Lecture 3 - A theorist's toolkit

1 Overview of this lecture

In this lecture we first discuss two useful bounds for random variables. Then we recall Lagrange's interpolation formula. This is followed by a construction of a set of k-wise independent random variables, and an application of these. We then give a negative result. Finally, we consider how a Hadamard matrix can be used to construct pairwise and 3-wise independent random variables.

2 Useful Bounds

The most basic of all probability bounds is the following due to Markov.

Theorem 1 (Markov). *Let X be a random variable on a space S and $f : S \to \mathbb{R}_+$. Then*

$$\Pr[f(X) \geq k] \leq \frac{\mathrm{E}[f(X)]}{k} \ .$$

Proof. Let Y be the indicator variable for the event $f(X) \geq k$. Since f is positive we have $Y \leq \frac{f(X)}{k}$. This implies that $\Pr[f(X) \geq k] = \mathrm{E}[Y] \leq \frac{\mathrm{E}[f(X)]}{k}$. □

Markov's bound can be used to derive a bound on the probability that a real valued random variable takes a value far from its expectation.

Theorem 2 (Chebychev). *Let X be a random variable on $\mathbb{R}$ and let σ be the variance of X. Then*

$$\Pr[|X - \mathrm{E}[X]| \geq s\sigma] \leq \frac{1}{s^2} \ .$$

Proof. Define $f(x) = (x - \mathrm{E}[x])^2 = x^2 - 2x\mathrm{E}[x] + \mathrm{E}[x]^2$. Then $\mathrm{E}[f(X)] = \mathrm{E}\left[X^2\right] - \mathrm{E}[X]^2 = \sigma^2$. From Markov's inequality now follows that

$$\Pr[|X - \mathrm{E}[X]| \geq s\sigma] = \Pr[f(X) \geq s^2\sigma^2] \leq \frac{\mathrm{E}[f(X)]}{s^2\sigma^2} = \frac{1}{s^2} \ .$$

□

2.1 An Interesting Example

Suppose we have n uniformly and independently distributed random variables $X_1, \ldots, X_n$ taking values in $\{1, -1\}$. What is the probability that we get at least t more ones than minus ones?

Define $X = \sum_{i=1}^{n} X_i$. The probability we are looking for is $\Pr[X \geq t]$. We have

$$\Pr[X \geq t] = \Pr[e^{\lambda X} \geq e^{\lambda t}] \leq \frac{\mathrm{E}\left[e^{\lambda X}\right]}{e^{\lambda t}} \ .$$

Here we use that e^x is an increasing function, i.e. $x \geq t$ is equivalent with $e^x \geq e^t$. Then we use the definition of X and independence to get

$$\mathrm{E}\left[e^{\lambda X}\right] = \mathrm{E}\left[\prod_{i=1}^{n} e^{\lambda X_i}\right] = \left\{\text{independent } X_i\right\} = \prod_{i=1}^{n} \mathrm{E}\left[e^{\lambda X_i}\right] \ .$$

Finally, we have $\mathrm{E}\left[e^{\lambda X_i}\right] = \frac{1}{2}(e^{\lambda} + e^{-\lambda})$. This can be bounded by

$$\begin{aligned} e^{\lambda} + e^{-\lambda} &= \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \frac{\lambda^4}{4!} + \ldots\right) + \left(1 - \lambda + \frac{\lambda^2}{2!} - \frac{\lambda^3}{3!} + \frac{\lambda^4}{4!} - \ldots\right) \\ &= 2\left(1 + \frac{\lambda^2}{2!} + \frac{\lambda^4}{4!} + \frac{\lambda^6}{6!} + \ldots\right) \leq 2e^{\lambda^2/2} \ , \end{aligned}$$

so $\mathrm{E}\left[e^{\lambda X_i}\right] \leq e^{\lambda^2/2}$. This means that $\Pr[X \geq t] \leq e^{n\lambda^2/2 - \lambda t}$. We find the minimum $\lambda = t/n$ to the quadratic expression, which gives the bound $\Pr[X \geq t] \leq e^{-\frac{t^2}{2n}}$.

Loosely speaking, it is extremely unlikely that there are more than $\sqrt{n}$ more ones than minus ones. The example can be viewed as a special case of the more general Chernoff-bounds. A nice source for such bounds is the appendix of the book *The Probabilistic Method* by Alon and Spencer.

