Physical Communication 19 (2016) 70–80

Contents lists available at ScienceDirect

Physical Communication

journal homepage: www.elsevier.com/locate/phycom

Full length article

Models, statistics, and rates of binary correlated sources

Marco Martalò [a,b,*], Riccardo Raheli [a]

[a] Department of Information Engineering, University of Parma, Italy
[b] E-Campus University, Novedrate (CO), Italy

ARTICLE INFO

Article history:
Received 16 June 2015
Received in revised form 16 December 2015
Accepted 6 April 2016
Available online 19 April 2016

Keywords:
Correlated sources
Binary sources
Statistical characterization
Entropy rate
Achievable region

ABSTRACT

This paper discusses and analyzes various models of binary correlated sources, which may be relevant in several distributed communication scenarios. These models are statistically characterized in terms of joint Probability Mass Function (PMF) and covariance. Closed-form expressions for the joint entropy of the sources are also overviewed. The asymptotic entropy rate for very large number of sources is shown to converge to a common limit for all the considered models. This fact generalizes recent results on the information-theoretic performance limit of communication schemes which exploit the correlation among sources at the receiver.

1. Introduction

The efficient transmission of correlated signals, observed at various nodes, to one or more collectors is of wide interest in various scenarios, such as sensor networks [1], and has been the subject of recent research attention. For instance, [2] discusses the spatial dependence between data according to the distribution of the nodes in the monitored area through empirical measurements. The design of efficient transmission schemes for correlated sources through orthogonal additive white Gaussian noise (AWGN) channels is a well established topic, see, e.g., [3]. In this case, the separation between source and channel coding is optimal and the ultimate performance can be achieved by means of distributed source coding (DSC) followed by independent capacity-achieving channel coding [3,4]. An alternative solution is based on the use of distributed joint source-channel coding (JSCC), where proper codes are used to encode the correlated sources. JSCC is based on the use of standard channel codes and may represent a practical and efficient solution. In both cases, knowledge of the statistical source correlation is exploited at the joint decoder, whereas source encoding is performed separately [5]. Orthogonal multiple access schemes with an arbitrary number of correlated sources have been recently addressed in [6]. In this contribution, the asymptotic achievable region for increasing number of sources has been characterized in terms of individual channel capacities, for a specific correlation model, and pragmatic JSCC schemes have been proposed.

In this paper, we discuss various correlation models for an arbitrary number of binary sources, which may be of interest in several realistic communication scenarios. With the exception of [6], there are not many papers in the literature which discuss correlation models for a possibly large numbers of sources. In [7], the authors proposed a correlation model based on a set of linear equations in the binary field. This model is shown to be related, under special conditions, to one of the binary symmetric channels (BSC)-based models discussed in this paper. The main contribution of this paper is twofold. First, we give a unified overview on the statistical characterization of such models in terms of joint probability mass function (PMF), covariance, and joint entropy of the sources. Then,

* Corresponding author at: Department of Information Engineering, University of Parma, Italy.
E-mail addresses: marco.martalo@unipr.it (M. Martalò), riccardo.raheli@unipr.it (R. Raheli).

http://dx.doi.org/10.1016/j.phycom.2016.04.002
1874-4907/

we derive the asymptotic entropy rate for large number of sources and show it is invariant to the considered model. This implies that the asymptotic achievable region, recently discussed in [6], which is characterized in terms of the source entropy rate, can be inferred to be independent of the specific correlation model and pragmatic joint source-channel coded schemes may be expected to have similar asymptotic behavior regardless of the model.

This paper is structured as follows. In Section 2, we present various binary source correlation models. In Section 3, we statistically characterize these models, by deriving the joint PMF of their output sequences and the corresponding covariance matrices. In Section 4, we statistically characterize these schemes in terms of their joint entropy. In Section 5, we use the joint entropy rate of these models to characterize the performance limit of orthogonal multiple access schemes transmitting correlated symbols. Finally, concluding remarks are given in Section 6.

2. Source correlation models

Consider N source nodes, possibly spatially distributed, which output (emit) binary information sequences $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)^T$, where $(\cdot)^T$ is the transpose operation. The binary information symbols are assumed to be marginally equiprobable, but correlated with each other according to a given PMF $P_{\boldsymbol{X}}(\boldsymbol{x})$, in which the N-element vector $\boldsymbol{x}$ describes a possible realization of $\boldsymbol{X}$. This scenario may be representative of a sensor network in which the sensors observe N correlated physical quantities of interest. In Fig. 1, possible correlation models are shown: (a) parallel, (b) serial, and (c) mixed. In the parallel model (a), the source symbols are the output of a set of parallel BSCs, with cross-over probability $1-\rho_\ell$, for $\ell = 1, 2, \ldots, N$, denoted as BSC(ρ_ℓ), whose input is a hidden common information bit B. The ℓ-th source symbol is given by

