$$\begin{aligned}2 \cdot \mathrm{E}\left[\mu(\mathbf{g}_{d+1})\right] &\leqslant \mathrm{E}\left[\mu\left(\mathbf{g}_d^0 \wedge x_{d+1}^0\right)\right] + \mathrm{E}\left[\mu\left(\mathbf{g}_d^0 \vee x_{d+1}^0\right)\right] + \\ &\quad \mathrm{E}\left[\mu\left(\mathbf{g}_d^1 \wedge x_{d+1}^1\right)\right] + \mathrm{E}\left[\mu\left(\mathbf{g}_d^1 \vee x_{d+1}^1\right)\right] \\ &\leqslant \mathrm{E}\left[\mu(\mathbf{g}_d^0)\right] + \mu(x_{d+1}^0) + \mathrm{E}\left[\mu(\mathbf{g}_d^1)\right] + \mu(x_{d+1}^1) \\ &\leqslant 2 \cdot \mathrm{E}\left[\mu(\mathbf{g}_d)\right] + 2 \\ &\leqslant 2d + 4.\end{aligned}$$

This completes the proof of (20.11). But this inequality only says that the expected value of $\mu(\mathbf{g}_n)$ does not exceed $n+1$ for a *random* function $\mathbf{g}_n$, whereas our goal is to give an upper bound on $\mu(f_n)$ for *each* function f_n. So, we must somehow "derandomize" this result. To achieve this goal, observe that every function $f_n \in F_n$ can be expressed in the form

$$f_n = (\mathbf{g}_n \wedge (\mathbf{g}_n \oplus f_n \oplus 1)) \vee ((\mathbf{g}_n \oplus 1) \wedge (\mathbf{g}_n \oplus f_n). \tag{20.16}$$

But $\mathbf{g}_n \approx \mathbf{g}_n \oplus f_n \oplus 1 \approx \mathbf{g}_n \oplus 1 \approx \mathbf{g}_n \oplus f_n$. So, applying to (20.16) the inequalities (20.8) and (20.9), averaging the result over $\mathbf{g}_n$ and applying (20.11) with $d = n$, we obtain $\mu(f_n) = \mathrm{E}\left[\mu(f_n)\right] \leqslant 4 \cdot \mathrm{E}\left[\mu(\mathbf{g}_n)\right] \leqslant 4n + 4$, as desired. □

20.9 Discrepancy

Let $X_1, \ldots, X_k$ be n-element sets, and $X = X_1 \times \cdots \times X_k$. A subset T of X is called a *cylinder* in the ith dimension if membership in T_i does not depend on the ith coordinate. That is, $(x_1, \ldots, x_i, \ldots, x_k) \in T_i$ implies that $(x_1, \ldots, x_i', \ldots, x_k) \in T_i$ for all $x_i' \in X_i$. A subset $T \subseteq X$ is a *cylinder intersection* if it is an intersection $T = T_1 \cap T_2 \cap \cdots \cap T_k$, where T_i is a cylinder in the ith dimension. The (normalized) *discrepancy* of a function $f : X \to \{-1, 1\}$ on a set T is defined by

$$\mathrm{disc}_T(f) = \frac{1}{|X|} \sum_{x \in T} f(x).$$

The *discrepancy* $\mathrm{disc}(f)$ of f is the maximum, over all cylinder intersections T, of the absolute value $|\mathrm{disc}_T(f)|$.

The importance of this measure stems from the fact that functions with small discrepancy have large *multi-party communication complexity.* (We will discuss this in Sect. 29.3.2 devoted to multi-party games.) However, this fact alone does not give immediate lower bounds for the multi-party communication complexity, because $\mathrm{disc}(f)$ is very hard to estimate. Fortunately, the discrepancy can be bounded from above using the following more tractable measure.

