Lower Bounds for Sampling Algorithms for Estimating the Average*

Ran Canetti† Guy Even ‡ Oded Goldreich§

October 6, 1994

Abstract

We show lower bounds on the number of sample points and on the number of coin tosses used by general sampling algorithms for estimating the average value of functions over a large domain. The bounds depend on the desired precision and on the error probability of the estimate.

Our lower bounds match upper bounds established by known algorithms, up to a multiplicative constant. Furthermore, we give a non-constructive proof of existence of an algorithm that improves the known upper bounds by a constant factor.

Key words: Theory of computation, Sampling, Estimating, Randomness, Lower bounds.

1 Introduction

Consider the following problem. An algorithm is needed, that given an oracle access to a function $f : \{0,1\}^n \to [0,1]$ estimates the average of $f(x)$ over all $x \in \{0,1\}^n$. We stress that given an integer n, the algorithm should work "properly" (in a sense defined later) on *any* function $f : \{0,1\}^n \to [0,1]$. We are of course interested in algorithms which are much faster than the obvious exhaustive algorithm.

Any such algorithm essentially consists of querying the oracle for f at some chosen points in $\{0,1\}^n$, and performing some internal computation based on the function value at these points. We refer to querying the oracle for the value of $f(x)$ as **sampling** the function f at point x, and say that x is a **sample point** of the algorithm. The value $f(x)$ is called the **sample value**. We call such an algorithm a **general sampling algorithm**, or a **sampler**.

Obviously, a sampler has to be randomized, if the estimate is to be of any practical value. In fact, a quantification of this intuitive statement is one of the contributions of this work. Thus, the estimate (to the average of a given function) made by the sampler is a random variable determined by the random choices of the sampler.

*Supported by grants No. 89-00312 and 92-00226 from the United States - Israel Binational Science Foundation (BSF), Jerusalem, Israel.

†Department of Applied Mathematics and Computer Science, Weizmann Institute, Israel.

‡Department of Computer Science, Technion, Israel

§Department of Applied Mathematics and Computer Science, Weizmann Institute, Israel.

We measure the *quality* of the estimate using two parameters: one is the precision of the estimate and the other is the error probability. Following Karp and Luby [KL] (see also [KLM]), we say that a random variable S estimates a value v with (additive) precision ϵ and error probability δ if $\text{Prob}(|S-v| \geq \epsilon) \leq \delta$. An $(\epsilon(n), \delta(n))$-sampler is a sampler that estimates the average of *any* function $f : \{0,1\}^n \rightarrow [0,1]$ with precision $\epsilon(n)$ and error probability $\delta(n)$. The parameters we use to measure the *cost* of the estimate are the number of sample points used, and the amount of randomness (number of coin tosses) used.

Lower bounds. We show lower bounds on the two cost parameters of an estimate as a function of the quality parameters of the estimate. For the lower bound on the number of samples, we first consider samplers that given n always make $t(n)$ samples at distinct places. We call such samplers $t(n)$-**regular**. We show that any $t(n)$-regular $(\epsilon(n), \delta(n))$-sampler satisfies

$$t(n) = \Omega(\frac{1}{\epsilon(n)^2} \log \frac{1}{\delta(n)}) \tag{1}$$

We then generalize this lower bound to non-regular samplers.

We remark that analogous lower bounds on the number of samples needed for *hypothesis testing* and *estimating,* with respect to memoryless sources, exist in different settings. We note in particular the Cramer-Rao Bound and its many variants (see Chapter 2.4 of [VT]). However, none of these bounds applies directly to our computational setting.

Next, we show a lower bound on the number of coin tosses, $r(n)$, used by a sampler for functions $f : \{0,1\}^n \rightarrow [0,1]$. The bound depends on n, on the number of sample points $t(n)$, and on the error probability $\delta(n)$. Specifically, we get

$$r(n) \geq n + \log \frac{1}{\delta(n)} - \log t(n) - O(1) \tag{2}$$

Combining bounds (1) and (2), we get that for samplers using a minimal number of samples (namely, $t(n) = O(\frac{1}{\epsilon(n)^2} \log \frac{1}{\delta(n)})$), the bound on the number of coin tosses is $r(n) \geq n + \Omega(\log \frac{1}{\delta(n)}) - O(\log \frac{1}{\epsilon(n)})$.

