Journal of Statistical Physics, Vol. 121, Nos. 3/4, November 2005 (© 2005)
DOI: 10.1007/s10955-005-7576-y

The Entropy of a Binary Hidden Markov Process

Or Zuk,[1] Ido Kanter,[2] and Eytan Domany[1]

Received January 31, 2005; accepted June 10, 2005

The entropy of a binary symmetric Hidden Markov Process is calculated as an expansion in the noise parameter ϵ. We map the problem onto a one-dimensional Ising model in a large field of random signs and calculate the expansion coefficients up to second order in ϵ. Using a conjecture we extend the calculation to 11th order and discuss the convergence of the resulting series.

KEY WORDS: Hidden Markov Process; entropy; random-field Ising model.

1. INTRODUCTION

Hidden Markov Processes (HMPs) have many applications, in a wide range of disciplines—from the theory of communication[(1)] to analysis of gene expression.[(2)] Comprehensive reviews on both theory and applications of HMPs can be found in refs.1 and 3. Recent applications to experimental physics are in refs. 4 and 5. The most widely used context of HMPs is, however, that of construction of reliable and efficient communication channels.

In a practical communication channel the aim is to reliably transmit source message over a noisy channel. Figure 1 shows a schematic representation of such a communication. The source message can be a stream of words taken from a text. It is clear that such a stream of words contains information, indicating that words and letters are not chosen randomly. Rather, the probability that a particular word (or letter) appears at a given point in the stream depends on the words (letters) that were previously transmitted. Such dependency of a transmitted symbol on the precedent stream is modeled by a Markov process.

[1]Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel; e-mail: fedomany@wisemail.weizmann.ac.il
[2]Department of Physics, Bar Ilan University, Ramat Gan, 52900, Israel.

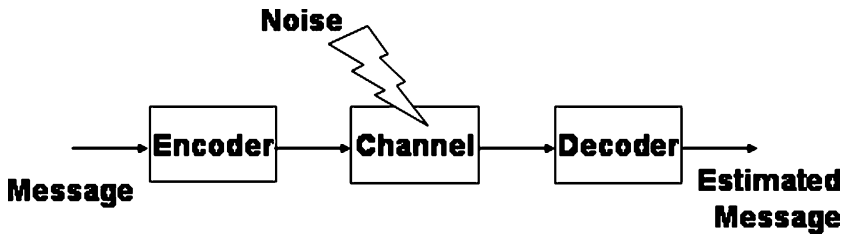

Fig. 1. Schematic drawing of message transmission through a noisy channel.

The Markov model is a finite state machine that changes state once every time unit. The manner in which the state transitions occur is probabilistic and is governed by a state-transition matrix, P, that generates the new state of the system. The Markovian assumption indicates that the state at any given time depends only on the state at the previous time step. When dealing with text, a state usually represents either a letter, a word or a finite sequence of words, and the state-transition matrix represents the probability that a given state is followed by another state. Estimating the state-transition matrix is in the realm of linguistics; it is done by measuring the probability of occurrence of pairs of successive letters in a large corpus.

One should bear in mind that the Markovian assumption is very restrictive and very few physical systems can expect to satisfy it in a strict manner. Clearly, a Markov process imitates some statistical properties of a given language, but can generate a chain of letters that is grammatically erroneous and lack logical meaning. Even though the Markovian description represents only some limited subset of the correlations that govern a complex process, it is the simplest natural starting point for analysis. Thus one assumes that the original message, represented by a sequence of N binary bits, has been generated by some Markov process. In the simplest scenario, of a binary symmetric Markov process, the underlying Markov model is characterized by a single parameter—the flipping rate p, denoting the probability that a 0 is followed by 1 (the same as a 1 followed by a 0). The stream of N bits is transmitted through a noisy communication channel. The received string differs from the transmitted one due to the noise. The simplest way to model the noise is known as the *Binary Symmetric Channel* (BSC), where each bit of the original message is flipped with probability ϵ. Since the observer sees only the received, noise-corrupted version of the message, and neither the original message nor the value of p that generated it are known to him, what he records is the outcome of a *Hidden Markov Process*. Thus, HMPs are double embedded stochastic

processes; the first is the Markov process that generated the original message and the second, which does not influence the Markov process, is the noise added to the Markov chain after it has been generated.

Efficient information transmission plays a central role in modern society, and takes a variety of forms, from telephone and satellite communication to storing and retrieving information on disk drives. Two central aspects of this technology are error correction and compression. For both problem areas it is of central importance to estimate Ω_R, the number of (expected) received signals.

In the noise free case this equals the expected number of transmitted signals Ω_S; when the Markov process has flipping rate $p = 0$, only two strings (all 1 or all 0) will be generated and $\Omega_S = 2$, while when the flip rate is $p = 1/2$ each string is equally likely and $\Omega_S = 2^N$.

In general, Ω_R is given, for large N, by 2^{NH}, where $H = H(p, \epsilon)$ is the *entropy* of the process. The importance of knowing Ω_R for compression is evident: one can number the possible messages $i = 1, 2, \ldots, \Omega_R$, and if $\Omega_R < 2^N$, by transmitting only the index of the message (which can be represented by $\log_2 \Omega_R < N$ bits) we compress the information. Note that we can get further compression using the fact that the Ω_R messages do not have equal probabilities.