A *cube* is defined to be a multi-set $D = \{a_1, b_1\} \times \{a_2, b_2\} \times \cdots \times \{a_k, b_k\}$, where $a_i, b_i \in X_i$ (not necessarily distinct) for all i. Being a multi-set means that one element can occur several times. Thus, for example, the cube $D = \{a_1, a_1\} \times \cdots \times \{a_k, a_k\}$ has 2^k elements. For a cube D, define its *sign* $f(D)$ to be

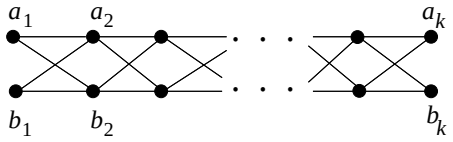

Fig. 20.1. A cube

$$f(D) = \prod_{x \in D} f(x).$$

Hence, $f(D) = 1$ if and only if $f(x) = 1$ for an even number of vectors $x \in D$. We choose a cube $\mathbf{D}$ at random according to the uniform distribution. This can be done by choosing $\mathbf{a}_i, \mathbf{b}_i \in X_i$ for each i according to the uniform distribution. Let

$$\mathcal{E}(f) = \mathrm{E}\,[f(\mathbf{D})]$$

be the expected value of the sign of a random cube $\mathbf{D}$. To stress the fact that the expectation is taken over a particular random object (this time, over $\mathbf{D}$) we will also write $\mathrm{E}_{\mathbf{D}}\,[f(\mathbf{D})]$ instead of $\mathrm{E}\,[f(\mathbf{D})]$.

The following result was proved in Chung (1990) and generalizes a similar result from Babai et al. (1992).

Theorem 20.11. *For every* $f : X \to \{-1, 1\}$,

$$\mathrm{disc}(f) \leqslant \mathcal{E}(f)^{1/2^k}.$$

The theorem is very useful because $\mathcal{E}(f)$ is a much simpler object than $\mathrm{disc}(f)$. For many functions f, it is very easy to compute $\mathcal{E}(f)$ exactly. In Chung and Tetali (1993), $\mathcal{E}(f)$ was computed for some explicit functions, resulting in highest known lower bounds for the multi-party communication complexity of these functions. A new example of such a function is given in Sect. 20.9.1. The example is due to Raz (2000) who also has found a new and easier proof of the theorem itself.

Proof (due to Raz 2000). We first give a lower bound for $\mathcal{E}(f)$ in terms of the "absolute discrepancy"

$$\Delta(f) = \mathrm{E}\,[f(\mathbf{x})]\,,$$

where $\mathbf{x}$ is a random vector uniformly distributed over X.

Claim 20.12. *For all* $f : X \to \{-1, 1\}$, $\mathcal{E}(f) \geqslant |\Delta(f)|^{2^k}$.

Proof. Let $\mathbf{D} = \{\mathbf{a}_1, \mathbf{b}_1\} \times \cdots \times \{\mathbf{a}_k, \mathbf{b}_k\}$ be a random cube. That is, for each i, $\mathbf{a}_i \in X_i$ and $\mathbf{b}_i \in X_i$ are chosen according to the uniform distribution. Let also $\mathbf{D}' = \{\mathbf{a}_1, \mathbf{b}_1\} \times \cdots \times \{\mathbf{a}_{k-1}, \mathbf{b}_{k-1}\}$. Then

$$\begin{aligned}\mathcal{E}(f) = \mathrm{E}\,[f(\mathbf{D})] &= \mathrm{E}\,[f(\mathbf{D}' \times \{\mathbf{a}_k, \mathbf{b}_k\})] \\ &= \mathrm{E}\,[f(\mathbf{D}' \times \{\mathbf{a}_k\}) \cdot f(\mathbf{D}' \times \{\mathbf{b}_k\})]\,.\end{aligned}$$

For any function g defined on X_k we have

$$\mathrm{E}\left[g(\mathbf{a}_k)\cdot g(\mathbf{b}_k)\right] = \mathrm{E}\left[g(\mathbf{a}_k)\right]\cdot \mathrm{E}\left[g(\mathbf{b}_k)\right] = \left(\mathrm{E}\left[g(\mathbf{a}_k)\right]\right)^2.$$

