Randomized Methods in Computation*
Lecture 4: Small Pairwise-Independent Sample Spaces (cont.) and Hash Functions

Notes taken by Olga Grinchtein and Denis Simakov

Spring 2001

Pluralitas non est ponenda sine neccesitate
(Entities should not be multiplied unnecessarily)
Ockham's Razor (William of Ockham)

Summary: A good construction of k-wise independent sample space should be small and convenient to use in applications. The construction presented in the previous lecture satisfies the first requirement (i.e., were small), but is rather complex, imposing a number of technical problems in application. In this lecture a second, much more convenient construction is presented, based on affine transformations. This construction yields small enough sample spaces, but only for pairwise independence ($k = 2$).

We also start discussing hash functions, show their relation to small pairwise independent sample spaces, and present a Hashing Lemma, connecting k-wise independency of hash functions to their "uniformness".

1 A second construction of small pairwise independent sample space

1.1 Motivation

Let's recall the problem we address. We have a finite set S and an integer m. We want to generate random strings of length m over S. So conceivably our sample space is S^m. But the size of this sample space is $|S|^m$ that is often much more than we can afford.

So naturally the task is to find set $\Omega \subset S^m$, which would be rather small and would possess the property of k-wise independency (k is one more integer parameter). This would enable us to use this small sample space Ω instead of huge S^m in problems, where k-wise independence suffices.

More specifically, *k-wise independence* means that if we watch strings through arbitrary k "windows", and these strings are uniformly chosen from Ω, then we will see in these windows uniform distribution, that is, for any distinct $i_1, \ldots, i_n$ and any $\alpha_1, \ldots, \alpha_k \in S$

$$\Pr_{e \in_R \Omega}\left[e_{i_1} = \alpha_1 \wedge \cdots \wedge e_{i_k} = \alpha_k\right] = \frac{1}{|S|^k}.$$

Concerning the size of Ω, typically we want it to be polynomial in $\max(|S|, m)^k$.

* Lecture Notes for a course given by Oded Goldreich at the Weizmann Institute of Science, Israel.

A first construction, solving this problem, we saw the last time. It is based on associating S with a finite field $\mathcal{F}$. Then k-wise independent Ω is constructed as the set of polynomials of degree less than k over the field $\mathcal{F}$. In explicit representation (in terms of elements of S^m) Ω is the set of values of such polynomials on some m distinct elements of the field:

$$\begin{aligned}\Omega &= \{\text{polynomials of degree} < k \text{ over field } \mathcal{F}\} \\ &= \Big\{\Big(P(1), \ldots, P(m)\Big) \;:\; P \text{ – such polynomials}\Big\}.\end{aligned}$$

This construction is quite satisfactory regarding independence (it works for any k) and size of Ω ($|\Omega| = |S|^k$, that is the smallest possible for k-wise independent space). But it is rather inconvenient for practical use due to a number of annoying technical difficulties that we have to overcome. We have to impose on S algebraic structure which can be complex (for example, finite fields), we have constraints on the size of S (it should be a prime power), and there is also inherent condition that $m \leq |S|$. Although we have demonstrated how to manage all these technicalities, we still would like more convenient methods.

In this lecture we will present *the second construction* of small pairwise independent space. It will have the following major differences, compared to *the first construction*:

+ less problematic;

− only for $k = 2$ (pairwise independence), and we don't know how to generalize it for arbitrary k.

So the second construction we present below is nice, and can be recommended for applications that require only pairwise independence. But whenever we need higher degree of independence, we still have our first construction at hand.

1.2 Construction

We'll assume that $S = \{0,1\}^n$ (if it's not the case, we can embed our sample space into set of binary strings). Unlike the first construction, we do not assume anything about the relation of m vs $|S|$.

Instead of polynomials over some field, in the second construction we will use affine transformations of boolean vectors. Affine transformation, given by pair (A, b), where A is $n \times l$ boolean matrix, b is n-dimensional boolean vector, maps all vectors $v \in \{0,1\}^l$ into vectors $Av + b \in \{0,1\}^n$ (all operations are done modulo 2).

