IV. The Poisson Model

Consider the number of defectives in a given interval N. Suppose that we partition N into n equally spaced subintervals such that the probability of each subinterval containing a defective is p and containing more than one defective is zero. Then Theorems 1 and 2 also apply to this case. If we let $n \to \infty$ and $np \to \lambda$, then we have Theorems 1 and 2 when the number of defectives is a Poisson variable.

The Poisson model has been widely used in the analysis of the capacity of a multiple access channel. In particular, Mikhailov and Tsybakov [2] proved a special case of Theorem 1 e), i.e., for $b = 1$ or 2 and $a_i = 2$, and used it to obtain an upper bound (the best so far) for the capacity. In a recent paper of Tsybakov, Mikhailov, and Likhanov [3], Theorem 1 b) was proved for the special case $a_i = k$ for a fixed k. This inequality was then crucially used to obtain upper bounds for the capabilities of channels which can simultaneously transmit up to $k-1$ messages.

V. The Case of Independent Random Variables

In this section $N = (X_1, \cdots, X_n)$ is a set of independent nonnegative integer-valued random variables with probability density function (pdf) $P_1, \cdots, P_n$. We first show that Theorem 1 b)–e) are no longer true even if $P_i = P$ for all i.

Let X and Y be independent identically distributed (i.i.d.) and

$$P(X=1) = P(X=9) = 0.30$$
$$P(X=2) = P(X=8) = 0.15$$
$$P(X=3) = P(X=7) = 0.04$$
$$P(X=4) = P(X=6) = 0.01.$$

Then

$$P(X \geq 2 | X+Y=7) = \frac{19}{34} < 1 = P(X \geq 2 | X+Y=6)$$

$$P(X \geq 2 | X+Y \geq 7) = \frac{2003}{2503} < \frac{1211}{1511}$$
$$= P(X \geq 2 | X+Y \geq 6)$$

$$P(X \geq 2 | X+Y \geq 6) = \frac{1211}{1511} < 1 = P(X \geq 2 | X+Y=6)$$

$$\frac{P(X+Y \geq 7 | X \geq 2)}{P(X+Y \geq 6 | X \geq 2)} = \frac{7509}{7555} < \frac{6009}{6055} = \frac{P(X+Y \geq 7)}{P(X+Y \geq 6)}.$$

All inequalities in Theorem 1 b)–e) are reversed.

Next we show that Theorem 2 remains true under the current probability model. Let $x = (x_1, \cdots, x_n)$ where x_i is the random value of X_i. Let

$$x \vee y = (\min\{x_1, y_1\}, \cdots, \min\{x_n, y_n\})$$
$$x \wedge y = (\max\{x_1, y_1\}, \cdots, \max\{x_n, y_n\}).$$

Let $P(x) = (P_1(x_1), \cdots, P_n(x_n))$. Then it is easily seen that

$$P(x)P(y) = P(x \wedge y) P(x \vee y)$$

where the multiplication stands for inner product. Furthermore, E and F are both increasing in x. Hence the FKG inequality still applies.

This proof can be extended to the continuous variable case by a standard "limit" argument.

VI. Concluding Remarks

Theorems 1 and 2 provide an interesting contrast in appreciating the power and the limitations of the FKG inequality. While the FKG inequality proves Theorem 2 effortlessly, saving a lengthy and painstaking argument of the type used in proving Theorem 1 a), there does not seem to be an easy way of applying it to those similar inequalities in Theorem 1 b) to e).

Acknowledgment

We are grateful to C. L. Mallows for the construction of the coupling in the proof of Theorem 1 and to R. Chen for the counterexamples in the i.i.d. case in Section V.

References

[1] C. M Fortuin, P. W. Kasteleyn, and J. Ginibre, "Correlation inequalities on some partially ordered sets," *Commun. Math. Phys.*, vol. 22, pp. 89–103, 1971.

[2] V. A. Mikhailov and B. S. Tsybakov, "Upper bound for the capacity of a random access system," *Prob. Peredach. Inform.*, vol. 17, pp. 90–95, 1981.

