Finite Fields and Their Applications 47 (2017) 46–63

Contents lists available at ScienceDirect

Finite Fields and Their Applications

www.elsevier.com/locate/ffa

Code generator matrices as RNG conditioners

A. Tomasi *, A. Meneghetti **, M. Sala **

Department of Mathematics, University of Trento, Via Sommarive 14, 38123 Povo, Italy

ARTICLE INFO

Article history:
Received 6 March 2017
Received in revised form 3 May 2017
Accepted 4 May 2017
Available online xxxx
Communicated by Arne Winterhof

MSC:
65C10
60B99
11T71
94B99

Keywords:
Random number generator
conditioning

ABSTRACT

We quantify precisely the distribution of the output of a binary random number generator (RNG) after conditioning with a binary linear code generator matrix by showing the connection between the Walsh spectrum of the resulting random variable and the weight distribution of the code. Previously known bounds on the performance of linear binary codes as entropy extractors can be derived by considering generator matrices as a selector of a subset of that spectrum. We also extend this framework to the case of non-binary codes.

1. Introduction

Our objective is to precisely quantify the result of applying any one specified code generator matrix a single time as conditioning function to the output of an entropy source. We follow the recommendations set out by NIST [1] for the precise meaning to be given to these terms.

* Principal corresponding author.
** Corresponding author.
E-mail addresses: twin.ion.engine@gmail.com (A. Tomasi), almenegh@gmail.com (A. Meneghetti), maxsalacodes@gmail.com (M. Sala).

http://dx.doi.org/10.1016/j.ffa.2017.05.005

Linear transformations based on codes have previously been applied to sources of entropy producing output that can be treated as independent but biased bit sequences; bounds on the statistical distance of such output from the uniform distribution have been shown in [2–4]. The application of these functions is generally presented as part of the framework of randomness extractors, as summarised for instance in [5], in the sense that random matrices are chosen with specific properties, such as minimum distance of the code, or an approximate distribution of the weights. The performance of these functions is usually presented in the form of bounds on the statistical (total variation) distance of the resulting conditioned output from the uniform distribution. The present work extends and complements the known results by quantifying precisely the statistical distribution of the output after conditioning with a specified generator matrix by showing the connection between the probability mass function of the resulting random variable and the weight distribution of the code; the known bounds can then be derived as special cases.

We treat binary streams in groups of k bits as discrete random variables X, in the sense that the number of possible outcomes is finite and the variables admit a discrete probability mass function $\mu_X(j) = \mathbb{P}(X = x_j)$; moreover, we begin by considering these variables to take values in a finite field $\mathbb{F}_p$ or a vector space $(\mathbb{F}_p)^k$, with particular regard to the special case of binary variables, $p = 2$. In Section 2 we show the connection between the Walsh spectrum of X and the bias of individual bits $X(j)$, and use this in Section 3 to show how previously known bounds can be derived by considering generator matrices as a selector of a subset of that spectrum. We then extend this framework to the case of output in non-binary finite fields by use of the Fourier transform in Section 5.

2. Total variation distance and the Walsh–Hadamard transform

We show in the following one way in which the Walsh–Hadamard transform may be used to bound the total variation distance of binary random variables with a known probability mass function. This may seem an unnecessary exercise since the TVD can simply be computed exactly from this knowledge, but aside from revealing some interesting structure to the calculation it will become more explicitly useful in the following section.

Consider a random vector $Y \in (\mathbb{F}_2)^k$ with probability mass function

$$\mu_Y \in \mathbb{R}^{2^k},$$

$$\mu_Y(j) = \mathbb{P}(Y = \mathbf{j})$$

where in writing j and $\mathbf{j}$ we use the binary representation of integers $a \in \mathbb{Z}_{2^k}$ as vectors

$$\mathbf{a} = (a_j)_{j=0,\cdots,k-1} \in (\mathbb{F}_2)^k .$$

The a-th order Walsh function evaluated at b is

$$h_a(b) = (-1)^{\mathbf{a}\cdot\mathbf{b}} \tag{1}$$

with $\cdot$ the dot product on $(\mathbb{F}_2)^k$. The jth Walsh characteristic function of Y as defined in [6] is

$$\begin{aligned} \chi_j(Y) &= \sum_{a=0}^{2^k-1} h_j(a)\mu_Y(a) \qquad (2)\\ &= \mathbb{E}[h_j(Y)]. \qquad (3) \end{aligned}$$

Note that the dot product of two binary vectors $\mathbf{b}\cdot\mathbf{v}$ is the bitwise sum, i.e. the linear combination, of those elements $\mathbf{v}(i)$ that correspond to the non-zero entries of $\mathbf{b}$; therefore, the random variable

$$h_b(Y) = (-1)^{\mathbf{b}\cdot Y}$$

will take value -1 if the linear combination of the selected elements of Y is equal to one, and 1 otherwise. The sum of the selected elements is itself a random variable, $B = \mathbf{b}\cdot Y$ following the Bernoulli distribution with probability $\mu_B(1)$ of being equal to 1; it follows that

$$h_b(Y) = 1 - 2B,$$

and hence we can conclude that

$$\begin{aligned} \chi_b(Y) &= \mathbb{E}[h_b(Y)] \\ &= 1 - 2\mu_B(1). \end{aligned}$$