3 Interpolation and Lagrange's Formula

Suppose that $f \in \mathbb{F}[x]$ is a $(k-1)$-degree polynomial, and that we are given the value of f at k distinct points $\alpha_1, \ldots, \alpha_k$, but no explicit description of f iself. Can we recover the polynomial f? We can consider the equation system

$$\begin{pmatrix} 1 & \alpha_1 & \alpha_1^2 & \cdots & \alpha_1^{k-1} \\ 1 & \alpha_2 & \alpha_2^2 & \cdots & \alpha_2^{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_k & \alpha_k^2 & \cdots & \alpha_k^{k-1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{k-1} \end{pmatrix} = \begin{pmatrix} f(\alpha_1) \\ f(\alpha_2) \\ \vdots \\ f(\alpha_k) \end{pmatrix} \ ,$$

and recover the coefficients $a_0, \ldots, a_{k-1}$ if the matrix has full rank. This would give us the polynomial $f(x) = \sum_{i=0}^{k-1} a_i x^i$. The matrix is a Vandermonde matrix and it has determinant $\prod_{1 \leq i < j \leq n} (\alpha_j - \alpha_i)$. Thus, the equation system is solvable, as long as all the α_i are different.

Another way to look at this is to try to find a candidate polynomial $f'(x)$ such that $f'(\alpha_i) = f(\alpha_i)$ for $i = 1, \ldots, k$. Then $(f - f')(x)$ is a degree $k-1$ polynomial over $\mathbb{F}_q[x]$ with k zeros in $\mathbb{F}_q$. From the fundamental theorem of algebra we know this to be impossible unless $f' \equiv f$.

There is a simple way to find the candidate f' due to Lagrange. Set

$$f'(x) = \sum_{i=1}^{k} f(\alpha_i) \prod_{l \neq i} \frac{x - \alpha_l}{\alpha_i - \alpha_l} \quad ,$$

where the product is over all $l = 1, \ldots, i-1, i+1, \ldots, k$. The product is clearly zero for $x = \alpha_j$ with $j \neq i$, but it is one for $x = \alpha_i$, so $f'(\alpha_i) = f(\alpha_i)$ for $i = 1, \ldots, k$. Thus, this is really a alternative description of the polynomial f we started with. We could in principle expand the expression and recover the coefficients $a_0, \ldots, a_{k-1}$ of f.

4 k-Wise Independence

For independent random variables $X_1, \ldots, X_n$ the expected value commutes with products, i.e., we have $\mathrm{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathrm{E}\left[X_i\right]$. This was a crucial step in the example above. Independence of all variables is a very strong property, and sometimes it suffices if all subsets of k variables are independent. A set of k-wise independent variables have exactly this property.

Definition 3 (k-Wise Independence). *The random variables $X_1, \ldots, X_n$ are k-wise independent if $X_{i_1}, \ldots, X_{i_k}$ are independent for every k-subset $\{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$.*

4.1 Construction of k-Wise Independent Variables

Suppose we wish to construct a list $X_1, \ldots, X_n$ of n uniformly distributed k-wise independent random variables taking values in a finite field $\mathbb{F}_q$ with q elements.

We do this as follows. If $a = (a_0, \ldots, a_{k-1}) \in \mathbb{F}_q^k$ we write $f_a(x) = \sum_{i=0}^{k-1} a_i x^i \in \mathbb{F}_q[x]$ for the corresponding degree $k-1$ polynomial. Then we let $\alpha_1, \ldots, \alpha_n$ be n distinct elements in $\mathbb{F}_q$ and define a map

$$\begin{aligned} g : \mathbb{F}_q^k &\rightarrow \mathbb{F}_q^n \\ g : a &\mapsto (f_a(\alpha_1), \ldots, f_a(\alpha_n)) \ . \end{aligned}$$

Note that we require that $q \geq n$, but we do not require that $\alpha_i \neq 0$.

