A Note on Convergence Rate of Constrained Capacity Estimation Algorithms over ISI Channels

Tolga M. Duman
Electrical Engineering Department
Arizona State University
Tempe, AZ 85287-5706, USA
Email: duman@asu.edu

Junshan Zhang
Electrical Engineering Department
Arizona State University
Tempe, AZ 85287-5706, USA
Email: junshan@asu.edu

***Abstract*— It has recently become popular to use simulation-based algorithms to empirically estimate achievable information rates over intersymbol interference (ISI) channels with inputs from specific input constellations. Such algorithms are guaranteed to converge by invoking the Shannon-McMillan-Brieman theorem provided that the output sequence is stationary and ergodic. In this note, we establish a central limit theorem result on the rate of convergence, and show that the variance of the estimates decreases like $1/N$ (where N is the sequence length employed) as N goes to infinity. This result indicates that it is possible to achieve estimation accuracy with any desired level by simply increasing the number of samples appropriately.**

I. MOTIVATION

Intersymbol interference channels appear in many important communication scenarios, including recording channels, wireless channels when the delay spreads are large, underwater acoustic channels, etc. With this motivation, the capacity of deterministic ISI channels with additive white Gaussian noise (AWGN) has been derived, and it is shown that channel capacity can be achieved with correlated Gaussian inputs [1], [2]. For the case of random ISI channels encountered in wireless communications, i.e., for frequency selective fading channels, results on channel capacity are also available [3]. Extensions to multi-input multi-output (MIMO) systems have also been considered in the literature, e.g. [4], [5].

Consider, for instance, the case of deterministic ISI channels with AWGN, channel capacity results indicate that one needs to use correlated Gaussian inputs to achieve optimal transmission. However, in a practical system, one is usually limited to signals from specific constellations such as phase shift keying (PSK), quadrature amplitude modulation (QAM), etc. Therefore, it is also of importance to characterize the information rates under these constraints. When independent inputs are picked from a specific signal constellation as the input, the resulting information rates are simply referred as achievable. If the inputs are limited to specific constellations, but memory is allowed (i.e., Markovian sequences are selected as inputs) and optimization is carried out over the Markov input statistics, the resulting information rates could be referred as constrained capacity.

When the inputs are picked from specific constellations such as binary PSK, the input-output relationship can be described by a state diagram, or equivalently by a channel trellis. It is this structure that allows simplifications in the calculation of the mutual information expressions in an efficient manner once a long sequence of channel inputs and outputs are produced. This approach is employed in simulation based techniques for estimation of the achievable information rates/constrained channel capacities over ISI channels [6]–[9].

The rest of the note is organized as follows: In the next section, we state the rate of convergence problem formally. In Section III, we give the main result and provide the proof. We discuss some generalizations and give an example in Sections IV and V, respectively. Finally, we conclude the note in Section VI.

II. PROBLEM DESCRIPTION

Consider N uses of an ISI channel corrupted by additive white Gaussian noise. The input-output relationship can be written as

$$y_n = \sum_{l=0}^{L-1} h_l x_{n-l} + \eta_n \quad \text{for } n = 1, 2, \ldots, N, \tag{1}$$

where the coefficients $h_0, h_1, \ldots, h_{L-1}$ denote the deterministic channel taps, $x_1, x_2, \ldots, x_N$ are the channel inputs, and $\{\eta_n\}$ are the independent zero mean Gaussian noise samples with variance σ^2 each.

The capacity of this ISI channel is given by

$$C = \lim_{N\to\infty} \frac{1}{N} \max_{p(x_1,\ldots,x_N)} I(X_1^N; Y_1^N) \tag{2}$$

where $X_1^N = \{X_1, X_2, \ldots, X_N\}$ is the entire input sequence, $Y_1^N = \{Y_1, Y_2, \ldots, Y_N\}$ is the entire output sequence, and $I(\cdot;\cdot)$ denotes the mutual information. If the maximization is done over the Markov input distribution (with a specific constellation), a constrained channel capacity is obtained.

Clearly, using the independence of the input and noise terms, we can write

$$C = \lim_{N\to\infty} \frac{1}{N} \max_{p(x_1,\ldots,x_N)} H(Y_1^N) - H(\eta_1^N) \tag{3}$$

where $H(Y_1^N)$ denotes entropy of the output sequence which is the only unknown in this expression.