[3] B. S. Tsybakov, V. A. Mikhailov, and N. B. Likhanov, "Bounds for packet transmission rate in a random-multiple-access system," *Prob. Peredach. Inform.*, vol. 19, pp. 61–81, 1983.

Measures of Mutual and Causal Dependence Between Two Time Series

JORMA RISSANEN AND MATI WAX, MEMBER, IEEE

***Abstract*—New measures are proposed for mutual and causal dependence between two time series, based on information theoretical ideas. The measure of mutual dependence is shown to be the sum of the measure of unidirectional causal dependence from the first time series to the second, the measure of unidirectional causal dependence from the second to the first, and the measure of instantaneous causal dependence. The measures are applicable to any kind of time series: continuous, discrete, or categorical.**

I. Introduction

In areas such as econometrics, engineering, medicine, ecology, biology, education, and psychology it is frequently of interest to measure dependence between two time series. For example, we may wish to measure, in some way, the amount by which the prime interest rate influences unemployment, or, as another example, the amount by which the existence of a death penalty influences the number of severe crimes.

For the special case that the variables of the two time series are continuous, Geweke [5] proposed certain measures that are based on autoregressive (AR) and autoregressive with external inputs (ARX) modeling of the two time series. The unemployment rate, for example, is modeled first as an AR process and then, using the values of the prime interest rate, with an ARX model. The measures were defined by Geweke in terms of the variances of the two associated innovation processes. This approach is based on the framework developed by Granger [6], who built on an earlier work of Wiener [14]. Related work on the problem of testing whether some sort of causal dependence exists between two time series is presented in Sims [12], Caines and Chan [1], and Caines *et al.* [2].

In this correspondence we propose new measures for mutual and causal dependence which are applicable to any kind of time series, be it continuous, discrete or categorical. Our measures are based on a class of probabilistic models which includes the AR and ARX classes of models used by Geweke as a particular case.

Manuscript received March 29, 1985; revised February 17, 1986.
J. Rissanen is with the IBM Research Laboratory, San Jose, CA 95193, USA.
M. Wax is with Rafael, 84, Box 2250, Haifa, 31021, Israel.
IEEE Log Number 8612818.

We measure dependence by predictability, with predictability interpreted in terms of the codelength required to encode the time series with predictive coding. The idea of using predictive coding to measure predictability or, equivalently, stochastic complexity was introduced by Rissanen [10], which in turn was inspired by the complexity theory of information developed by Solomonoff [13], Kolmogorov [7], and Chaitin [4].

II. Measure of Unidirectional Causal Dependence

We plan to introduce a measure for the amount of unidirectional causal dependence from one observed time series $y = y^n = y_1, \cdots, y_n$ to another $x = x^n = x_1, \cdots, x_n$ relative to a selected class of models. By unidirectional causal dependence we mean the difference in predictability of x from its past samples provided by the added knowledge of the past samples of y. We interpret predictability in terms of the codelength required to encode the time series. Intuitively, the more predictable a time series is, the smaller will be the number of bits required to describe it using predictive coding. Predictive coding is a coding scheme whereby the "next" sample x_{t+1} is encoded using the "past" samples $x^t = x_1, \cdots, x_t$. Let $P_{\theta^{(k)}}(x_{t+1}|x^t)$ denote the probability distribution of x_{t+1} given the past samples x^t, where $\theta^{(k)}$ denotes a k-dimensional parameter vector that parameterizes this distribution. It is well-known (see Shannon [11]) that the ideal codelength for encoding x_{t+1} from its past is given by $-\log P_{\theta^{(k)}}(x_{t+1}|x^t)$. Thus the total codelength for encoding the time series x from its past is given by $\sum_{t=0}^{n-1} -\log P_{\theta^{(k)}}(x_{t+1}|x)$. Similarly, let $P_{\theta^{(r)}}(x_{t+1}|x^t, y^t)$ denote the probability distribution of x_{t+1} given the past samples of x and the past samples of y. The codelength required for encoding the time series x using its past and the past of y is given by $\sum_{t=0}^{n-1} -\log P_{\theta^{(r)}}(x_{t+1}|x^t, y^t)$.