Error correcting codes are commonly used in methods of information transmission to compensate for noise corruption of the data during transmission. These methods require the use of additional transmitted information, i.e., redundancy, together with the data itself. That is, one transmits a string of $M > N$ bits; the percentage of additional transmitted bits required to recover the source message determines the coding efficiency, or channel capacity, a concept introduced and formulated by Shannon. The channel capacity for the BSC and for a random i.i.d. source was explicitly derived by Shannon in his seminal paper of 1948.[6] The calculation of channel capacity for a Markovian source transmitted over a noisy channel is still an open question.

Hence, calculating the entropy of a HMP is an important ingredient of progress towards deriving improved estimates of both compression and channel capacity, of both theoretical and practical importance for modern communication. In this paper we calculate the entropy of a HMP as a power series in the noise parameter ϵ.

In Section 2 we map the problem onto that of a one-dimensional nearest neighbor Ising model in a field of fixed magnitude and random signs (see ref. 7 for a review on the Random Field Ising Model). Expansion in ϵ corresponds to working near the infinite field limit.

Note that the object we are calculating is *not* the entropy of an Ising chain in a quenched random field, as shown in Eq. (17) and in the

discussion following it. In technical terms, here we set the replica index to $n = 1$ after the calculation, whereas to obtain the (quenched average) properties of an Ising chain one works in the $n \to 0$ limit.

In Section 3 we present exact results for the expansion coefficients of the entropy up to second order. While the zeroth and first order terms were previously known,[8,9] the second order term was not.[10] In Section 4 we introduce bounds on the entropy that were derived by Cover and Thomas;[11] we have strong evidence that these bounds actually provide the exact expansion coefficients. Since we have not proved this statement, it is presented as a conjecture; on it's basis the expansion coefficients up to 11th order are derived and listed. We conclude in Section 5 by studying the radius of convergence of the low-noise expansion, and summarize our results in Section 6.

2. A HIDDEN MARKOV PROCESS AND THE RANDOM-FIELD ISING MODEL

2.1. Defining the Process and its Entropy

Consider the case of a binary signal generated by the source. Binary valued symbols, $s_i = \pm 1$ are generated and transmitted at fixed times $i\Delta t$. Denote a sequence of N transmitted symbols by

$$S = (s_1, s_2, \ldots, s_N) \tag{1}$$

The sequence is generated by a Markov process; here we assume that the value of s_{i+1} depends only on s_i (and not on the symbols generated at previous times). The process is parametrized by a transition matrix P, whose elements are the transition probabilities

$$P_{+,-} = Pr(s_{i+1} = +1 | s_i = -1) \quad P_{-,+} = Pr(s_{i+1} = -1 | s_i = +1) \tag{2}$$

Here we treat the case of a *symmetric* process, i.e., $P_{+,-} = P_{-,+} = p$, so that we have

$$s_{i+1} = \begin{cases} s_i & \text{prob.} = 1 - p \\ -s_i & \text{prob.} = p \end{cases} \tag{3}$$

The first symbol s_1 takes the values ± 1 with equal probabilities, $Pr(s_1 = +1) = Pr(s_1 = -1) = 1/2$. The probability of realizing a particular

sequence S is given by

$$Pr(S) = \frac{1}{2}\prod_{i=2}^{N} Pr(s_i|s_{i-1}) \tag{4}$$

The generated sequence S is "*sent*" and passes through a noisy channel; hence the *received* sequence,

$$R = (r_1, r_2, \ldots, r_N) \tag{5}$$

is not identical to the transmitted one. The noise can flip a transmitted symbol with probability ϵ:

$$Pr(r_i = -s_i|s_i) = \epsilon \tag{6}$$

Here we assumed that the noise is generated by an independent identically distributed (i.i.d.) process; the probability of a flip at time i is independent of what happened at other times $j < i$ and of the value of i. We also assume that the noise is symmetric, i.e., the flip probability does not depend on s_i.

Once the underlying Markov process S has been generated, the probability of observing a particular sequence R is given by

$$Pr(R|S) = \prod_{i=1}^{N} Pr(r_i|s_i) \tag{7}$$

and the joint probability of any particular S and R to occur is given by

$$Pr(R, S) = Pr(R|S)Pr(S) \tag{8}$$

The original transmitted signal, S, is "hidden" and only the received (and typically corrupted) signal R is "seen" by the observer. Hence it is meaningful to ask—what is the probability to observe any particular received signal R? The answer is

$$Q(R) = \sum_{S} Pr(R, S) \tag{9}$$

Furthermore, one is interested in the Shannon entropy H of the observed process,[3]

$$H_N = -\sum_R Q(R) \log Q(R) \tag{10}$$

and in particular, in the *entropy rate*, defined as

$$H = \lim_{N\to\infty} \frac{H_N}{N} \tag{11}$$

2.2. Casting the Problem in Ising Form

It is straightforward to cast the calculation of the entropy rate onto the form of a one-dimensional Ising model. The conditional Markov probabilities (3), that connect the symbols from one site to the next, can be rewritten as

$$Pr(s_{i+1}|s_i) = e^{J s_{i+1} s_i}/(e^J + e^{-J}) \qquad \text{with} \qquad e^{2J} = (1-p)/p \tag{12}$$

and similarly, the flip probability generated by the noise, Eq. (6), is also recapitulated by the Ising form

$$Pr(r_i|s_i) = e^{K r_i s_i}/(e^K + e^{-K}) \qquad \text{with} \qquad e^{2K} = (1-\epsilon)/\epsilon \tag{13}$$

