Randomized Methods in Computation*
Lecture 7: Small Bias Sample Spaces

Notes taken by Yehuda Lindell and Alon Rosen

Spring 2001

Summary: In this lecture we introduce the notion of ϵ-bias sample spaces. Informally, a random variable over $\{0,1\}^n$ is ϵ-bias if for every subset of bits, the difference between the probability that the parity of the bits is zero and the probability it is one is at most ϵ. We show a connection between ϵ-bias and the statistical distance to the uniform distribution. Next, we present an application of ϵ-bias sample spaces for generating "approximately" k-wise independent distributions (i.e., distributions over $\{0,1\}^n$ in which every k bits look almost uniformly distributed over $\{0,1\}^k$).

1 Introduction

In previous lectures, the notion of k-wise independent distributions was considered. Such distributions are a relaxation of the uniform distribution in that any subset of size k is uniformly distributed (and yet nothing can be said of subsets of size larger than k). The motivation for considering such relaxations is the desire to obtain (efficiently constructible) "small" sample spaces (where by small, we mean of size polynomial in n where, for example, the space is distributed over $\{0,1\}^n$).

In this lecture, we introduce a different relaxation of uniform distributions, called ϵ-bias. Loosely speaking, we say that a random variable (over $\{0,1\}^n$) is ϵ-bias if for *every* subset of bits, the difference between the probability that the parity of the bits is 0 and 1 is at most ϵ. This is a natural relaxation of the uniform distribution, for which the probability that any subset of bits has parity 0 equals exactly 1/2 (likewise for parity 1). In other words, the uniform distribution has zero bias. As we shall see in this lecture, the converse is also true and thus a random variable is uniformly distributed *if and only if* it has zero bias. Therefore, the notion of bias can be used to fully characterize the uniform distribution.

As we have mentioned, a zero-bias sample space is uniform. A natural question to ask is how to characterize the distribution of an ϵ-bias sample space, for ϵ that is strictly greater than zero. Indeed, the bulk of this lecture focuses on understanding the "distance" of an ϵ-bias sample space from the uniform distribution (where the notion of distance used is the classical one of statistical difference).

We also present an application of ϵ-bias sample spaces (another more "direct" application is presented in the next lecture). Loosely speaking, we show that "almost" k-wise independent spaces can be derived from ϵ-bias sample spaces, where almost k-wise independent spaces are such that when considering any subset of size k, the resulting distribution is "approximately" uniform (rather

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

than being truly uniform as in the case of k-wise independence). We note that such spaces can be made much smaller than is possible for (fully) k-wise independent spaces.

2 Definitions

2.1 Statistical Difference

We begin by defining a natural measure of the distance between distributions. This measure, called the statistical difference (or variation distance) is the traditional way of measuring the distance between distributions. We note that, as discussed in Lecture 1, we abuse notation and view a random variable as a distribution (or alternatively, a sample space), rather than as a function mapping n-bit strings to $[0,1]$.

Definition 1 (statistical difference): *Let $X, Y \in \{0,1\}^n$ be random variables. The* statistical difference *between X and Y is defined by the function:*

$$\Delta(X,Y) \stackrel{\text{def}}{=} \frac{1}{2} \cdot \sum_{\alpha \in \{0,1\}^n} |\Pr[X = \alpha] - \Pr[Y = \alpha]|$$

Observe that removing the absolute value from the definition would render it meaningless (as it would always sum to zero). A seemingly strange part of this definition is the constant factor of 1/2. Firstly, this factor provides a normalization of the statistical difference in that for every X and Y, it holds that $0 \leq \Delta(X,Y) \leq 1$. Secondly, including this factor results in the above definition being equivalent to the following one:

Definition 2 (statistical difference – alternative definition): *Let $X, Y \in \{0,1\}^n$ be random variables. The* statistical difference *between X and Y is defined by,*

$$\Delta(X,Y) \stackrel{\text{def}}{=} \max_{S \subseteq \{0,1\}^n} \{\Pr[X \in S] - \Pr[Y \in S]\}$$

Proposition 2.1 *Definitions 1 and 2 are equivalent. That is, for all random variables X and Y, $\Delta_1(X,Y) = \Delta_2(X,Y)$, where Δ_i denotes the statistical difference by Definition i.*

Proof: We begin by observing that the set S for which the value $(\Pr[X \in S] - \Pr[Y \in S])$ is maximum has the following property:

1. For every $\alpha \in S$, it holds that $\Pr[X = \alpha] \geq \Pr[Y = \alpha]$, and
2. For every $\alpha \notin S$, it holds that $\Pr[X = \alpha] \leq \Pr[Y = \alpha]$.

Intuitively, the above property holds because otherwise an α contradicting it could be moved from S to $\overline{S}$ (or the reverse) and a "larger than maximum" value would be obtained. That is, if there exists an $\alpha \in S$ for which $\Pr[X = \alpha] < \Pr[Y = \alpha]$, then it is easy to verify that

$$\Pr[X \in S \setminus \{\alpha\}] - \Pr[Y \in S \setminus \{\alpha\}] > \Pr[X \in S] - \Pr[Y \in S]$$

in contradiction to the maximality of the value $(\Pr[X \in S] - \Pr[Y \in S])$ induced by S. An analogous argument holds for item (2) of the above property.

