Proposition 4: Let \( a \) be a \( q \)-ary finite-field sequence of period \( P_a \) and let \( N \) be a positive integer such that \( 1 \leq N \leq P_a - 1 \). Let the minimum polynomial of \( a \) be an irreducible polynomial of degree \( n \). Then the linear complexities of \( a \) and \( a^{(N)} \) are equal if any of the two following conditions is satisfied.

1) \( q \equiv 2 \) and \( n \) is a prime.

2) \( P_a = q^n - 1 \) and \( N = q^m - 1 \mod P_a \), where \( n \) and \( m \) are positive integers.

Namely, the first condition immediately follows from Proposition 3, because if \( q \equiv 2 \), then \( 1 \leq N \leq q - 1 \), and the degree of \( f(N) \) must be equal to \( n \) as its period is necessarily bigger than one. The second condition is a direct consequence of the fact from number theory that \( n^q - 1 \mod n \), where \( p \) is a prime and \( n, m, k \) are positive integers, see [10, Lemma 2.2.4]. In the second condition, \( a \) is a \( q \)-ary maximum-length sequence and it is implied that \( N \geq 1 \Rightarrow n \mod m \). Observe that the first condition does not hold for \( q > 2 \), because the period of \( f(N) \), although bigger than one, can then divide \( q - 1 \) in which case its degree is equal to one. Nevertheless, if \( a \) is a \( q \)-ary maximum-length sequence, \( q > 2 \), \( n \) is a prime, and \( N \equiv 1 \leq N \leq q - 2 \), then \( a \) is a prime and \( n, m, k \) are positive integers, see [10, Lemma 2.2.4]. In the second condition, \( a \) is a \( q \)-ary maximum-length sequence and it is implied that \( N \geq 1 \Rightarrow n \mod m \). Observe that the first condition does not hold for \( q > 2 \), because the period of \( f(N) \), although bigger than one, can then divide \( q - 1 \) in which case its degree is equal to one. Nevertheless, if \( a \) is a \( q \)-ary maximum-length sequence, \( q > 2 \), \( n \) is a prime, and \( N \equiv 1 \leq N \leq q - 2 \), then \( a \) is a prime and \( n, m, k \) are positive integers, see [10, Lemma 2.2.4].

Since any two maximum-length sequences generated by a given primitive polynomial are translates of each other [8], the linear complexity condition from Proposition 2 is also necessary if \( a \) is a maximum-length finite-field sequence. In general, this is not likely to be the case. An interesting problem is then to find other sufficient conditions for the uniformly decimated sequences

\[
\{a(i + Nt)\}_{i=0}^{\infty}, \quad 0 \leq i \leq P_a - 1
\]

to be distinct, apart from the one given in Proposition 2.

V. CONCLUSIONS

Interleaving and nonuniform decimation play an important role in generation of pseudorandom sequences for various applications. Interleaving is particularly interesting for the parallel generation of high-speed sequences. Nonuniform decimation yields large sequence complexities which is especially interesting for stream cipher cryptography and spread-spectrum communications. In this correspondence, the periods of interleaved and nonuniformly decimated sequences are analyzed.

A general theorem characterizing the period of an interleaved sequence in terms of the constituent periodic integer sequences is first established. This theorem may be used to determine the period of various pseudorandom sequences, because every periodic integer sequence can be regarded as an interleaved one. This is demonstrated by deriving the period of a nonuniformly decimated sequence obtained from any periodic integer sequence by an arbitrary periodic difference decimation sequence, thus generalizing the result of Blakley and Purdy [1]. Some practically useful conditions for the period to be guaranteed are also pointed out, thus extending the results on the period of clock-controlled linear feedback shift registers. The achieved results can also be used to analyze the period of the well-known multiplexed sequences [10].
The character of the input in the second example is entirely different. We consider a black and white image coded into $n = 2^{19} = 524288$ bits (coding is based on 256 levels of grayness for each of $256 \times 256$ pixels, requiring $2^8$ bits per pixel for the $2^{16}$ pixels). In Fig. 1 we present the image and its distortion.

II. SIMONS’ ESTIMATOR

A. Basic Assumption

In [8], a strongly consistent estimate of the parameter $p$ was obtained based on the assumption that certain finite sequences of zeros and ones do not appear, or do appear relatively rarely, as blocks in the input $X$. To formulate precisely this basic assumption, let $F_{n,m}^{w}(u)$ denote the relative frequency of appearances of a block $u$ of length $m$, $m$-block for short, among the first $n$ $m$-blocks of a sequence $w = \{w_i\} \in \{0,1\}^N$, i.e.,

$$F_{n,m}^{w}(u) = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{w_{k+1}^{k+m}}(u)$$

(1)

where

$$w_{k+1}^{k+m} = \{w_{k+1}, \ldots, w_{k+m}\}$$

is the $k$th $m$-block in $w$, and $\delta_u$ is the probabilistic measure concentrated on a one-element set $\{u\}$. The function $F_{n,m}^{w}$, viewed as a convex combination of such probability measures, is also a probability measure on $\{0,1\}^m$. The rarest $m$-block among the first $n$ $m$-blocks of $w$ has relative frequency

$$\rho_{n,m}^{w} = \min_{u \in \{0,1\}^m} F_{n,m}^{w}(u).$$

Assumption 1: There exists an $m \in \mathbb{N}$ such that

$$\lim_{n \to \infty} \rho_{n,m}^{X} = 0.$$  

(2)

In other words, we assume here that in the input $X$ there are some patterns of bits for which the relative frequency of appearances converges to zero as the length of the considered part of $X$ grows without bound. Some theoretical aspects of necessity of Assumption 1 with relation to the problem of consistent estimation of $p$ are discussed in the Appendix.

If a block of the length $m$ is rare in $X \in \{0,1\}^N$, then each larger block containing it should also be rare. Consequently, if (2) is satisfied for some $m$, then it holds for $m + 1$. This fact invites the following definition.
Definition 1: The smallest \( m \) for which (2) holds is called the complexity of \( X \). If there is no such number, then we say the complexity of \( X \) is infinite.

For finite sequences the complexity of the input, in the strict sense of Definition 1, is not well-defined. By analogy to an infinite input, we shall look for the smallest number among such positive integers \( m \) that in the input there are \( m \)-blocks which are relatively rare. For \textit{Aeneid}, there exist two reasonable choices for the value of complexity. One can be the smallest \( m \) such that some \( m \)-blocks never appear in the input, which is 9 and there are three absent \( 9 \)-blocks. Alternative choice is the number 8, since there is an 8-block which appears with the frequency about 0.0000004—relatively small compared to the other frequencies (see Fig. 2). For the coded black and white image, there is an 11-block which never appears in the input, but also there are relatively rare 9- and 10-blocks. The plots of ordered frequencies for each of the both inputs, and for the mentioned values of \( m \), are presented in Fig. 2.