So we define our new small sample space as follows:

$$\begin{aligned}\Omega &\stackrel{\text{def}}{=} \left\{\begin{array}{l}\text{set of all affine transformations} \\ \text{from } [m] = \{0,1\}^l \text{ to } S = \{0,1\}^n\end{array}\right\} \\ &\equiv \left\{(A, b) \;:\; \begin{array}{l}A \text{ is } n \times l \text{ boolean matrix,} \\ b \text{ is } n\text{-dimentional boolean vector}\end{array}\right\} \\ &= \left\{(Av_1 + b, \ldots, Av_m + b) \;:\; \begin{array}{l}\forall (A, b), \quad v_1, \ldots, v_m \text{ are fixed} \\ \text{distinct } l\text{-dimensional vectors}\end{array}\right\}.\end{aligned}$$

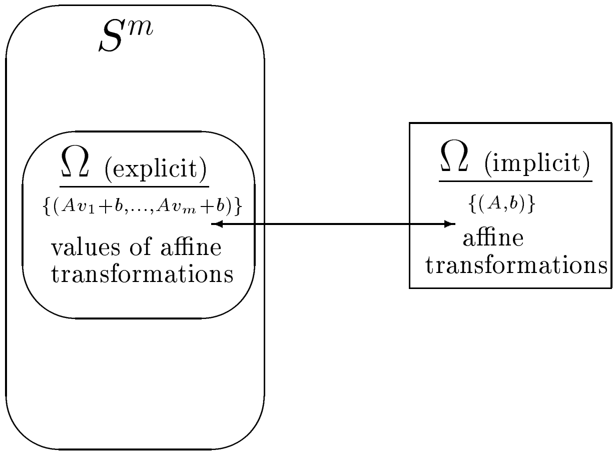
Remark. In this definition (and likewise in the 1st construction) Ω is given in two representations:

1. *Implicit representation* (as the set of affine transformations; or polynomials as before).

This is how we define Ω, but in this representation Ω is not a subset of the initial sample space S^m. Implicit representation is a compact way of writing elements of Ω, and it allows us, for example, to generate all elements in Ω.

2. *Explicit representation* (as affine transformations, evaluated at m points; or evaluated polynomials). It is our implicit representation, now embedded into S^m. For Ω, written in this way, we can speak about pairwise independence.

Both representations are schematically presented in the picture:



Let's return to our new construction. Matrix A has n rows and l columns, vector b has length n. The sequence of m distinct vectors $\{v_i\}_{i=1}^m$ (each of length l) can be chosen, for example, to include binary numbers in canonical order (i.e. v_i is binary representation of number i).

So the sequence $(Av_1 + b, \ldots, Av_m + b)$ consists of m vectors of length n, and lies in $S^m = (\{0,1\}^n)^m$.

The parameter l is typically chosen to be $\lceil \lg_2 m \rceil$. We do not assume here any relation between m and n, so both cases $l \geq n$ and $l < n$ can happen.

Let's calculate the size of Ω. The number of binary $n \times l$ matrices A is 2^{nl}, the number of binary n-dimensional vectors b is 2^n, so

$$|\Omega| = \text{number of pairs } (A,b) = 2^{nl+n} \cong 2^{n(1+\lg m)},$$

$$|S^m| = (2^n)^m = 2^{nm} \gg |\Omega|.$$

This shows that we have achieved our goal of finding *small* sample space. Yet in the first construction we had size of k-wise independent sample space $|\Omega| = 2^{kn} = 2^{2n}$ for $k = 2$, which is considerably better. We'll address the issue of improving $|\Omega|$ in Sect. 2, but first let's check whether we indeed obtained a pairwise independent space.

1.3 Analysis: pairwise independence

Let's formulate the condition of pairwise independence as applied to the constructed Ω (from here on A is assumed to be $l \times n$ boolean matrix and b a boolean vector of length n):

$$\forall i \neq j \ \forall \alpha, \beta \in S \quad \Pr_{(A,b)}[Av_i + b = \alpha \wedge Av_j + b = \beta] = \frac{1}{|S|^2}.$$

It's convenient for us to formulate and proof another claim (clearly a more general one).