We remark that both lower bounds hold even in a non-uniform setting, in which a different algorithm is allowed for each integer n. Furthermore, both bounds hold even for samplers that operate only on *Boolean* functions (namely, functions $f : \{0,1\}^n \rightarrow \{0,1\}$).

Tightness of the bounds. The lower bound on the number of samples is trivially tight, as demonstrated by the following straightforward sampler. On input n, output the average of the function value at $t(n)$ mutually independent and uniformly chosen sample points. Using a Chernoff-type bound [Ho], we get that $t(n) = O(\frac{1}{\epsilon(n)^2} \log \frac{1}{\delta(n)})$ sample points are sufficient for an $(\epsilon(n), \delta(n))$-sampler. However, this sampler is very wasteful in randomness: it uses $n \cdot t(n)$ coin tosses.

Several more randomness-efficient samplers are known in the literature. We refer the reader to [BGG] for a survey. Of special interest is the sampler presented in [BGG]: this sampler is simultaneously tight with respect to *both* bounds (up to a multiplicative constant). Namely, in order to achieve an $(\epsilon(n), \delta(n))$ estimate for functions $f : \{0,1\}^n \rightarrow [0,1]$, this sampler uses $O(\frac{1}{\epsilon(n)^2} \log \frac{1}{\delta(n)})$ samples, and tosses $2n + O(\log \frac{1}{\delta(n)})$ coins.

Finally, we give a *non-constructive* proof of existence of a sampler that is more efficient than the [BGG] sampler, both in the number of samples and in the number of coins tossed.

Namely, we show that given precision ϵ, error probability δ, and any number $n \in \mathbf{N}$, there exists an (ϵ, δ)-sampler (for functions $f : \{0,1\}^n \rightarrow [0,1]$) that uses $\frac{2}{\epsilon^2} \ln \frac{4}{\delta}$ samples, and tosses only $n + \log n + 2 \log \frac{2}{\delta} + 2$ coins.

Recently, Goldreich and Wigderson have presented an explicit (and efficient) sampler that given ϵ, δ, and n as above uses $O(\frac{1}{\epsilon^2} \log \frac{1}{\delta})$ samples and tosses $n + O(\log \frac{1}{\epsilon}) + O(\log \frac{1}{\delta})$ coins [GW].

2 The setting

We use the following notational conventions. Let $a \in_{\mathrm{R}} A$ denote a random variable, a, uniformly distributed over the set A. Let $\mathrm{Prob}_{e \in_{\mathrm{R}} D}\ (T(e))$ denote the probability of event $T(e)$ when element e is chosen uniformly at random from domain D. Let $\mathrm{E}_{e \in_{\mathrm{R}} D}\ (X(e))$ denote the expected value of random variable $X(e)$ when element e is chosen uniformly at random from domain D.

For a function $f : \{0,1\}^n \rightarrow [0,1]$, let $\bar{v}_f$ be the average of the function f:

$$\bar{v}_f \;=\; \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)$$

An $(\epsilon(n), \delta(n))$-**sampler** is a randomized oracle Turing machine, S, that for every n and for every function $f : \{0,1\}^n \rightarrow [0,1]$ satisfies:

$$\mathrm{Prob}_{r \in_{\mathrm{R}} \mathcal{R}} \left(|S^f(n,r) - \bar{v}_f| > \epsilon(n) \right) < \delta(n) \tag{3}$$

where $\mathcal{R}$ is the domain of the random inputs of S, and $S^f(n,r)$ denotes the output of S with oracle f on input n and random input r. We stress that samplers can be adaptive, that is the location of a sample point may depend on the function value at previous sample points. Furthermore, the *number* of sample points may vary as well. Given a sampler, let $t(f)$ be the random variable describing the number of samples made when given oracle access to function f.