For a given input distribution, one way to estimate the entropy of the corresponding output sequence is to generate a long sequence of channel input and output realizations, and compute the probability of the resulting output sequence. Due to the well-defined trellis structure of the channel description, this computation can be efficiently done using the forward recursion of the well-known BCJR algorithm [6]–[9]. For a Markov input sequence, the output sequence is stationary and ergodic, and the result is guaranteed to converge by invoking the Shannon-McMillan-Breiman Theorem. If one is after the actual channel capacity, then an optimization over the input probability mass function is needed.

In summary, for the simulation based algorithms, we need to compute the following

$$S_N = -\log p(Y_1^N), \tag{4}$$
$$= -\sum_{n=1}^{N} Z_n, \tag{5}$$

where $p(\cdot)$ is the joint probability density function of the output sequence, and

$$Z_n = \log p(Y_n | Y_1^{n-1}) \tag{6}$$

with $p(\cdot|\cdot)$ denoting conditional probability density function.

III. MAIN RESULTS

Assume that the input signal is Markov with a given order. Our main result is that Var$(S_N)/N \rightarrow \sigma_s^2$ where σ_s^2 is a constant, and $S_N/\sqrt{N}$ converges to a normal random variable with variance σ_s^2. Therefore, the variance of the estimation error in computing the mutual information supported by the channel is proportional to $1/N$ and hence approaches 0 as N goes to infinity.

Before proceeding with the proof, it is illuminating to comment on a trivial special case. When the channel has no memory (i.e., $L = 1$), the input that maximizes the mutual information (achieves the channel capacity) is independent identically distributed (i.i.d.). In this case, the sequence $\{Y_1, Y_2, \ldots, Y_N\}$ becomes independent, hence $\{Z_1, Z_2, \ldots, Z_N\}$ are i.i.d. It is easy to verify that any moment of Z_n is finite as the random variable Z_n can be sandwiched between quadratic forms, and the expectations can be carried out over the Gaussian p.d.f. since the noise is Gaussian. Therefore, the claim in the previous paragraph follows directly from the central limit theorem. The point of this note is to establish a central limit theorem when the channel has memory, i.e. $L > 1$, and the inputs are Markov.

Before we proceed with the proof, two definitions, and a related result from probability theory will be useful [10].

Definition 1: A random sequence $\{W_n,\ n = 1, 2, 3, \ldots\}$, is *m-dependent* if the random variables W_{n_1} and W_{n_2} are independent for all n_1 and n_2 such that $n_1 - n_2 > m$ for a fixed $m < \infty$. That is, the samples of the random sequence taken sufficiently far apart are independent. Sometimes, i.i.d. sequences are referred as 0-dependent.

Definition 2: A random sequence W_n is α-mixing if the random samples are approximately independent with increasing separation between them. More precisely, let α_n be a number such that

$$|P(A \cap B) - P(A)P(B)| \leq \alpha_n$$

where

$$A \in \sigma(W_1, W_2, \ldots, W_k)$$

and

$$B \in \sigma(W_{n+k}, W_{n+k+1}, \ldots),$$

for $k \geq 1$ and $n \geq 1$ ($\sigma(\cdot)$ denotes the σ-field generated by random variables given in the argument). If $\alpha_n \rightarrow 0$ as $n \rightarrow \infty$, then $\{W_n\}$ is α-mixing.

The following result is a central limit theorem for α-mixing sequences [10].

Theorem 1: If W_n is α-mixing with $\alpha(n) = O(n^{-5})$ and $E(W_n^{12})$ is finite, then the central limit theorem holds, that is, $\frac{1}{\sqrt{N}} S_N = \frac{1}{\sqrt{N}}(W_1 + W_2 + \cdots + W_N)$ converges to a Gaussian random variable as N goes to infinity.

Clearly, m-dependent sequences form a special case of α-mixing random processes, hence the central limit theorem holds for this case as well. Also, note that the stated conditions on the convergence are sufficient, but not necessary. Based on this theorem, we now restate and prove our main result.

Theorem 2: Consider S_N in Equation 4. Var$(S_N)/N$ goes to a constant as N goes to ∞.

Proof:

Let us first consider the case of i.i.d. inputs where the dependence on the input, output and z_n sequences are illustrated in Figure 1. In this case, clearly, the channel output at time n is independent of the channel outputs prior to the time $n - L$. Hence, Z_n can be written as

$$Z_n = \log p(Y_n | Y_{n-L+1}^{n-1}).$$

It is also clear that the sequence Z_n is m-dependent since Z_n and Z_k are independent when n and k are sufficiently far apart (for $n - k \geq 2L - 1$). Since the central limit theorem applies to m-dependent sequences, the result follows.