We now claim that for every set S for which the above property holds, we have that

$$\Pr[X \in S] - \Pr[Y \in S] = \frac{1}{2} \cdot \sum_{\alpha \in \{0,1\}^n} |\Pr[X = \alpha] - \Pr[Y = \alpha]| = \Delta_1(X, Y) \tag{1}$$

Eq. (1) follows from the following simple calculation:

$$\begin{aligned}
\Pr[X \in S] - \Pr[Y \in S] &= \frac{1}{2} \cdot (\Pr[X \in S] - \Pr[Y \in S] - \Pr[Y \in S] + \Pr[X \in S]) \\
&= \frac{1}{2} \cdot (\Pr[X \in S] - \Pr[Y \in S] + \Pr[Y \notin S] - \Pr[X \notin S]) \\
&= \frac{1}{2} \cdot \left(\sum_{\alpha \in S} (\Pr[X = \alpha] - \Pr[Y = \alpha]) + \sum_{\alpha \notin S} (\Pr[Y = \alpha] - \Pr[X = \alpha]) \right) \\
&= \frac{1}{2} \cdot \sum_{\alpha \in \{0,1\}^n} |\Pr[X = \alpha] - \Pr[Y = \alpha]|
\end{aligned}$$

where the last equality is due to the fact that the above property holds with respect to S. This concludes the proof of the proposition because we have that for S for which $\Pr[X \in S] - \Pr[Y \in S]$ is maximum, Eq. (1) holds. This implies that $\Delta_1(X, Y) = \Delta_2(X, Y)$. ∎

Notation: We denote by U_n the uniform distribution over n-bit strings. In this lecture, we are interested in distributions X such that $\Delta(X, U_n)$ is small (clearly, $\Delta(X, U_n) = 0$ if and only if $X = U_n$).

> **An advanced comment:** The sets S in Definition 2 can be viewed as (not necessarily efficiently computable) tests for the statistical difference between X and Y. That is, for each set S, associate a distinguisher D_S who upon input α outputs 1 if and only if $\alpha \in S$. Then, it follows that
>
> $$\Pr[D_S(X) = 1] - \Pr[D_S(Y) = 1] = \Pr[X \in S] - \Pr[Y \in S]$$
>
> However, notice that such a distinguisher may not be efficient (i.e., polynomial-time). This is because there may not necessarily be efficient procedures for recognizing whether or not $\alpha \in S$.
>
> The notion of computational indistinguishability, in a specific incarnation that states that no polynomial-size circuit can distinguish between two distribution ensembles $\{X_n\}_{n \in \mathbb{N}}$ and $\{Y_n\}_{n \in \mathbb{N}}$ (where for every n, X_n and Y_n range over strings of length polynomial in n), is a relaxed notion of statistical closeness. (We say that $\{X_n\}$ and $\{Y_n\}$ are statistically close if $\Delta(X_n, Y_n)$ is a negligible function in n.) To see this, first notice that for every circuit C_n and every n, it holds that C_n can distinguish X_n from Y_n with probability at most $\Delta(X_n, Y_n)$. This is because we can define $S_n = \{\alpha \mid C_n(\alpha) = 1\}$ and thus it follows that for every n,
>
> $$|\Pr[C_n(X_n) = 1] - \Pr[C_n(Y_n) = 1]| = |\Pr[X_n \in S_n] - \Pr[Y_n \in S_n]| \leq \Delta(X_n, Y_n)$$
>
> From this, we can conclude that if $\{X_n\}$ and $\{Y_n\}$ are statistically close, then they are also computationally indistinguishable. (Notice that the notion of computational indistinguishability, as described above, refers to polynomial-size circuits. Nevertheless, this

implies computational indistinguishability by probabilistic polynomial-time algorithms, since one may hardwire into a circuit the algorithm's "best" coin-tosses. Thus, we are guaranteed that the distinguishing gap of the circuit is at least as good as that of the algorithm.)

2.2 ϵ-Bias Sample Spaces

We introduce the notion of ϵ-bias sample spaces or random variables. (We will move freely between the notions of sample spaces and random variables, with the intention being that the random variable is defined according to the distribution of the corresponding sample space.) Loosely speaking, the bias of a random variable, with respect to a subset of indices I, equals the difference between the probability that the parity of the bits indexed by I equals zero and the probability that it equals one. An ϵ-bias random variable is one for which the bias (over all subsets of indices) is at most ϵ. In some sense this has the flavor of uniformity in that the parity of any subset of bits of a uniformly distributed string is unbiased to 0 or 1. In fact, the following two conditions are equivalent:

1. The bits of the random variable $X = X_1, \ldots, X_n$ are independent and uniformly distributed.
2. For every subset of bits of X, it is equally likely that the parity of the subset is either zero or one.

(The fact that the above two properties are equivalent is proven in the next section, see Corollary 6.) Therefore, ϵ-bias can be viewed as some measure of distance from the uniform distribution. We stress that although there is some connection between ϵ-bias and statistical distance from uniform, the notions are not at all equivalent.

Definition 3 (bias): *Let $X = X_1, \ldots, X_n \in \{0,1\}^n$ be a random variable, and let $I \subseteq [n]$ (where $[n]$ denotes the set $\{1, \ldots, n\}$). The* bias *of the random variable X, relative to the subset I, is defined by,*

$$\mathrm{bias}_I(X) \stackrel{\mathrm{def}}{=} \Pr\left[\bigoplus_{i \in I} X_i = 0\right] - \Pr\left[\bigoplus_{i \in I} X_i = 1\right]$$

where by convention for any $(x_1, \ldots, x_n)$, it holds that $\bigoplus_{i \in \emptyset} x_i \equiv 0$.

An equivalent definition of bias is given by:

$$\mathrm{bias}_I(X) \stackrel{\mathrm{def}}{=} 2 \cdot \Pr\left[\bigoplus_{i \in I} X_i = 0\right] - 1 \tag{2}$$

The equivalence is derived from the trivial fact that $\Pr[\bigoplus_{i \in I} X_i = 1] = 1 - \Pr[\bigoplus_{i \in I} X_i = 0]$. Notice that $-1 \leq \mathrm{bias}_I(X) \leq 1$ always holds.