Theorem 4. *If $A_0, \ldots, A_{k-1}$ are uniformly and independently distributed random variables taking values in $\mathbb{F}_q$, then $(X_1, \ldots, X_n) = g(A_0, \ldots, A_{k-1})$ is a list of uniformly and k-wise independently distributed random variables taking values in $\mathbb{F}_q$.*

Proof. For every k-subset $I = \{i_1, \ldots, i_k\} \subset \{1, \ldots, n\}$ we define

$$\begin{aligned} g_I : \mathbb{F}_q^k &\rightarrow \mathbb{F}_q^k \\ g_I : a &\mapsto (f_a(\alpha_{i_1}), \ldots, f_a(\alpha_{i_k})) \ , \end{aligned}$$

i.e., this is the restriction of g to the indices in I. If g_I is a bijection we are done, since that means that $X_{i_1}, \ldots, X_{i_k}$ are identically distributed to $A_0, \ldots, A_{k-1}$. It suffices to show that g_I is injective, i.e. that for any $(x_{i_1}, \ldots, x_{i_k}) \in \mathbb{F}_q^k$, there exists a unique $a \in \mathbb{F}_q$ such that $(x_{i_1}, \ldots, x_{i_k}) = g_I(a)$, but this is exactly the problem of interpolating a polynomial considered above so we are done. □

The above is nice, but what if we want n k-wise independent binary variables? Naively, we would like to set $q = 2$, but this is not possible due to the restriction $q \geq n$. Instead we set $q = 2^t$ as the smallest power of 2 greater than n. Then we do the construction above, but in the final step we define X_i to equal the ith bit in the string $(f_a(\alpha_1), \ldots, f_a(\alpha_n))$ of nq bits. It is easy to see that all bits in the same word $f_a(\alpha_i)$ are independent, and k-wise independence follows from the theorem. From the minimality of t follows that $q \leq 2n$. Thus, the size of the sample space is bounded by $2^k n^k = O(n^k)$, so we can get away with a polynomial sized sample space.

4.2 Approximation of Max-3-SAT and 3-Wise Independence

Consider an instance ϕ of Max-3-SAT with n variables $x_1, \ldots, x_n$ and m clauses $C_1, \ldots, C_m$, where each clause C_j is on the form $l_{i_{j,1}} \vee l_{i_{j,2}} \vee l_{i_{j,3}}$ and each literal $l_{i_{j,k}}$ is either the variable $x_{i_{j,k}}$ or its negation. We can view ϕ as a degree 3 polynomial $\phi(x_1, \ldots, x_n) = \sum_{j=1}^m C_j(x_{i_{j,1}}, x_{i_{j,2}}, x_{i_{j,3}})$, with

$$C_j(x_{i_{j,1}}, x_{i_{j,2}}, x_{i_{j,3}}) = 1 - (x_{i_{j,1}} - b_{j,1})(x_{i_{j,2}} - b_{j,2})(x_{i_{j,3}} - b_{j,3})$$

where $b_{j,k} = 1$ if $x_{i_{j,k}}$ is negated in C_j and $b_{j,k} = 0$ otherwise. It is easy to see that the two representations are equivalent.

Proposition 5. *The expected number of clauses in ϕ satisfied by a uniformly and independently distributed assignment is $\frac{7}{8}m$.*

Proof. Let $X_1, \ldots, X_n$ be uniformly and independently distributed binary random variables. We then have

$$\mathrm{E}\left[\phi(X_1, \ldots, X_n)\right] = \sum_{j=1}^{m} \mathrm{E}\left[C_j(X_{i_{j,1}}, X_{i_{j,2}}, X_{i_{j,3}})\right] \ .$$

The claim now follows since we have

$$\begin{aligned}
\mathrm{E}\left[C_j(X_{i_{j,1}}, X_{i_{j,2}}, X_{i_{j,3}})\right] &= 1 - \mathrm{E}\left[(X_{i_{j,1}} - b_{j,1})(X_{i_{j,2}} - b_{j,2})(X_{i_{j,3}} - b_{j,3})\right] \\
&= 1 - \mathrm{E}\left[X_{i_{j,1}} - b_{j,1}\right]\mathrm{E}\left[X_{i_{j,2}} - b_{j,2}\right]\mathrm{E}\left[X_{i_{j,3}} - b_{j,3}\right] \\
&= 1 - \mathrm{E}\left[X_{i_{j,1}}\right]\mathrm{E}\left[X_{i_{j,2}}\right]\mathrm{E}\left[X_{i_{j,3}}\right] = 1 - \frac{1}{8} = \frac{7}{8} \ .
\end{aligned}$$

The second inequality holds since the variables are independent and the third from uniformity. □