B. Largest Bernoulli Component (lbc)

To define the estimator proposed in [8], we need the notion of largest Bernoulli component, which was also explored in a somewhat different context in [5] and [7]. Let \( B^m_p \) denote the Bernoulli distribution on \( \{0, 1\}^m \) with parameter \( p \in [0, 1/2] \), i.e., for \( u \in \{0, 1\}^m \), \( B^m_p(u) = p^{|u|}(1 - p)^{m-|u|} \), where \(|u|\) is the number of ones in \( u \). It is convenient to consider \( B^m_p \) for \( p \in \mathbb{R} \), even though it cannot be interpreted as a probability distribution, when \( p \not\in [0, 1] \). For two arbitrary functions \( F, G \) on \( \{0, 1\}^m \), not necessarily probability distributions, we define their convolution \( F \ast G \) by

\[
(F \ast G)(u) = \sum_{v \in \{0, 1\}^m} F(u \oplus v)G(v).
\]

Definition 2: For a probability distribution \( F \) on \( \{0, 1\}^m \), the largest Bernoulli component \( p_F \) is the largest number \( p \in [0, 1/2] \) for which

\[
F = G + B^m_p
\]

for some probability distribution \( G \).

It is easily seen, by means of a simple continuity argument, that the largest Bernoulli component must exist (but might equal zero).

We often refer to (3) as a decomposition, or representation, of \( F \) with a Bernoulli component equal to \( p \). For each real \( p_1 \) and \( p_2 \), we have

\[
B^m_{p_1} + B^m_{p_2} = B^m_{p_1+p_2-2p_1p_2}.
\]

This elementary identity is essential for studying Bernoulli components. In particular, it implies that \( F \) has a decomposition with the Bernoulli component \( p \) for each \( p \in [0, p_F) \). Indeed, for \( p' = (p_F - p)/(1 - 2p) \), which is in \( [0, 1/2] \), we have

\[
F = G + B^m_{p'}, \quad (G + B^m_{p'}) \ast B^m_p)
\]

The same identity implies that

\[
p_F = \sup \{p \in [0, 1/2) : F \ast B^m_{p/(1 - 2p)}(u) \geq 0 \text{ for all } u \in \{0, 1\}^m \}.
\]

C. Estimating the Distortion by lbc

The estimator \( \hat{p}_{n,m} \) of distortion \( p \) is defined as the largest Bernoulli component of the empirical distribution of the output: \( F^n_{z,m} \). Since \( Z \) is random, so is \( \hat{p}_{n,m} \). Under Assumption 1, if \( m \) is fixed and greater or equal to the complexity of \( X \), then \( \hat{p}_{n,m} \) is a strongly consistent estimator of \( p \) (cf. [8]).

Without going into technical details, let us explain some rationales behind this result. The Law of Large Numbers gives us an approximate, in the almost sure sense, decomposition

\[
F^n_{z,m} \approx E F^n_{z,m} = F^X_n + B^m_p.
\]

From Definition 2, there exists a probability distribution \( G \) such that \( F^n_{z,m} = G + B^m_{p_{n,m}} \), with \( p_{n,m} \) being the largest number in \([0, 1/2]\) which enables such a decomposition. Thus in view of (5), we would expect that \( \hat{p}_{n,m} - p \) has a lower limit greater or equal to zero. To see, on the one hand, that the upper limit should not exceed zero, we refer to Definition 1, that by (4) and (5), we have

\[
F^n_{z,m} = (F^X_n + B^m_p) \ast B^m_{p/(1 - 2p)} \approx F^X_n + B^m_{p/(1 - 2p)}
\]

This gives the approximate representation of \( F^n_{z,m} \) with the Bernoulli component equal to \( (\hat{p}_{n,m} - p)/(1 - 2p) \). However, the following relations between \( p_{n,m} \) and the largest Bernoulli component \( r_{n,m} \) of \( F^n_{z,m} \) were proved in [8]:

\[
r_{n,m} \leq \hat{p}_{n,m} \leq r_{n,m}.
\]

Thus if the upper limit of \( \hat{p}_{n,m} - p \) was positive, then the largest Bernoulli component \( r_{n,m} \) would be separated from zero which, by (6), would violate the basic assumption.

D. Complexity and lbc

The relation (6) tells us that the complexity can be equivalently defined as the smallest \( m \) for which the largest Bernoulli component \( r_{n,m} \) converges in \( n \) to zero. This fact can be helpful in a choice of the value of complexity for a finite sequence of bits. As we have seen (cf. Fig. 2), for finite inputs it can happen that in spite of all \( m \)-patterns being present in the input some of them are relatively rare. However, note that values of frequencies by themselves are not adequate indicators of the complexity—something small for, say, \( m = 8 \) can be large for \( m = 20 \) (the number of relative frequency values increases as \( 2^m \), while their total for every \( m \) adds to one). The values of the largest Bernoulli component are free of this flaw, and thus it seems to be more appropriate for defining the complexity of finite sequences.

Fig. 3 shows the largest Bernoulli components \( r^X_{n,m} \) of the \textit{Aeneid} input computed for various \( m \) and for two different sample sizes. Because of the relation (see Section II-B)

\[
\hat{p}_{n,m} = p + (1 - 2p)r^X_{n,m} + o(n^{1/m})
\]

we should consider values \( r^X_{n,m} \) relatively to the value \( p \). Here, it seems reasonable to view the complexity as \( 8 \) when, for example, \( p = 0.1 \) or \( p = 0.03 \), while the value \( 9 \) should be entertained when \( p = 0.005 \) \( (r^X_{n,8} \approx 0.002) \). On the other hand, if we consider the complexity as an asymptotic characteristic of the input, then in all cases \( 9 \) should be chosen since the largest Bernoulli components exhibit stability with respect to \( n \). (The graphs in Fig. 3 for \( n = 220178 \) and \( n = 2201785 \) are almost identical, while Definition 1 implies convergence of \( r_n \) to zero with \( n \).)

E. Negative Bias

Despite the consistence result, \( \hat{p}_{n,m} \) often exhibits a negative bias, increasing in size with the value of \( p \). But the size of the negative bias depends also on the choice of \( m \). In the top graph of Fig. 4, there are shown for the \textit{Aeneid} example the computed values of the estimate for \( p = 0.1, 0.03, 0.005 \) and for various \( m \) (to illustrate the dependence of estimation on size of data we have included also the analogous graphs obtained for 10\% of the original data). With increasing \( m > 9 \) we observe a significant increase of negative bias. A clue to understanding this phenomenon is given in [6], where it is pointed out, and demonstrated through examples, that the size of the (negative) bias also increases with the total number of different rare \( m \)-blocks present in the input. The latter effect is linked to the
Fig. 2. Ordered frequencies of $m$-blocks in the input (for Aeneid $m = 8, 9$; for the control image $m = 9, 10, 11$).