The parameter vectors of the probability distributions are unknown and hence have to be estimated from the data. One would like to pick $\theta^{(k)}$ so as to minimize the codelength of the encoded sample: $\min_{\theta^{(k)}}\{-\log P_{\theta^{(k)}}(x_{t+1}|x^t)\}$. However, such a minimization would make the solution dependent on the value of x_{t+1}, which in turn would make decoding impossible, since it would mean that the parameter vector used for decoding the value of x_{t+1} would depend on this value. Thus, to enable decodability, the parameter vector must depend only on the past samples. Assuming ergodicity, the probability distribution governing the behavior of the next sample is like that governing the past samples. Thus, instead of minimizing the codelength required to describe the next sample, we may choose to estimate the parameter vector so as to minimize the sum of codelengths required to encode the past samples, i.e.,

$$\min_{\theta^{(k)}} \left\{ \sum_{i=0}^{t-1} -\log P_{\theta^{(k)}}\left(x_{i+1}|x^i \right) \right\} = \max_{\theta^{(k)}} \log P_{\theta^{(k)}}\left(x^{t-1} \right),$$

which is the maximum likelihood estimator of $\theta^{(k)}$ from the past samples, to be denoted by $\hat{\theta}^{(k)}(t)$. Notice that we apply here a principle of inductive inference—we pick that value which has worked best in the past, hoping it will work for the future as well.

The codelength required to encode x with predictive coding using the maximum likelihood estimator is therefore, with a slight abuse of notation, given by

$$L\left(x_{t+1}|x^t, k \right) = \sum_{t=0}^{n-i} -\log P_{\hat{\theta}^{(k)}(t)}\left(x_{t+1}|x^t \right). \tag{1}$$

The index t in the left-hand side is dummy and is introduced to denote the way the coding is done. Notice that (1) requires calculating $\hat{\theta}^{(k)}(t)$ for $t < k$, which leads to an ill-conditioned optimization. To avoid this we can either determine $\hat{\theta}^{(k)}(t)$ from prior knowledge or alternatively, when no such knowledge is available, never estimate more parameters than there are data points; that is, we begin by predicting the first sample by a

Then, having seen the first sample we can either estimate one parameter or, if not, we continue predicting the next sample by the initial constant. In this way we gradually increase the number of estimated parameters one by one until all the k parameters are estimated.

The codelength (1) was conditioned on knowing the dimension k of the parameter vector. However, k is unknown in general and thus has to be estimated from the data and then encoded as well. The code for k will be placed as a preamble in the code string. Since the preamble must be a prefix code, it follows from Rissanen [9] that it requires $\log^* k + \log c$ binary digits, where $\log^* k = \log k + \log\log k + \cdots$, the sum including all the positive iterates, and c is a constant, about 2.865. Thus, to minimize the total codelength, we should pick k so as to minimize $L(x_{t+1}|x^t, k) + \log^* k$. With this choice, the total codelength required to encode x is given by

$$L\left(x_{t+1}|x^t \right) = \min_k \left\{ L\left(x_{t+1}|x^t, k \right) + \log^* k \right\}. \tag{2}$$

Similarly, the total codelength required to encode x using the past samples of y is given by

$$L\left(x_{t+1}|x^t, y^t \right) = \min_r \left\{ L\left(x_{t+1}|x^t, y^t, r \right) + \log^* r \right\} \tag{3a}$$

where $L(x_{t+1}|x^t, y^t, r)$ is the predictive codelength conditioned on the parameter vector having dimension r

$$L\left(x_{t+1}|x^t, y^t, r \right) = \sum_{t=0}^{n-1} -\log P_{\hat{\theta}^{(r)}(t)}\left(x_{t+1}|x^t, y^t \right). \tag{3b}$$