Claim 1.1 (pairwise independence of the construction)

$$\forall u \neq v \in \{0,1\}^l \quad \forall \alpha, \beta \in S = \{0,1\}^n$$
$$\Pr_{(A,b)}[Av + b = \alpha \wedge Au + b = \beta] = \frac{1}{|S|^2}.$$

Proof:

Step 1. We present a simplified expression for $\Pr_{(A,b)}[Av + b = \alpha \wedge Au + b = \beta]$.

Let's rewrite our probability in the following form:

$$\begin{aligned} &\Pr_{(A,b)}[Au + b = \alpha \wedge Av + b = \beta] \\ &\quad = \Pr_{(A,b)}[Av + b = \alpha] \cdot \Pr_{(A,b)}[Au + b = \beta | Av + b = \alpha] \end{aligned} \tag{1}$$

Consider the multipliers in the expression (1) one by one.

1. b is chosen uniformly, A and b are independent, whereas v and α are fixed, so $Av + b$ is uniformly distributed as well. So $\Pr_{(A,b)}[Av + b = \alpha] = \frac{1}{|S|}$. Let's show this more formally in two steps.
   First, for any fixed matrix A we have $\Pr_b[b = \alpha - Av] = \frac{1}{|S|}$ (because b is uniformly distributed on S).
   Second, A and b are independent, so from the fact that $\Pr_b[Av + b = \alpha] = \frac{1}{|S|}$ for any fixed A it follows, that this probability is also $\frac{1}{|S|}$ when A is random. This follows immediately by expanding as follows:
   $$\begin{aligned} \Pr_{(A,b)}[Av + b = \alpha] &= \sum_{\mathcal{A}} \Pr_{(A,b)}[b = \alpha - Av \mid A = \mathcal{A}] \cdot \Pr_A[A = \mathcal{A}] \\ &= \sum_{\mathcal{A}} \Pr_b[b = \alpha - \mathcal{A}v] \cdot \Pr_A[A = \mathcal{A}] \\ &= \frac{1}{|S|} \sum_{\mathcal{A}} \Pr_A[A = \mathcal{A}] \\ &= \frac{1}{|S|} \end{aligned}$$

2. In $\Pr_{(A,b)}[Au + b = \beta | Av + b = \alpha]$ the event $Au + b = \beta$ is considered only under the condition $Av + b = \alpha$, so we can eliminate b by substitution $b = \alpha - Av$:
   $$\begin{aligned} &\Pr_{(A,b)}[Au + b = \beta | Av + b = \alpha] \\ &\quad = \Pr_{(A,b)}[A(u - v) = \beta - \alpha | b = \alpha - Av] \\ &\quad = \Pr_{(A,b)}[Aw = \gamma | b = \alpha - Av] \end{aligned} \tag{2}$$
   where $w \stackrel{\text{def}}{=} u - v \neq 0$ and $\gamma \stackrel{\text{def}}{=} \beta - \alpha$ are fixed vectors.
   Now we see that event $b = \alpha - Av$ in (2) is totally irrelevant to $Aw = \gamma$. Hence $\Pr_{(A,b)}[Aw = \gamma | b = \alpha - Av] = \Pr_A[Aw = \gamma]$ (this can be shown by using Bayesian Rule[1]).

Thus, we obtained the following expression:

$$\Pr_{(A,b)}[Au + b = \alpha \wedge Av + b = \beta] = \frac{1}{|S|} \Pr_A[Aw = \gamma] \tag{3}$$

[1] $\Pr[Aw = \gamma \mid b = \alpha - Av] = \frac{\Pr[b = \alpha - Av \mid Aw = \gamma] \cdot \Pr[Aw = \gamma]}{\Pr[b = \alpha - Av]} = \Pr[Aw = \gamma]$,
because for every A $\Pr_b[b = \alpha - Av \mid Aw = \gamma] = \Pr_b[b = \alpha - Av] = 2^{-n}$

where w and γ are completely arbitrary, except for the fact that $w \neq 0$ (this is of course crucial). *Notice, that this result was obtained without using any assumptions about the internal structure of A (e.g., the distribution of its elements and so on).*

Now in order to prove our claim, we need only to show that $\Pr_A[Aw = \gamma] = \frac{1}{|S|}$.

Step 2. Prove that $\Pr_A[Aw = \gamma] = \frac{1}{|S|}$.