We note now that the bias of a Bernoulli random variable $B \in \mathbb{F}_2$ is commonly defined as

$$\begin{aligned} \frac{\varepsilon_B}{2} &= \frac{1}{2}\left|\mathbb{P}(B=1) - \mathbb{P}(B=0)\right| \\ &= \frac{1}{2}\left|2\mathbb{P}(B=1) - 1\right| \\ &= \frac{1}{2}\left|2\mathbb{E}[B] - 1\right|, \end{aligned}$$

and we observe that the Walsh characteristic of $\mathbf{b}\cdot Y$ leads to the bias of the bth linear combination of the elements of Y via the relation

$$|\chi_b(Y)| = \varepsilon_{\mathbf{b}\cdot Y}\,. \tag{4}$$

In particular, the combinations corresponding to exact powers of two, $b = 2^j$, lead to the bias of each individual element of Y; and the zeroth Walsh characteristic, corresponding to $b = 0$, will be equal to 1 in all entries, in all cases.

The set $\{ h_i \}$ is known to correspond to the rows of a Hadamard matrix H of size 2^k; the set of all Walsh characteristics of Y can thus be written compactly in matrix notation as

$$\chi(Y) = H\mu_Y \, . \tag{5}$$

As a matter of notation, for a uniformly distributed random variable $U \in (\mathbb{F}_2)^k$ we have

$$\mu_U = \frac{\mathbf{1}}{2^k}$$
$$\chi(U) = \mathbb{I}_{2^k}(\cdot, 1)$$

with $\mathbf{1}$ a column vector of ones and $\mathbb{I}_{2^k}(\cdot, 1)$ the first column of the identity matrix of size 2^k. We may use this to estimate the total variation distance of Y from uniform as follows.

Theorem 1. *The total variation distance of a random vector $Y \in (\mathbb{F}_2)^k$ from uniform U,*

$$\begin{aligned} \mathrm{TVD}(Y,\mathcal{U}) &= \frac{1}{2} \left\| \mu_Y - \mu_U \right\|_1 \\ &= \frac{1}{2} \delta_Y \end{aligned}$$

is bounded by the sum of the bias of all non-trivial linear combinations of the output bits,

$$\delta_Y \leq \sum_{\mathbf{b} \in (\mathbb{F}_2)^k \setminus \mathbf{0}} \varepsilon_{\mathbf{b} \cdot Y} \, .$$

Proof.

$$\begin{aligned} \left\| \mu_Y - \mu_U \right\|_2 &= \left\| \frac{H^T H}{2^k} (\mu_Y - \mu_U) \right\|_2 && (6) \\ &= \frac{1}{2^{k/2}} \left\| \frac{H^T}{2^{k/2}} (\chi(Y) - \chi(U)) \right\|_2 && (7) \\ &\leq \left\| \chi(Y) - \chi(U) \right\|_1 && (8) \\ &= \sum_{\mathbf{b} \in (\mathbb{F}_2)^k \setminus \mathbf{0}} \varepsilon_{\mathbf{b} \cdot Y} \, . && (9) \end{aligned}$$

Here H^T is the transpose of the Hadamard matrix H. Eq. (6) follows from $H^T/\sqrt{2^k}$ being the unitary inverse Hadamard transform; Eq. (7) uses the definition of $\chi(Y)$ in Eq. (5); and lastly, the bound in Eq. (8) stems from the ℓ_1 bound on q-dimensional vector spaces, $\| \cdot \|_1 \leq q^{1/2} \| \cdot \|_2$. □

Corollary 1. *If the bits $Y(j) \in \mathbb{F}_2$ are i.i.d. with known bias ε_y, then*

$$\begin{aligned}\delta_Y &\leq \sum_{l=1}^{k} A_l \varepsilon_y^l \\ &= \sum_{l=1}^{k} \binom{k}{l} \varepsilon_y^l\end{aligned}$$

where A_l is the number of $\mathbf{b} \in (\mathbb{F}_2)^k$ with Hamming weight $w(\mathbf{b}) = l$.

3. W–H bound on binary generator matrices as extractors

We now consider the previous bound as applied to random variables $Y = GX$, with $X \in (\mathbb{F}_2)^n$ a sequence of n Bernoulli random variables with known probability mass $\mathbb{P}(X = b) = \mu_X(b)$ and identical bias $\varepsilon_X = |1 - 2\mathbb{P}(X(i) = 1)|$ for each bit, and $G \in (\mathbb{F}_2)^{k\times n}$ the generator matrix of an (n, k, d) linear code C with weight distribution $\{ A_l \}$; in other words, C is a subspace of $(\mathbb{F}_2)^n$ and A_l is the number of $\mathbf{c} \in C$ with Hamming weight $w(\mathbf{c}) = l$.

