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
1071-5797/

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

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