Since $w \neq 0$ there is a non-zero entry somewhere in w. Let the i'th element of w be not zero (i.e. be equal to 1).

Look at the i'th column of A. A is totally random, and its i'th column is also random and independent of all other columns of A. This column will add to Aw, consequently Aw is also fully random (uniformly distributed).

More formally we can write it in the following way. Let $A = (C_1, \ldots, C_l)$, where $C_1, \ldots C_l$ are columns of A. They are selected independently and uniformly of one another.

$$\Pr_A[Aw = \gamma] = \Pr_{C_1,\ldots,C_l}\Big[\sum_j C_j w_j = \gamma\Big],$$

$$\forall C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_l \quad \Pr_{C_i}\Big[C_i = \gamma - \sum_{j \neq i} C_j w_j\Big] = \frac{1}{|S|}.$$

Now the same argument as in Step 1 works: the probability for every fixed tuple of values of non-i'th columns is $\frac{1}{|S|}$, hence it is not changed when $\{C_j\}_{j \neq i}$ are random.

Step 1 together with Step 2 complete the proof of the claim. We also noticed that conclusions of Step 1 were made without exploiting information about internal structure of matrix A, while Step 2 relies on the fact that the columns of A are independent.

■

So now we know that Ω is pairwise independent. Next we will consider improvement of the construction, that enables us to pass from $|\Omega| \sim 2^{nl}$ to $|\Omega| \sim 2^{n+l}$.

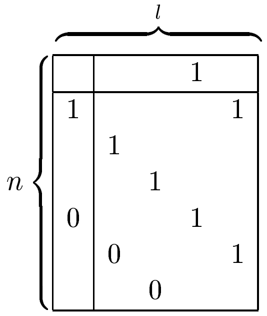
2 Improvement of initial construction (Toeplitz matrices)

2.1 Improved construction

To achieve smaller size of Ω we want to constrain the matrix space. For this purpose let's consider Toeplitz matrices, i.e. matrices with equal elements on all diagonals. The construction will be the same, but we will use a small subspace of matrices.

Definition 1 *$n \times l$ matrix $T = \{t_{i,j}\}$ is called Toeplitz if for every i, j $(1 \leq i \leq n-1, 1 \leq j \leq l-1)$ it holds that $t_{i,j} = t_{i+1,j+1}$.*

The following picture illustrates the notion of Toeplitz matrix.



A Toeplitz matrix is fully specified by $n + l - 1$ bits (for example, by values in the 1st row and the 1st column). So there are 2^{n+l-1} such matrices, which gives the size of our sample space $|\Omega| = 2^{l+2n+1}$ — comparable with 2^{2n} that we had in the first construction, and much less than 2^{nl+n} in the case of arbitrary affine transformations.

2.2 Analysis: pairwise independence

The proof of pairwise independence is very similar to the analogous proof for the case of arbitrary matrices (Claim 1.1). Part of it will be literally the same, and the rest will require some extra elaboration.

Here and below we will denote by T $n \times l$ boolean Toeplitz matrix and by b boolean vector of length n.

Claim 2.1 (pairwise independence of the Toeplitz matrices construction)

$$\forall u \neq v \in \{0,1\}^l \quad \forall \alpha, \beta \in S = \{0,1\}^n$$
$$\Pr_{(T,b)}[Tv + b = \alpha \wedge Tu + b = \beta] = \tfrac{1}{|S|^2}.$$

Proof:

Step 1. As we notices in the proof of Claim 1.1, Step 1 doesn't rely on the kind of matrices we deal with. Thus the results obtained there remain true if arbitrary matrix A is replaced by Toeplitz matrix T. So, we have "for free" the following conclusion (Eq. (3) of Claim 1.1):

$$\Pr_{(T,b)}[Tu + b = \alpha \wedge Tv + b = \beta] = \frac{1}{|S|}\Pr_T[Tw = \gamma]$$

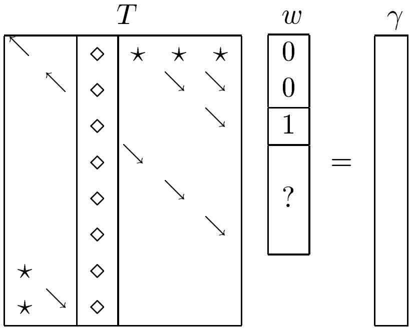
where w and γ are completely arbitrary, but $w \neq 0$.