If the input is Markov with a certain memory order, say M, the channel input at time n, conditioned on the previous M inputs $X_{n-1}, X_{n-2}, \ldots, X_{n-M}$, is independent of $\{X_{n-M-k}, k \geq 1\}$. Assume that the initial probabilities of the states (there are finitely many of them) are set to the stationary probabilities. Then, the input and output sequences are stationary. In this case, the sequence $\{X_n\}$ is α-mixing, where the dependence in time reduces exponentially for X_n and X_k (with $n - k$) (see Example 27.6 on page 363 of [10]). It is clear that (see Figure 1), Y_n is fully determined by the inputs $X_n, X_{n-1}, \ldots, X_{n-L+1}$ and independent noise terms, and accordingly term Z_n is described by the inputs $X_n, X_{n-1}, \ldots, X_{n-2L+2}$.

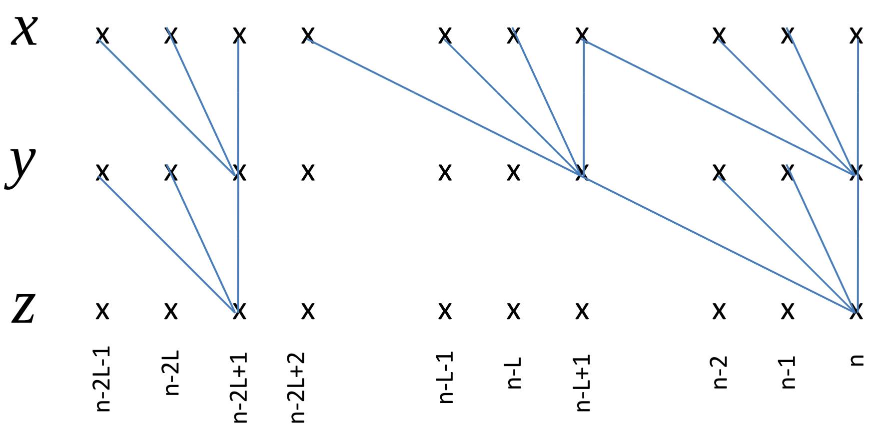

Fig. 1. Illustration of dependence of the random sequences of interest.

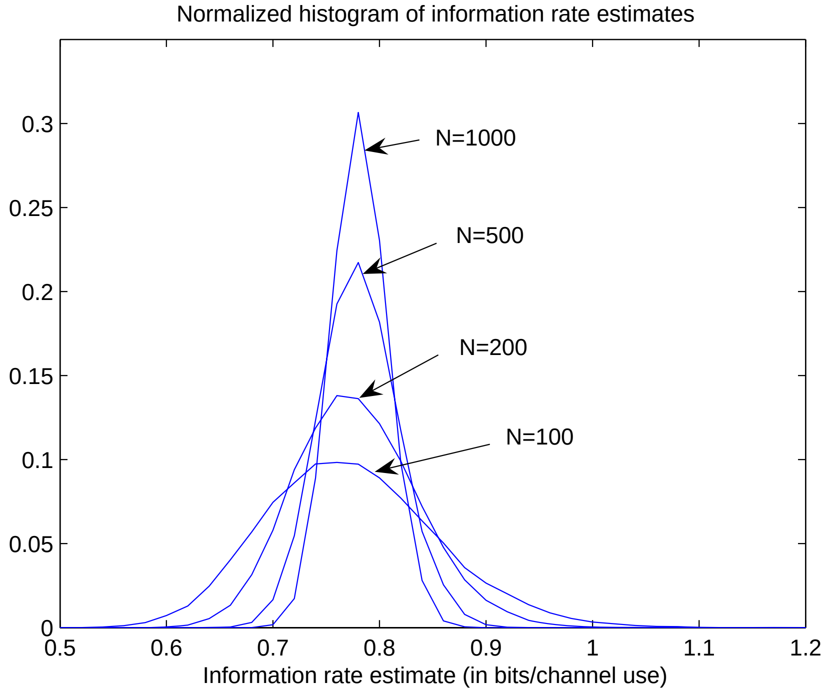

Fig. 2. Histogram of the information rate estimates for the given channel for different simulation lengths.