The number of rare $m$-blocks increases with $m$—at least doubling when $m$ increasing to $m + 1$. Thus a cost can be anticipated when the chosen value of $m$ strictly exceeds the complexity.

This can be observed clearly in the bottom part of Fig. 4. There density estimates of distribution of $\hat{p}_{n,8}$, $\hat{p}_{n,9}$, and $\hat{p}_{n,10}$ are drawn based on Monte Carlo simulations. We observe, with increase of $m$. 

and describe an estimator of distortion which is computable without using the asymptotic properties of the introduced estimators, we quality of estimation instead of increasing the negative bias. Finally, the role in our new approach to estimating the distortion.

A. Generalized Largest Bernoulli Components (glbc)

Note that for \( m = 8 \) we see even positive bias which is due to the presence of the nonzero lbc in the input which shifts the distribution to the right by approximately 0.0016. After correction by this value (dashed line) the distribution becomes centered at the true parameter which confirms that in the case of one rare \( m \)-block the estimator is not biased (for \( m = 8 \) only one \( m \)-block can be considered rare).

In the next section, we introduce and study a generalized notion of the Bernoulli component. With this notion, we define estimators of distortion which reduce the negative bias, and even the standard deviation, from that of \( \hat{p}_{n,m} \) (without giving up the latter’s consistency). In our approach, we exploit the fact that a larger number of rare \( m \)-blocks can, in fact, if appropriately used, increase the quality of estimation instead of increasing the negative bias. Finally, using the asymptotic properties of the introduced estimators, we describe an estimator of distortion which is computable without prior determination (estimation) of the complexity. This effectively eliminates the practical issue of having to choose the parameter \( m \) judiciously.

III. GENERALIZED BERNULLI COMPONENTS AND DISTORTION ESTIMATES

A. Generalized Largest Bernoulli Components (glbc)

The following generalization of the concept of lbc plays a decisive role in our new approach to estimating the distortion.

Definition 3: For a probabilistic distribution \( F \) on \( \{0, 1\}^m \), the generalized largest Bernoulli components \( r_s(F) \), for \( s = 1, \ldots, 2^m \), are defined by

\[
r_s(F) = \sup \{ p \in [0, 1/2] : F * B^m_{p/(1-2p)}(S) \geq 0, \text{ if } \text{card}(S) = s \}.
\]

In the following result, we list more important properties of glbc. Some of them are closely related to the properties of lbc which, in consequence, will allow later to establish the analogous consistency result.

Proposition 1: Let \( F, \hat{F} \) be probability measures on \( \{0, 1\}^m \) and let

\[
\|F - \hat{F}\|_\infty = \max_{u \in \{0, 1\}^m} |F(u) - \hat{F}(u)|.
\]

The following conditions hold:

i) For each \( p \in [0, 1/2] \)

\[
r_s(F * B^m_p) = r_s(F)(1 - 2p) + p.
\]

ii) If \( r_s(F) < 1/2 \), then there exists an additive function \( G \) on subsets of \( \{0, 1\}^m \) such that \( F = G * B_{r_s} \) and

\[
\min \{ G(S) : \text{card}(S) = s \} = 0.
\]

iii)

\[
[r_s(F)]^m = \frac{1}{\min \{ F(S) : \text{card}(S) = s \}} \leq r_s(F).
\]

iv) For \( s \in \{1, \ldots, 2^m - 1\} \)

\[
r_s(F) \leq r_{s+1}(F)
\]

and \( r_1(F) \) corresponds to the largest Bernoulli component while \( r_{2^m}(F) = 1/2 \).

v)

\[
|r_s(F) - r_s(\hat{F})|^m \leq \|F - \hat{F}\|_\infty.
\]

Proof: If for some \( r \in [0, 1/2] \) we have \( F \ast B^m_{\frac{r}{(1-2r)}}(S) \geq 0 \), whenever \( \text{card}(S) = s \), then the same condition is satisfied for any \( r' \in [0, r] \), which follows from

\[
F \ast B^m_{\frac{r'}{(1-2r')}} = F \ast B^m_{\frac{r}{(1-2r)}} \ast B^m_p
\]

Fig. 3. Lbc of the input for various block sizes.
Fig. 4. Performance of the estimator for the *Aeneid* input. Top: Computed values of the estimator for various $m$ and for three values of the distortion: $p = 0.005, 0.03, 0.1$. Bottom: The density estimates as results of Monte Carlo simulations of the estimator in the case of distortion $p = 0.1$ and $m = 8, 9, 10$. 
where \( p = (r - r')/(1 - 2r') > 0 \). Consequently, the supremum in (8) is taken either over \([0, r_s(F)]\) or \([0, 1/2] \). Now, by the above equality,

\[
(F + B_{r_s}(F)) + B_{(1-2r_s,1)}(S) \geq 0
\]

if and only if \( F + B_{r_s}(F) + \) \( B_{(1-2r_s,1)}(S) \geq 0 \), where \( r' = r(1 - 2p) + p \) and thus i) follows.

Let \( r_s < 1/2 \). Setting \( G = F + B_{r_s}(F) \) we have

\[
\min \{ G(S) : \text{card}(S) = s \} \geq 0.
\]

By continuity arguments, \( G \) cannot be strictly positive on all subsets of cardinality \( s \) as \( r_s \) could not be maximal. Consequently, ii) holds.

Let \( S \) be of cardinality \( s, r_s = r_s(F) < 1/2 \), and \( G \) be as in ii).

Then iii) follows, by ii), from the following relations:

\[
F(S) = \sum_{v \in [0,1]^m} G(v \oplus S) + B_{r_s}^m(v)
\]

\[
\geq r_s^m \sum_{v \in [0,1]^m} G(v \oplus u) = sr_s^m
\]

\[
F(S) \leq r_s \sum_{v \in [0,1]^m} G(v \oplus S) + (1 - r_s)^m G(S)
\]

\[
= sr_s + G(S)(1 - r_s)^m - r_s
\]

and by the fact that there exists an \( S \) such that \( G(S) = 0 \). In the case \( r_s = 1/2 \), we can use the analogous relation for \( r < 1/2 \), and then obtain iii) by passing with \( r \) to the limit \( 1/2 \).