The definition of Walsh characteristic functions as expected values in Eq. (3) directly leads to

$$\begin{aligned}\chi_b(GX) &= \mathbb{E}[h_b(GX)] \\ &= \sum_{x=0}^{2^n-1} (-1)^{\mathbf{b}\cdot G\mathbf{x}} \mu_X(x).\end{aligned}$$

We note here that the dot product in the Walsh function can equivalently be expressed using the transpose $\mathbf{b}^T$ as

$$\mathbf{b} \cdot G\mathbf{x} = \mathbf{b}^T G\mathbf{x}\,,$$

and in particular the product

$$\mathbf{c}^T = \mathbf{b}^T G$$

is a linear combination of the rows of G: since G is the generator matrix of a linear code C the rows of G form a basis of C, and hence any linear combination of them is again a word $\mathbf{c} \in C$. Consequently, just as the Walsh characteristic led to a measure of bias in Eq. (4), we can conclude that

$$|\chi_b(Y)| = \varepsilon_{\mathbf{c}\cdot X}\,. \tag{10}$$

In other words, the bias of the bth element of Y is equal to the bias of a linear combination of $w(\mathbf{c})$ bits of X (compare to Eq. (4)). This leads directly to the following bound.

Theorem 2. *Let $Y = GX$, with $X \in (\mathbb{F}_2)^n$ a sequence of n independent but not necessarily identically distributed Bernoulli random variables, and $G \in (\mathbb{F}_2)^{k\times n}$ the generator matrix of an (n, k, d) linear code C. The total variation distance of the random variable $Y \in (\mathbb{F}_2)^k$ from uniform,*

$$\mathrm{TVD}(Y,\mathcal{U}) = \frac{\delta_Y}{2}$$

is bounded by the sum of the bias of all linear combinations of the bits in X defined by the codewords of C, in the following measure:

$$\begin{aligned}\delta_Y &\leq \sum_{\mathbf{b}\in(\mathbb{F}_2)^k\setminus\mathbf{0}} \varepsilon_{\mathbf{b}^T G\cdot X} \\ &= \sum_{\mathbf{c}\in C\setminus\mathbf{0}} \varepsilon_{\mathbf{c}\cdot X}\,.\end{aligned}$$

Corollary 2. *If the bits $X(j) \in \mathbb{F}_2$ are i.i.d. with known bias ε_x, then*

$$\delta_Y \leq \sum_{l=d}^{n} A_l \varepsilon_x^l\,,$$

with $\{ A_l \}$ the weight distribution of C.

Note that Corollary 1 is closely related to Corollary 2 if we consider that in this context the $\{ A_l \}$ in the former correspond precisely to the weight distribution of the trivial code given by the message space itself, $(\mathbb{F}_2)^k$. A particular case of Corollary 2 for strictly binomial $\{ A_l \}$ was proved in [4], Theorem 6. We can thus recover the following known bound (see [2], Theorem 1):

Corollary 3. *Considering only the minimum distance d rather than the full weight distribution,*

$$\delta_Y \leq (2^k - 1)\varepsilon_x^d\,.$$

4. Total variation distance and the Fourier transform

The Hadamard transform is a special case of the Fourier transform constructed with primitive 2-nd root of unity $p = 2$, $\omega_p = -1$, and the Hadamard matrix of size 2^k is constructed by the Kronecker product $H_{2^k} = H_2 \otimes H_{2^{k-1}}$, so the binary case considered in Section 3 can be seen as a special case. Employing the Fourier transform is natural in this setting and closely follows well-established techniques for the sum of continuous random variables, which have their own convolution theorem and proofs of convergence to a limiting distribution.

Given an integer $a \in \mathbb{Z}_{p^k}$, we denote its p-ary representation by

$$\mathbf{a} = (a_j)_{j=0,\cdots,k-1} \in (\mathbb{F}_p)^k \ .$$

We shall use this notation interchangeably in the following as a natural indexing of the elements of $(\mathbb{F}_p)^k$.

Consider a random variable $Z \in \mathbb{F}_p$ with probability mass function

$$\begin{aligned} &\mu_Z \in \mathbb{R}^p, \\ &\mu_Z(j) = \mathbb{P}(Z = \mathbf{j}) \, . \end{aligned}$$

Note that this implicitly assumes an ordering of the mass function μ_Z corresponding to the representation of elements $\beta_j \in \mathbb{F}_p$ as integers. The discrete Fourier transform of μ_Z may then be written in matrix form as

$$F_p \mu_Z = \lambda_Z \ ,$$

where λ is the set of eigenvalues of the circulant matrix generated by μ_Z. Indeed, the above can be restated in terms of the unitary DFT,

$$\hat{F}_p = \frac{F_p}{\sqrt{p}} \qquad \hat{F}_p^* = \frac{F_p^*}{\sqrt{p}}$$

with F_p^* the conjugate transpose of F_p, diagonalising the circulant matrix C_Z generated by μ_Z:

$$\begin{aligned} C_Z &= circ(\mu_Z) \\ \hat{F}_p C_Z \hat{F}_p^* &= \Lambda_Z \end{aligned}$$

with Λ_Z the $p \times p$ diagonal matrix containing all eigenvalues of C_Z. Note that for a uniformly distributed random variable $U \in \mathbb{F}_p$, we have

$$\begin{aligned} \mu_U &= \frac{\mathbf{1}}{p} \\ \lambda_U &= F_p \mu_U = \mathbb{I}_p(\cdot, 1) \end{aligned}$$

where $\mathbf{1}$ is a vector of ones of length p, and the only non-zero eigenvalue is the zeroth one, so the full set λ_U corresponds to the first column of the identity matrix of size p.

