

Computation of Information Rates by Particle Methods

Justin Dauwels and Hans-Andrea Loeliger

Dept. of Information Technology and Electrical Engineering, ETH, CH-8092 Zürich, Switzerland

{dauwels, loeliger}@isi.ee.ethz.ch

Abstract — **Prior work on the computation of information rates of channels with memory is extended to continuous state spaces by means of sample-based numerical integration (“particle filtering”).**

I. INTRODUCTION

We consider the problem of computing the information rate

$$I(X; Y) \triangleq \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1, \dots, X_n; Y_1, \dots, Y_n) \quad (1)$$

between the input process $X = (X_1, X_2, \dots)$ and the output process $Y = (Y_1, Y_2, \dots)$ of a time-invariant discrete-time channel with memory. Let $x_k^n \triangleq (x_k, x_{k+1}, \dots, x_n)$ and $x^n \triangleq (x_1, x_2, \dots, x_n)$. We will assume that there is an ergodic stochastic process $S = (S_0, S_1, S_2, \dots)$ such that

$$p(x^n, y^n, s_0^n) = p(s_0) \prod_{k=1}^n p(x_k, y_k, s_k | s_{k-1}) \quad (2)$$

for all $n > 0$ and with $p(x_k, y_k, s_k | s_{k-1})$ not depending on k .

For *finite* input alphabet \mathcal{X} (= range of X_k) and *finite* state space \mathcal{S} (= range of S_k), a practical method for computing the information rate (1) was proposed earlier (e.g., see [1]). In [2], this method was described in greater generality and extended to the computation of upper and lower bounds on the information rate of very general channels. In this paper, we extend the methods of [1] and [2] to continuous state spaces \mathcal{S} and input alphabets \mathcal{X} .

This paper is structured as follows. In Section II, we review the basic idea of [1] as presented in [2]. In Section III, we briefly show how particle methods allow to deal with a continuous state space.

II. REVIEW OF BASIC METHOD

We first note that the sequence $-\frac{1}{n} \log p(X^n)$ converges with probability 1 to the entropy rate $H(X)$, the sequence $-\frac{1}{n} \log p(Y^n)$ converges with probability 1 to the differential entropy rate $h(Y)$, and $-\frac{1}{n} \log p(X^n, Y^n)$ converges with probability 1 to $H(X) + h(Y|X)$. From these observations, the quantity $I(X; Y) = h(Y) - h(Y|X)$ can be computed as follows:

1. Sample two “very long” sequences x^n and y^n .
2. Compute $\log p(x^n)$, $\log p(y^n)$, and $\log p(x^n, y^n)$. If $h(Y|X)$ is known analytically, then it suffices to compute $\log p(y^n)$.
3. Conclude with the estimate

$$\hat{I}(X; Y) \triangleq \frac{1}{n} \log p(x^n, y^n) - \frac{1}{n} \log p(x^n) - \frac{1}{n} \log p(y^n) \quad (3)$$

or, if $h(Y|X)$ is known analytically, $\hat{I}(X; Y) \triangleq -\frac{1}{n} \log p(y^n) - h(Y|X)$.

The computations in Step 2 can be carried out by forward sum-product message passing through the factor graph of (2). Consider, for example, the computation of the term $-\frac{1}{n} \log p(y^n)$. The state metric $\mu_k(s_k) \triangleq p(s_k, y^k)$ is computed by the recursion

$$\mu_k(s_k) = \lambda_k \int_{x_k} \int_{s_{k-1}} \mu_{k-1}(s_{k-1}) p(x_k, y_k, s_k | s_{k-1}), \quad (4)$$

where λ_k are scale factors. The quantity $-\frac{1}{n} \log p(y^n)$ is computed as the average of the logarithms of these scale factors, i.e., $\frac{1}{n} \sum_{k=1}^n \log \lambda_k = -\frac{1}{n} \log p(y^n)$, which converges (almost surely) to $h(Y)$. If necessary, the quantities $\log p(x^n)$ and $\log p(x^n, y^n)$ can be computed by the same method, see [2].

III. A PARTICLE METHOD

If both the input alphabet \mathcal{X} and the state space \mathcal{S} are finite sets, then the method of the previous section is a practical algorithm. However, we are now interested in the case where \mathcal{S} (and perhaps also \mathcal{X}) are continuous, as stated in the introduction. In this case, the computation of (4) is a problem.

This problem can be addressed by Monte Carlo methods known as particle filtering [3]. Such algorithms may be viewed as message passing algorithms where the messages (which represent probability distributions) are represented by a list of samples (“particles”) from the distribution [4]. In particular, we will represent the message μ_k by a list $\{\hat{s}_{k,\ell}\}_{\ell=1}^N$ of N samples and we will represent the distribution $\mu_{k-1}(s_{k-1}) p(x_k, s_k | s_{k-1})$ by a list of three-tuples $(\hat{s}_{k-1,\ell}, \hat{x}_{k,\ell}, \hat{s}_{k,\ell})$. From (4), we then obtain

$$\lambda_k^{-1} \approx \frac{1}{N} \sum_{\ell=1}^N p_{Y_k | X_k, S_k, S_{k-1}}(y_k | \hat{x}_{k,\ell}, \hat{s}_{k,\ell}, \hat{s}_{k-1,\ell}). \quad (5)$$

We applied this technique to a simple channel model and will present the results we obtained.

ACKNOWLEDGMENTS

The authors wish to thank Nicolas Cedraschi for assistance with the simulations. The helpful discussions with Daniel Hösli and Volker Koch are also gratefully acknowledged.

REFERENCES

- [1] D. Arnold and H.-A. Loeliger, “On the information rate of binary-input channels with memory,” *Proc. 2001 IEEE Int. Conf. on Communications*, Helsinki, Finland, June 11–14, 2001, pp. 2692–2695.
- [2] D. Arnold, H.-A. Loeliger, and P. O. Vontobel, “Computation of information rates from finite-state source/channel models,” *Proc. 40th Annual Allerton Conference on Communication, Control, and Computing*, (Allerton House, Monticello, Illinois), October 2–4, 2002, pp. 457–466. Available from <http://www.isi.ee.ethz.ch/~loeliger>.
- [3] Djuric et al., “Particle filtering,” *IEEE Signal Proc. Mag.*, September 2003, pp. 19–38.
- [4] H.-A. Loeliger, “An introduction to factor graphs,” *IEEE Signal Proc. Mag.*, Jan. 2004, pp. 28–41.