To prove iv), let \( 1 \leq s < 2^m, r_s = r_s(F) < 1/2 \), \( G \) be as in ii), and \( S \subseteq \{0,1\}^m \) be of cardinality \( s + 1 \). There exists \( u_0 \in S \) such that

\[
G(u_0) = F + B_{r_s(1-2r_s)}(u_0) \geq 0
\]

otherwise, in contradiction with ii), \( G \) would not be nonnegative on \( s \)-element subsets of \( S \) Thus

\[
G(S) = G(S \setminus \{u_0\}) + G(u_0) \geq 0
\]

with both terms on the right-hand side being nonnegative. This proves that \( r_{s+1} \geq r_s \). The case of \( r_s = 1/2 \) can be obtained from these arguments as a limiting case. Finally, \( r_1(F) \) is the largest Bernoulli component by its definition, while \( r_{2m}(F) = 1/2 \) since

\[
F + B_{r_s(1-2r_s)}(\{0,1\}^m) = 1, \quad \text{for any } r \in [0, 1/2).
\]

To prove v), assume \( r_s = r_s(F) < r_s = r_s(F) < 1/2 \), and define

\[
G = F + B_{r_s(1-2r_s)}
\]

\[
\tilde{G} = F + B_{r_s(1-2r_s)}
\]

Then

\[
\tilde{G} + B_{r_s}^m - G = (F - \tilde{F}) + B_{r_s(1-2r_s)}.
\]

For \( S \) of cardinality \( s \), we have

\[
\tilde{G} + B_{r_s}^m(S) \geq \sum_{v \in [0,1]^m} \tilde{G}(v \oplus S) \geq sq^m.
\]

From ii), there exists \( S \) of cardinality \( s \) such that \( G(S) = 0 \). For this particular \( S \), by the preceding relations, we have

\[
\frac{s}{1 - 2r_s} \geq \tilde{G} + B_{r_s}^m(S) = (F - \tilde{F}) + B_{r_s(1-2r_s)} + G(S)
\]

\[
= \sum_{v \in [0,1]^m} \frac{(F - \tilde{F})(v, -r)}{(1 - r_s)^m - r} \leq \frac{(1 - 2r_s)^m}{1 - 2r_s} \sum_{v \in [0,1]^m} (r_v)^m (1 - r_s)^m - r
\]

and, since the last sum is equal to one, v) is proven.

### Remark 1

Define

\[
\|F - \tilde{F}\|_{\infty} = \max \{ |F(S) - \tilde{F}(S)| : S \subseteq \{0,1\}^m, \text{card}(S) = s \}
\]

From the proof of v), it follows that

\[
|r_s(F) - r_s(\tilde{F})| \leq \|F - \tilde{F}\|_{\infty}.
\]

Obviously, this bound can be much smaller than \( \|F - \tilde{F}\|_{\infty} \), and thus the above inequality is usually stronger than v). This fact partially explains the observed reduction of variance of estimators based on generalized Bernoulli components that occurs with increasing \( s \). More generally, \( \|F - \tilde{F}\|_{\infty} \leq \|F - \tilde{F}\|_{\infty+1} \). (See Sections II-D and II-E for details.)

### B. Numerical Computing gIbc

The definition of gIbc is somewhat implicit, but there exists an explicit method of finding it, adopted from the one proposed for hbc in [8]. It is based on the fact that \( F = G + B_{r_s}^m \) is equivalent to \( G = F \ast B_{r_s(1-2r_s)} \) and on ii) of Proposition 1, which imply that, for fixed \( s, p < 1/2 \) is gIbc of \( F \) if and only if

\[
h(p) = \min \{ F \ast B_{r_s(1-2r_s)}(S) : \text{card}(S) = s \}.
\]

One can verify that \( h(p) \) is strictly decreasing and continuous on \([0, 1/2] \) with range \((\infty, \min \{ F(S) : \text{card}(S) = s \}) \). Thus standard numerical methods of finding zeros of a continuous and monotone function apply. Also considering \( F \ast B_{r_s(1-2r_s)} \) as a vector with \( 2^m \) coordinates, computation of \( h(p) \) simplifies to adding up its smallest order statistics. To conclude, a numerical solution to (9) can be effectively computed by rather standard computer algorithms.

### C. GIbc of Empirical Distributions

Having established the basic properties of gIbc, we exploit them in estimating the distortion, extending the ideas behind the properties of \( \hat{p}_{nm} \).

Recall that in Assumption 1 we require that in the input, for some \( m \), some \( m \)-blocks have to be rare among first \( n \) blocks with increase of \( n \). Note that if a block, say, \( w \), of the length \( m \) is rare in \( X \), then at least two blocks of the length \( m + 1 \), namely, \( w00 \) and \( w10 \), have to be rare (usually there will be more than just two as \( 0w \) and \( 1w \) will be rare as well). This guarantees that by selecting appropriately large \( m \) we can find subsets of \( m \)-blocks of any size \( s \) on which \( F_{nm}^X \) takes small values, i.e., \( m \)-blocks in such subsets are rare. It follows then from iii) of Proposition 1 that also \( r_s(F_{nm}^X) \) has to converge to zero, with \( n \) increasing, and, by similar arguments as in [8], \( \hat{p}_{nm}(s) = r_s(F_{nm}^X) \) should converge to the distortion. However now, \( F_{nm}^X \ast B_{r_s(1-2r_s)}(\{0,1\}^m) \) is nonnegative only on sets of cardinality \( s \) or more, and, in contrast to the original approach, we admit negative values on subsets of smaller cardinality than \( s \). As a result, we obtain two desirable effects: averaging random fluctuations which reduce the variance of the estimator, and, at the same time, reduction of the negative bias, since, by iv) of Proposition 1, \( \hat{p}_{nm}(s) \geq F_{nm} \). Below, we work out these ideas in formal detail.

We have previously observed that the complexity of \( X \) is equivalently defined as the smallest integer \( m \) for which \( r_{nm} = r_1(F_{nm}^X) \) converges to zero in \( n \). In the present context, the following definition is a natural extension of this idea.
Definition 4: For \( s \in \mathbb{N} \), the complexity \( m^X_s \) of \( X \) with respect to sets of cardinality \( s \) is the smallest number \( m \) for which

\[
\lim_{n \to \infty} \min \{ F^X_{n,m}(S) : \card S = s \} = 0
\]

or, equivalently, by iii) of Proposition 1

\[
m^X_s = \min \left\{ m > \log_2 s : \lim_{n \to \infty} r^X_{n,m}(s) = 0 \right\}
\]

where

\[
r^X_{n,m}(s) \equiv r_x(F^X_{n,m}).
\]

Note the following properties of complexities \( m^X_s \) and generalized Bernoulli components \( r^X_{n,m}(s) \).