For the lower bound on the number of samples we also consider a special type of samplers. These samplers are both **regular** and **weak**. These two unrelated properties are defined as follows. A $t(n)$-**regular** sampler, given n, always sample the function at $t(n)$ (adaptively chosen) places, where all sample points are distinct. A **weak** sampler has an additional special output symbol, $\perp$, meaning "I don't know". A weak sampler may output $\perp$ with any probability that is strictly less than 1. Furthermore, it has to satisfy the condition (3) only when its output is not $\perp$. That is, for all n and for all $f : \{0,1\}^n \rightarrow [0,1]$ a weak (ϵ, δ)-sampler S satisfies:

$$\begin{array}{ll} \text{(a)} & \mathrm{Prob}_{r \in_{\mathrm{R}} \mathcal{R}} \left(S^f(n,r) = \perp \right) < 1 \\ \text{(b)} & \mathrm{Prob}_{r \in_{\mathrm{R}} \mathcal{R}} \left(|S^f(n,r) - \bar{v}_f| > \epsilon(n) \mid S^f(n,r) \neq \perp \right) < \delta(n) \end{array}$$

In the sequel, we use t, δ, ϵ, r to shorthand $t(n), \delta(n), \epsilon(n), r(n)$, respectively.

3 A lower bound on the number of samples

In this section we show lower bounds on the number of samples made. First we show, in Theorem 1 below, that a t-regular, (ϵ, δ)-sampler satisfies $t = \Omega(\frac{1}{\epsilon^2} \log(\frac{1}{\delta}))$, even in the case where the sampler is weak. Namely, the sampler satisfies the lower bound even though it may almost always output $\perp$. We prove this result only for samplers that use at most $2^{\frac{n}{2}-2}$ sample points. However, we hardly lose generality by this restriction: any sampler that uses as many as $2^{\frac{n}{2}-2}$ samples is highly inefficient and is not likely to be interesting. In Corollary 1 we generalize the lower bound to non-regular samplers. The corollary uses the fact that Theorem 1 holds also for weak samplers (actually, it is for this corollary that we have introduced the notion of weak samplers).

Theorem 1 *Let $\epsilon \leq \frac{1}{8}$, $\delta \leq \frac{1}{6}$, and $t \leq 2^{n/2-2}$. Let S be a t-regular, weak (ϵ, δ)-sampler. Then, for all large enough n,*

$$t \ \geq \frac{1}{4\epsilon^2} \ln \left(\frac{1}{8\mathrm{e}\sqrt{\pi}\delta} \right) \tag{4}$$

where e *is the natural base of logarithms.*

Corollary 1 *Let $\epsilon \leq \frac{1}{8}$, $\delta \leq \frac{1}{6}$, and $t \leq 2^{n/2-2}$. Let S be an (ϵ, δ)-sampler. Then, for all large enough n there exists a function $f : \{0,1\}^n \to \{0,1\}$ such that for all $0 < \alpha < 1$,*

$$\mathrm{E}_{r \in_{\mathrm{R}} \mathcal{R}}(t(f)) \geq \frac{\alpha}{4\epsilon^2} \ln \left(\frac{1-\alpha}{8\mathrm{e}\sqrt{\pi}\delta} \right)$$

Remark: The best value for α is determined by δ. For constant α, say $\alpha = \frac{1}{2}$, we get $\mathrm{E}_{r \in_{\mathrm{R}} \mathcal{R}}(t(f)) = \Omega(\frac{1}{\epsilon^2} \ln \frac{1}{\delta})$.