The minimization in (3) is over an index set which includes that of (2), implying that

$$L\left(x_{t+1}|x^t, y^t \right) \leq L\left(x_{t+1}|x^t \right). \tag{4}$$

The codelengths (2) and (3) measure, respectively, the predictability of x from its past and the predictability of x from its past and the past of y. Their difference, therefore, is a measure of the added predictability of x provided by the knowledge of the past of y, and consequently, of the unidirectional causal dependence between the past of y and the present of x. In this light we define the measure of unidirectional causal dependence between y and x to be

$$\mu_{y \to x} = 1/n \left[L\left(x_{t+1}|x^t \right) - L\left(x_{t+1}|x^t, y^t \right) \right]. \tag{5}$$

Notice that it follows from (4) that this measure is always positive.

III. Example

To demonstrate the applicability of our measure we apply it to the special case where the two time series are continuous and modeled by AR and ARX models. We then compare it with the measure proposed by Geweke for this special case.

Let x and y be two continuous time series. To measure the causal dependence between y and x we have to compute the codelength required to encode x from its past and then compare it with the codelength required to encode x from its past and the past of y. To encode x from its past, we use an AR model

$$x_{t+1} = \sum_{i=0}^{k-1} a_i x_{t-i} + u_{t+1} \tag{6a}$$

and then, to take advantage of the past of y, we use an ARX model

$$x_{t+1} = \sum_{i=0}^{p-1} a_i x_{t-i} + \sum_{i=0}^{q-1} b_i y_{t-i} + v_{t+1} \tag{6b}$$

where $\{u_t\}$ and $\{v_t\}$ are modeled as independent zero mean

The predictive codelength required to encode the time series x, from (2), is given by

$$L\left(x_{t+1}|x^t\right)=\min_k\left\{L\left(x_{t+1}|x^t,k\right)+\log^* k\right\} \tag{7a}$$

where $L(x_{t+1}|x^t,k)$ is straightforwardly computed from (1) and (6):

$$L\left(x_{t+1}|x^t,k\right)=\sum_{t=0}^{n-1}\left[\log\hat{\sigma}_u(t)+\left(x_{t+1}-\hat{x}_{t+1}(t)\right)^2/2\hat{\sigma}_u^2(t)\right]. \tag{7b}$$

Here

$$\hat{x}_{t+1}=\sum_{i=0}^{t-1}\hat{a}_{i,k}(t)\,x_{t-i} \tag{7c}$$

and

$$\hat{\sigma}_u^2(t)=1/t\sum_{i=0}^{t-1}\left(x_i-\hat{x}_i\right) \tag{7d}$$

with $\hat{a}_{i,k}(t)$ denoting the maximum likelihood estimates obtained by fitting an AR model of order k to the past data x^t. Notice that the order of the AR model is obtained as an integral part of the computation. The codelength required to describe x by using the ARX model (6b) is obtained analogously.

It is of interest to investigate the asymptotic behavior of the measure when the sample size grows. It can be shown (see [9]) that asymptotically the mean of the codelength is given by

$$EL\left(x_{t+1}|x^t,k\right)\to n\log\sigma_u+1/2(k+1)\log n. \tag{8}$$

Thus the mean of the causality measure is asymptotically given by

$$E\mu_{y\to x}\to\log\sigma_u/\sigma_v+(k-p-q)\log n/2n. \tag{9}$$

The first term is the measure proposed by Geweke (Pierce [7] proposed a measure that is simply a one-to-one transformation of this measure into the unit interval), while the second term, which approaches zero as the sample size grows, can be interpreted as reflecting the difference in the complexities of the AR and ARX models.

In the studied case our measure approaches asymptotically the measure proposed by Geweke. However, the two measures are fundamentally different. Ours admits its own interpretation and is well defined regardless of the stochastic mechanism that generated the data. In contrast, Geweke's measure relies on the assumption that the generating processes are jointly stationary, purely nondeterministic, have autoregressive representations, and that the orders of the models are correctly estimated.