Proposition 2:

i) The following inequalities hold:

\[
r^X_{n,m+1}(2s) \leq r^X_{n,m}(s) \leq r^X_{n,m}(s + 1) \tag{10}
\]

and

\[
m^X_s \leq m^X_{s+1}. \tag{11}
\]

ii) If for some \( s \) we have \( m^X_{s+1} > m^X_s \), then \( m^X_{s+1} = m^X_s + 1 \), and \( s_m = s \), where \( m = m^X_s \) is the maximal number of \( m \)-blocks for which

\[
\lim_{n \to \infty} r^X_{n,m}(s) = 0.
\]

iii) If \( s_m \) is defined as in ii), then

\[
s_{m+1} \geq 2s_m.
\]

Proof: Let

\[
\pi : \{0, 1\}^{m+1} \to \{0, 1\}^m
\]

be defined for \( u \in \{0, 1\}^m \) by \( \pi(u) = \pi(0u) = u \). Note the following direct relations:

\[
F^X_{n,m} = F^X_{n,m+1} \circ \pi^{-1}
\]

\[
B^m_{p} = B^{m+1}_p \circ \pi^{-1}
\]

\[
(F \circ G) \circ \pi^{-1} = (F \circ \pi^{-1}) \circ (G \circ \pi^{-1})
\]

where \( p \in \mathbb{R} \), \( F \) and \( G \) are additive functions on subsets of \( \{0, 1\}^m \).

To prove (10), it is enough to prove the middle inequality since the others follow directly from iv) of Proposition 1. Without loss of generality, we can assume that \( r^X_{n,m}(s) < 1/2 \). Let \( p = r^X_{n,m+1}(2s) \), and \( S \subseteq \{0, 1\}^m \) has cardinality \( s \). Then

\[
F^X_{n,m} \ast B^{m+1}_{p/(1-2p)}(S)
\]

\[
= (F^X_{n,m+1} \circ \pi^{-1}) \ast (B^{m+1}_{p/(1-2p)} \circ \pi^{-1})(S)
\]

\[
= (F^X_{n,m} \ast B^{m+1}_{p/(1-2p)} \circ \pi^{-1})(S)
\]

and the last expression \( (s) \) is nonnegative since \( \pi^{-1} S \) has cardinality \( 2s \). Consequently, \( r^X_{n,m}(s) \geq p \).

Now, if \( m = m^X_{s+1} - 1 \) then \( r^X_{n,m}(s) \) does not converge to zero. This, by (10), gives that also \( r^X_{n,m}(s + 1) \) does not converge to zero which proves \( m^X_{s+1} \geq m^X_s \) thus finishing the proof of i).

To prove ii), note that if \( r^X_{n,m}(s) \) converges to zero and thus, again by (10), \( r^X_{n,m + 1}(2s) \) converges to zero, too. Thus

\[
m^X_{s+1} \leq m^X_s + 1
\]

and since \( m^X_s \) is nonincreasing in \( s \), we have

\[
m^X_{s+1} = m^X_s + 1
\]

which gives the first part of ii).

Moreover, if \( m^X_{s+1} = m^X_s + 1 \) and for \( m = m^X_s \) we have

\[
\lim_{n \to \infty} r^X_{n,m}(s + 1) = 0
\]

then

\[
m \geq m^X_{s+1} \geq m^X_s + 1
\]

which leads to contradiction.

Finally, if

\[
\lim_{n \to \infty} r^X_{n,m}(s_m) = 0
\]

then, by i), also

\[
\lim_{n \to \infty} r^X_{n,m+1}(2s_m) = 0
\]

thus \( s_{m+1} \geq 2s_m \) which concludes the whole proof.

D. Application to Consistent Estimating the Distortion

Now, for a pair \((m, s)\) such that \( s \leq 2^m \), we define \( \hat{\mu}_n(m, s) = \bar{r}^X_{n,m}(s) \). We have the following consistency result.

Theorem 1: For each \((s, m)\) such that \( m \geq m^X_s \) and under Assumption 1 the estimator \( \hat{\mu}_n(m, s) \) consistently estimates the distortion \( p \). The consistency does not hold if \( m \) is strictly less than \( m^X_s \).

Proof: The basic idea is based on the proof of consistency presented in [8]. First, recall that with probability one

\[
\lim_{n \to \infty} \left| \frac{F^X_{n,m} - F^X_{n,m} \ast B^m_p}{n} \right| = 0 \tag{12}
\]

and that \( E F^X_{n,m} = F^X_{n,m} \ast B^m_p \).

By v) of Proposition 1

\[
|\hat{\mu}_n(m, s) - r_x(F^X_{n,m} \ast B^m_p)\| = \left\| F^X_{n,m} - F^X_{n,m} \ast B^m_p \right\|_{1/m}
\]

and thus consistency of \( \hat{\mu}_n(m, s) \) is equivalent to convergence of \( r_x(F^X_{n,m} \ast B^m_p) \) to \( p \). But, by i) of Proposition 1

\[
r_x(F^X_{n,m} \ast B^m_p) = r^X_{n,m}(s)(1 - 2p) + p.
\]

Thus consistency of the estimators is equivalent to convergence of \( r^X_{n,m}(s) \) to zero with \( n \) which is true if and only if \( m \geq m^X_s \).

Let us consider some aspects of the last two results. Fig. 5 presents a complementary graphical illustration to our discussion.

Having an infinite output \( Z \), for each pair \((m, s)\), \( s \leq 2^m \), we can compute the sequence of generalized Bernoulli components \( \bar{F}_n(m, s) \). By the above result, there is an infinite number of pairs for which this sequence estimates the distortion consistently. We can also infer that those pairs occupy a region which has a typical form shown in Fig. 5. Our theoretical and numerical studies indicate that moving down or right in this region of consistent estimation results in more biased estimates with bigger variances. Thus for the best estimates one should look to the boundaries of this region with possibly "best" choices as presented on Fig. 5. At present, we do not know any selection criterion for these "best" estimators but probably it should be based on an \( m \) which reveals clear distinction between \( m \)-blocks which are and which are not rare. It is also worth noting that, by i) of Proposition 2, we cannot reduce the negative bias just by increasing \( m \) by one and, thereby, doubling the number of rare states. Indeed, we always have \( r^X_{n,m}(s) \geq r^X_{n,m+1}(2s) \). This can be seen as well in Fig. 5, where a vertical part of the boundary between consistent and nonconsistent pairs \((m, s)\) has length twice the sum of all the vertical lengths for parts corresponding to smaller values of block size \( m \).
E. Reduction of the Negative Bias