The joint probability of realizing a pair of transmitted and observed sequences (S, R) takes the form[12,13]

$$Pr(R, S) = A \exp\left(J \sum_{i=1}^{N-1} s_{i+1} s_i + K \sum_{i=1}^{N} r_i s_i \right) \tag{14}$$

where the constant A is the product of two factors, $A = A_0 A_1$, given by

$$A_0 = \tfrac{1}{2}\left(e^J + e^{-J}\right)^{-(N-1)} \qquad A_1 = \left(e^K + e^{-K}\right)^{-N} \tag{15}$$

The first sum in Eq. (14) is the Hamiltonian of a chain of Ising spins with open boundary conditions and nearest neighbor interactions J; the

[3]The Shannon entropy is defined using $\log_2$; we use natural log for simplicity.

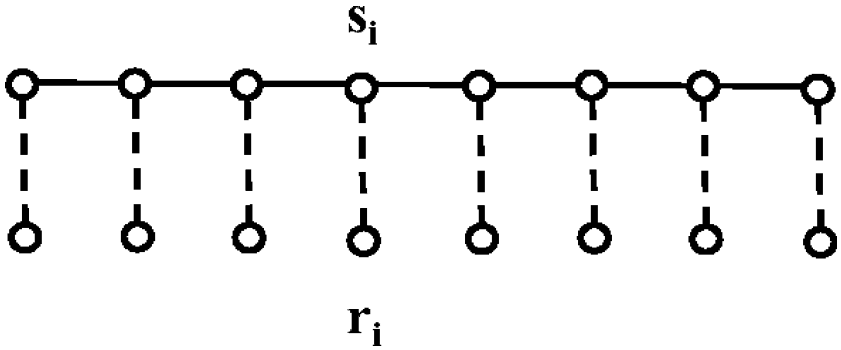

Fig. 2. An Ising model in a random field. The solid lines represent interactions of strength **J** between neighboring spins $\mathbf{S_i}, \mathbf{S_{i+1}}$ while the dashed lines represent local fields $\mathbf{Kr_i}$ acting on the spin $\mathbf{S_i}$.

interactions are ferromagnetic ($J > 0$) for $p < 1/2$. The second term corresponds, for small noise $\epsilon < 1/2$, to a strong ferromagnetic interaction K between each spin s_i and another spin, r_i, connected to s_i by a "dangling bond" (see Fig. 2).

Denote the summation over the hidden variables by $Z(R)$:

$$Z(R) = \sum_S \exp\left(J \sum_{i=1}^{N-1} s_{i+1} s_i + K \sum_{i=1}^{N} r_i s_i \right) \tag{16}$$

so that the probability $Q(R)$ becomes (see Eq. (9)) $Q(R) = A\ Z(R)$. Substituting in Eq. (10), the entropy of the process can be written as

$$H_N = -\sum_R A\ Z(R) \log[AZ(R)] = -\left[\frac{d}{dn} \sum_R [A\ Z(R)]^n \right]_{n=1} \tag{17}$$

The interpretation of this expression is obvious: an Ising chain is submitted to local fields $h_i = Kr_i$, with the sign of the field at each site being $\pm$ with equal probabilities, and we have to average $Z(h_1, \ldots, h_N)^n$ over the field configurations. This is precisely the problem one faces in order to calculate properties of a well-studied model, of a nearest neighbor Ising chain in a quenched random field of uniform strength and random signs at the different sites (there one is interested, however, in the limit $n \to 0$). This problem has not been solved analytically, albeit a few exactly solvable simplified versions of the model do exist,[14–17] as well as expansions (albeit in the weak field limit).[18]

One should note that here we calculate the entropy associated with the observed variables R . In the Ising language this corresponds to an entropy associated with the randomly assigned signs of the local fields, and *not* to the entropy of the spins S. Because of this distinction the entropy H_N has no obvious physical interpretation or relevance, which explains why the problem has not been addressed yet by the physics community. The two entropies are, however, related, and our results can be used to calculate the entropy of the spins.

We are interested in calculating the entropy rate *in the limit of small noise*, i.e., $\epsilon \ll 1$. In the Ising representation this limit corresponds to $K \gg 1$ and hence an expansion in ϵ corresponds to expanding near the infinite field limit of the Ising chain.