Proof of Corollary 1: Let $\tau = \frac{1}{\alpha} \max_{f:\{0,1\}^n \to \{0,1\}} \mathrm{E}(t(f))$. Consider the τ-regular sampler S' that runs S with the exception that at most τ samples are made. If S outputs an estimate before the $(\tau+1)$st sample, then S' outputs this estimate and halts. Otherwise S' outputs $\perp$ and halts. By Markov inequality, S' outputs $\perp$ with probability at most α. Therefore the probability that S' outputs a wrong estimate, given that S' does not output $\perp$, is at most $\frac{\delta}{1-\alpha}$ (the worst case is when all wrong estimates occur when S makes less than τ samples). That is, S' is a τ-regular, weak $(\epsilon, \frac{\delta}{1-\alpha})$-sampler. The corollary follows by applying Theorem 1 to sampler S'. $\square$

Proof of Theorem 1: The proof follows an idea of Johan Hastad which greatly simplifies our original proof. Consider a run of sampler S with oracle access to some function f, and let $\vec{a} = a_1, \ldots, a_t$ be the function values at the points sampled by S. Then, the output of S depends on *only* $\vec{a}$ and the random choices of S (the sample points are determined by the random choices and the previous sample values). We exploit this fact as follows. We define two sets of functions, called $\mathcal{F}_{\frac{1}{2}+\epsilon}$ and $\mathcal{F}_{\frac{1}{2}-\epsilon}$, and proceed in two steps roughly described below. First, we show that in order for S to be a weak (ϵ, δ)-sampler the distribution of $\vec{a}$ when $f \in_{\mathrm{R}} \mathcal{F}_{\frac{1}{2}+\epsilon}$ should differ considerably from the distribution of $\vec{a}$ when $f \in_{\mathrm{R}} \mathcal{F}_{\frac{1}{2}-\epsilon}$. Next, we show that in order for the distributions of $\vec{a}$ to differ as required, the number of samples has to be at least as claimed in the theorem.

Consider the following two sets of functions:

$$\mathcal{F}_{\frac{1}{2}+\epsilon} \ \stackrel{\triangle}{=} \ \{f : \{0,1\}^n \to \{0,1\} | \bar{v}_f = \frac{1}{2} + \epsilon\}$$
$$\mathcal{F}_{\frac{1}{2}-\epsilon} \ \stackrel{\triangle}{=} \ \{f : \{0,1\}^n \to \{0,1\} | \bar{v}_f = \frac{1}{2} - \epsilon\}$$

Consider the execution of S given oracle access to a function drawn uniformly from $\mathcal{F}_{\frac{1}{2}+\epsilon}$ (respectively, $\mathcal{F}_{\frac{1}{2}-\epsilon}$). Let $\vec{A}$ be the random variable describing the function values at the places sampled by S. That is, $\vec{A} = \vec{a} = a_1, \ldots, a_t$ means that the function value at the ith sampled place is $a_i \in \{0, 1\}$. Note that $\vec{A}$ depends both on the choice of f and the random choices of S. Let $\vec{A}_r$ (respectively, $\vec{A}^f$) have the distribution of $\vec{A}$ when the random choices of S are fixed to be r (respectively, when the sampled function is f).

Since S is a weak (ϵ, δ)-sampler, we have

$$P_{\frac{1}{2}+\epsilon} \triangleq \mathrm{Prob}_{r\in_{\mathrm{R}}\mathcal{R}, f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}} \left(S^f(n,r) \leq \frac{1}{2} |\not\perp \right) \leq \delta \tag{5}$$

$$P_{\frac{1}{2}-\epsilon} \triangleq \mathrm{Prob}_{r\in_{\mathrm{R}}\mathcal{R}, f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}-\epsilon}} \left(S^f(n,r) \geq \frac{1}{2} |\not\perp \right) \leq \delta \tag{6}$$

where $\not\perp$ denotes the event that S does not output $\perp$ (note that these inequalities are meaningful as long as S outputs $\perp$ with probability strictly less than 1). The value $P_{\frac{1}{2}+\epsilon}$ can be written as

$$\begin{aligned} P_{\frac{1}{2}+\epsilon} &= \sum_{\vec{a}\in\{0,1\}^t} \mathrm{Prob}_{r\in_{\mathrm{R}}\mathcal{R}, f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}}(\vec{A} = \vec{a}) \cdot \mathrm{Prob}_{r\in_{\mathrm{R}}\mathcal{R}, f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}} \left(S^f(n,r) \leq \frac{1}{2} \middle| \vec{A} = \vec{a}, \not\perp \right) \\ &= \sum_{\vec{a}\in\{0,1\}^t} \mathrm{E}_{r\in_{\mathrm{R}}\mathcal{R}} \left(\mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}}(\vec{A}_r = \vec{a}) \right) \cdot \mathrm{E}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}} \left(\mathrm{Prob}_{r\in_{\mathrm{R}}\mathcal{R}}(S^f(n,r) \leq \frac{1}{2} \middle| \vec{A}^f = \vec{a}, \not\perp) \right) \end{aligned}$$