Lemma 1. *The mass function μ_Z of a random variable $Z \in \mathbb{F}_p$ satisfies*

$$\|\mu_Z - \mu_U\|_2 = \frac{1}{\sqrt{p}} \|\lambda_Z - \lambda_U\|_2$$

where $\lambda_Z = F_p \mu_Z$ is the discrete Fourier transform of Z.

Proof.

$$\|\mu_Z - \mu_U\|_2 = \left\| \frac{\hat{F}^*}{\sqrt{p}} (\lambda_Z - \lambda_U) \right\|_2 \tag{11}$$

$$= \frac{1}{\sqrt{p}} \|\lambda_Z - \lambda_U\|_2 \tag{12}$$

$$= \frac{1}{\sqrt{p}} \left(\sum_{j=1}^{p-1} (\lambda_Z(j))^2 \right)^{1/2} . \tag{13}$$

Eq. (12) follows from the unitary Fourier transform preserving ℓ_2 distance. □

We can obtain a first, crude bound on the TVD by considering the largest non-trivial eigenvalue, defined as follows for future reference.

Definition 1. Given a random variable $Z \in \mathbb{F}_p$ with mass function μ_Z and eigenvalues $F_p\mu_Z = \lambda_Z$, denote the greatest non-trivial eigenvalue by

$$\lambda_{Z*} = \max_{1 \le j \le p-1} |\lambda_Z(j)|.$$

Theorem 3. *The total variation distance of a random variable $Z \in \mathbb{F}_p$ from uniform,*

$$\mathrm{TVD}(Z,\mathcal{U}) = \frac{\delta_Z}{2}$$

is bounded by

$$\delta_Z \le (p-1)^{1/2} \lambda_{Z*}$$

with λ_{Z*} as in Definition 1.

Proof. This follows from considering the worst-case scenario in which all eigenvalues λ_Z in Lemma 1, except the zeroth eigenvalue $\lambda_Z(0) = 1$, are equal to the greatest non-trivial eigenvalue λ_{Z*} by applying the bound on p-dimensional vector spaces $\|x\|_1 \le p^{1/2}\|x\|_2$. □

We can now consider how this affects the distribution of a sum of two variables, $S_2 = X_0 + X_1 \in \mathbb{F}_p$, which is the discrete convolution of the two probability masses,

$$\mathbb{P}(S_2 = r) = \sum_{j \in \mathbb{F}_p} \mathbb{P}(Z_0 = j)\mathbb{P}(Z_1 = r - j)$$

$$= \sum_{j=0}^{p-1} \mu_{Z_1}(r-j)\mu_{Z_0}(j).$$

The distribution of S_2 may then be expressed in matrix notation as

$$\mu_{S_2} = C_{Z_1}\mu_{Z_0}\,, \tag{14}$$

where C_{Z_1} is the circulant matrix uniquely defined by μ_{Z_1}. In other words, the entry r, j of C_{Z_1} contains the measure under Z_1 of the element $\beta_r - \beta_j \in \mathbb{F}_p$, which we denote in matrix form by

$$C_{Z_1} = \mu_{Z_1}(B)\,,$$
$$B(r,j) = \beta_r - \beta_j.$$

Considering the particular case of summing two identical variables Z with probability mass function μ_Z, the distribution of $S_2 = Z + Z$ may be written as

$$S_2 \sim C_Z\mu_Z,$$

where C_Z is the circulant matrix defined uniquely by μ_Z itself. By induction,

$$\begin{aligned} S_n &\sim C_Z^{n-1}\mu_Z \\ &\sim \left(\prod_{j=1}^{n-1} \frac{F_p^*}{p}\Lambda_Z F_p\right)\mu_Z \\ &\sim \frac{\hat{F}_p^*}{\sqrt{p}}\Lambda_Z^{n-1}F_p\mu_Z \\ &\sim \frac{1}{\sqrt{p}}\hat{F}_p^*\lambda_Z^n. \end{aligned} \tag{15}$$

As well as being conceptually equivalent to using the convolution theorem, this may also be seen as considering S_n as a Markov chain

$$S_0 = Z$$
$$S_j = S_{j-1} + Z$$

with transition matrix C_Z.

Lemma 1 may be extended as follows.

Lemma 2. *The probability mass function μ_{S_n} of a random variable $S_n = \sum_{j=1}^{n} Z$, $Z \in \mathbb{F}_p$ satisfies*

$$\|\mu_{S_n} - \mu_U\|_2 = \left(\frac{p-1}{p}\right)^{1/2} \|\left(\lambda_{S_n} - \lambda_U\right)^n\|_2.$$

Proof. The proof is substantially the same as that of Lemma 1, using Eq. (15). □

Lemma 3. *The total variation distance of $S_n = \sum_{j=1}^{n} Z$, $Z \in \mathbb{F}_p$ from uniform may be bounded by*

$$\delta_{S_n} \leq (p-1)^{1/2}\, \lambda_{Z*}^{n} \tag{16}$$

with λ_{Z*} as in Definition 1.