Step 2. We show that $\Pr_T[Tw = \gamma] = \frac{1}{|S|}$, where $w \neq 0$ and γ are some fixed vectors (if $w = 0$ then $Tw = 0$ independently of random choice of T, so condition that $w \neq 0$ is crucial).

We cannot simply repeat the argument we had for the case of arbitrary random matrices A, because the columns of T are no longer independent.

Now instead of using an arbitrary non-zero element in w we select a particular one. Let it be the first one, let its index be i.

According to this given i (recall that w and γ are fixed, and only T is random) we choose specific representation of T: the last $i-1$ elements of the 1st column, all elements of the i'th column, and the last $l-i$ elements of the 1st row, as described in the picture (arrows describe how values "spread" over Toeplitz matrix):



So we can describe T by $\diamond$ and $\star$ parameters: $T \equiv (\diamond, \star)$, where $\diamond$ is the i'th column in T, $\star$ – the other part of the representation. These two parts of representation are independent from each other (for uniformly chosen T). We will show below a procedure, which for a fixed $\star$ determines $\diamond$ uniquely (from the equation $T(\diamond, \star)w = \gamma$). This implies that for any value of $\star$, the conditional probability $\Pr[Tw = \gamma | \star]$ equals $\frac{1}{|S|}$ (because $Tw = \gamma$ for given $\star$ determines a unique column $\diamond$ out of the $2^l = |S|$ possible ones in the residual probability spcae). This fact combined with the independence of $\star$ and $\diamond$ will give us what we need: $\Pr[Tw = \gamma] = \frac{1}{|S|}$ (by the same argument as in the Claim 1.1).

To complete the proof we only need to describe the procedure of solving equation $T(\diamond, \star)w = \gamma$ with respect to $\diamond$.

We will solve this equation row-by-row from top to bottom. First $i-1$ columns of T don't influence Tw (because they correspond to elements of w that are zero), so we can simply ignore them. Also taking into account that $w_i = 1$, we can write equation $Tw = \gamma$ in the following detailed form:

$$\begin{cases} t_{1,i} + t_{1,i+1}w_{i+1} + \ldots + t_{1,l}w_l = \gamma_1 \\ t_{2,i} + t_{2,i+1}w_{i+1} + \ldots + t_{2,l}w_l = \gamma_2 \\ \cdots \\ t_{j,i} + t_{j,i+1}w_{i+1} + \ldots + t_{j,l}w_l = \gamma_j \\ \cdots \\ t_{n,i} + t_{n,i+1}w_{i+1} + \ldots + t_{n,l}w_l = \gamma_n \end{cases} \tag{4}$$

The leftmost terms of these equations form precisely our unknown $\diamond = \{t_{1,i}, t_{2,i}, \ldots, t_{n,i}\}$.

Let's now look at the first row in (4). All values $t_{1,i+1}, \ldots, t_{1,l}$ are known (they are from $\star$, fixed in advance), hence we determine $t_{1,i}$ (the first element of $\diamond$) in the obvious way: $t_{1,i} = \gamma_1 - t_{1,i+1}w_{i+1} - \ldots - t_{1,l}w_l$.

But now, since we know the value of $t_{1,i}$, we can "propagate" it down the diagonal (i.e., it determines $t_{2,i+1}$). Then in the second row equation in (4) will remain only one unknow $t_{2,i}$.

In general, suppose we have found a unique specification for $t_{1,i}, \ldots, t_{j,i}$, $j < n$. Since we work with a Toeplitz matrix, this gives us values of all elements of T in the $(j+1)$'s row except for $t_{j+1,i}$ (black circles $\bullet$ in the picture below represent already known values of $\diamond$):

$$\begin{array}{rc|cccc|}
\cline{3-6}
1 \rightarrow & \bullet & \star & \star & \star & \star \\
 & \bullet & \searrow & \searrow & & \\
 & \bullet & \searrow & \searrow & \searrow & \\
j+1 \rightarrow & \diamond & \searrow & \searrow & \searrow & \searrow \\
 & \diamond & & & & \\
 & \diamond & & & &
\end{array}$$

So we find a unique specification for $t_{j+1,i}$ from $(j+1)$'s equation in (4):

$$t_{j+1,i} = \gamma_{j+1} - t_{j+1,i+1}w_{i+1} - \ldots - t_{j+1,l}w_l$$

(we showed that all values on the right-hand side are known from the previous steps).