We have also carried out Monte Carlo simulations for the example with the Aeneid input. To decide on the number \( s \) we have analyzed the frequencies \( F_{n,m}^X \), depicted also in Fig. 2. (In practice, since the input is not available, we recommend preliminary analysis based on the empirical distribution of the output deconvoluted by its largest Bernoulli component.) In Fig. 6, we present density estimates of the distributions of \( \hat{p}_{n,s} \) for selected values of \( s \) in this range, based on Monte Carlo simulation of 250 samples for each estimate (the true value of the parameter is \( p = 0.1 \)). The picture clearly illustrates the superiority of the modified approach to the estimating distortion. Suitable choice of \( s \) not only significantly reduces the bias but also results in a smaller value of variance. For \( s = 15 \), for example, we observe a very small positive bias and the variance even smaller than for the original estimate for the block size \( m = 8 \) (cf. Fig. 4). This suggests that the appropriate choice of \( s \) can be more important for the quality of estimation than the choice of block size, which is so troublesome anyway. This is a very important feature of our modified method, especially if we take into account the slow asymptotics of the estimates of the distortion (see Section IV). In particular, slight overestimation of the complexity leads to a larger set of forbidden states which, if their number is accurately estimated, can reduce both the bias and the variance of the modified estimate.

IV. ASYMPTOTICS AND MODIFIED ESTIMATES OF THE DISTORTION

A. Asymptotics of Empirical Distribution

Below, we examine the asymptotic behavior of \( \hat{p}_{n}(m, s) \). The crucial point for establishing the strong consistency of these estimators was to use the fact that with probability one

\[
\lim_{n \to \infty} \left[ F_{n,m}^Z - E F_{n,m}^Z \right] = 0.
\]

On account of the Law of Iterated Logarithm, the actual rate of convergence is, in fact, \( O(\sqrt{\log \log n/n}) \). The next theorem establishes a somewhat slower rate of convergence but uniformly in \( m, 1 \leq m \leq m_n \), where \( m_n \to \infty \) with \( n \). This uniformity is essential for consistency of the estimator which is proposed at the end of this section.

**Theorem 2:** Let \( \alpha \in (0, 1/2) \) and \( (m_n) \) be a sequence of positive integers such that \( m_n = O(\log n) \) as \( n \to \infty \). Then with probability one

\[
\lim_{n \to \infty} \max_{1 \leq m \leq m_n} \left| F_{n,m}^{Z} - E F_{n,m}^{Z} \right| \leq 0.
\]

The notationally troublesome although standard proof of the result is given in the Appendix.

B. Asymptotics of \( gibc \)

Having the asymptotic behavior of \( F_{n,m}^{Z} \), we can relatively easily derive it from the asymptotics for \( \hat{p}_{n}(m, s) \).

**Theorem 3:** For each \( (m, s) \), \( s \leq 2^m \), and for each \( \alpha \in (0, 1/2) \) with probability one

\[
\left| \hat{p}_{n,m}(s) - p \right| \leq o(n^{-\alpha/m}) + (1 - 2p) r_{n,m}(s). \tag{13}
\]

**Proof:** Using the same relations as in Theorem 1, we obtain

\[
\begin{align*}
\left| \hat{p}_{n,m}(s) - p \right| & \leq \left| F_{n,m}^{Z} - E F_{n,m}^{Z} \right|^{1/m} + (1 - 2p) r_{n,m}(s) \\
& \leq o(n^{-\alpha/m}) + (1 - 2p) r_{n,m}(s).
\end{align*}
\]

**Remark 2:** To establish the asymptotics of \( \hat{p}_{n}(m, s) - p \) by this result, we need the rate of the convergence \( r_{n,m}(s) \). It is evident that some requirement on the asymptotic behavior of \( F_{n,m}^{X} \) has to be imposed since \( r_{n,m}(s) \) is due to the nonrandom component of \( \hat{p}_{n}(m, s) \) related to the input \( X \). This component inevitable must influence the rate of convergence.

To illustrate the asymptotics of the estimator of the distortion, we have computed values of the estimate \( \hat{p}_{n,m} \) for fixed \( m = 9 \) in the case of the Aeneid input, \( n \) varying from 22017 (1% of data) to 2201785, and for three various parameters of distortion: \( p = 0.1 \), \( p = 0.03 \), and \( p = 0.005 \). The results are presented on Fig. 7. As we have already mentioned, \( r_{n,0} = 0 \) for any available \( n \). Thus by Theorem 3, with probability one

\[
\left| \hat{p}_{n,m} - p \right| \leq o(n^{-\alpha/m})
\]

for each \( \alpha \in (0, 1/2) \). It should be realized that this asymptotic rate is rather slow, as can be seen from Fig. 7, where we drew, with a dashed line, a curve with asymptotics \( 0.1 - an^{-\beta} \), with \( \beta = 1/20 \).
Fig. 6. Comparison of the original estimate with the modified ones for various values of $s$ and for $m = 9$ ($p = 0.01$).

Fig. 7. Asymptotics of the estimator.
which corresponds to $\alpha = 9/20$, and some positive irrelevant scale $a$. Note the negative bias, observed for all three values of $p$.

C. Distortion Estimation Without Knowing Complexity

In order to use glb $\hat{p}_n(m, s)$ as estimators of the distortion, we have first to decide on a pair $(m, s)$ for which consistency holds with possibly “best” other properties of the corresponding estimator (small variance and reduced bias). However, the consistency condition $m \geq m^X_n$ cannot be verified directly from the output as the complexity $m^X_n$ is a characteristics of the unknown input $X$. Below we introduce a consistent estimator of the distortion which can be determined with no prior knowledge of $m^X_n$ needed.

Suppose that $s$ is fixed. The simplest estimator which does not require $m^X_n$ can be defined as $\bar{p}_n(k_n, s)$, where a sequence $k_n$, with $n \to \infty$, satisfies

$$k_n \to \infty \quad \frac{\log n}{k_n} \to \infty. \quad (14)$$

Indeed, then we have

$$|\bar{p}_n(k_n, s) - p| \leq |\bar{p}_n(k_n, s) - r_x(E(F^Z_n, k_n))| \leq \left(\frac{F^Z_n}{F^Z_n, k_n}\right)^{1/k_n} + (1 - 2p)r_x(m, X(s))$$

with both terms converging to zero: the first by Theorem 2, the second by Assumption 1 and the definition of $m^X_n$. However, this estimator, although consistent, is obtained from generalized Bernoulli components by considering pairs $(k_n, s)$ which are moving up away from the boundary which correspond to estimators with better properties (see Fig. 2). Therefore, we propose a different estimator that is defined by pairs $(\tilde{m}_n, s)$ which stay close to this boundary.