A key observation is that when the sample values are $\vec{a}$, the output of the sampler is a (deterministic) function of $\vec{a}$ and the random input, r. Let $V(\vec{a}, r)$ denote this function. Then, for any function $f \in \mathcal{F}_{\frac{1}{2}+\epsilon} \cup \mathcal{F}_{\frac{1}{2}-\epsilon}$ the probability $\mathrm{Prob}_{r\in_{\mathrm{R}}\mathcal{R}}(S^f(n,r) < \frac{1}{2} | \vec{A}^f = \vec{a}, \not\perp)$ equals $\mathrm{Prob}_{r\in_{\mathrm{R}}\mathcal{R}}(V(\vec{a}, r) < \frac{1}{2} | \not\perp)$; this last probability depends only on V, $\vec{a}$ and r, and is, in particular, the same for all functions f. In addition, the probability $\mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}}(\vec{A}_r = \vec{a})$ is the same for all $r \in \mathcal{R}$ (this is so since, when $f \in_{\mathrm{R}} \mathcal{F}_{\frac{1}{2}+\epsilon}$, all sequences of t distinct sample points are equally likely to result in sample values $\vec{a}$). Fix some arbitrary $r_0 \in \mathcal{R}$. The previous equation now becomes

$$P_{\frac{1}{2}+\epsilon} = \sum_{\vec{a}\in\{0,1\}^t} \mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}}(\vec{A}_{r_0} = \vec{a}) \cdot \mathrm{Prob}_{r\in_{\mathrm{R}}\mathcal{R}} \left(V(\vec{a}, r) \leq \frac{1}{2} |\not\perp \right) \tag{7}$$

Similarly,

$$P_{\frac{1}{2}-\epsilon} = \sum_{\vec{a}\in\{0,1\}^t} \mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}-\epsilon}}(\vec{A}_{r_0} = \vec{a}) \cdot \mathrm{Prob}_{r\in_{\mathrm{R}}\mathcal{R}} \left(V(\vec{a}, r) \geq \frac{1}{2} |\not\perp \right) \tag{8}$$

Let $Q(\vec{a}) \triangleq \mathrm{Prob}_{r\in_{\mathrm{R}}\mathcal{R},}(V(\vec{a}, r) \leq \frac{1}{2} | \not\perp)$. Then, $1 - Q(\vec{a}) \leq \mathrm{Prob}_{r\in_{\mathrm{R}}\mathcal{R},}(V(\vec{a}, r) \geq \frac{1}{2} | \not\perp)$. Adding Inequalities (5) and (6), and using Equations (7)and (8), we get

$$\begin{aligned} 2\delta &\geq P_{\frac{1}{2}+\epsilon} + P_{\frac{1}{2}-\epsilon} \\ &\geq \sum_{\vec{a}\in\{0,1\}^t} \left(Q(\vec{a}) \cdot \mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}}(\vec{A}_{r_0} = \vec{a}) + (1 - Q(\vec{a})) \cdot \mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}-\epsilon}}(\vec{A}_{r_0} = \vec{a}) \right) \\ &\geq \sum_{\vec{a}\in\{0,1\}^t} \min\{\mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}}(\vec{A}_{r_0} = \vec{a}), \mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}-\epsilon}}(\vec{A}_{r_0} = \vec{a})\} \end{aligned} \tag{9}$$