Proceeding in this manner to the n's row, we determine the whole $\diamond$ element by element.

■

Remark: The construction of small pairwise independent sample spaces, presented here, is certainly not limited to binary strings ($S = \{0,1\}^n$). We can be easily extended to sequences over any field $\mathcal{F}$ (for example, over field $\mathcal{F} = \{0, \ldots, p-1\}$ with usual modular multiplication and addition, where p is a prime). We can simply replace 2-element field $\mathcal{F}_2 = \{0,1\}$ by an arbitrary field $\mathcal{F}$ (then $S = \mathcal{F}^n$), and all our arguments will remain valid (because they only depend on the distinction between zero and non-zero elements). The same remark refers to the first construction.

3 Hash Functions

3.1 Small pairwise independent sample spaces viewed as hash functions

Consider sequences of m elements over S as functions from $[m]$ to S. Then we can write our small sample space Ω as space of hash functions:

$$\Omega = \{h : [m] \rightarrow S\} = H$$

or, in our case (i.e. the second construction), $\Omega = \left\{h : \{0,1\}^l \rightarrow \{0,1\}^n\right\}$. We obtain hash functions (i.e. shrinking of input) when $l > n$.

Even in the first construction of small pairwise independent sample space, using procedure of expansion to the case $m > |S|$ (recall that the basic case was $m \leq |S|$), we get hash functions. If we have an embedding μ of large space S' (identified with a finite field), $|S'| \geq m$ into our small initial space S, and polynomial P over field of size m, we get hash function $h = \mu \circ P$ from S' to S.

This gives us the way of obtaining hash functions when we need k-wise independence, $k \neq 2$ (in this case the less problematic second construction doesn't work).

3.2 Good hash functions

Hash functions map long strings into short ones. We expect from hash functions a sort of "justice": each short string is hit by approximately the same number of long strings.

Definition 2 *Hash function h is called k-wise independent, if for any distinct $i_1, \ldots, i_k$ and any $\alpha_1, \ldots, \alpha_k$*

$$\Pr_{h \in_R H}\left[h(i_1) = \alpha_1 \wedge \cdots \wedge h(i_k) = \alpha_k\right] = \frac{1}{|S|^k}$$

In the first construction, the representation of k-wise independent hash function is of size $k \cdot l$. When considering $k = 2$ this representation is compatible to $l + 2n + 1$, obtained in our second construction.

Hash functions can be computed rather efficiently, in polynomial time or even better.

3.3 Applications of hash functions

Typically in applications we want to map large space $\{0,1\}^l$ into small space $\{0,1\}^n$ with hash functions, so that this mapping looks like random.

To define an arbitrary (boolean) hash function we need 2^l bits – to specify value at every point. To define a k-wise independent hash function we need only kl bits.

But usually we want even more: we have a subset $I \subset \{0,1\}^l$, and wish our mapping to be uniform also with respect to this subset I. If $|I| < 2^n$ we want that there will be no collisions, and even when I is bigger than the target set, we still want the mapping to distribute I evenly.

This case (I is a bit larger than $\{0,1\}^n$) often appears in complexity theory.

It turns out, that even pairwise independence gives us a considerable degree of uniformness in distribution, as states the following Lemma 3.1.

3.4 Hashing Lemma

Lemma 3.1 *Let $H = \{h : [m] \to \{0,1\}^n\}$ be a family of pairwise independent hash functions* (we can think of $m > 2^n$ although the Lemma holds for general case).

Let $I \subseteq [m]$ (typically $|I| > 2^n$) *and $\mu \stackrel{\text{def}}{=} \frac{|I|}{2^n}$* (the number of hits expected at a particular point, usually $\mu > 1$). THEN *for every $\alpha \in \{0,1\}^n$*

$$\Pr_{h \in H}\left[\Big| |\{i \in I \ : \ h(i) = \alpha\}| - \mu \Big| > \epsilon\mu\right] < \frac{1}{\epsilon^2 \mu}.$$

(Recall: μ is the expected number of hits; this is meaningful when $\epsilon > \frac{1}{\sqrt{\mu}}$ in which case the right-hand side is less than 1.)