3. EXPANSION TO ORDER ϵ^2: EXACT RESULTS

We are interested in calculating the entropy rate

$$H = -\lim_{N\to\infty}\left[\frac{1}{N}\sum_{R} A\ Z(R)\log A\ Z(R)\right] \tag{18}$$

to a given order in ϵ. A few technical points are in order. First, we will actually use

$$e^{-2K} = \epsilon/(1-\epsilon) \tag{19}$$

as our small parameter and expand to order ϵ^2 afterwards. Second, we will calculate H_N and take the large N limit. Therefore we can replace the open boundary conditions with periodic ones (setting $s_{N+1} = s_1$)—the difference is a surface effect of order $1/N$. The constant A_0 becomes

$$A_0 = \left(e^J + e^{-J}\right)^{-N} \tag{20}$$

and the interaction term Js_1s_N is added to the first sum in Eq. (14), which contains now N pairs of neighbors.

Expanding $Z(R)$. Consider $Z(R)$ from Eq. (16). For any fixed $R = (r_1, r_2, \ldots r_N)$ the leading order is obtained by the S configuration with $s_i = r_i$ for all i. For this configuration each site contributes K to the "field term" in Eq. (16). The contribution of this configuration to the summation over S in Eq. (16) is

$$Z(R)^{(0)} = e^{NK}\exp\left(J\sum_{i=1}^{N} r_{i+1}r_i\right) \tag{21}$$

The next term we add consists of the contributions of those S configurations which have $s_i = r_i$ at *all but one position*. The field term of such a configuration is K from $N-1$ sites and $-K$ from the single site with $s_j = -r_j$. There are N such configurations, and the total contribution of these terms to the sum (16) is

$$Z(R)^{(1)} = e^{NK} e^{-2K} \exp\left(J \sum_{i=1}^{N} r_{i+1} r_i \right) \sum_{j=1}^{N} \exp[-2J r_j (r_{j-1} + r_{j+1})] \quad (22)$$

The next term is of the highest order studied in this paper; it involves configurations S with all but two spins in the state $s_i = r_i$; the other two take the values $s_j = -r_j$, $s_k = -r_k$, i.e., are flipped with respect to the corresponding local fields. These S configurations belong to one of two classes. In class a the two flipped spins are located on nearest neighbor sites, e.g., $k = j+1$; there are N such configurations. To the second class, b, belong those configurations in which the two flipped spins are *not* neighbors—there are $N(N-3)/2$ such terms in the sum (16), and the respective contributions are[4]

$$Z(R)^{(2a)} = e^{NK} e^{-4K} \exp\left(J \sum_{i=1}^{N} r_{i+1} r_i \right) \times \sum_{j=1}^{N} \exp[-2J(r_j r_{j-1} + r_{j+1} r_{j+2})] \quad (23)$$

$$Z(R)^{(2b)} = e^{NK} e^{-4K} \exp\left(J \sum_{i=1}^{N} r_{i+1} r_i \right) \times \frac{1}{2} \sum_{j=1}^{N} \sum_{k \neq j, j \pm 1} \exp[-2J r_j (r_{j-1} + r_{j+1}) - 2J r_k (r_{k-1} + r_{k+1})] \quad (24)$$

Calculation of H is now straightforward, albeit tedious: substitute AZ into Eq. (18), expand everything in powers of ϵ, to second order, and for each term perform the summation over all the r_i variables. These summations

[4]We use the obvious identifications imposed by periodic boundary conditions, e.g., $r_{N+1} = r_1$, $r_{N+2} = r_2$.

involve two kinds of terms. The first is of the "partition-sum-like" form

$$\sum_R e^{\mathcal{H}(R)} \quad \text{where} \quad \mathcal{H}(R) = \sum_j \Delta_j J r_j r_{j+1} \quad \text{with} \quad \Delta_j = \pm 1 \tag{25}$$

For the case studied here we encounter either all bonds $\Delta_j J > 0$, or two have a flipped sign (corresponding to Eqs. (22) and (23)), or four have flipped signs (corresponding to Eq. (24)). These "partition-sum-like" terms are independent of the signs of the Δ_j; in fact we have for all of them

$$A_0 \sum_R e^{\mathcal{H}(R)} = 1 \tag{26}$$

The second type of term that contributes to H is of the "energy-like" form:

$$\sum_R e^{\mathcal{H}(R)} r_k r_{k+1} \tag{27}$$

The absolute value of these terms is again independent of the Δ_j, but one has to keep track of their signs. Finally, one has to remember that the constant A_1 also has to be expanded in ϵ. The calculation finally yields the following result (here we switch from J to the "natural" variable p using Eq. (12)):

$$H(p, \epsilon) = \sum_{k=0}^{\infty} H^{(k)}(p) \epsilon^k \tag{28}$$

with the coefficients H_k given by:

$$H^{(0)} = -p \log p - (1-p) \log(1-p) \qquad H^{(1)} = 2(1-2p) \log \left[\frac{1-p}{p} \right] \tag{29}$$

$$H^{(2)} = -2(1-2p) \log \left[\frac{1-p}{p} \right] - \frac{(1-2p)^2}{2p^2(1-p)^2} \tag{30}$$

The zeroth and first order terms (29) were known,[8,9] while the second order term is new.[10]