The above definition of bias refers to a *specific* subset of bits of X. We now present the definition of max-bias, which extends this to all subsets of X.

Definition 4 (max-bias and ϵ-bias sample spaces): *Let $X \in \{0,1\}^n$ be a random variable. The* max-bias *of X is defined by,*

$$\mathrm{maxbias}(X) \stackrel{\mathrm{def}}{=} \max_{\substack{I \subseteq [n] \\ I \neq \emptyset}} \{|\mathrm{bias}_I(X)|\}$$

We say that X is an ϵ-bias sample space *if* $\mathrm{maxbias}(X) \leq \epsilon$.

Clearly, the max-bias of a uniform space equals 0 (i.e., $\mathrm{maxbias}(U_n) = 0$). As we shall see, the converse is also true. However, as we have mentioned above, it is not true that an ϵ-bias sample space X satisfies $\Delta(X, U_n) \le \epsilon$. Despite this, there is a connection between these notions (as stated in Theorem 5 below).

3 Statistical Difference versus Maxbias

The following theorem demonstrates the connection between ϵ-bias and statistical difference. In particular, it implies that if X is ϵ-bias, then X is at most $\frac{\epsilon \cdot \sqrt{2^n}}{2}$-far from uniform.

Theorem 5 *Let $X \in \{0,1\}^n$ be a random variable. Then,*

$$\mathrm{maxbias}(X) \le 2 \cdot \Delta(X, U_n) \le \sqrt{2^n} \cdot \mathrm{maxbias}(X)$$

We remark that an immediate corollary of Theorem 5 is that a random variable X is uniformly distributed (i.e., $\Delta(X, U_n) = 0$) if and only if $\mathrm{maxbias}(X) = 0$. Thus, the notion of bias offers an alternative characterization of the uniform distribution.

Corollary 6 *A random variable X is uniform if and only if* $\mathrm{maxbias}(X) = 0$.

3.1 Proof of Theorem 5 (Main Result)

We first show that

$$\mathrm{maxbias}(X) \le 2 \cdot \Delta(X, U_n) \tag{3}$$

Let $\epsilon = \mathrm{maxbias}(X)$. Then, there exists a non-empty subset $I \subseteq [n]$ such that $\mathrm{bias}_I(X) = \epsilon$. Now, define $S_I = \{\alpha \mid \bigoplus_{i \in I} \alpha_i = 0\}$. Intuitively, this subset of strings (for which X is biased towards zero) constitutes a "test" for the statistical difference by Definition 2. Specifically,

$$\Pr\left[X \in S_I\right] = \Pr\left[\bigoplus_{i \in I} \alpha_i = 0\right] = \frac{1}{2} + \frac{\epsilon}{2}$$

where the second equality is from the alternative definition of ϵ-bias as shown in Eq. (2). On the other hand, clearly $\Pr[U_n \in S_I] = \frac{1}{2}$. Therefore, we have found a set S for which

$$\Pr[X \in S] - \Pr[U_n \in S] = \frac{\epsilon}{2}$$

Eq. (3) now follows. (We note that it is not necessarily true that $\mathrm{maxbias}(X) = 2 \cdot \Delta(X, U_n)$ because there may be other sets S, for which the difference $\Pr[X \in S] - \Pr[U_n \in S]$ is even greater than $\epsilon/2$.)

We now show that

$$2 \cdot \Delta(X, U_n) \le \sqrt{2^n} \cdot \mathrm{maxbias}(X) \tag{4}$$

In general, probability distributions over $\{0,1\}^n$ are viewed as functions f mapping n-bit strings to $[0,1]$ (and satisfy $\sum_\alpha f(\alpha) = 1$). An alternative view that is instructive for this proof, is to represent such functions as vectors in $\mathbb{R}^N$, where $N = 2^n$ (and $\mathbb{R}$ denotes the set of real numbers). According to this view, the length-N vector representation of a function f is such that the coordinate indexed by $x \in \{0,1\}^n$ contains the value $f(x)$. Specifically, the set $f : \{0,1\}^n \to \mathbb{R}$ of real functions on $\{0,1\}^n$ forms an N-dimensional real vector space where vectors in this space are of the form:

$$v_f \stackrel{\text{def}}{=} \langle f(x) \rangle_{x \in \{0,1\}^n}$$

We call such a vector a probability vector if all its entries are non-negative and they sum-up to 1. From now on, we associate functions f with vectors v_f of the above form. The inner-product of two vectors (functions) f and g is defined by:

$$\langle f, g \rangle = \sum_{x \in \{0,1\}^n} f(x) \cdot g(x)$$

The above way of viewing probability distributions turns out to be useful when analyzing sample spaces. As we will see, there is a correspondence between different bases of such vector spaces and the notions of max-bias and statistical difference. Then, using properties of linear algebra, we derive Eq. (4). We consider two different N-dimensional bases for the above vector space:

1. **Kronecker Basis:** The Kronecker Basis, denoted $\mathcal{K} = \{k_\alpha\}_{\alpha \in \{0,1\}^n}$ is defined as follows:

$$k_\alpha(x) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } x = \alpha \\ 0 & \text{if } x \neq \alpha \end{cases}$$

   This is the standard basis for vector spaces (in vector notation this is the basis in which the i^{th} element of the basis is the vector with a single 1 in the i^{th} place and zeros elsewhere).

