland, College Park, Jan. 1997.

[14] B. G. Leroux, "Maximum likelihood estimation for hidden Markov models," *Stochastic Processes Their Applic.*, vol. 40, no. 1, Feb. 1992.

[15] C.-C. Liu and P. Narayan, "Order estimation and sequential universal data compression of a hidden Markov source by the method of mixtures," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1167–1180, July 1994.

[16] N. Merhav, "The estimation of model order in exponential families," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1109–1113, Sept. 1989.

[17] N. Merhav, M. Gutman, and J. Ziv, "On the estimation of the order of a Markov chain and universal data compression," *IEEE Trans. Inform. Theory*, vol. 35, pp. 1014–1019, Sept. 1989.

[18] H. Michiel and K. Laevens, "Teletraffic engineering in a broadband era," *Proc. IEEE*, vol. 85, pp. 2007–2033, 1997.

[19] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*. Singapore: World Scientific, 1989.

[20] S. M. Ross, *Stochastic Processes*. New York: Wiley, 1983.

[21] J. Ziv and N. Merhav, "Estimating the number of states of a finite state source," *IEEE Trans. Inform. Theory*, vol. 38, pp. 61–65, Jan. 1992.

Universal Codes for Finite Sequences of Integers Drawn From a Monotone Distribution

Dean P. Foster, Robert A. Stine, *Member, IEEE*, and Abraham J. Wyner, *Member, IEEE*

Invited

***Abstract*—We offer two noiseless codes for blocks of integers $X^n = (X_1, \ldots, X_n)$. We provide explicit bounds on the relative redundancy that are valid for any distribution F in the class of memoryless sources with a possibly infinite alphabet whose marginal distribution is monotone. Specifically, we show that the expected code length $L(X^n)$ of our first universal code is dominated by a linear function of the entropy of X^n. Further, we present a second universal code that is efficient in that its length is bounded by $nH_F + o(nH_F)$, where H_F is the entropy of F which is allowed to vary with n. Since these bounds hold for any n and any monotone F we are able to show that our codes are strongly minimax with respect to relative redundancy (as defined by Elias). Our proofs make use of the elegant inequality due to Aaron Wyner.**

***Index Terms*—Elias codes, relative redundancy, strongly minimax, universal noiseless coding of integers, Wyner's inequality.**

I. INTRODUCTION

Consider the problem of lossless compression of a finite collection of n positive integers $X^n = (X_1, \ldots, X_n)$ into a prefix code of shortest expected length. The $X_i \ge 1$ are independent, integer-valued random

Manuscript received March 20, 2000; revised October 17, 2001. The material in this correspondence was presented in part at the IEEE International Symposium on Information Theory, Sorrento, Italy, June 2000.

The authors are with the Department of Statistics, The Wharton School of the University of Pennsylvania, Philadelphia, PA 19104-6302 USA (e-mail: foster@wharton.upenn.edu; stine@wharton.upenn.edu; ajw@wharton.upenn.edu).

Publisher Item Identifier S 0018-9448(02)04033-6.