Suppose that for $m > \log_2 s$ we consider $\tilde{p}_n(m - 1, s) - \tilde{p}_n(m, s)$. By Theorem 2, these differences should be uniformly small for $m \in \{m^X_n + 1, \ldots, k_n\}$, where $k_n$ satisfies (14). On the other hand,

$$\tilde{p}_n(m^X_n - 1, s) - \tilde{p}_n(m^X_n, s) \geq o(1) + (1 - 2p)r_x(m^X_n - 1, s)$$

where the right-hand side can be relatively large ($r_x(m^X_n - 1, s)$ does not converge to zero by the definition of $m^X_n$). Our method seeks to exploit these properties of $\tilde{p}_n(m - 1, s) - \tilde{p}_n(m, s)$ to define $\tilde{m}_n$, which stays possibly closely to $m^X_n$. Here is a formal description of the method.

Let $\alpha \in (0, 1/2]$ be fixed and integers $k_n$ satisfy (14) so $n^{1/k_n} \to \infty$. Define

$$\tilde{m}_n = \begin{cases} \max\{m > \log_2 s : m \leq k_n, \tilde{p}_n(m - 1, s) - \tilde{p}_n(m, s) > \tilde{p}_n(k_n, n^{1/k_n})\} \\
\{k_n : \text{if the above maximum is over the empty set} \}
\end{cases}$$

and

$$\tilde{p}_n(s) = \tilde{p}_n(\tilde{m}_n, s).$$

Remark 3: Because of slow asymptotics, in a concrete application ($k_n$) should be carefully chosen with relation to size of available data as well as predicted range of the complexity. Two examples of a possible choice are $k_n = [\log \log n]$ or $k_n = \lfloor \log n \rfloor$, with $\beta < 1$.

As an application of Theorem 2 we obtain the following consistency result for $(\tilde{m}_n, \tilde{p}_n(s))$.

Theorem 4: Under Assumption 1, with probability one $\tilde{p}_n(s)$ is a strongly consistent estimator of the distortion $p$ and

$$\lim_{n \to \infty} \sup \tilde{m}_n = m^X_n. \quad (15)$$

Proof: By Theorem 2 with probability one and for sufficiently large $n$, we have

$$\max_{m \leq m^X_n} (\tilde{p}_n(m - 1, s) - \tilde{p}_n(m, s)) \leq \frac{2}{m^X_n} \max_{m \leq m^X_n} |\tilde{p}_n(m - 1, s) - \tilde{p}_n(m, s)| \leq \frac{2}{m^X_n} \left(\max_{m \leq m^X_n} |F^Z_n - FE^Z_n|\right)^{1/k_n} \leq o(n^{-\alpha/k_n}).$$

Moreover, as it has been already shown, $\tilde{p}_n(k_n, s)$ is convergent to positive $p$. Thus for sufficiently large $n$, we also have

$$\max_{m \leq m^X_n} (\tilde{p}_n(m - 1, s) - \tilde{p}_n(m, s)) \leq \tilde{p}_n(k_n, s) n^{-\alpha/k_n}$$

which implies that $\lim \inf \tilde{m}_n \leq m^X_n$. But 4

$$\tilde{p}_n(s) \leq \tilde{p}_n(m^X_n, s) + \sum_{m = m^X_n + 1} \tilde{p}_n(m - 1, s) - \tilde{p}_n(m, s) \leq \tilde{p}_n(m^X_n, s) + m^X_n n^{-\alpha/k_n} \tilde{p}_n(k_n, s)$$

and thus

$$\lim \sup \tilde{p}_n(s) \leq \lim \inf \tilde{p}_n(m^X_n, s) = p. \quad \square$$

Our method of estimating the distortion is illustrated in Fig. 8. In the procedure we compare values of differences $\tilde{p}_n(m, s) - \tilde{p}_n(m - 1, s)$ to the value $\tilde{p}_n(k_n, s) n^{-\alpha/k_n}$, with $m$ decreasing and starting from $m = k_n$. For our example, we have chosen $s = 1$, $k_n = [\lfloor \log n \rfloor]$ with $\beta = 0.9$ and $\alpha = 0.49$. In Fig. 8, we see the graphs of the differences. The values $p_n, 1 < n < k_n$ are illustrated by the horizontal dashed lines. Thus $\tilde{m}_n$ is equal to 5 for $p = 0.1$, to 8 for $p = 0.03$, and to 11 for $p = 0.005$. To illustrate the improvement resulting from the increase of data size, we have also included in Fig. 8 the graphs of the differences computed for 10% of the data.

Note that the asymptotics of $k_n$, used in the estimation procedure is very slow ($k_n = [\lfloor \log n \rfloor]$ with $\beta = 0.9$, yields, for $n = 2, 201, 785$, the value 11). We consider only $p_n, m$ for below $k_n$, thus if the complexity $m_0$ is large, to achieve $k_n \geq m_0$ we need a huge number of observations. On the other hand, the computation time for finding the largest Bernoulli components increases rapidly with $m$ ($F_m + B_{m-1}$ is defined over $2^m$ points). Thus our methodology has unavoidable practical limitations with respect to possible values of the complexity. Nevertheless, in many cases, the value of the complexity is in the range of computational power of personal computers. For example, a black-and-white image has complexity not exceeding 11, which is within the range of feasible computations.

APPENDIX

A. Necessity of the Basic Assumption

In [8], the problem of characterizing the sets $\Lambda \subseteq \{0, 1\}^N$ was posed for which there exists a consistent sequence of estimators $\tilde{\pi}_n$, of $p$, for each input $X$ in $\Lambda$. One such $\Lambda$ is the set of $X$ that satisfy the basic assumption. We denote this set as $\Lambda_0$. Some restrictions on $X$ are needed, i.e., $\Lambda$ cannot be equal to $\{0, 1\}^N$. Heuristic arguments supporting this statement were given in [8]. A more formal argument is provided by the following stronger result.

Proposition 3: If a set $\Lambda$ contains $X_0 \subseteq \Lambda_4$ with $X_0 \in \Lambda$ and $\Lambda_4 \subseteq \{0, 1\}^N$ of positive probability with respect to a product
measure $Q^N$ on $\{0, 1\}^N$ with $Q(\{1\}) \in (0, 1/2)$, then there does not exist $\hat{p}_n$ which consistently estimates the distortion $p$ for each $p \in (0, 1/2)$ and for each $X \in \Lambda$.

Proposition 3 does not imply that $\Lambda_0$ is the maximal set for which consistency result is possible. (Here maximal means that there does not exist a larger one.) Nevertheless, it gives some insight into restrictiveness of Assumption 1. For clarity of arguments, we consider the class of estimators which are consistent at least in the case when the input consists of zeros, and thus in Proposition 3, we take $X_0 = 0$.