Note that $\mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}}(\vec{A}_{r_0}=\vec{a})$ depends only on the number of '1' entries in $\vec{a}$, denoted $w(\vec{a})$. Also, if a vector $\vec{b}$ is the bit complement of $\vec{a}$ (i.e., $\vec{b}\oplus\vec{a}=1^t$) then $\mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}}(\vec{A}_{r_0}=\vec{a})=\mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}-\epsilon}}(\vec{A}_{r_0}=\vec{b})$. Thus (9) becomes:

$$2\cdot\sum_{\vec{a}:\ w(\vec{a})<\lceil\frac{t}{2}\rceil}\mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}}(\vec{A}_{r_0}=\vec{a})\leq 2\delta$$

Namely,

$$\mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}}\left(w(\vec{A}_{r_0})<\lceil\frac{t}{2}\rceil\right)\leq\delta \tag{10}$$

We remark that similarily we may obtain $\mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}-\epsilon}}\left(w(\vec{A}_{r_0})>\lfloor\frac{t}{2}\rfloor\right)\leq\delta$. Thus, the distributions of $\vec{A}_{r_0}$ when $f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}$ and when $f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}-\epsilon}$ differ substantially. This concludes the first step of our proof.

In the second step of the proof we bound the left hand side of (10) from below. Let $\mathcal{A}(\vec{a})$ denote the set of functions f in $\mathcal{F}_{\frac{1}{2}+\epsilon}$ such that $\vec{a}$ is the sequence of sample values obtained by $S^f(n,r_0)$. It follows that on random input r_0, sampler S samples all functions in $\mathcal{A}(\vec{a})$ at the same places. Thus, for each $\vec{a}\in\{0,1\}^t$, we have

$$\begin{aligned}\mathrm{Prob}_{f\in_{\mathrm{R}}\mathcal{F}_{\frac{1}{2}+\epsilon}}(\vec{A}_{r_0}=\vec{a}) &= \frac{|\mathcal{A}(\vec{a})|}{|\mathcal{F}_{\frac{1}{2}+\epsilon}|}\\ &= \frac{\binom{2^n-t}{2^n(\frac{1}{2}+\epsilon)-w(a)}}{\binom{2^n}{2^n(\frac{1}{2}+\epsilon)}}\\ &\geq 2^{-t-1}\cdot(1+2\epsilon)^{w(\vec{a})}\cdot(1-2\epsilon)^{t-w(\vec{a})}\end{aligned}$$

(where the last inequality holds for all sufficiently large n and for $t\leq\frac{1}{4}\sqrt{2^n}$. For details see our Technical Report [CEG]). Plugging this bound in (10), we get

$$\begin{aligned}\delta &\geq 2^{-t-1}\sum_{i=0}^{\lceil t/2\rceil-1}\binom{t}{i}\cdot(1+2\epsilon)^i\cdot(1-2\epsilon)^{t-i}\\ &> 2^{-t-1}\cdot(1+2\epsilon)^{\lceil t/2\rceil-\lceil\sqrt{t/2}\rceil}\cdot(1-2\epsilon)^{\lceil t/2\rceil+\lceil\sqrt{t/2}\rceil}\cdot\sum_{i=\lceil t/2\rceil-\lceil\sqrt{t/2}\rceil}^{\lceil t/2\rceil-1}\binom{t}{i}\end{aligned} \tag{11}$$

Using the Stirling approximation, (11) yields (for $\epsilon<\frac{1}{8}$):

$$\begin{aligned}\delta &> \frac{1}{4\mathrm{e}\sqrt{\pi}}\cdot(1-4\epsilon^2)^{\lceil t/2\rceil+\lceil\sqrt{t/2}\rceil}\\ &> \frac{1}{8\mathrm{e}\sqrt{\pi}}\cdot\mathrm{e}^{-4\epsilon^2 t}\end{aligned}$$

The theorem follows. $\square$

4 A lower bound on the number of coin tosses

In this section we show a lower bound on the amount of randomness used by an (ϵ, δ)-sampler, as a function of the input length n, the precision ϵ, the error probability δ, and the number of samples t.