There is a derandomization technique to transform the above into a deterministic algorithm. The idea is simple and based on conditional expected values. Note that we have

$$\begin{aligned}
&\mathrm{E}\left[\phi(X_1, \ldots, X_n) \mid (X_1, \ldots, X_s) = (a_1, \ldots, a_s)\right] \\
&= \frac{1}{2}\mathrm{E}\left[\phi(X_1, \ldots, X_n) \mid (X_1, \ldots, X_{s+1}) = (a_1, \ldots, a_s, 0)\right] \\
&\quad + \frac{1}{2}\mathrm{E}\left[\phi(X_1, \ldots, X_n) \mid (X_1, \ldots, X_{s+1}) = (a_1, \ldots, a_s, 1)\right]
\end{aligned}$$

for $s = 0, \ldots, n-1$. We iteratively choose $a_1, a_2, \ldots, a_n$ to maximize the conditional expectation. The above equality implies that in doing so the conditional expectation never goes below the starting point $\frac{7}{8}m$ guaranteed by Proposition 5. Thus, we have constructed a deterministic algorithm. More on derandomization techniques can be found in *The Probabilistic Method* by Alon and Spencer, or in *Randomized Algorithms* by Motwani and Raghavan.

Our algorithm is sequential in nature. Suppose we are looking for a parallel algorithm. We observe that the proof of Proposition 5 would go through with any distribution, as long as each X_i is uniformly distributed in $\{0, 1\}$ and the X_i are 3-wise independent. Thus, we have the following stronger version of the proposition.

Proposition 6. *The expected number of clauses in ϕ satisfied by a uniformly and 3-wise independently distributed assignment is $\frac{7}{8}m$.*

The importance of this proposition is clear from the following application where we sketch a parallel random access machine (PRAM) running in $O(\log n)$ time using a polynomial number of processors, and which outputs an assignment that satisfies at least $\frac{7}{8}m$ clauses. For each possible assignment a in the sample space of the list $X_1, \ldots, X_n$ of 3-wise independent variables and each clause C we have a processor $P_{a,C}$ which simply checks if the assignment a satisfies C, and outputs 1 or 0 accordingly. This is done in constant time, since $P_{a,C}$ only checks three literals. We also have a polynomial number of adding processors for each clause C organized in a tree which simply adds the outputs of the processors $P_{a,C_1}, \ldots, P_{a,C_m}$ handling the assignment a. Denote by P_a the final processor in this tree. Finally, we have a tree of processors that outputs the assignment a that maximizes the output of P_a.

Note that a necessary condition for the construction to work is that the sample space of $(X_1, \ldots, X_n)$ is polynomial size. The construction would fail terribly if we had chosen each X_i uniformly and independently in $\{0, 1\}$ as we would need 2^n processors.

4.3 More On 3-Wise Independence for Binary Variables

The application in the previous section shows the importance of having a small sample space. We have showed how to construct a list $(X_1, \ldots, X_n)$ of k-wise independent uniformly distributed binary variables X_i such that the sample space of the list have size $O(n^k)$. It is natural to ask if we can get away with a smaller sample space by a more clever construction.

We consider an example. We list all strings of 4 bits that have even parity and assume that we have a uniform distribution on this space.

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Thus, the possible outcomes of X_i corresponds to the ith column. Note that if we assign X_i some value b_i we still have probability $1/2$ for each outcome of X_j for all $j \neq i$. Thus, the variables X_i are at least pairwise independent. On the other hand they are not 4-wise independent, since we always have $X_1 \oplus X_2 = X_3 \oplus X_4$.

We can rephrase our argument for 2-wise independence as saying that for any $i \neq j$, if we pick the ith and jth columns all patterns of two bits have the same frequency. This leads us to the following more general statement.

Lemma 7. *Let $(X_1, \ldots, X_n)$ be a list of uniformly distributed pairwise independent binary random variables. Denote by $S \subset \{0, 1\}^n$ the sample space of $(X_1, \ldots, X_n)$. Then $|S| \geq n$.*

Proof. Consider the table consisting of all elements in the sample space of $(X_1, \ldots, X_n)$,

$$\begin{bmatrix} b_{1,1} & b_{1,2} & \cdots & b_{1,n} \\ b_{2,1} & b_{2,2} & \cdots & b_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m,1} & b_{m,2} & \cdots & b_{m,n} \end{bmatrix}.$$

If we pick any two columns C_i and C_j all patterns must be equally frequent so m must at least be even. Furthermore, there must be precisely $m/2$ ones if we add any two columns C_i and C_j as vectors in $\mathbb{F}_2^m$. We have an isomorphism of vectorspaces $h : \mathbb{F}_2^m \to \{-1,1\}^m$, $h : (b_i) \to ((-1)^{b_i})$, that turns addition modulo 2 into multiplication. Our requirement in multiplicative notation is that if we "add" $h(C_i)$ and $h(C_j)$ we must have a vector with the same number of ones and minus ones. This means that if we consider $h(C_i)$ and $h(C_j)$ as vectors with the usual addition, their inner product must be zero, i.e. they are orthogonal.