For fixed D', take $g(\mathbf{a}_k) = f(D'\times\{\mathbf{a}_k\})$ to get

$$\begin{aligned}\mathcal{E}(f) = \mathrm{E}_{\mathbf{D}}\left[f(\mathbf{D})\right] &= \mathrm{E}_{\mathbf{D}'}\mathrm{E}_{\mathbf{a}_k,\mathbf{b}_k}\left[f(\mathbf{D}'\times\{\mathbf{a}_k,\mathbf{b}_k\})\right]\\ &= \mathrm{E}_{\mathbf{D}'}\left(\mathrm{E}_{\mathbf{a}_k}\left[f(\mathbf{D}'\times\{\mathbf{a}_k\})\right]\right)^2.\end{aligned}$$

By the Cauchy–Schwarz inequality, for any random variable ξ, $\mathrm{E}\left[\xi^2\right] \geqslant \mathrm{E}\left[\xi\right]^2$. Therefore,

$$\mathcal{E}(f) \geqslant \left(\mathrm{E}_{\mathbf{D}'}\mathrm{E}_{\mathbf{a}_k}\left[f(\mathbf{D}'\times\{\mathbf{a}_k\})\right]\right)^2 = \left(\mathrm{E}\left[f(\mathbf{D}'\times\{\mathbf{a}_k\})\right]\right)^2.$$

Repeat the same argument k times to get

$$\begin{aligned}\mathcal{E}(f) &\geqslant \left(\mathrm{E}\left[f(\{\mathbf{a}_1\}\times\{\mathbf{a}_2\}\times\cdots\times\{\mathbf{a}_k\})\right]\right)^{2^k}\\ &= \left(\mathrm{E}\left[f(\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_k)\right]\right)^{2^k} = \left|\Delta(f)\right|^{2^k}.\end{aligned}$$

□

Given two functions f and g, we denote by $f\cdot g$ their pointwise product, i.e., $f\cdot g(x) = f(x)\cdot g(x)$. A function g is *cylindrical* if it does not depend on at least one of its input variables, i.e., if there is an i such that $g(x_1,\ldots,x_i,\ldots,x_k) = g(x_1,\ldots,x_i',\ldots,x_k)$ for all $(x_1,\ldots,x_i,\ldots,x_k)\in X$ and $x_i'\in X_i$.

Claim 20.13. *For all $f,g: X\to\{-1,1\}$, if g is cylindrical then*

$$\mathcal{E}(f\cdot g) = \mathcal{E}(f).$$

Proof. By the definition of the sign function, for every cube D we have $f\cdot g(D) = f(D)\cdot g(D)$. We can assume that g does not depend on x_k. Then for all $x_1,\ldots,x_{k-1},a_k,b_k$ we have $g(x_1,\ldots,x_{k-1},a_k) = g(x_1,\ldots,x_{k-1},b_k)$. Therefore, g takes the value 1 on an even number of elements of D (remember that D is a multi-set). Hence $g(D)=1$ and so $f\cdot g(D) = f(D)$. Since this holds for every cube D, we are done. □

Claim 20.14. *For every $f: X\to\{-1,1\}$ there exists $h: X\to\{-1,1\}$ such that $\mathcal{E}(h) = \mathcal{E}(f)$ and $|\Delta(h)|\geqslant \mathrm{disc}(f)$.*

Proof. Take a cylinder intersection $T = T_1\cap T_2\cap\cdots\cap T_k$ with $|\mathrm{disc}_T(f)| = \mathrm{disc}(f)$. The idea is to define a random function $\mathbf{g}: X\to\{-1,1\}$ such that $\mathrm{E}\left[\mathbf{g}(x)\right] = \mathrm{E}_{\mathbf{g}}\left[\mathbf{g}(x)\right]$ is the characteristic function of T.

For every $i=1,\ldots,k$, define $\mathbf{g}_i: X\to\{-1,1\}$, as a random function, in the following way: with probability $1/2$, $\mathbf{g}_i$ is the constant function 1, and with probability $1/2$, $\mathbf{g}_i(x)$ takes the value 1 on all elements $x\in T_i$, and -1 otherwise. Then for $x\in T_i$, $\mathbf{g}_i(x)=1$ with probability 1, while for $x\notin T_i$, $\mathbf{g}_i(x)=1$ with probability $1/2$ and $\mathbf{g}_i(x)=-1$ with probability $1/2$.