$$X_\ell = B \oplus Z_\ell \tag{1}$$

where B is an equiprobable binary random variable, Z_ℓ are independent binary random variables with $P(Z_\ell = 0) = \rho_\ell$, $1/2 \leq \rho_\ell \leq 1$ for $\ell = 1, 2, \ldots, N$, and $\oplus$ denotes a modulo-2 sum. The random variables B and $\{Z_\ell\}_{\ell=1}^N$ are independent. Note that the random variables $\{Z_\ell\}_{i=1}^N$ and $\{X_\ell\}_{i=1}^N$ form stochastic processes in the spatial domain. Obviously, if $\rho_\ell = 0.5$ there is no correlation among the binary information symbols $\{X_\ell\}_{\ell=1}^N$, whereas if $\rho_\ell = 1$ they are identical with probability 1.

The parallel model in (1) can be seen as a special case of a parallel model with binary asymmetric channels characterized by two probabilities:

$$P(X_\ell = 0|B = 0) = \rho_{\ell,0}$$
$$P(X_\ell = 1|B = 1) = \rho_{\ell,1}.$$

In the network information theory realm, the parallel model in Fig. 1(a) is a particular instance of the Chief Executive Problem (CEO) [8]. Note, however, that in the CEO problem the detection of the hidden source B is of interest, whereas in this paper we focus on the statistical characterization of the correlated sources $\{X_\ell\}_{\ell=1}^N$. The asymmetric model is representative of wireless sensor networking with asymmetric false alarm and missed detection probabilities, see, e.g., [9–11]. In the following, we focus on the balanced case, but a generalization of the statistical characterization presented in Sections 3 and 4 for the asymmetric parallel model is provided in Appendix A.

In Fig. 1(b), a possible serial correlation model is shown, in which the source symbols are correlated by a cascade of BSCs[1] with cross-over probability $1-\rho_\ell$. Again the output symbols are uncorrelated for $\rho_\ell = 0.5$, whereas they are equal with probability 1 for $\rho_\ell = 1$. This model may arise in multihop relay networks, where a source symbol is delivered to a final destination through intermediate relays [12]. However, note that the paper's focus is on the correlation structure of the data at the output of the intermediate BSCs.

A more general "mixed" case with a number m of serial branches is shown in Fig. 1(c), in which $1-\rho_{ij}$ denotes the cross-over probability of the ith BSC on the jth branch. In this case, the correlated data at the ℓ-th, $\ell = 1, 2, \ldots, N$, source on the jth branch, $j = 1, 2, \ldots, M$, can be expressed as

$$X_{\ell j} = B \oplus \underbrace{\sum_{i=1}^{\ell}\!\!\oplus\, Z_{ij}}_{\triangleq Z'_{\ell j}} = B \oplus Z'_{\ell j}$$

where the symbol $\sum\oplus$ denotes modulo-2 sums. The random variable $Z'_{\ell j}$ can be easily characterized by its distribution [13, Lemma 4.1]

$$p_{\ell j} \triangleq P(Z'_{\ell j} = 0) = \frac{1}{2}\left[1 + \prod_{i=1}^{\ell}(2\rho_{ij} - 1)\right]. \tag{2}$$

Note that for $M = 1$, this mixed model reduces to the serial one of Fig. 1(b) and the index j in (2) can be dropped.

A model based on a set of binary linear equations was considered in [7]:

$$\boldsymbol{A}\boldsymbol{X} = \boldsymbol{Z} \tag{3}$$

where $\boldsymbol{A}$ is a binary matrix (whose entries are equal to either 0 or 1) defining the set of equations and $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_N)^T$ is a vector of independent Bernoulli binary random variables with parameters $\rho_\ell = P(Z_\ell = 0)$. Note that matrix operations are performed in the binary field.

If $\boldsymbol{A}$ is such that $A_{\ell\ell} = 1$, $A_{\ell,\ell-1} = 1$, $A_{\ell k} = 0$ for $k > \ell$, and $\rho_1 = 1/2$, (3) is equivalent to the serial model in Fig. 1(b).[2] In fact, as a special case, (3) may encompass a set of recursive equations of the form

$$\sum_{i=0}^{\min\{D,\ell-1\}}\!\!\oplus\, A_{\ell,\ell-i}X_{\ell-i} = Z_\ell \tag{4}$$

[1] The first BSC does not play any role in the serial model and could be omitted—it is kept for notational consistency with the other models.

[2] Note that the parallel model (1) cannot be directly described by (3) since no structure of $\boldsymbol{A}$ can be found. However, the parallel model could be described by a modified system $\boldsymbol{AX} \oplus \boldsymbol{B} = \boldsymbol{Z}$, where $\boldsymbol{B}$ is a size-N vector with all elements equal to B and $\boldsymbol{A} = \boldsymbol{I}_N$, being $\boldsymbol{I}_N$ the identity matrix of size N. Similar considerations hold for the mixed model, which is based on a combination of parallel and serial models.