Proof. The proof follows by applying Lemma 2 in the same way as Lemma 1 was applied to Theorem 3, i.e. assuming each of the $p-1$ eigenvalues in Lemma 2 that are of magnitude less than 1 to be bounded by λ_*^n, using Eq. (15). □

Lemma 3 is a slight improvement on a known bound on the convergence rates of Markov chains on Abelian groups; see e.g. [7], Fact 7.

We have so far assumed an ordering of the mass function μ_X of a random variable $X \in \mathbb{F}_p$ according to the representation of the elements of $\mathbb{F}_p$ as integers. Similarly for vector spaces $X \in (\mathbb{F}_p)^k$ we may assume an ordering of μ_X by least significant digit. Generalising to the distribution of the sum S_2 of two random variables $X_0, X_1 \in (\mathbb{F}_p)^k$, this may still be expressed in a form such as Eq. (14), but the matrix C_{X_1} is a level k block circulant with circulant base blocks of size $(p \times p)$. Concretely, while considering all coefficients of B as elements of $(\mathbb{F}_p)^k$, we may write

$$\begin{aligned}
B_p &= circ([\mathbf{0}, \mathbf{1}, \ldots \mathbf{p-1}]) \\
B_p^{\circ 2} &= circ(B_p, \mathbf{p} + B_p, \ldots (\mathbf{p-1})\mathbf{p} + B_p) \\
B_p^{\circ k} &= circ(B_p^{\circ k-1}, \mathbf{p^{k-1}} + B_p^{\circ k-1}, \ldots (\mathbf{p-1})\mathbf{p^{k-1}} + B_p^{\circ k-1})
\end{aligned}$$

with the *circ* function defined column-wise following [8], and $B_p^{\circ k}$ used as short-hand to indicate a matrix therein defined as belonging to the class $\mathcal{BCCB}(p, p, \ldots p)$, k times.

As shown in [8], matrices with this structure are diagonalised by $\mathbb{F}_p^{\otimes k}$. This naturally extends the known structure for binary random variables, since as discussed in Section 2 the convolution matrix for variables in $(\mathbb{F}_2)^k$ is diagonalised by the Hadamard matrix H_{2^k}, which by construction is equal to $H_2^{\otimes k}$.

The Fourier matrix of size p can be written as a Vandermonde matrix of a primitive p-th root of unity as

$$F_p = \begin{pmatrix} \omega_p^{0\cdot 0} & \omega_p^{0\cdot 1} & \ldots & \omega_p^{0\cdot (p-1)} \\ \omega_p^{1\cdot 0} & \omega_p^{1\cdot 1} & \ldots & \omega_p^{0\cdot 0} \\ \ldots & \ldots & \ldots & \ldots \\ \omega_p^{(p-1)\cdot 0} & \omega_p^{(p-1)\cdot 1} & \ldots & \omega_p^{(p-1)^2} \end{pmatrix}.$$

In other words, the entry in row r, column s is

$$F_p(r, s) = \omega_p^{rs}, \qquad r, s \in \mathbb{Z}_p.$$

By definition of the Kronecker product of two $(p \times p)$ matrices,

$$\begin{aligned} K &= M_1 \otimes M_2 \\ K(u,v) &= M_1(r_1,s_1)M_2(r_2,s_2) \quad u,v,r_i,s_i \in \mathbb{Z}_p \\ u &\equiv r_1 p + r_2 \qquad (17) \\ v &\equiv s_1 p + s_2 \qquad (18) \\ (F_p \otimes F_p)(u,v) &= \omega_p^{r_1 s_1} \omega_p^{r_2 s_2} \end{aligned}$$

which extends to the k-fold Kronecker product by induction using the p-ary representation of integers

$$F_p^{\otimes k}(u,v) = \omega_p^{\mathbf{r}\cdot\mathbf{s}}$$

for the specific $\mathbf{r}, \mathbf{s}$ satisfying a polynomial in p such as (17) and (18) of degree $k-1$. In general, keeping either the row or column index fixed and iterating over the other means iterating over every element of $(\mathbb{F}_p)^k$; concretely, when evaluating the eigenvalues of a probability mass $\mu \in \mathbb{R}^{p^k}$, the b-th eigenvalue corresponds to

$$\begin{aligned} \lambda(b) &= F_p^{\otimes k}(b,\cdot)\mu \\ &= \sum_{j=0}^{p^k-1} F_p^{\otimes k}(b,j)\mu(j) \\ &= \sum_{j=0}^{p^k-1} \omega_p^{\mathbf{b}\cdot\mathbf{j}}\mu(j)\,. \qquad (19) \end{aligned}$$

Generalising from the case of the Walsh transform, this suggests the definition of the a-th order Fourier function evaluated at b as

$$f_a(b) = \omega_p^{\mathbf{b}\cdot\mathbf{a}}$$

(compare to Eq. (1)), so that if $Y \in (\mathbb{F}_p)^k$ is the random variable with mass function μ, the eigenvalues may be written as

$$\lambda_Y(b) = \mathbb{E}\left[f_b(Y)\right] \qquad (20)$$

(compare to Eq. (3)).