IV. Measures of Instantaneous Causal Dependence

We shall now extend the ideas we have already introduced to measure the instantaneous causal dependence between two time series. We want the measure of instantaneous causal dependence to reflect the dependence between the present of x and the present of y. It should therefore be symmetric in terms of x and y and should reflect both the added predictability of the present of x offered by the present of y and the added predictability of the present of y offered by the present of x. Let $L(x_{t+1}|x^t,y^t)$ denote the codelength required to encode the present of x from the past samples of x and the past samples of y, and let $L(x_{t+1}|x^t,y^{t+1})$ denote the codelength required to encode the present of x from the past samples of x and the past and present samples of y. The difference $L(x_{t+1}|x^t,y^t)-L(x_{t+1}|x^t,y^{t+1})$ reflects the added predictability of the present of x offered by the present of y. Similarly, let $L(y_{t+1}|y^t,x^t)$ denote the codelength required to encode the present of y from the past samples of y and the past samples of x, and let $L(y_{t+1}|y^t,x^{t+1})$ denote the codelength required to encode the present of y from the past samples of y and the past and present samples of x. The difference $L(y_{t+1}|y^t,x)-L(y_{t+1}|y^t,x^{t+1})$ reflects the added predictability of the present of y offered by the present of x. We therefore define the measure of instantaneous dependence to be

$$\mu_{x\cdot y}=1/2n\left\{\left[L\left(x_{t+1}|x^t,y^t\right)-L\left(x_{t+1}|x^t,y^{t+1}\right)\right]+\left[L\left(y_{t+1}|y^t,x^t\right)-L\left(y_{t+1}|y^t,x^{t+1}\right)\right]\right\} \tag{10a}$$

where $L(x_{t+1}|x^t,y^t)$ is given by (3a) and

$$L\left(x_{t+1}|x^t,y^{t+1}\right)=\min_q\left\{\sum_{t=0}^{n-1}-\log P_{\hat{\theta}(q)(t)}\left(x_{t+1}|x^t,y^{t+1}\right)+\log^* q\right\}. \tag{10b}$$

V. Measure of Mutual Dependence

So far we have introduced two measures of causal dependence, measuring different aspects of the dependence. Now we introduce yet another measure, which tries to capture the whole mutual dependence, and point out its relation to the two previously defined measures.

We want the measure of mutual dependence to reflect the predictability of each of the two time series from the other. Interpreting predictability in terms of codelength, we define the measure to be the saving in the codelength obtained in encoding x and y using the added knowledge of each other. Now, the encoding of x and y using the added knowledge of each other can be done in two different ways. Assuming that we have encoded the past samples of both x and y, we can either first encode x_{t+1} from the past samples x^t and y^t and then encode y_{t+1} from y^t and x^{t+1}, for which the resulting codelength is $L(x_{t+1}|x^t,y^t)+L(y_{t+1}|y^t,x^{t+1})$ or, alternatively, first encode y_{t+1} from x^t and y^t and then encode x_{t+1} from x^t and y^{t+1}, for which the resulting codelength is $L(x_{t+1}|x^t,y^t)+L(y_{t+1}|y^t,x^{t+1})$. Since we want the measure to be symmetric in terms of x and y, we take the average of the two codelengths to represent the codelength required to encode x and y with the added knowledge of each other. Now, since the codelength required to encode x and y separately is given by $L(x_{t+1}|x^t)$ and $L(y_{t+1}|y^t)$, respectively, our measure of mutual dependence is given by

$$\mu_{xy}=1/n\left\{\left[L\left(x_{t+1}|x^t\right)+L\left(y_{t+1}|y^t\right)\right]-1/2\left[L\left(x_{t+1}|x^t,y^t\right)+L\left(y_{t+1}|y^t,x^{t+1}\right)+L\left(y_{t+1}|y^t,x^t\right)+L\left(x_{t+1}|x^t,y^{t+1}\right)\right]\right\}. \tag{11}$$

Intuitively, we would expect the measure of mutual dependence to be decomposed, in some way, into the measures of unidirectional and instantaneous causal dependences. Indeed, it can be readily verified from (5) and (11) that the following decomposition holds:

$$\mu_{xy}=\mu_{y\to x}+\mu_{x\to y}+\mu_{x\cdot y}; \tag{12}$$

that is, the measure of mutual dependence between x and y is the sum of the measure of unidirectional causal dependence from y to x, the measure of unidirectional causal dependence from x to y, and the measure of instantaneous causal dependence.