Theorem 2 *Let $\epsilon < \frac{1}{2}$, and let $\delta \leq \frac{1}{2}$. Let S be an (ϵ, δ)-sampler that on input n tosses r coins, and uses at most t sampling points so that $t \leq \frac{1}{2}2^n$. Then,*

$$r \ \geq \ n \ - \ \log t \ + \ \log \frac{1}{\delta} \ + \ \log(1 - 2\epsilon) \ - \ 2. \tag{12}$$

Remarks: (Recall that ϵ is the precision of the estimate, and δ is the error probability.)

- Note that a non-trivial lower bound on the number of coin-tosses can hold only for $t \leq \frac{1}{2}2^n$ and $\epsilon < \frac{1}{2}$. If $t > \frac{1}{2}2^n$ then the average over t fixed sample points is an $(\epsilon, 0)$-sampler that tosses no coins (for some $\epsilon < \frac{1}{2}$, where ϵ depends on t). For $\epsilon \geq \frac{1}{2}$, the algorithm that always outputs $\frac{1}{2}$ constitutes an $(\epsilon, 0)$-sampler.
- For $\epsilon < c < \frac{1}{2}$ (where $c > 0$ is a constant), the above bound takes the form $r \geq n - \log t + \log \delta^{-1} - O(1)$.
- For samplers using a minimal number of samples (namely, $t(n) = O(\frac{1}{\epsilon(n)^2} \log \frac{1}{\delta(n)})$), the bound on the number of coin tosses is $r \geq n + \log \frac{1}{\delta} - 2 \log \frac{1}{\epsilon} - O(\log \log \frac{1}{\delta})$.

Proof. Let S be an (ϵ, δ)-sampler as in the theorem. Construct a function $f : \{0,1\}^n \to \{0,1\}$ in the following manner. Enumerate all the 2^r possible sequences of the coin tosses of S. Let T denote the set of $\lceil 2\delta 2^r \rceil$ first sequences. Let $D \subseteq \{0,1\}^n$ denote the set of all points that are sampled by the sampler S when its sequence of coin tosses is in T. Clearly, $|D| \leq t \cdot |T|$. Let

$$f(x) = \begin{cases} 1 & x \in D \\ 0 & \text{otherwise} \end{cases}$$

Let E denote the event that all the sample values are 1. When sampling function f, event E has probability at least 2δ (over the coin tosses of the algorithm).

Note that when the sampled function is the constant function $g \equiv 1$, event E has probability 1. The estimate made by S given oracle access to function g must be at least $1 - \epsilon$, with probability at least $1 - \delta$. Thus, given event E, the answer of algorithm S must be at least $1 - \epsilon$ with probability at least $1 - \delta$. Returning to function f, we now have

$$\begin{aligned} \mathrm{Prob}(S^f(n) \geq 1 - \epsilon) \ &\geq \ \mathrm{Prob}(S^f(n) \geq 1 - \epsilon \,|E) \cdot \mathrm{Prob}(E) \\ &\geq \ (1 - \delta) \cdot 2\delta \\ &\geq \ \delta \end{aligned}$$

when the probability is taken over the coin tosses of sampler S. Consequently, since S is an (ϵ, δ)-sampler, the average of function f must satisfy

$$\bar{v}_f \ \geq \ 1 - 2\epsilon. \tag{13}$$

On the other hand, we bound $\bar{v}_f$ from above. Clearly,

$$\begin{aligned}\bar{v}_f &= 2^{-n} \cdot |f^{-1}(1)| \\ &= 2^{-n} \cdot |D| \\ &\leq 2^{-n} \cdot |T| \cdot t \\ &= 2^{-n} \cdot \lceil 2\delta 2^r \rceil \cdot t.\end{aligned}$$