Lemma 4. *The probability mass function μ_Y of a random variable $Y \in (\mathbb{F}_p)^k$ satisfies*

$$\|\mu_Y - \mu_U\|_2 = \frac{1}{p^{k/2}}\|\lambda_Y - \lambda_U\|_2.$$

Proof.

$$\|\mu_Y - \mu_U\|_2 = \left\| \frac{F_p^{\otimes k}}{p^{k/2}} (\mu_Y - \mu_U) \right\|_2 \tag{21}$$

$$= \frac{1}{p^{k/2}} \|\lambda_Y - \lambda_U\|_2 . \qquad \square \tag{22}$$

Corollary 4. *If the elements $Y(j) \in \mathbb{F}_p$ are independent but not necessarily identically distributed,*

$$\|\mu_Y - \mu_U\|_2 = \frac{1}{p^{k/2}} \left\| \bigotimes_{j=0}^{k-1} \lambda_{Y(j)} - \lambda_U \right\|_2 .$$

Proof. Since the $X(j)$ are independent, the probability mass function of $Y \in (\mathbb{F}_p)^k$ is

$$\begin{aligned} \mu_Y &= \mu_{Y(0)} \otimes \mu_{Y(1)} \otimes \dots \mu_{Y(k-1)} \\ &= \bigotimes_{j=0}^{k-1} \mu_{Y(j)} , \end{aligned}$$

and the eigenvalues will be

$$\begin{aligned} \lambda_Y &= F_p^{\otimes n} \mu_Y \\ &= \bigotimes_{j=0}^{k-1} F_p \mu_{Y(j)} \\ &= \bigotimes_{j=0}^{k-1} \lambda_{Y(j)} \end{aligned}$$

where the second step follows by the mixed-product property of the Kronecker product. □

We can now extend Theorem 1 as follows.

Theorem 4. *The total variation distance of a random vector $Y \in (\mathbb{F}_p)^k$ from uniform,*

$$\begin{aligned} \mathrm{TVD}(Y, \mathcal{U}) &= \frac{1}{2} \|\mu_Y - \mu_U\|_1 \\ &= \frac{1}{2} \delta_Y \end{aligned}$$

is bounded by

$$\delta_Y \leq \sum_{\mathbf{b}\in(\mathbb{F}_p)^k\setminus\mathbf{0}} \left| \prod_{u=0}^{k-1} \lambda_Y(\mathbf{b}(u)) \right| . \tag{23}$$

Proof. Each eigenvalue may be written as

$$\lambda_Y(b) = \prod_{u=0}^{k-1} \lambda_Y(\mathbf{b}(u)) .$$

The result follows directly from Lemma 4 and the known bound on vector spaces. □

We can also extend Corollary 1 to establish a connection with the number of vectors of a specific Hamming weight, but in the non-binary case we can also go into more detail if the full composition of each vector in the space is known, as in the following definition.

Definition 2. Let $s(\mathbf{b})$ be the composition of $\mathbf{b} \in (\mathbb{F}_p)^k$ such that $s_j(\mathbf{b})$ is the number of components of $\mathbf{b}$ equal to j.

$$\begin{aligned} s(\mathbf{b}) &= (s_0, s_1, \dots s_{p-1}) \\ s_j &= |\{\, i \mid \mathbf{b}(i) = j \,\}| . \end{aligned}$$

Let $W_{(\mathbb{F}_p)^k}(t)$ be the enumerator of the elements $\mathbf{b}$ having composition equal to t, with t being a p-tuple summing to k:

$$\begin{aligned} W_{(\mathbb{F}_p)^k}(t) &= |\,\{\, \mathbf{b} \in (\mathbb{F}_p)^k \mid s(\mathbf{b}) = t \,\}\,| \\ t &\in T \subset \mathbb{N}^p \\ \sum_j t_j &= k ; \end{aligned}$$

then the number of $\mathbf{b}$ with Hamming weight equal to l is

$$A_l = \sum_t W(t) \qquad t \in \{\, t_0 = k - l \,\} .$$

In particular, if instead of $\mathbf{b} \in (\mathbb{F}_p)^k$ we consider a set of codewords $\mathbf{c} \in C$, the enumerator W_C is the complete weight enumerator of C, and A_l its weight distribution, as defined in [9], ch. 5, §6.

Corollary 5. *If each $Y(j)$ is i.i.d., the total variation distance of a random vector $Y \in (\mathbb{F}_p)^k$ from uniform is bounded by*

$$\delta_Y \leq \sum_t W(t) \prod_{u=0}^{p-1} \left(\lambda_{Y(j)}(u)\right)^{t_u} \qquad t \in \{\, t_0 < k \,\} . \tag{24}$$

Without knowledge of the full spectrum of $Y(j)$ one may obtain a coarser bound using the second largest eigenvalue is λ_{Y*}, as in Definition 1:

$$\delta_Y \leq \sum_{l=1}^{k} A_l \lambda_{Y*}^l \,, \tag{25}$$

where A_l is the number of $\mathbf{b} \in (\mathbb{F}_p)^k$ with Hamming weight $w(\mathbf{b}) = l$.