Define $\mathbf{g} = \mathbf{g}_1\cdots\mathbf{g}_k$. Then for $x\in T$, $\mathbf{g}(x)=1$ with probability 1, while for $x\notin T$, $\mathbf{g}(x)=1$ with probability $1/2$ and $\mathbf{g}(x)=-1$ with probability

$1/2$ (this is so because the functions $\mathbf{g}_i$ are independent of each other, and $x \notin T$ iff $x \notin T_i$ for at least one i). Thus, the expectation $\mathrm{E}[\mathbf{g}(x)]$ takes the value 1 on all $x \in T$, and takes the value 0 on all $x \notin T$, i.e., $\mathrm{E}[\mathbf{g}(x)]$ is the characteristic function of the set T.

Let now $\mathbf{x}$ be a random vector uniformly distributed in X. Then

$$\begin{aligned}\mathrm{E}_{\mathbf{g}}[\Delta(f \cdot \mathbf{g})] &= \mathrm{E}_{\mathbf{g}}\mathrm{E}_{\mathbf{x}}[f \cdot \mathbf{g}(\mathbf{x})] = \mathrm{E}_{\mathbf{g}}\mathrm{E}_{\mathbf{x}}[f(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x})] \\ &= \mathrm{E}_{\mathbf{x}}\mathrm{E}_{\mathbf{g}}[f(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x})] = \mathrm{E}_{\mathbf{x}}[f(\mathbf{x}) \cdot \mathrm{E}_{\mathbf{g}}[\mathbf{g}(\mathbf{x})]] = \mathrm{disc}_T(f),\end{aligned}$$

and by convexity,

$$\mathrm{E}[|\Delta(f \cdot \mathbf{g})|] \geqslant |\mathrm{E}[\Delta(f \cdot \mathbf{g})]| = |\mathrm{disc}_T(f)| = \mathrm{disc}(f).$$

By the pigeonhole property of the expectation, there exists a function $g = g_1 \cdots g_k$ for which $|\Delta(f \cdot g)| \geqslant \mathrm{disc}(f)$. On the other hand, since $g_1, \ldots, g_k$ are cylindrical, by the previous claim we have $\mathcal{E}(f \cdot g) = \mathcal{E}(f)$. Claim 20.14 follows by taking $h = f \cdot g$. □

Now, Theorem 20.11 is an immediate consequence of Claims 20.12 and 20.14:

$$\mathcal{E}(f) = \mathcal{E}(h) \geqslant |\Delta(h)|^{2^k} \geqslant \mathrm{disc}(f)^{2^k}.$$

□

20.9.1 Example: matrix multiplication

Let $X = X_1 \times \cdots X_k$, where each X_i is the set of all $m \times m$ matrices over the field $\mathbb{F}_2$; hence, $|X_i| = n = 2^{m^2}$. For $x_1 \in X_1, \ldots, x_k \in X_k$, denote by $x_1 \cdots x_k$ the product of $x_1, \ldots, x_k$ as matrices over $\mathbb{F}_2$. Let $F(x_1, \ldots, x_k)$ be a boolean function whose value is the element in the first row and the first column of the product $x_1 \cdots x_k$. Define the function $f : X \to \{-1, 1\}$ by

$$f(x_1, \ldots, x_k) = (-1)^{F(x_1, \ldots, x_k)} = 1 - 2F(x_1, \ldots, x_k).$$

Theorem 20.15 (Raz 2000). $\quad \mathrm{disc}(f) \leqslant \left(\dfrac{k-1}{\sqrt{\log_2 n}}\right)^{1/2^k}.$

Proof. For every cube $D = \{a_1, b_1\} \times \cdots \times \{a_k, b_k\}$,

$$f(D) = \prod_{x \in D} f(x) = \prod_{x \in D} (-1)^{F(x)} = (-1)^{\bigoplus_{x \in D} F(x)},$$

where $\oplus$ denotes the addition over $\mathbb{F}_2$. Since F is linear in each variable,

$$f(D) = (-1)^{F(a_1 \oplus b_1, \ldots, a_k \oplus b_k)} = 1 - 2F(a_1 \oplus b_1, \ldots, a_k \oplus b_k),$$