2. **Fourier Basis:** The Fourier Basis, denoted $\mathcal{F} = \{f_I\}_{I \subseteq [n]}$ is defined as follows:

$$f_I(x) \stackrel{\text{def}}{=} \frac{1}{\sqrt{N}} \cdot (-1)^{\sum_{i \in I} x_i}$$

In the proof we switch freely between viewing a random variable X as a distribution on $\{0,1\}^n$ (that is, a sample space) and viewing it as a probability function $\pi : \{0,1\}^n \to [0,1]$ (representing the probability that a specific string in $\{0,1\}^n$ is chosen in the sample space).

We now show that both the Kronecker and Fourier bases are orthonormal (this technicality is needed later in the proof).

Fact 3.1 *The Kronecker basis is* orthonormal. *That is, for every $\alpha \neq \beta \in \{0,1\}^n$:*

$$\langle k_\alpha, k_\alpha \rangle = 1 \text{ and } \langle k_\alpha, k_\beta \rangle = 0$$

Proof: Note that for every $\alpha, \beta \in \{0,1\}^n$:

$$\begin{aligned} \langle k_\alpha, k_\beta \rangle &= \sum_{x \in \{0,1\}^n} k_\alpha(x) \cdot k_\beta(x) \\ &= k_\alpha(\beta) \end{aligned}$$

where the last equality is due to the fact that $k_\beta(\beta) = 1$, whereas for every $x \neq \beta$, $k_\beta(x) = 0$. This completes the proof, because $k_\alpha(\alpha) = 1$ and $k_\alpha(\beta) = 0$ for every $\beta \neq \alpha$. ■

Fact 3.2 *The Fourier basis is* orthonormal. *That is, for every $I \neq J \subseteq [n]$:*

$$\langle f_I, f_I \rangle = 1 \text{ and } \langle f_I, f_J \rangle = 0$$

Proof: Let $I \subseteq [n]$. Then:

$$\begin{aligned}
\langle f_I, f_I \rangle &= \sum_{x\in\{0,1\}^n} f_I(x)\cdot f_I(x) \\
&= \tfrac{1}{N}\cdot \sum_{x\in\{0,1\}^n} (-1)^{\sum_{i\in I} x_i}\cdot(-1)^{\sum_{i\in I} x_j} \\
&= \tfrac{1}{N}\cdot \sum_{x\in\{0,1\}^n} (-1)^{2\sum_{i\in I} x_i} \\
&= \tfrac{1}{N}\cdot \sum_{x\in\{0,1\}^n} 1 \\
&= \tfrac{1}{N}\cdot N
\end{aligned}$$

as required. On the other hand, for $I \neq J \subseteq [n]$, we have:

$$\begin{aligned}
\langle f_I, f_J \rangle &= \sum_{x\in\{0,1\}^n} f_I(x)\cdot f_J(x) \\
&= \tfrac{1}{N}\cdot \sum_{x\in\{0,1\}^n} (-1)^{\sum_{i\in I} x_i}\cdot(-1)^{\sum_{j\in J} x_j} \\
&= \tfrac{1}{N}\cdot \sum_{x\in\{0,1\}^n} (-1)^{\sum_{i\in I\nabla J} x_i} \qquad (5)
\end{aligned}$$

where $I\nabla J$ denotes the symmetric difference between sets I and J (that is, $I\nabla J \stackrel{\text{def}}{=} (I\setminus J)\cup(J\setminus I)$). Notice that $I\nabla J \neq \emptyset$ (since $I \neq J$), and that the sum in Eq. (5) ranges over all $x \in \{0,1\}^n$. For, $b \in \{0,1\}$, consider the sets $S^b_{I\nabla J} = \{x \mid \oplus_{i\in I\nabla J}\, x_i = b\}$. Using this notation, we have:

$$\begin{aligned}
\tfrac{1}{N}\cdot \sum_{x\in\{0,1\}^n} (-1)^{\sum_{i\in I\nabla J} x_i} &= \tfrac{1}{N}\cdot\left(\sum_{x\in S^0_{I\nabla J}} (-1)^{\sum_{i\in I\nabla J} x_i} + \sum_{x\in S^1_{I\nabla J}} (-1)^{\sum_{i\in I\nabla J} x_i}\right) \\
&= \tfrac{1}{N}\cdot\left(\left|S^0_{I\nabla J}\right|\cdot 1 - \left|S^1_{I\nabla J}\right|\cdot 1\right)
\end{aligned}$$

The key observation is that if $I\nabla J \neq \emptyset$, then the size of $S^0_{I\nabla J}$ equals the size of $S^1_{I\nabla J}$. This can be seen by showing a bijection between the two sets. A simple such bijection is a function that flips the i^{th} bit of $x \in S^0_{I\nabla J}$, where $i \in I\nabla J$. That is, for an arbitrary $i \in I\nabla J$ define,

$$F_i(x_1\cdots x_n) \stackrel{\text{def}}{=} x_1\cdots x_{i-1}\cdot \overline{x}_i\cdot x_{i+1}\cdots x_n$$

Clearly, for $x \in S^0_{I\nabla J}$, we have $F_i(x) \in S^1_{I\nabla J}$ (the fact that F_i is a bijection is trivial). We conclude that,

$$\begin{aligned}
\langle f_I, f_J \rangle &= \tfrac{1}{N}\cdot\left(\left|S^0_{I\nabla J}\right| - \left|S^1_{I\nabla J}\right|\right) \\
&= \tfrac{1}{N}\cdot 0
\end{aligned}$$

as required. ■

In the following fact, we show that the bias of a probability function π, relative to the subset I, can be stated in terms of the inner-product of π with f_I.