<u>*Remark*</u> *For $\epsilon = \mu^{-\frac{1}{3}}$ we obtain particular case of the formula,*

$$\Pr_{h \in H}\left[\Big| |\{i \in I \ : \ h(i) = \alpha\}| - \mu \Big| > \mu^{\frac{2}{3}}\right] < \mu^{-\frac{1}{3}}$$

which is non-trivial for $\mu > 1$ (then $\mu^{\frac{2}{3}}$ is smaller than μ).

Proof: The main thing in this proof is to define an appropriate random variable and makes some simple observations. After that the conclusion comes out naturally.

Fixing $\alpha \in \{0,1\}^n$, define for each $i \in I$ a random variable X_i:

$$X_i \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } h(i) = \alpha \\ 0 & \text{otherwise} \end{cases}$$

(the sample space here is H, and so X_i is actually $X_i(h)$).

Notice the following facts:

1. $\Pr[X_i = 1] = 2^{-n}$, because for fixed i and random $h \in H$ the value $h(i)$ is uniformly distributed in $\{0,1\}^n$. Therefore $\mathbf{E}(X_i) = 1 \cdot \Pr[X_i = 1] + 0 \cdot \Pr[X_i = 0] = 2^{-n}$.

2. The X_i's are pairwise independent (they are functions of pairwise independent values, i.e. the values $h(i)$'s).

3. $\mathrm{VAR}(X_i) = \mathbf{E}(X_i^2) - \mathbf{E}(X_i)^2 = \mathbf{E}(X_i) \cdot (1 - \mathbf{E}(X_i)) \leq \mathbf{E}(X_i)$ *(valid for any $\{0,1\}$-valued random variables).*

4. For a random choice of h the number of elements from I that hit a particular element from $\{0,1\}^n$ equals $|\{i \in I \,:\, h(i) = \alpha\}| = \sum_{i \in I} X_i$.

5. $\mathbf{E}\Big(\sum_{i \in I} X_i\Big) = \sum_{i \in I} \mathbf{E}(X_i) = \frac{|I|}{2^n} = \mu$.

6. $\mathrm{VAR}\Big(\sum_{i \in I} X_i\Big) = \sum_{i \in I} \mathrm{VAR}(X_i) \leq \sum_{i \in I} \mathbf{E}(X_i) \leq \mu$.

These facts together with Chebishev's inequality (applied to $\sum_{i \in I} X_i$) imply:

$$\Pr\left[\left|\sum_{i \in S} X_i - \mathbf{E}\Big(\sum_{i \in S} X_i\Big)\right| > \epsilon\mu\right] < \frac{1}{(\epsilon\mu)^2} \cdot \mathrm{VAR}\Big(\sum_{i \in I} X_i\Big) \leq \frac{\mu}{\epsilon^2\mu^2} = \frac{1}{\epsilon^2\mu}$$

and the lemma follows. ■

Oded's Note: *Using a k-wise independent family of hash functions allows to prove a stronger bound than the one provided by Lemma 3.1. Specifically, for such a family H and notations as in Lemma 3.1 the following holds:* for every $\alpha \in \{0,1\}^n$

$$\Pr_{h \in H}\Big[\Big|\,|\{i \in I \,:\, h(i) = \alpha\}| - \mu\Big| > \epsilon\mu\Big] < \left(\frac{k^2}{\epsilon^2\mu}\right)^{k/2}$$

The proof follows by using the stronger bounds provided in Lecture 2 for the sum of k-wise independent random variables. In particular, using the union bound and $k = n$, we get

$$\Pr_{h \in H}\Big[\exists \alpha \in \{0,1\}^n \text{ s.t. } \Big|\,|\{i \in I \,:\, h(i) = \alpha\}| - \mu\Big| > \epsilon\mu\Big] < \left(\frac{4k^2}{\epsilon^2\mu}\right)^{k/2}$$

which will be used in Lecture 6.