Now, suppose that consistent estimation of the distortion is possible for some set $\Lambda$ which is larger than $\Lambda_0$. By Proposition 3, $\Lambda$ has to be of measure zero for each product measure $Q^N$. At the same time, for each $w \in \Lambda \setminus \Lambda_0$ and each $m$, the largest Bernoulli components of $F^w_{n,m}$ are separated asymptotically from zero. On the other hand, if a subset has positive measure with respect to some $Q^N$ (thus not allowing consistent estimation), then, for almost all its elements $w$, the largest Bernoulli components of $F^w_{n,m}$ are separated asymptotically from zero uniformly with respect to $m$. Thus in this sense, there is not much “space” to enlarge $\Lambda_0$ to some, say, $\Lambda$ for which a consistent estimation procedure is still possible.

Finding a complete solution remains an interesting and open theoretical problem. Possibly an answer could be found in topological, rather than measure-theoretical terms, but this possibility will not be investigated in the present correspondence.

Proposition 3 might suggest that sets for which some consistent estimation is possible cannot be large in measure-theoretical sense. The following example demonstrates that this is not necessarily true if we allow some structural restrictions on the input.

**Example 1:** Let $q$ be a fixed number in $(0, 1)$ and define

$$\hat{p}_n = \frac{|Z|_1^2/n - q}{1 - 2q}.$$

The set of input sequences for which this estimator is consistent have measure one with respect to $Q^N$. Indeed, if the input will be a realization of the Bernoulli sequence with the parameter $q$, then, by the Law of Large Number, for almost all such realizations, $\frac{|Z|_1^2}{n}$ converges to $p + q - 2pq$. This illustrates how the problem of estimation can be approached when a distributional assumption is placed on the input (here the input is a result of the Bernoulli noise with the known parameter $q$). A more general Bayesian approach was explored in [4].

**B. Proofs**

Proofs of Theorem 2 and Proposition 3 are given below.

**Proof of Theorem 2:** Let first list some useful properties of the empirical distribution $F^Z_{n,m}$.

Note that

$$E\delta_{(X \oplus Y)^{i+m}} = \delta_{X^{i+m}} B^m_p$$

and

$$\text{Var} \delta_{(X \oplus Y)^{i+m}} = \left(\delta_{X^{i+m}} B^m_p\right) \left(1 - \delta_{X^{i+m}} B^m_p\right)$$

where multiplication is understood in a component-wise sense. Consequently, we have

$$EF^Z_{n,m} = \left(\frac{1}{n} \sum_{i=0}^n \delta_{X^{i+m}}\right) * B^m_p = F^X_{n,m} * B^m_p$$

and, for $p \in (0, 1/2]$, we have

$$0 < p^m(1 - (1 - p)^m) \leq \text{Var} \delta_{(X \oplus Y)^{i+m}} \leq (1 - p)^m(1 - p^m) < 1.$$

We exploit limit theorems for sums of independent random variables. However, $F^Z_{n,m}$ is a sum of random variables which are dependent unless their indices differ by $m$ or more. The following
largely circumvents difficulties by segregating \( Y_{r_1/m} \), \( Y_{r_2/m} \) when \( |i - j| \geq m \). If \( N = [n/m] \), i.e., \( Nm \leq n < (N + 1)m \), then
\[
\left\| (F_{m,n} - EF_{m,n}) - (F_{m,N,m} - EF_{m,N,m}) \right\|_\infty \leq \frac{2m}{n} \tag{16}
\]
and
\[
F_{m,N,m} - EF_{m,N,m} = \frac{1}{m} \sum_{k=0}^{m-1} \left( \frac{N}{N-1} \sum_{i=0}^{N-1} \left( \delta_{\{X_{i+k}\}} k_{r_1} + \frac{m}{m} + \frac{m}{m} - \delta_{\{X_{i+k}\}} k_{r_2} + \frac{m}{m} \right) \right)
\]
where the last equality should serve as a definition of random variables \( \xi_{k,i} \). This representation, for fixed \( k \), random variables \( \xi_{k,i}, i = 0, \ldots, N - 1 \) are independent, with values in \( \mathbb{R}^{m} \) with mean zero, and finite variance having coordinates in the interval \([p^m(1 - (1 - p)^m), (1 - p)^m, (1 - p)^m] \).

With these facts in place, the proof of the theorem becomes more evident. By (16), we have
\[
n^{\alpha} \max_{1 \leq m \leq m_n} \left\| F_{m,n} - EF_{m,n} \right\|_\infty \leq n^{\alpha} \max_{1 \leq m \leq m_n} \left\| F_{m,N,m} - EF_{m,N,m} \right\|_\infty + 2m_n n^{-1}
\]
where \( N = [n/m] \). Thus it is enough to prove that the first summand on the right-hand side tends to zero almost surely. We follow the idea of the proof of the Glivenko–Cantelli theorem (see, for example, [2]).

Proof of Proposition 3: Assume, to the contrary, that there exists a sequence \( \hat{p}_n \) of estimates of \( p \) with appropriate properties, i.e., for each \( X \in \Lambda \) and for each Bernoulli sequence \( Y \) with a parameter \( p \) with probability one \( \lim_{n \to \infty} \hat{p}_n(X \oplus Y) = p \).

Let \( q = Q(\{1\}) \) and let us fix \( p \in (q, 1/2) \) and take independent Bernoulli sequences \( Y_1, Y_2 \) with parameters \( q \) and \( (p - q)/(1 - 2q) \), respectively. Since \( Y_1 \oplus Y_2 \) is a Bernoulli sequence with parameter \( p \), we have with probability one
\[
\lim_{n \to \infty} \hat{p}_n(X_0 \oplus Y_1 + Y_2) = p
\]
and, since \( X_0 \oplus \Lambda_+ \subseteq \Lambda \), for each \( y \in \Lambda_+ \) with probability one
\[
\lim_{n \to \infty} \hat{p}_n(X_0 \oplus Y_1 + Y_2) = (p - q)/(1 - 2q).
\]
Define \( \Omega_0 \) as a set of probability one on which (18) holds and \( \Omega_\Psi = \{ \omega \in \Omega_0 : y_1(\omega) = y \} \). Note that the above convergence conditions can hold only if for each \( y \in \Lambda_+ \) we have \( P(\Omega_\Psi) = 0 \). But then noting that the distribution of \( Y_1 \) is equal to \( Q^N \) and using Fubini theorem, we are led to the following contradiction:
\[
1 = P(\Omega_0) = \int P(\Omega_\Psi) dQ^N(y) \leq \int_{\Lambda_+} P(\Omega_\Psi) dQ^N(y) + Q^N(\Lambda_+) = 1 - Q^N(\Lambda_+) < 1. \quad \square
\]

References