Fact 3.3 *For every $I \subseteq [n]$ and for every probability function $\pi : \{0,1\}^n \to [0,1]$,*

$$\mathrm{bias}_I(\pi) = \sqrt{N} \cdot \langle \pi, f_I \rangle$$

Proof: The key observation for proving the fact is that for $b \in \{0,1\}$, the probability that the parity of x equals b can be computed by summing $\pi(x)$ for all strings x for which the parity of x is indeed b. Denote the set of strings x, for which the parity of the bits of x indexed by I equals b, by S_I^b. That is, $S_I^b \stackrel{\text{def}}{=} \{x \mid \bigoplus_{i\in I} x_i = b\}$. Then, for $b \in \{0,1\}$ and random variable $X = (X_1, \ldots, X_n)$ with distribution π:

$$\Pr\left[\bigoplus_{i\in I} X_i = b\right] = \sum_{x \in S_I^b} \pi(x)$$

We thus have,

$$\begin{aligned} \mathrm{bias}_I(\pi) &= \left(\sum_{x\in S_I^0} \pi(x)\right) - \left(\sum_{x\in S_I^1} \pi(x)\right) \\ &= \sum_{x\in\{0,1\}^n} \pi(x) \cdot (-1)^{\sum_{i\in I} x_i} \end{aligned}$$

where the second equality is due to the fact that for $x \in S_I^b$, it holds that $(-1)^{\sum_{i\in I} x_i} = (-1)^b$, and thus $\pi(x)(-1)^{\sum_{i\in I} x_i} = \pi(x)(-1)^b$. Now, by the definition of the Fourier basis (and the inner-product), it follows that,

$$\begin{aligned} \mathrm{bias}_I(\pi) &= \sum_{x\in\{0,1\}^n} \pi(x) \cdot \sqrt{N} \cdot f_I(x) \\ &= \sqrt{N} \cdot \langle \pi, f_I \rangle \end{aligned}$$

as required. ■

Given Fact 3.3, we are ready to show a connection between maxbias(π) and the $L_\infty^{\mathcal{F}}$ norm (i.e., the L_∞ norm with respect to the Fourier basis $\mathcal{F}$). Following this, we show a connection between $\Delta(\pi, U_n)$ and the $L_1^{\mathcal{K}}$ norm (i.e., the L_1 norm with respect to the Kronecker basis $\mathcal{K}$). We then conclude by presenting a fact from linear algebra connecting between the L_1 and L_∞ norms. We first recall the definition of $L_1^{\mathcal{B}}$ and $L_\infty^{\mathcal{B}}$, for a basis $\mathcal{B}$:

$$\begin{aligned} \| v \|_1^{\mathcal{B}} &\stackrel{\text{def}}{=} \sum_{b\in\mathcal{B}} |\langle v, b\rangle| \\ \| v \|_\infty^{\mathcal{B}} &\stackrel{\text{def}}{=} \max_{b\in\mathcal{B}} \{|\langle v, b\rangle|\} \end{aligned}$$

In general for any integer p (e.g., $p = 2$) the $L_p^{\mathcal{B}}$ norm is defined as:

$$\| v \|_p^{\mathcal{B}} \stackrel{\text{def}}{=} \left(\sum_{b\in\mathcal{B}} |\langle v, b\rangle|^p\right)^{1/p}$$

Now, define the constant function $\mu : \{0,1\}^n \to \mathbb{R}$ such that for every $x \in \{0,1\}^n$, $\mu(x) = \frac{1}{N}$. The function μ is the probability function associated with the uniform distribution U_n. The following claim establishes the connection between the maxbias of a probability function π and the $L_\infty^{\mathcal{F}}$ norm of π (or, actually of $\pi - \mu$).

Claim 3.4 *For every probability function* $\pi : \{0,1\}^n \to \mathbb{R}$,

$$\text{maxbias}(\pi) = \sqrt{N} \cdot \| \pi - \mu \|_\infty^{\mathcal{F}}$$

Proof: By Fact 3.3 we have that

$$\text{maxbias}(\pi) \stackrel{\text{def}}{=} \max_{\substack{I \subseteq [n] \\ I \neq \emptyset}} \{|\text{bias}_I(\pi)|\} = \sqrt{N} \cdot \max_{\substack{I \subseteq [n] \\ I \neq \emptyset}} \{|\langle \pi, f_I \rangle|\} \tag{6}$$

We now make the following two observations regarding the Fourier basis:

1. For every non-empty subset $I \subseteq [n]$, $\langle \mu, f_I \rangle = 0$.

   This is due to the fact that $\langle \mu, f_I \rangle = \frac{1}{N\sqrt{N}} \sum_x (-1)^{\sum_{i \in I} x_i}$. As we have previously seen (in the proof of Fact 3.2), for any non-empty subset I, the above sum equals zero.

2. For every probability vector v, $\langle v, f_\emptyset \rangle = \frac{1}{\sqrt{N}}$ (by the definition of the Fourier basis and the fact that the sum of the elements of v equals 1).

This implies that for every $I \subseteq [n]$,

$$\langle \pi - \mu, f_I \rangle = \langle \pi, f_I \rangle - \langle \mu, f_I \rangle = \begin{cases} \langle \pi, f_I \rangle & \text{if } I \neq \emptyset \\ 0 & \text{if } I = \emptyset \end{cases} \tag{7}$$

Then, combining Eq. (6) and(7) we obtain that

$$\text{maxbias}(\pi) \; = \; \sqrt{N} \cdot \max_{\substack{I \subseteq [n] \\ I \neq \emptyset}} \{|\langle \pi - \mu, f_I \rangle|\} \; = \; \sqrt{N} \cdot \| \pi - \mu \|_\infty^{\mathcal{F}}$$

as required. ■

We now turn to establish the connection between the statistical difference of π from U_n and the $L_1^{\mathcal{K}}$ norm of $\pi - \mu$.