Since the input sequence is Markov with order M, we can define a new Markov process that describes the sequence of interest Z_n with a memory order of at most $2L+M-2$, and the sequence $\{Z_n\}$ is Hidden Markov. Also, the dependence between samples of $\{Z_n\}$ reduces exponentially, hence the sequence is $\alpha-$mixing with $\alpha_n = \rho^n$ with $\rho < 1$ (see Example 27.6 on page 363 of [10]). Therefore, for this particular α-mixing process we have $\alpha_n = O(n^{-5})$. Also, it is straightforward to use the total probability theorem, and upper and lower bound random variables Z_n by quadratic forms. Thus, any moment of Z_n is finite due the noise being Gaussian. Therefore, the central limit theorem holds, thereby proving the Theorem 2. ■

IV. EXTENSIONS OF ISI CHANNEL MODEL

So far, we have assumed that the channel is a deterministic ISI channel. Let us now consider the scenario where the channel is fading, and the channel coefficients evolve as a stationary ergodic Markov process. Assume that the channel coefficients are available at the receiver, hence the same entropy expression conditioned on the channel coefficients is needed to estimate the relevant information rates, or constrained channel capacity. The same type of proof also holds in this case with slight modifications (where M should be selected as the maximum of the Markov input memory and the memory of the channel coefficients). Therefore, the estimation error goes to zero as the sequence length N goes to infinity ($\sim 1/N$).

We further note that the results also hold for MIMO ISI channels and for the (constrained) ergodic capacity of MIMO frequency selective fading channels. The proof follows along the same lines.

V. EXAMPLE

As a brief illustration of the main result, let us consider a specific channel model. Assume that the channel transfer function is given by $H(D) = [\frac{1}{2} \quad \frac{1}{2} \quad -\frac{1}{2} \quad -\frac{1}{2}]$. Figure 2 shows the normalized histogram of the estimates of the information rates for i.i.d. BPSK inputs ($+1$ and -1 with uniform distribution) for several values of the sequence length used in the estimation process for a signal to noise ratio $1/\sigma^2$ of 5 dB. The estimate is 0.78 bits per channel use, and the variances of the estimates are calculated to be (from numerical data) 0.0066, 0.0033, 0.0013 and 0.000653 for simulation lengths of $N = 100$, $N = 200$, $N = 500$ and $N = 1000$, respectively.

This example clearly corroborates our result above.

VI. CONCLUSIONS

We have shown that simulation based methods for estimating constrained capacities of ISI channels result in an estimate variance that behaves like $\sim 1/N$ where N is the sequence length used in the estimation process, and that the estimates converge to a Gaussian random variable. Hence, the method, now widely used, over a variety of ISI channel models is sound, and can yield results with any desired accuracy by selecting appropriate simulations lengths.

ACKNOWLEDGMENT

The authors would like to Jun Hu for producing the figure in the example.

REFERENCES

[1] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inform. Theory*, vol. 34, no. 3, pp. 380–388, May 1988.

[2] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, N.J, USA: Wiley, 2006.

[3] E. Telatar and D. N. C. Tse, "Capacity and mutual information of wideband multipath fading channels," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1384–1400, 2000.

[4] H. El Gamal, A. R. H. Jr., Y. Liu, M. P. Fitz, and O. Y. Takeshita, "On the design of space-time and space-frequency codes for MIMO frequency-selective fading channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 9, pp. 2277–2292, Sept. 2003.

[5] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Viswanath, "Capacity limits for MIMO channels," *IEEE J. Select. Areas Commun.*, vol. 21, no. 5, pp. 684–702, 2003.

[6] D. Arnold and H.-A. Loeliger, "On the information rate of binary-input channels with memory," in *IEEE International Conference on Communications (ICC)*, Helsinki, June 2001, pp. 2692–2695.

[7] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite-state isi channels," in *IEEE GLOBECOM*, San Antonio, TX, Nov. 2001, pp. 2992–2996.

[8] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[9] Z. Zhang, T. M. Duman, and E. M. Kurtas, "Achievable information rates and coding for MIMO systems over ISI channels and frequency-selective fading channels," *IEEE Trans. Commun.*, vol. 52, no. 10, pp. 1698–1710, Oct. 2004.

[10] P. Billingsley, *Probability and Measure*, 3rd ed. Wiley-Interscience, 1995.