Proof. Each eigenvalue may further be written as

$$\begin{aligned}\lambda_Y(b) &= \prod_{u=0}^{k-1} \lambda_Y(\mathbf{b}(u)) \\ &= \prod_{u=0}^{k-1} \sum_{v=0}^{p-1} \omega_p^{\mathbf{b}(u)\cdot v} \mu_{Y(u)}(v)\,,\end{aligned}$$

so all the $\mathbf{b}$ with identical composition t will correspond to equal eigenvalues, leading directly to Eq. (24). If the Hamming weight $w(\mathbf{b}(u)) = 0$, then the u-th term of the product will be equal to 1; Eq. (25) follows by considering the worst case $\lambda_Y(j) = \lambda_{Y*}\ \forall j > 0$. □

5. Fourier bound on entropy extractors

In order to arrive at a bound involving the distribution of weights, we begin by showing there is an unique association between code words and eigenvalues, just as there was with the bias of individual bits in the binary case (see Theorem 2).

If $Y = GX$, with X a random vector in $(\mathbb{F}_p)^n$, G a generator matrix of an (n, k, d) code over $\mathbb{F}_p$, we can establish a direct link between eigenvalues of Y and codewords of G using Eq. (20):

$$\begin{aligned}\lambda_Y(b) &= \mathbb{E}\left[f_b(GX)\right] \\ &= \sum_{j=0}^{p^n-1} \omega_p^{\mathbf{b}^T G\mathbf{j}} \mu_X(j) \\ &= \sum_{j=0}^{p^n-1} \omega_p^{\mathbf{c}\cdot\mathbf{j}} \mu_X(j)\end{aligned} \tag{26}$$

with $\mathbf{c} = \mathbf{b}^T G$ a particular word of the code. Note that choosing a particular $(k \times n)$ matrix G is equivalent to selecting the specific p^k rows specified by all the k codewords $\mathbf{c}$ that forms a subset of the p^n rows of $F_p^{\otimes n}$ by which to multiply μ_X.

Having noted this fundamental link in principle in the same manner as for the binary case (see Eq. (10)), and having developed the required tools in Section 4, we can immediately state some more specific results for particular cases of practical interest, beginning with an extension of Theorem 4.

Theorem 5. *Let $Y = GX$, where $X \in (\mathbb{F}_p)^n$ is a random vector of length n, with each entry being an independent but not necessarily identically distributed variable $X(j) \in \mathbb{F}_p$ with mass function $\mu_{X(j)} \in \mathbb{R}^p$, and G is the generator matrix of an (n,k,d) linear code over $\mathbb{F}_p$. Then the b-th eigenvalue of the distribution of Y is*

$$\lambda_Y(b) = \prod_{j=0}^{n-1} \lambda_{X(j)}(\mathbf{c}(j)) \tag{27}$$

where $\mathbf{c}(j) \in \mathbb{F}_p$ is the j-th symbol in the codeword $\mathbf{c}^T = \mathbf{b}^T G$.

Proof. The specific combination corresponding to a word $\mathbf{c}$ is from Eq. (26):

$$\begin{aligned} \lambda_Y(b) &= \sum_{j=0}^{p^n-1} \omega_p^{\mathbf{c}\cdot\mathbf{j}} \mu_{X(j)} \\ &= \prod_{u=0}^{n-1} \sum_{v=0}^{p-1} \omega_p^{\mathbf{c}(u)\cdot\mathbf{v}} \mu_{X(u)}(v) \,. \qquad \square \end{aligned}$$

Corollary 6. *If all $X(j)$ are also i.i.d., the total variation distance of a random vector $Y \in (\mathbb{F}_p)^k$ from uniform is bounded by*

$$\delta_Y \leq \sum_t W_C(t) \prod_{u=0}^{p-1} \left(\lambda_{X(j)}(u)\right)^{t_u} \qquad t \in \{t_0 < n\}\,. \tag{28}$$

Without knowledge of the full spectrum of $X(j)$ one may obtain a coarser bound using the second largest eigenvalue is λ_{X*}, as in Definition 1:

$$\delta_Y \leq \sum_{l=d}^{n} A_l \lambda_{X*}^l \,. \tag{29}$$

Here W_C and A_l are the complete weight enumerator and weight distribution of C, respectively, as in Definition 2.

Proof. The proof follows in the same manner as for Corollary 5. □

The above can be viewed as a statement regarding the sum of n random variables, each in $\mathbb{F}_p$: if only $w(\mathbf{c})$ symbols are non-zero, this corresponds to a sum of $w(\mathbf{c})$ terms.

Corollary 7. *Using the minimum distance d, one may obtain the bound*

$$\delta_Y \leq (p^k - 1)\lambda^d_{X*}\,.$$

Note that all the results in this section extend to random vectors $X \in (\mathbb{F}_{p^m})^n$, that is to sequences of random vectors taking values in $\mathbb{F}_{p^m}$ by using the right matrix to diagonalise the convolution matrix of the sum of two such variables in order to compute the eigenvalues, and assuming the symbols of the generator matrix are taken in the same field, i.e. the code is chosen over $\mathbb{F}_{p^m}$. Following Section 4, this may be done using the Kronecker product $F_p^{\otimes mn}$.

Comparing Corollaries 3 and 7, it appears that a bound based solely on the minimum distance quickly risks becoming far from sharp as the dimension of the underlying random variable $X(j)$ increases.