Claim 3.5 *For every probability function* $\pi : \{0,1\}^n \to \mathbb{R}$*:*

$$2 \cdot \Delta(\pi, U_n) \; = \| \pi - \mu \|_1^{\mathcal{K}}$$

Proof: Recall that for every vector v, $\langle v, k_\alpha \rangle = v(\alpha)$. Thus:

$$\begin{aligned} 2 \cdot \Delta(\pi, U_n) &= \sum_{x \in \{0,1\}^n} |\pi(x) - \mu(x)| \\ &= \sum_{x \in \{0,1\}^n} |\langle \pi, k_x \rangle - \langle \mu, k_x \rangle| \\ &= \sum_{x \in \{0,1\}^n} |\langle \pi - \mu, k_x \rangle| \\ &= \| \pi - \mu \|_1^{\mathcal{K}} \end{aligned}$$

as required. ■

We conclude the proof by combining Claims 3.4 and 3.5 with the following technical lemma from linear algebra (proven below).

Lemma 3.6 *For every two orthonormal bases $\mathcal{A}$ and $\mathcal{B}$, and for every vector v:*

$$\| v \|_1^{\mathcal{A}} \leq N \cdot \| v \|_\infty^{\mathcal{B}}$$

Using Lemma 3.6 we have:

$$\begin{aligned} 2 \cdot \Delta(\pi, U_n) &= \| \pi - \mu \|_1^{\mathcal{K}} \\ &\leq N \cdot \| \pi - \mu \|_\infty^{\mathcal{F}} \\ &= \sqrt{N} \cdot \mathrm{maxbias}(\pi) \end{aligned}$$

Eq. (4) follows. Thus, all that remains to complete the proof of Theorem 5 is to prove Lemma 3.6.

3.2 Proof of Lemma 3.6 (Technical Lemma from Linear Algebra)

In order to prove Lemma 3.6 we will make use of the following three facts:

Fact 3.7 *For every basis $\mathcal{A}$:*

$$\| v \|_1^{\mathcal{A}} \leq \sqrt{N} \cdot \| v \|_2^{\mathcal{A}}$$

Proof: We start by recalling the Cauchy-Schwartz inequality. Let $\{\alpha_i\}_{i=1}^N$ and $\{\beta_i\}_{i=1}^N$ be two sequences of real numbers. Then:

$$\left(\sum_{i=1}^N \alpha_i \cdot \beta_i \right)^2 \leq \left(\sum_{i=1}^N \alpha_i^2 \right) \cdot \left(\sum_{i=1}^N \beta_i^2 \right)$$

Now, let $\mathcal{A} = \{a_i\}_{i=1}^N$ be a basis. Then, by the definition of the $L_1^{\mathcal{A}}$ and $L_2^{\mathcal{A}}$ norms and by the Cauchy-Schwartz inequality:

$$\begin{aligned} \left(\| v \|_1^{\mathcal{A}} \right)^2 &= \left(\sum_{i=1}^N |\langle v, a_i \rangle| \right)^2 \\ &\leq \left(\sum_{i=1}^N 1^2 \right) \cdot \left(\sum_{i=1}^N |\langle v, a_i \rangle|^2 \right) \\ &= N \cdot \left(\| v \|_2^{\mathcal{A}} \right)^2 \end{aligned}$$

as required. ■

Fact 3.8 *For every two orthonormal bases $\mathcal{A}$ and $\mathcal{B}$:*

$$\| v \|_2^{\mathcal{A}} = \| v \|_2^{\mathcal{B}}$$

Proof: Intuitively, the proof follows from the fact that for *any* orthonormal basis, $\mathcal{A}$, the $L_2^{\mathcal{A}}$ norm of the vector v is equal to the square root of the inner product of v with itself (regardless of the choice of the orthonormal basis $\mathcal{A}$). This is proved formally as follows. Let $\mathcal{A} = \{a_i\}_{i=1}^N$ be an orthonormal basis and let v be a vector. Then there exist a sequence of real numbers $\{\alpha_i\}_{i=1}^N$

(specifically, $\alpha_i = \langle v, a_i \rangle$) so that v can be written as $\sum_{i=1}^{N} \alpha_i \cdot a_i$. By the definition of the $L_2^{\mathcal{A}}$ norm and by the orthonormality of $\mathcal{A}$ we have:

$$\begin{aligned}
\langle v, v \rangle &= \langle \sum_{i=1}^{N} \alpha_i \cdot a_i, \sum_{j=1}^{N} \alpha_j \cdot a_j \rangle \\
&= \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \cdot \alpha_j \cdot \langle a_i, a_j \rangle \\
&= \sum_{i=1}^{N} \alpha_i^2 \qquad (8) \\
&= \sum_{i=1}^{N} \langle v, a_i \rangle^2 \\
&= \left(\| v \|_2^{\mathcal{A}} \right)^2
\end{aligned}$$

where Eq. (8) follows from the fact that for every $i \neq j$, $\langle a_j, a_i \rangle = 0$ and $\langle a_i, a_i \rangle = 1$. It follows that for every two orthonormal bases $\mathcal{A}$ and $\mathcal{B}$:

$$\left(\| v \|_2^{\mathcal{A}} \right)^2 = \langle v, v \rangle = \left(\| v \|_2^{\mathcal{B}} \right)^2$$

as required. ■

Fact 3.9 *For every basis $\mathcal{B}$:*

$$\| v \|_2^{\mathcal{B}} \leq \sqrt{N} \cdot \| v \|_\infty^{\mathcal{B}}$$

Proof: Let $\mathcal{B} = \{b_i\}_{i=1}^{N}$ be a basis. By definition of the $L_2^{\mathcal{B}}$ and $L_\infty^{\mathcal{B}}$ norms:

$$\begin{aligned}
\left(\| v \|_2^{\mathcal{B}} \right)^2 &= \sum_{i=1}^{N} |\langle v, b_i \rangle|^2 \\
&\leq N \cdot \max_{i \leq i \leq N} \{ |\langle v, b_i \rangle|^2 \} \\
&= N \cdot \left(\max_{i \leq i \leq N} \{ |\langle v, b \rangle| \} \right)^2 \\
&= N \cdot \left(\| v \|_\infty^{\mathcal{B}} \right)^2
\end{aligned}$$

as required. ■

By combining Facts 3.7, 3.8 and 3.9 we obtain:

$$\begin{aligned}
\| v \|_1^{\mathcal{A}} &\leq \sqrt{N} \cdot \| v \|_2^{\mathcal{A}} \\
&= \sqrt{N} \cdot \| v \|_2^{\mathcal{B}} \\
&\leq N \cdot \| v \|_\infty^{\mathcal{B}}
\end{aligned}$$

which completes the proof of Lemma 3.6.

4 (ϵ, k)-Approximations of the Uniform Distribution

4.1 (ϵ, k)-approximation of the uniform distribution

In this section, we introduce the notion of "approximately" uniform distributions and show how they can be derived from ϵ-bias sample spaces. Intuitively, we say that a distribution X is an ϵ-approximation of U_n if $\Delta(X, U_n) \leq \epsilon$. We combine this ϵ-approximation with the k-wise independent relaxation that we have seen in previous lectures. That is, we say that X is an (ϵ, k)-approximation of U_n if, when projecting X on any subset of size k of it coordinates, the resulting distribution is an ϵ-approximation of U_k. Formally,

Definition 7 ((ϵ, k)-approximation of the uniform distribution): *Let $X \in \{0,1\}^n$ be a random variable, let $\epsilon > 0$ and let $k \in [n]$ be an integer. We say that $X = (X_1, \ldots, X_n)$ is an (ϵ, k)-* approximation *of the uniform distribution, if for every subset $I = \{i_1, \ldots, i_k\} \subseteq [n]$ (of size k) the distribution $X_I = \{X_{i_1}, \ldots, X_{i_k}\}$ is ϵ-far from uniform (that is, $\Delta(X_I, U_k) \leq \epsilon$).*

Clearly, an $(0, k)$-approximation of the uniform distribution is a k-wise independent sample space. Furthermore, an (ϵ, n)-approximation of the uniform distribution is what we informally called an ϵ-approximation above. It is easily verified (see Proposition 4.1 below) that for every $\epsilon > 0$ and every $k > 1$, an (ϵ, k)-approximation of the uniform distribution is also $(\epsilon, k-1)$-approximation of the uniform distribution. Essentially, an (ϵ, k)-approximation of uniform is a relaxation of k-wise independence. As we shall see, this further approximation enables us to construct sample spaces that are far smaller than k-wise independent spaces.

Proposition 4.1 *For every $\epsilon > 0$ and every $k > 1$, an (ϵ, k)-approximation of the uniform distribution is also an $(\epsilon, k-1)$-approximation of the uniform distribution.*

Proof: Let $X \in \{0,1\}^n$ be a random variable, let $\epsilon > 0$, let $k > 1$ be an integer and suppose that X is an (ϵ, k)-approximation of the uniform distribution. If X is not an $(\epsilon, k-1)$-approximation of the uniform distribution, then there exists a set $I \subseteq [n]$ of size $k-1$ (for simplicity, assume that $I = \{1, \ldots, k-1\}$) so that the distribution $X_I = \{X_1, \ldots, X_{k-1}\}$ is not ϵ-far from U_{k-1} (that is, $\Delta(X_I, U_{k-1}) > \epsilon$). By the alternative definition of statistical difference (Definition 2) this means that there exists a set $S \subseteq \{0,1\}^{k-1}$ so that:

$$\Pr[X_I \in S] - \Pr[U_{k-1} \in S] > \epsilon \tag{9}$$

For $b \in \{0,1\}$ consider now the set $S_b \subseteq \{0,1\}^k$ that is obtained by concatenating each element $\alpha \in S$ with the bit b (that is, $S_b = \{\alpha \circ b \mid \alpha \in S\}$). Notice that $\Pr[U_k \in S_0 \cup S_1] = \Pr[U_{k-1} \in S]$. Similarly, $\Pr\left[X_{I\cup\{k\}} \in S_0 \cup S_1\right] = \Pr[X_I \in S]$ (where $X_{I\cup\{k\}} = \{X_1, \ldots, X_k\}$ is the distribution that is induced by the k-sized subset $I \cup \{k\}$). Thus, for the set $I \cup \{k\} \subseteq [n]$ of size k, we have:

$$\begin{aligned}
\Delta(X_{I\cup\{k\}}, U_k) &= \max_{S\subseteq\{0,1\}^k} \left\{\Pr\left[X_{I\cup\{k\}} \in S\right] - \Pr[U_k \in S]\right\} \\
&\geq \Pr\left[X_{I\cup\{k\}} \in S_0 \cup S_1\right] - \Pr[U_k \in S_0 \cup S_1] \\
&= \Pr[X_I \in S] - \Pr[U_{k-1} \in S]
\end{aligned}$$

which by Eq. (9) is greater than ϵ. This is in contradiction to the assumption that X is an (ϵ, k)-approximation of the uniform distribution. ∎

4.2 Achieving (ϵ, k)-approximations of uniform

The following corollary of Theorem 5 establishes the relation between the max-bias of X and the quality of the (ϵ, k)-approximation of the uniform distribution achieved by X.