In the special case that x and y are continuous and modeled by AR and ARX models, it can be shown, in complete analogy to the measure of causal dependence, that both the measures of instantaneous and mutual dependence approach asymptotically those proposed by Geweke [5] for which the decomposition (12) holds as well. Notice, however, that all the fundamental differences between our measures and those of Geweke, mentioned in the discussion of the measure of unidirectional causal dependence, hold for these measures as well.

The measure of mutual dependence we propose is similar in spirit to the measure of statistical dependence proposed by Fine [3] on the basis of the Solomonoff–Kolmogorov–Chaitin complexity measure. However, Fine's measure is not computable and hence not operational.

VI. Concluding Remarks

We have presented a conceptually new framework for measuring dependence between two time series. Unlike the framework developed by Geweke, which is limited to the AR and ARX class of models, and the framework developed by Fine, which is not operational, our framework is both very general in its applicability and straightforward operationally. All one needs to compute, say, the measure of unidirectional causal dependence from y to x are the two predictive densities $P(x_{t+1}|x^t)$ and $P(x_{t+1}|x^t, y^t)$. These densities need not be fully specified in advance; the parameters of these densities and even the number of parameters is determined as an integral part of the computation of the measure. Obviously, the value of the resulting measure depends critically on the class of predictive densities selected. A good selection gives a sharp measure of causal dependence while a bad one masks a possible causal dependence, which of course is just as it should be.

References

[1] P. E. Caines and C. W. Chan, "Feedback between stationary processes," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 498–508, 1975.
[2] P. E. Caines, C. W. Keng, and S. P. Sethi, "Causality analysis and multivariate autoregression modeling with application to supermarket sales analysis," *J. Economics Dynamics Contr.*, vol. 3, pp. 267–298, 1981.
[3] T. Fine, *Theories of Probability*. New York: Academic, 1973.
[4] G. J. Chaitin, "A theory of program size formally identical to information theory," *J. Assoc. Comput. Mach.*, vol. 22, pp. 329–340, 1981.
[5] J. Geweke, "Measurements of linear dependence and feedback between multiple time series," *J. Amer. Statist. Assoc.*, vol. 77, pp. 304–313, 1982.
[6] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, pp. 424–438, 1969.
[7] A. M. Kolmogorov, "Three approaches to quantitative definition of information," *Problems Inform. Transmission*, vol. 1, pp. 4–7, 1965.
[8] D. A. Pierce, "R measure for time series," *J. Amer. Statist. Assoc.*, vol. 74, pp. 901–910, 1979.
[9] J. Rissanen, "A universal prior for the integers and estimation by the minimum description length," *Ann. Statist.*, vol. 11, pp. 416–431, 1983.
[10] ——, "Universal coding, information, prediction, and estimation," *IEEE Trans. Inform. Theory*, vol. IT–30, pp. 629–636, 1984.
[11] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–425, 523–656, 1948.
[12] C. A. Sims, "Money and causality," *Amer. Economic Rev.*, vol. 62, pp. 540–552, 1972.
[13] R. J. Solomonoff, "A formal theory of inductive inference," *Inform. Contr.*, vol. 7, pp. 1–22, 224–254, 1964.
[14] N. Wiener, "The theory of prediction," in *Modern Mathematics for Engineers*, E. F. Beckenbach, Ed. New York: McGraw-Hill, 1956.