4. UPPER BOUNDS DERIVED USING A SYSTEM OF FINITE LENGTH

When investigating the limit H, it is useful to study the quantity $C_N = H_N - H_{N-1}$, which is also known as the *conditional* entropy. C_N can be interpreted as the average amount of uncertainty we have on r_N, assuming that we know $(r_1, \ldots, r_{N-1})$. Provided that H exist, it easily follows that

$$H = \lim_{N\to\infty} C_N \tag{31}$$

Moreover, according to ref. 11, $C_N \geqslant H$, and the convergence is monotone:

$$C_N \searrow H \quad (N \to \infty) \tag{32}$$

We can express C_N as a function of p and ϵ by using Eq. (17). For this, we represent Z using the original variables p, ϵ (note that from this point of we work with open boundary conditions on the Ising chain of N spins):

$$\begin{aligned} Z(R) = \sum_{S} (1-p)^{\sum_{i=1}^{N-1} 1_{S_i,S_{i+1}}} p^{N-1-\sum_{i=1}^{N-1} 1_{S_i,S_{i+1}}} \\ \times (1-\epsilon)^{\sum_{i=1}^{N} 1_{S_i,R_i}} \epsilon^{N-\sum_{i=1}^{n} 1_{S_i,R_i}} \end{aligned} \tag{33}$$

where we denote $1_{s,s'} = (1+ss')/2$. Equation (33) gives $Z(R)$ explicitly as a polynomial in p and ϵ with maximal degree N, and can be represented as :

$$Z(R) = \sum_{i=0}^{N} Z_i(R)\epsilon^i \tag{34}$$

Here $Z_i = Z_i(R)$ are functions of p only.

Substituting this expansion in Eq. (17), and expanding $\log Z(R)$ according to the Taylor series $\log(a+x) = \log(a) - \sum_{n=1}^{\infty} \frac{(-x)^n}{na^n}$, we get

$$H_N = -\sum_{R} \left[\sum_{i=0}^{N} Z_i(R)\epsilon^i \right] \left[\log Z_0(R) - \sum_{j=1}^{k} \frac{(-\sum_{i=1}^{n} Z_i(R)\epsilon^i)^j}{j Z_0(R)^j} \right] + O(\epsilon^{k+1}) \tag{35}$$

When extended to terms of order ϵ^k, this equation gives us precisely the expansion of the upper-bound C_N up to the *k-th* order,

$$C_N = \sum_{i=0}^{k} C_N^{(i)} \epsilon^i + O(\epsilon^{k+1}) \tag{36}$$

For example, stopping at order $k=2$ gives

$$H_N = -\sum_R \left\{ Z_0(R)\log Z_0(R) + [Z_1(R)(1+\log Z_0(R))]\epsilon + \left[\frac{Z_1(R)^2}{2Z_0(R)} + Z_2(R)(1+\log Z_0(R))\right]\epsilon^2 \right\} + O(\epsilon^3) \tag{37}$$

The zeroth and first order terms can be evaluated analytically for any N; beyond first order, we can compute the expansion of H_N symbolically[5] (using Maple[19]), for any finite N. This was actually done, for $N \leqslant 8$ and $k \leqslant 11$. For the first order we have proved[10] that $C_N^{(1)}$ is independent of N (and equals $H^{(1)}$). The symbolic computation of higher order terms yielded similar independence of N, provided that N is large enough. So, $C_N^{(k)} = C^{(k)}$ for large enough N. For example, $C_N^{(2)}$ is independent of N for $3 \leqslant N \leqslant 8$ and equals the exact value of $H^{(2)}$ as given by Eq. (30). Similarly, $C_N^{(4)}$ settles, for $N \geqslant 4$, at some value denoted by $C^{(4)}$, and so on. For the values we have checked, the settling point for $C_N^{(k)}$ turned out to be at $N = \lceil \frac{k+3}{2} \rceil$. This behavior is, however, unproved for $k \geqslant 2$, and, therefore, we refer to it as a

Conjecture. For any order k, there is a critical chain length $N_c(k) = \lceil \frac{k+3}{2} \rceil$ such that for $N > N_c(k)$ we have $C_N^{(k)} = C^{(k)}$.

It is known that $C_N \to H$, and C_N and H are analytic functions of ϵ at $\epsilon = 0$[6], so that we can expand both sides around $\epsilon = 0$, and conclude that $C_N^{(k)} \to H^{(k)}$ for any $k \geqslant 1$ when $N \to \infty$. Therefore, if our conjecture is true, and $C_N^{(k)}$ indeed settles at some value $C^{(k)}$ independent of N (for $N > N_c(k)$), it immediately follows that this value equals $H^{(k)}$. Note that the settling is rigourously supported for $k = 0, 1$, while for $k = 2$ we showed that indeed $C^{(2)} = H^{(2)}$, supporting our conjecture.

The first orders up to $H^{(11)}$, obtained by identifying $H^{(k)}$ with $C^{(k)}$, are given in the Appendix, as functions of $\lambda = 1 - 2p$, for better readability. The values of $H^{(0)}$, $H^{(1)}$ and $H^{(2)}$ coincide with the results that were

[5]The computation we have done is exponential in N, but the complexity can be improved.
[6]See next Section on the Radius of Convergence.

derived rigorously from the low-temperature/high-field expansion, thus giving us support for postulating the above Conjecture.