6. Non-linear codes

As shown in [2], it is possible to construct ad-hoc non-linear maps with better properties than linear ones for specific cases; it was also noted that for a given compression ratio k/n of the output, there may exist non-linear codes with a greater minimum distance than any linear code. Since non-linear codes do not have a generator matrix G they are not straightforward to cover using the tools developed thus far, but we may use some of them to frame the fundamental issue with non-linear maps, as we see it, in terms of examining the distribution of the product of random variables. Consider the special case $X_1, X_0 \in \mathbb{F}_2$, and let their product be P_2; its mass function may be written as follows:

$$\begin{aligned} P_2 &= X_1 X_0 \\ \mu_{P_2} &= \begin{pmatrix} 1 & \mu_{X_1}(0) \\ 0 & \mu_{X_1}(1) \end{pmatrix} \begin{pmatrix} \mu_{X_0}(0) \\ \mu_{X_0}(1) \end{pmatrix}. \end{aligned}$$

As long as neither X_0 nor X_1 follow the categorical distribution with $\mathbb{P}(1) = 1$, the probability of their product being zero is strictly greater than either of the initial probabilities. By induction,

$$\lim_{j\to\infty} \mu_{P_j} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

We may conclude that, while increasing the number of linear operations will lead to the uniform distribution in expectation, increasing the number of non-linear operations will in general lead to a worsening of the output distribution, except in very specific cases. By way of example, consider the two Bernoulli variables

$$\begin{aligned} B_+ &\sim \mathcal{B}(2^{-1/2}), & B_- &\sim \mathcal{B}(1 - 2^{-1/2}) \\ \mu_{B_+} &= \begin{pmatrix} 1 - 2^{-1/2} \\ 2^{-1/2} \end{pmatrix} & \mu_{B_-} &= \begin{pmatrix} 2^{-1/2} \\ 1 - 2^{-1/2} \end{pmatrix}. \end{aligned}$$

Note that the bias of these two random variables is identical; however, the distributions of their products are quite different:

$$\mu_{B_+B_+} = \begin{pmatrix} 2^{-1} \\ 2^{-1} \end{pmatrix} \qquad \mu_{B_-B_-} = \begin{pmatrix} 2^{1/2} - \frac{1}{2} \\ \frac{3}{2} - 2^{1/2} \end{pmatrix}.$$

While it is possible to find non-linear maps that are optimal in some specific cases, we observe that not only does repeated processing by nonlinear maps in general lead to a worsening of the output, but it is also necessary to know or assume a specific distribution of the sequence to be processed to even attempt to find such a processing; even under the assumption of i.i.d. binary variables, knowledge of the bias of each bit is not sufficient.

7. Conclusions

We have shown new bounds on the statistical distance from the uniform distribution of random number sequences conditioned by linear transformations chosen from the generator matrices of linear codes, based on the assumption of independent generator output in $\mathbb{F}_{p^m}$; we have also shown how these bounds are natural generalisations of known bounds in $\mathbb{F}_{2^m}$ once the structure behind the known bounds is made clear. If the weight distribution or the complete weight enumerator of the code is known, this allows one to determine the distribution of the conditioned output exactly. This is of especial importance whenever a matrix is chosen once, possibly based on a random seed, and then seldom changed, if ever.

Acknowledgments

The authors would like to thank the anonymous referee of a previous work for suggesting an approach to the proof of Theorem 2.

The second author would like to thank his PhD supervisor (the third author).

This research was funded by the Autonomous Province of Trento, Call “Grandi Progetti 2012”, project “On silicon quantum optics for quantum computing and secure communications - SiQuro”.

References

[1] M. Sönmez Turan, E. Barker, J. Kelsey, K.A. McKay, M.L. Baish, M. Boyle, Recommendation for the Entropy Sources Used for Random Bit Generation (Second DRAFT), 01 2016, http://dx.doi.org/10.6028/NIST.SP.XXX, http://csrc.nist.gov/publications/drafts/800-90/sp800-90b_second_draft.pdf.

[2] P. Lacharme, Post-processing functions for a biased physical random number generator, in: Fast Software Encryption, in: Lect. Notes Comput. Sci., vol. 5086, Springer, 2008, pp. 334–342.

[3] P. Lacharme, Analysis and construction of correctors, IEEE Trans. Inf. Theory 55 (10) (2009) 4742–4748, http://dx.doi.org/10.1109/TIT.2009.2027483.

[4] H. Zhou, J. Bruck, Linear extractors for extracting randomness from noisy sources, in: International Symposium on Information Theory (ISIT) Proceedings, IEEE, 2011, pp. 1738–1742.

[5] R. Shaltiel, An introduction to randomness extractors, in: Proceedings of the 38th International Conference on Automata, Languages and Programming, ICALP, Springer, 2011, pp. 21–41.

[6] J. Pearl, Application of Walsh transform to statistical analysis, IEEE Trans. Syst. Man Cybern. SMC-1 (2) (1971) 111–119, http://dx.doi.org/10.1109/TSMC.1971.4308267.

[7] J.S. Rosenthal, Convergence rates for Markov chain, SIAM Rev. 37 (3) (1995) 387–405, http://www.jstor.org/stable/2132659.

[8] P.J. Davis, Circulant Matrices, 2nd edition, AMS Chelsea, 1994.

[9] F.J. MacWilliams, N.J.A. Sloane, The Theory of Error Correcting Codes, N.-Holl. Math. Libr., vol. 16, North-Holland Publishing Company, 1977.