A Note on the Shannon Capacity of Run-Length-Limited Codes

JONATHAN J. ASHLEY AND PAUL H. SIEGEL, MEMBER, IEEE

***Abstract*—It is proven that 100-percent efficient fixed-rate codes for run-length-limited (RLL) (d,k) and RLL charge-constrained $(d,k;c)$ channels are possible in only two cases, namely $(d,k;c)=(0,1;1)$ and (1,3;3). Specifically, the binary Shannon capacity of RLL (d,k) constrained systems is shown to be irrational for all values of $(d,k), 0 \le d < k$. For RLL charge-constrained systems with parameters $(d,k;c)$, the binary capacity is irrational for all values of $(d,k;c), 0 \le d < k, 2c \ge k+1$, except (0,1;1) and (1,3;3), which both have binary capacity 1/2.**

Manuscript received June 9, 1986; revised October 30, 1986.
The authors are with IBM Research K61/802, Almaden Research Center, 650 Harry Road, San Jose, CA 95120-6099, USA.
IEEE Log Number 8612812.

I. Introduction and Background

In this correspondence we show that 100-percent efficient fixed-rate codes are impossible for most run-length-limited (d,k) and run-length-limited charge-constrained $(d,k;c)$ channels. More precisely, we prove that, among these channels, only the (0,1;1) and (1,3;3) have Shannon capacities which permit 100-percent efficient codes.

Run-length-limited (RLL) codes are widely used in magnetic and optical data recording channels. These constrained codes are characterized by two parameters $(d,k), 0 \le d < k$, which specify the minimum and maximum allowable run-lengths of zeros between consecutive ones in the constrained binary sequences.

RLL charge-constrained systems combine run-length constraints with the bounded running digital sum (RDS_0) constraint to ensure a spectral null at dc, as we now describe in more detail.

In data recording, constrained sequences $\{b_j\}, j \ge 0$, with symbols $\{0,1\}$ are typically converted to a two-level channel input signal $\{s_j\}, j \ge 0$, via a precoding convention called non-return-to-zero-index (NRZI), defined by

$$s_j = s_{j-1}(-1)^{b_j}, \qquad j \ge 0, \tag{1}$$

assuming $s_{-1} = 1$. The average power spectrum of the constraint is given by

$$\Phi(f) = \lim_{N \to \infty} E\left\{ \frac{\left| \sum_{j=0}^{N} s_j e^{-i2\pi jf} \right|^2}{N} \right\}, \qquad i = \sqrt{-1} \tag{2}$$

where the expected value is taken with respect to the measure of maximal entropy on the ensemble of allowable sequences, and the limit is interpreted in the distribution sense. The power spectrum of maxentropic RLL sequences can be expressed in a simple closed form. See Immink [6]. Specifically, for the RLL (d,k) constraint, let λ be the largest real root of the polynomial $p(x) = x^{k+1} - x^{k-d} - \cdots - x - 1$. Then the spectrum (2) is given by

$$\Phi(f) = \frac{1}{\sin(\pi f)^2 \bar{L}} \frac{1 - |G(2\pi f)|^2}{|1 + G(2\pi f)|^2} \tag{3}$$

where

$$\bar{L} = \sum_{j=d+1}^{k+1} j\lambda^{-j} \tag{4}$$

and

$$G(2\pi f) = \sum_{j=d+1}^{k+1} \lambda^{-j} \exp(2\pi i j f), \qquad i = \sqrt{-1}. \tag{5}$$

From (3)–(5), it can be seen that $\Phi(0) \ne 0$ for all $(d,k), 0 \le d < k$. In certain applications, however, it is required that the code power spectrum $\Phi(f)$ have a null at $f=0$ (dc), that is $\Phi(0)=0$. It is well-known that the code spectrum has a null at dc if for each finite code sequence $\{b_0, b_1, \cdots, b_N\}$ the running digital sum of the associated channel input sequence $\{s_0, s_1, \cdots, s_N\}$, denoted $\mathrm{RDS}_0(s_0, s_1, \cdots, s_N)$, is bounded by a fixed constant. That is,

$$\mathrm{RDS}_0(s_0, s_1, \cdots, s_N) = \left| \sum_{j=0}^{N} s_j \right| \le B, \tag{6}$$

for some constant B, and all finite allowable input strings $\{s_j\}$.

0018-9448/87/0700-0601$01.00 ©1987 IEEE