Corollary 8 *Let $X \in \{0,1\}^n$ be a random variable such that $\text{maxbias}(X) \leq \epsilon'$. Then for every $k \in [n]$, X is an (ϵ, k)-approximation of the uniform distribution, where $\epsilon \stackrel{\text{def}}{=} \frac{\sqrt{2^k}}{2} \cdot \epsilon'$.*

Proof: Let $I \subseteq [n]$ be any subset of indices of size k; denote $I = \{i_1, \ldots, i_k\}$. Then, as with k-wise independence, consider the distribution $X_I = \{X_{i_1}, \ldots, X_{i_k}\}$ (as induced by I). Note that, by definition, it holds that $\text{maxbias}(X_I) \leq \text{maxbias}(X)$, implying that $\text{maxbias}(X_I) \leq \epsilon'$. Then, by applying Theorem 5 we have that

$$2 \cdot \Delta(X_I, U_k) \leq \sqrt{2^k} \cdot \text{maxbias}(X_I) \leq \sqrt{2^k} \cdot \epsilon'$$

Therefore, for every $I \subseteq [n]$ of size k, the distribution X_I is $\frac{\sqrt{2^k}}{2} \cdot \epsilon'$-far from uniform. ■

Corollary 8 implies that in order to obtain (ϵ, k)-approximations of uniform, it is enough to construct an ϵ'-bias sample space (for the appropriate ϵ'). In the next lecture, such a construction will be presented. That is, the following theorem will be proven:

Theorem 9 *There exist ϵ-bias sample spaces of size $\text{poly}(n/\epsilon)$. Moreover, such sample spaces can be constructed in $\text{poly}(n/\epsilon)$-time.*

Recall that k-wise independent spaces must be of size at least $\Omega(n^{\lfloor k/2 \rfloor})$. Therefore, efficiently constructible spaces of size $\text{poly}(n)$ can only be k-wise independent for *constant* values of k. On the other hand, combining Corollary 8 and Theorem 9, it is possible to efficiently construct (ϵ, k)-approximations of uniform, with size just $\text{poly}(n \cdot 2^{k/2}/\epsilon)$. An immediation ramification of this is that, unlike for k-wise independence, efficiently constructible polynomial-size $(\frac{1}{\text{poly}(n)}, k)$-approximations of uniform exist for *non-constant* values of k (i.e., for $k = O(\log n)$).

4.3 An application of (ϵ, k)-approximations of uniform

We conclude by briefly showing an application of (ϵ, k)-approximations of the uniform distribution to the problem of kCNF. Suppose that we would like to construct a deterministic algorithm, that given a kCNF formula φ as input finds an assignment that satisfies at least $1 - 2^{-k}$ of the clauses of φ. As we have already seen in Lecture 3, one can use a 3-wise independent sample space in order to solve this problem for $k = 3$ (i.e., in the case of 3CNF).

Specifically, we saw a deterministic algorithm that receives a 3CNF formula φ as input and finds an assignment satisfying at least $7/8$ of the clauses of φ. The algorithm described worked by first constructing a 3-wise independent sample space (of size polynomial in the length of φ). The key observation was that for any clause, the probability that it is satisfied by a random element of this sample space *equals* the probability that it is satisfied by a uniformly chosen assignment. Thus, the probability that a given clause is satisfied is exactly $7/8$. (The derandomization step consisted of simply traversing the entire 3-wise independent sample space and checking the number of clauses satisfied by each assignment from the space.)

By using a k-wise independent sample space (rather than a 3-wise independent space) one can generalize this approach to construct an algorithm that, given a kCNF formula φ, finds an assignment that satisfies at least $1 - 2^{-k}$ of the clauses of φ. The problem is that the size of any k-wise independent space must be at least $\Omega(n^{\lfloor k/2 \rfloor})$. Therefore, the resulting algorithm is polynomial-time only for constant values of k and also in that case its running time is $\Omega(n^{\lfloor k/2 \rfloor})$ (which is more than the alternative presented below). However, by replacing the k-wise independent space with an appropriately chosen (ϵ, k)-approximation of the uniform distribution one can do significantly better. In particular, one can construct a polynomial-time algorithm that generates the required assignment even if k is non-constant (specifically, it will work as long as $k = O(\log n)$). Details follow.

When using an (ϵ, k)-approximation of the uniform distribution (instead of a k-wise independent sample space), the probability that any given clause is satisfied is at least $(1 - 2^{-k}) - \epsilon$, and the expected number of clauses that are satisfied is at least $(1 - 2^{-k}) \cdot n - \epsilon \cdot n$ (where n equals the number of clauses in φ). Now, if we set ϵ such that $\epsilon \cdot n < 2^{-k}$ (e.g., $\epsilon = 1/(10 \cdot 2^k \cdot n)$), then we have that there exists a satisfying assignment in the sample space satisfying $1 - 2^{-k}$ of the clauses. By Corollary 8, we have that an (ϵ, k)-approximation of the uniform distribution can be obtained from a $(2 \cdot \epsilon/\sqrt{2^k})$-bias space. Thus, for $\epsilon = 1/(10 \cdot 2^k \cdot n)$ we need to construct (and scan) an ϵ'-bias sample space with $\epsilon' = 2/(10 \cdot 2^{3 \cdot k/2} \cdot n)$. By Theorem 9 this can be done in time $\text{poly}(n/\epsilon') = \text{poly}(n, 2^k)$, giving us the desired result. We note that $O((n \cdot 2^k)^c)$ is preferrable to $O(n^{\lfloor k/2 \rfloor})$ for every $k > 2 \cdot c$ and sufficiently large n.