Interestingly, the nominators have a simpler expression when considered as a functions of λ, which is the second eigenvalue of the Markov transition matrix P. Note that only even powers of λ appear. Another interesting observation is that the free element in $[p(1-p)]^{2(k-1)}H^{(k)}$ (when treated as a polynomial in p), is $\frac{(-1)^k}{k(k-1)}$, which might suggest some role for the function $\log(1+\frac{\epsilon}{[2p(1-p)]^2})$ in the first derivative of H. All of the above observations led us to conjecture the following form for $H^{(k)}$ (for $k \geqslant 3$):

$$H^{(k)} = \frac{2^{4(k-1)} \sum_{j=0}^{d_k} a_{j,k} \lambda^{2j}}{k(k-1)(1-\lambda^2)^{2(k-1)}} \tag{38}$$

where $a_{j,k}$ and d_k are integers that can be seen in the Appendix for $H^{(k)}$ up to $k=11$.

5. THE RADIUS OF CONVERGENCE

If one wants to use our expansion around $\epsilon=0$ for actually estimating H at some value ϵ, it is important to ascertain that ϵ lies within the radius of convergence of the expansion. The fundamental observation made here is that for $p=0$, the function $H(\epsilon)$ is not an analytic function at $\epsilon=0$, since its first derivative diverges. As we increase p, the singularity point "moves" to negative values of ϵ, and hence the function is analytic at $\epsilon=0$, but the radius of convergence is determined by the distance of $\epsilon=0$ from this singularity. Denote by $\rho(p)$ the radius of convergence of $H(\epsilon)$ for a given p; we expect that $\rho(p)$ grows when we increase p, while for $p \to 0$, $\rho(p) \to 0$.

It is useful to first look at a simpler model, in which there is no interaction between the spins. Instead, each spin is in an external field which has a uniform constant component J, and a site-dependent component of absolute value K and a random sign. For this simple i.i.d. model the entropy rate takes the form

$$H = h_b[p(1-\epsilon)+\epsilon(1-p)] \tag{39}$$

where $h_b[x] = -[x\log x + (1-x)\log(1-x)]$ is the binary entropy function. Note that for $\epsilon=0$ the entropy of this model equals that of the Ising

chain. Expanding Eq. (39) in ϵ (for $p > 0$) gives:

$$H = -(p\log p + (1-p)\log(1-p)) + (1-2p)\log\left(\frac{1-p}{p}\right)\epsilon + \sum_{k=2}^{\infty}\frac{1}{k(k-1)}\left[\frac{(2p-1)^k}{p^k} + \frac{(1-2p)^k}{(1-p)^{k-1}}\right]\epsilon^k \tag{40}$$

The radius of convergence here is easily shown to be $p/(1-2p)$; it goes to 0 for $p \to 0$ and increases monotonically with p.

Returning to the HMP, the orders $H^{(k)}$ are (in absolute value) usually larger than those of the simpler i.i.d. model, and hence the radius of convergence may be expected to be smaller. Since we could not derive $\rho(p)$ analytically, we estimated it using extrapolation based on the first 11 orders. We use the fact that $\rho(p) = \lim_{k\to\infty}\frac{H^{(k)}}{H^{(k+1)}}$ (provided the limit exists). The data was fitted to a rational function of the following form (which holds for the i.i.d. model):

$$\frac{H^{(k)}}{H^{(k+1)}} \sim \frac{ak+b}{k+c}, \tag{41}$$

For a given fit, the radius of convergence was simply estimated by a. The resulting prediction is given in Fig. 3 for both the i.i.d. model (for which it is compared to the known exact $\rho(p)$) and for the HMP. While quantitatively the predicted radius of the HMP is much smaller than that of the i.i.d. model, it has the same qualitative behavior, of starting at zero for $p = 0$, and increasing with p.

We compared the analytic expansion to estimates of the entropy rate based on the lower and upper bounds, for two values of ϵ (see Fig. 4). First we took $\epsilon = 0.01$, which is realistic in typical communication applications. For p less than about 0.1 this value of ϵ exceeds the radius of convergence and the series expansion diverges, whereas for larger p the series converges and gives a very good approximation to $H(p, \epsilon = 0.01)$. The second value used was $\epsilon = 0.2$; here the divergence happens for $p \leqslant 0.37$, so the expansion yields a good approximation for a much smaller range. We note that, as expected, the approximation is much closer to the upper bound than to the lower bound, given in ref. 11.

6. SUMMARY

Transmission of a binary message through a noisy channel is modeled by a Hidden Markov Process. We mapped the binary symmetric HMP