where $a_i \oplus b_i$ denotes the sum of matrices a_i and b_i over $\mathbb{F}_2$. If we choose $\mathbf{D}$ at random according to the uniform distribution, then $(\mathbf{a}_1 \oplus \mathbf{b}_1, \ldots, \mathbf{a}_k \oplus \mathbf{b}_k)$ is a random vector $(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ uniformly distributed over X. Therefore,

$$\begin{aligned}\mathcal{E}(f) &= \mathrm{E}\,[f(\mathbf{D})] = \mathrm{E}\,[1 - 2F(\mathbf{a}_1 \oplus \mathbf{b}_1, \ldots, \mathbf{a}_k \oplus \mathbf{b}_k)] \\ &= \mathrm{E}\,[1 - 2F(\mathbf{x}_1, \ldots, \mathbf{x}_k)] = \mathrm{E}\,[f(\mathbf{x}_1, \ldots, \mathbf{x}_k))] = \Delta(f).\end{aligned}$$

To estimate $\Delta(f)$, let E_d denote the event that the first row of the matrix $\mathbf{x}_1 \cdots \mathbf{x}_d$ contains only 0's. Define $p_d = \mathrm{Prob}\,(E_d)$. Since p_1 is determined by $\mathbf{x}_1$ and since $\mathbf{x}_1$ is uniformly distributed, we have $p_1 = \mathrm{Prob}\,(E_1) = 2^{-m}$. Clearly we also have $\mathrm{Prob}\,(E_{d+1} \mid E_d) = 1$. On the other hand, since $\mathbf{x}_{d+1}$ is uniformly distributed, $\mathrm{Prob}\,(E_{d+1} \mid \neg E_d) = 2^{-m}$ (see Exercise 17.2). Therefore, for all $1 \leqslant d < k$,

$$\begin{aligned}p_{d+1} &= \mathrm{Prob}\,(E_{d+1} \mid E_d) \cdot \mathrm{Prob}\,(E_d) + \mathrm{Prob}\,(E_{d+1} \mid \neg E_d) \cdot \mathrm{Prob}\,(\neg E_d) \\ &= p_d + (1 - p_d) \cdot 2^{-m} \leqslant p_d + 2^{-m},\end{aligned}$$

implying that $p_d \leqslant d \cdot 2^{-m}$ for all $d = 1, \ldots, k$.

If E_{k-1} occurs then $F(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ is always 0, and hence, $f(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ is always 1. If E_{k-1} does not occur then, since the first column of $\mathbf{x}_k$ is uniformly distributed, $F(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ is uniformly distributed over $\{0, 1\}$, and hence, $f(\mathbf{x}_1, \ldots, \mathbf{x}_k)$ is uniformly distributed over $\{-1, 1\}$. Therefore,

$$\mathcal{E}(f) = \Delta(f) = p_{k-1} \leqslant (k-1) \cdot 2^{-m},$$

and Theorem 20.11 yields the desired upper bound on $\mathrm{disc}(f)$. □

Exercises

20.1.$^-$ We have n letters going to n different persons and n envelopes with their addresses. We insert each letter into an envelope independently from each other at random (several letters may go in the same envelope). What is the expected number of correct matches? (Answer: $E = 1$.)

20.2.$^-$ There are k people in a lift at the ground floor. Each wants to get off at a random floor of one of the n upper floors. What is the expected number of lift stops?

Hint: Consider the indicator random variables X_i for the events that at least one person is off at the ith floor, and apply the linearity of expectation. Answer: $E = n(1 - (1 - 1/n)^k)$.

20.3. Let Ω be a uniform sample space, and let $X : \Omega \to \{0, 1, \ldots, M\}$ be a random variable with the expectation $\mu = M - a$ for some a. Prove that then, for any $1 \leqslant b \leqslant M$, $\mathrm{Prob}\,(X \geqslant M - b) \geqslant (b - a)/b$.

Sketch: Let B be the set of those points $\omega \in \Omega$ for which $X(\omega) < M - b$. Then $\mathrm{Prob}\,(B) \cdot (M - b) + \mathrm{Prob}\,(\overline{B}) \cdot M \geqslant M - a$, or $\mathrm{Prob}\,(B) \leqslant a/b$.