We can clearly have at most m orthogonal vectors in an m-dimensional vectorspace, so $m \geq n$. □

A more general theorem follows.

Theorem 8. *Let $(X_1, \ldots, X_n)$ be a list of uniformly distributed k-wise independent binary random variables. Denote by $S \subset \{0,1\}^n$ the sample space of $(X_1, \ldots, X_n)$. Then $|S| \geq \binom{n}{\lfloor k/2 \rfloor}$.*

Proof. Consider all $\lfloor k/2 \rfloor$-subsets $A_1, \ldots, A_s \subset \{1, \ldots, n\}$, and define the random variables $Y_l = \bigoplus_{i \in A_l} X_i$. This gives $\binom{n}{\lfloor k/2 \rfloor}$ variables Y_l. The new variables are pairwise independent, for if Y_l and $Y_{l'}$ are dependent, the variables $\{X_i\}_{i \in A_l \cup A_{l'}}$ are dependent and $A_l \cup A_{l'} \leq k$. From Lemma 7 we conclude $m \geq \binom{n}{\lfloor k/2 \rfloor}$. □

The theorem can be tightend slightly by considering all subsets of size at most k instead of only all subsets of size k.

5 Independence and Hadamard Matrices

In this section we discuss the relation between Hadamard matrices and pairwise and 3-wise independent uniformly distributed binary random variables.

5.1 Definition and Construction of Hadamard Matrices

Recall the definition of a Hadamard matrix.

Definition 9. *An $n \times n$-matrix H taking values in $\{1, -1\}$ such that $H^\top H$ is called a Hadamard matrix.*

We construct a Hadamard matrix H_{2^t} for any $t > 1$ as follows. Define

$$H_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} .$$

It is easy to verify that H_2 is a Hadamard matrix. Note that if we index the columns and rows by 0 and 1, the (α, β)th element in H_2 is given by

$(-1)^{\alpha\cdot\beta}$, where α and β are bits and $\alpha \cdot \beta$ is the inner product in $\mathbb{F}_2$. In this simple case the inner product is really the product, but we construct a larger Hadamard matrix we keep this property.

Suppose now we have a $2^t \times 2^t$-Hadamard matrix H_t. We construct a H_{t+1} Hadamard matrix by forming the block-matrix

$$H_{t+1} = \begin{pmatrix} H_t & H_t \\ H_t & -H_t \end{pmatrix} .$$

It is easy to see that this is a Hadamard matrix. Suppose that we indexed the rows and columns of H_t by binary strings $\alpha_0, \ldots, \alpha_{2^t-1}$ such that the (α, β)th element is given by $(-1)^{\alpha\cdot\beta}$. Then we index the rows and columns of H_{t+1} by $((0, \alpha_0), \ldots, (0, \alpha_{2^t-1}), (1, \alpha_0), \ldots, (1, \alpha_{2^t-1}))$. This implies that the (α, β)th element of H_{t+1} is given by $(-1)^{\alpha\cdot\beta}$.

5.2 Pairwise Independence from Hadamard Matrices

Our task is to construct $n = 2^t - 1$ pairwise independent uniformly distributed binary random variables $X_{\alpha_1}, \ldots, X_{\alpha_n}$. We do this as follows. Consider the Hadamard matrix $H_t = (C_\alpha)_{\alpha\in\{0,1\}^t}$, where C_α are the columns. We denote by S the set of rows in the submatrix $(C_\alpha)_{\alpha\in\{0,1\}^t,\alpha\neq 0}$. We have the following proposition.

Proposition 10. *If $(X_1, \ldots, X_n)$ is uniformly distributed with sample space S, then the X_is are uniformly and pairwise independently distributed in $\{1, -1\}$.*

Proof. Pick any two columns C_α and C_β, with $\alpha, \beta \neq 0$. We need to argue that each two-bit row is equally frequent and that each column contains as many ones as minus ones. Denote by $\#(d_\alpha, d_\beta)$ the number of rows (d_α, d_β) in the matrix (C_α, C_β).