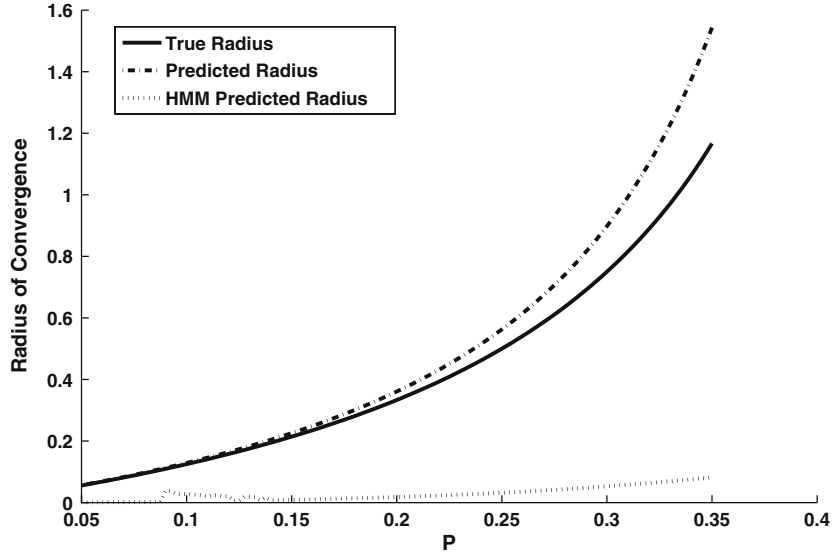

Fig. 3. Radius of convergence for the i.i.d. model (estimated and exact, see text), and HMP (estimated) for $0.05 \leqslant p \leqslant 0.35$.

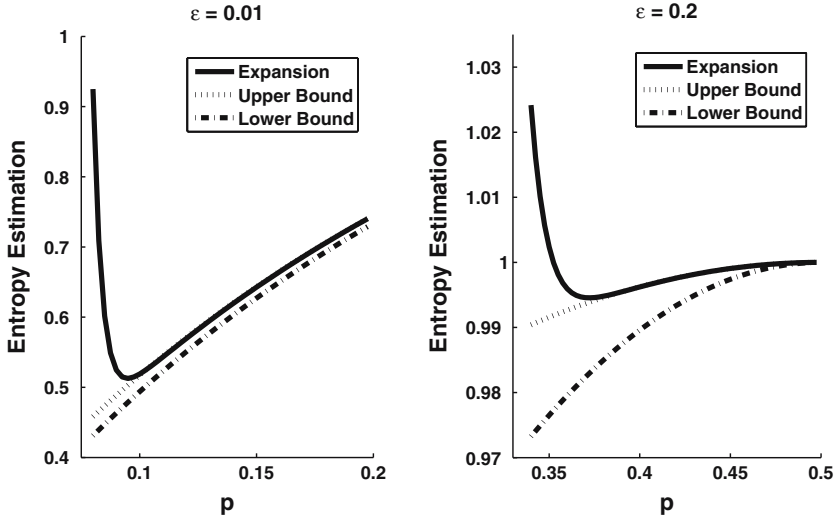

Fig. 4. Approximation using the first 11 orders in the expansion, for $\epsilon = 0.01$ (left) and $\epsilon = 0.2$ (right), for various values of p. For comparison, upper and lower bounds (using $N = 2$ from ref. 11) are displayed. For each ϵ there is some critical p below which the series diverges and the approximation is poor. For larger p the approximation becomes better.

onto an Ising chain in a random external field in thermal equilibrium. Using a low-temperature/high-random-field expansion we calculated the entropy of the HMP to second order $k=2$ in the noise parameter ϵ. We have shown for $k \leqslant 11$ that when the known upper bound on the entropy rate is expanded in ϵ, using finite chains of length N, the expansion coefficients settle, for $N_c(k) \leqslant N \leqslant 8$, to values that are independent of N. Posing a conjecture, that this continues to hold for any N, we identified the expansion coefficients of the entropy up to order 11. The radius of convergence of the resulting series was studied and the expansion was compared to the known upper and lower bounds.

By using methods of Statistical Physics we were able to address a problem of considerable current interest in the problem area of noisy communication channels and data compression.

APPENDIX

Orders 3–11, as function of $\lambda = 1-2p$. (Orders 0–2 are given in Eqs. (29)–(30)):

$$H^{(3)} = \frac{-16(5\lambda^4 - 10\lambda^2 - 3)\lambda^2}{3(1-\lambda^2)^4}$$

$$H^{(4)} = \frac{8(109\lambda^8 + 20\lambda^6 - 114\lambda^4 - 140\lambda^2 - 3)\lambda^2}{3(1-\lambda^2)^6}$$

$$H^{(5)} = \frac{-128(95\lambda^{10} + 336\lambda^8 + 762\lambda^6 - 708\lambda^4 - 769\lambda^2 - 100)\lambda^4}{15(1-\lambda^2)^8}$$

$$\begin{aligned} H^{(6)} = 128(&125\lambda^{14} - 321\lambda^{12} + 9525\lambda^{10} + 16511\lambda^8 - 7825\lambda^6 \\ &-17995\lambda^4 - 4001\lambda^2 - 115)\lambda^4/15(1-\lambda^2)^{10} \end{aligned}$$

$$\begin{aligned} H^{(7)} = -256(&280\lambda^{18} - 45941\lambda^{16} - 110888\lambda^{14} + 666580\lambda^{12} + 1628568\lambda^{10} \\ &-270014\lambda^8 - 1470296\lambda^6 - 524588\lambda^4 - 37296\lambda^2 - 245)\lambda^4/105(1-\lambda^2)^{12} \end{aligned}$$

$$\begin{aligned} H^{(8)} = 64(&56\lambda^{22} - 169169\lambda^{20} - 2072958\lambda^{18} - 5222301\lambda^{16} + 12116328\lambda^{14} \\ &+35666574\lambda^{12} + 3658284\lambda^{10} - 29072946\lambda^8 - 14556080\lambda^6 \\ &-1872317\lambda^4 - 48286\lambda^2 - 49)\lambda^4/21(1-\lambda^2)^{14} \end{aligned}$$