First we note that each column C_α with $\alpha \neq \alpha_0$ contains precisely 2^{t-1} ones and 2^{t-1} minus ones. This follows since its inner product with C_0 over $\mathbb{R}$ is zero, and C_0 is the all ones vector. Thus, we have

$$\begin{aligned} \#(1,1) + \#(1,-1) &= \#(-1,1) + \#(-1,-1) \\ \#(1,1) + \#(-1,1) &= \#(1,-1) + \#(-1,-1) \ , \end{aligned}$$

which implies $\#(1,-1) = \#(-1,1)$ and $\#(1,1) = \#(-1,-1)$. Furthermore, the vectors C_α and C_β are orthogonal so

$$\#(1,1) + \#(-1,-1) = \#(-1,1) + \#(1,-1)$$

which implies $\#(1,1) = \#(1,-1) = \#(-1,1) = \#(-1,-1)$. □

Alternative Proof. Another way to show that each two-bit row is equally frequent is as follows.

We know that there are 2^t rows, so it suffices to show that exactly $2^t/4$ rows equal $((-1)^{b_\alpha}, (-1)^{b_\beta})$ for every pair $(b_\alpha, b_\beta) \in \{0,1\}^2$. This is equivalent to show that there exists exactly $2^t/4$ solutions to the equation system

$$\begin{pmatrix} \alpha_0 & \alpha_2 & \cdots & \alpha_n \\ \beta_0 & \beta_2 & \cdots & \beta_n \end{pmatrix} \begin{pmatrix} z_0 \\ z_1 \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} b_\alpha \\ b_\beta \end{pmatrix} \bmod 2 \ .$$

This in turn follows as soon as the left most matrix has full rank, which is the case since α and β are distinct and non-zero. $\square$

5.3 3-Wise Independence from Hadamard Matrices

The ideas in the previous section can be generalized to give a 3-wise independent distribution. To do this we define S to be the set of rows in the submatrix $(C_\alpha)_{\alpha\in\{0,1\}^t,\alpha_0=1}$. This corresponds to the columns in the matrix consisting of two Hadamard matrices stacked on top of each other

$$\begin{pmatrix} H_{t-1} \\ -H_{t-1} \end{pmatrix} \ .$$

Proposition 11. *If $(X_1, \ldots, X_n)$ is uniformly distributed with sample space S, then the X_is are uniformly and 3-wise independently distributed in $\{1, -1\}$.*

Proof. The first part of the proof is unaltered, i.e. we argue that each column has equal number of ones as minus ones to show that each X_{α_i} is uniformly distributed in $\{0, 1\}$. Note that the all ones column C_α corresponding to the all zero index α is not included in the submatrix $(C_\alpha)_{\alpha\in\{0,1\}^t,\alpha_0=1}$.

In the next step we pick three columns C_α, C_β, and C_γ. We know that there are 2^t rows, so it suffices to show that exactly $2^t/8$ rows equal $((-1)^{b_\alpha}, (-1)^{b_\beta}, (-1)^{b_\gamma})$ for every triple $(b_\alpha, b_\beta, b_\gamma) \in \{0,1\}^3$. This is equivalent to show that there exists exactly $2^t/8$ solutions to the equation system

$$\begin{pmatrix} \alpha_0 & \alpha_2 & \cdots & \alpha_n \\ \beta_0 & \beta_2 & \cdots & \beta_n \\ \gamma_0 & \gamma_2 & \cdots & \gamma_n \end{pmatrix} \begin{pmatrix} z_0 \\ z_1 \\ \vdots \\ z_n \end{pmatrix} = \begin{pmatrix} b_\alpha \\ b_\beta \end{pmatrix} \bmod 2 \ .$$

This in turn follows as soon as the left most matrix has full rank, i.e. if the vectors α, β, and γ are independent. They are pairwise independent since they are distinct and non-zero. By construction $\alpha_0 = \beta_0 = \gamma_0 = 1$, so we can not get the zero vector by adding all vectors modulo 2. Thus, the matrix has full rank and we are done. $\square$

6 Summary On k-Wise Independent Distributions

We have showed using polynomials how to construct n uniformly and k-wise independently distributed binary random variables X_i such that the sample space of $(X_1, \ldots, X_n)$ is of size n^k. On the other hand we have showed that for any construction of such variables the sample space of $(X_1, \ldots, X_n)$ must have size roughly $n^{\lfloor k/2 \rfloor}$. For the special cases where $k = 2, 3$ we have given constructions where the size of the sample space is linear in n, i.e. these constructions are optimal up to a constant factor.