$$H^{(9)}=2048(37527\lambda^{22}+968829\lambda^{20}+8819501\lambda^{18}+20135431\lambda^{16}-23482698\lambda^{14}$$
$$-97554574\lambda^{12}-30319318\lambda^{10}+67137630\lambda^{8}+46641379\lambda^{6}+8950625\lambda^{4}$$
$$+495993\lambda^{2}+4683)\lambda^{6}/63(1-\lambda^{2})^{16}$$

$$H^{(10)}=-2048(38757\lambda^{26}+1394199\lambda^{24}+31894966\lambda^{22}+243826482\lambda^{20}$$
$$+571835031\lambda^{18}-326987427\lambda^{16}-2068579420\lambda^{14}-1054659252\lambda^{12}$$
$$+1173787011\lambda^{10}+1120170657\lambda^{8}+296483526\lambda^{6}+26886370\lambda^{4}$$
$$+684129\lambda^{2}+2187)\lambda^{6}/45(1-\lambda^{2})^{18}$$

$$H^{(11)}=8192(98142\lambda^{30}-1899975\lambda^{28}+92425520\lambda^{26}+3095961215\lambda^{24}$$
$$+25070557898\lambda^{22}+59810870313\lambda^{20}-11635283900\lambda^{18}$$
$$-173686662185\lambda^{16}-120533821070\lambda^{14}+74948247123\lambda^{12}$$
$$+102982107048\lambda^{10}+35567469125\lambda^{8}+4673872550\lambda^{6}+217466315\lambda^{4}$$
$$+2569380\lambda^{2}+2277)\lambda^{6}/495(1-\lambda^{2})^{20}$$

ACKNOWLEDGMENTS

I.K. thanks N. Merhav for very helpful comments, and the Einstein Center for Theoretical Physics for partial support. This work was partially supported by grants from the Minerva Foundation and by the European Community's Human Potential Programme under Contract HPRN-CT-2002-00319, STIPCO.

REFERENCES

1. Y. Ephraim and N. Merhav, Hidden Markov processes, *IEEE Trans. Inform. Theory* **48**:1518–1569 (2002).
2. A. Schliep, A. Schönhuth, and C. Steinhoff, Using hidden Markov models to analyze gene expression time course data, *Bioinformatics* **19**(Suppl. 1):i255–i263 (2003).
3. L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proc. IEEE* **77**:257–286 (1989).
4. I. Kanter, A. Frydman, and A. Ater, Is a multiple excitation of a single atom equivalent to a single ensemble of atoms? *Europhys. Lett.* **69**(6):874–878 (2005).
5. I. Kanter, A. Frydman, and A. Ater, Utilizing hidden Markov processes as a new tool for experimental physics, *Europhys. Lett.* **69**(5):798–804 (2005).
6. C. E. Shannon, A mathematical theory of communication, *Bell Sys. Tech. J.* **27**:379–423 and 623–656 (1948).
7. T. Nattermann, Theory of the random field Ising model, in *Spin Glasses and Random Fields*, A. P. Young, ed. (World Scientific, 1997).

8. P. Jacquet, G. Seroussi, and W. Szpankowski, On the Entropy of a Hidden Markov Process, Data Compression Conference, Snowbird (2004).
9. E. Ordentlich and T. Weissman, New Bounds on the Entropy Rate of Hidden Markov Processes, San Antonio Information Theory Workshop (Oct. 2004).
10. A preliminary presentation of our results is given in O. Zuk, I. Kanter, and E. Domany, Asymptotics of the Entropy Rate for a Hidden Markov Process, Data Compression Conference, Snowbird (2005).
11. T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
12. L. K. Saul and M. I. Jordan, *Boltzmann Chains and Hidden Markov Models, Advances in Neural Information Processing Systems*, Volume 7 (MIT Press, 1994).
13. D. J. C. MacKay, Equivalence of Boltzmann chains and Hidden Markov Models, *Neural Comput.* **8**(1):178–181 (1996).
14. B. Derrida, M. M. France, and J. Peyriere, Exactly solvable one-dimensional inhomogeneous models, *J. of Stat. Phys*. **45**(3–4):439–449 (1986).
15. D. S. Fisher, P. Le Doussal, and P. Monthus, Nonequilibrium dynamics of random field Ising spin chains: Exact results via real space renormalization group, *Phys. Rev. E* **64**(6):066–107 (2001).
16. G. Grinstein and D. Mukamel, Exact solution of a one-dimensional ising-model in a random magnetic field, *Phys. Rev. B* **27**:4503–4506 (1983).
17. T. M. Nieuwenhuizen and J. M. Luck, Exactly soluble random field Ising models in one-dimension, *J. Phys. A: Math. Gen.* **19**:1207–1227 (May 1986).
18. B. Derrida and H. J. Hilhorst, Singular behavior of certain infinite products of random 2×2 matrices, *J. Phys. A* **16**:2641–2654 (1983).
19. http://www.maplesoft.com/