It can be seen that $\lceil 2\delta 2^r \rceil \leq 4\delta 2^r$. (Otherwise, $\delta < 2^{-r}$. As error probabilities can take only multiples of 2^{-r}, the sampler S must have error probability $\delta = 0$. This requires either $\epsilon \geq \frac{1}{2}$ or $t > \frac{1}{2}2^n$, in contradiction with the provisions of the theorem.) We thus have,

$$\bar{v}_f \leq 2^{-n} \cdot 4 \cdot \delta 2^r \cdot t. \tag{14}$$

Combining (13) and (14), we get

$$4 \cdot \delta 2^r \cdot t \cdot 2^{-n} \;\geq\; 1 - 2\epsilon.$$

The theorem follows. □

5 On the existence of randomness-efficient samplers

Bellare, Goldreich and Goldwasser show a randomness-efficient sampling algorithm that uses an optimal number of samples, up to a constant factor. Specifically, the [BGG] algorithm uses $O(\frac{1}{\epsilon^2}\log\frac{1}{\delta})$ samples, and tosses $2n + O(\log\frac{1}{\delta})$ coins.

In this section we show, in a non-constructive manner, that given a number $n \in \mathbf{N}$, and *any* (ϵ, δ)-sampler S, it is possible to reduce the number of coin tosses used by S to $n + 2\log\frac{1}{\delta} + \log\log\frac{1}{\epsilon}$, without increasing the number of samples taken. We stress that the resulting samplers are non-uniform.

We apply this result to the straightforward sampler described in the introduction (i.e., the one that takes the average of the function value at some predefined number of uniformly and independently chosen sample points) to show the existence of an (ϵ, δ)-sampler that uses only $\frac{2}{\epsilon^2}\ln\frac{4}{\delta}$ samples and tosses $r \leq n + 2\log\frac{1}{\delta} + \log\log\frac{1}{\epsilon}$ coins. Since in our model $\epsilon \geq 2^{-n}$ always holds (if $\delta < \frac{1}{2}$ and $t < 2^n$), we in fact have $r \leq n + \log n + 2\log\frac{1}{\delta} + 1$.

Theorem 3 *Let $n \in \mathbf{N}$. Let S be an (ϵ, δ)-sampler that on input n uses t samples. Then there exists:*

(a). *An $(\epsilon, 2\delta)$-sampler, S', for functions $f : \{0,1\}^n \to \{0,1\}$, that on input n uses t samples, and tosses only $n + 2\log\frac{1}{\delta}$ coins.*

(b). *A $(2\epsilon, 2\delta)$-sampler, S'', for functions $f : \{0,1\}^n \to [0,1]$, that on input n uses t samples, and tosses only $n + 2\log\frac{1}{\delta} + \log\log\frac{1}{2\epsilon}$ coins.*

Proof. The proof adapts a technique of Newman [N] to this context.

(a). Let r be the number of coin tosses used by sampler S on input n. Consider the $2^{2^n} \times 2^r$ table where a row corresponds to a function $f : \{0,1\}^n \to \{0,1\}$, and a column corresponds to a coin-toss sequence of the sampler S. A table-entry (f,ρ) is **bad** if the estimate of sampler S, on random input ρ, is *at least* ϵ away from the average of f. Since S is an (ϵ,δ) sampler, there are at most a δ-fraction of bad entries in each row.

We use a probabilistic argument to show that there exists a small subset, denoted K, of columns (i.e., coin-toss sequences), with the following property: reducing the table to the columns of this subset, there are at most a 2δ-fraction of bad entries in each row. Sampler S' consists of choosing at random a column $\rho \in_{\mathrm{R}} K$, and running S on random input ρ. Thus, S' is a $(\epsilon, 2\delta)$-sampler, and tosses only $\log|K|$ coins.

We show the existence of such a subset, K, in the following way. Consider the following choice of a set of columns: k independent selections of a column are made, where each selection is made with uniform distribution. Say that a chosen set of columns is **unfortunate** for a row, if more than $2\delta k$ of the chosen entries in this row are bad. By the Hoefding Inequality [Ho], the probability that a chosen set is unfortunate for a row is less than $2\mathrm{e}^{-2k\delta^2}$. Thus, the probability that there *exists* a row for which the choice is unfortunate is at most $2^{2^n}\cdot 2\mathrm{e}^{-2k\delta^2}$.

Setting $k=\frac{2^n}{\delta^2}$, we have that the probability (over the choices of k columns) that there *exists* a row for which the choice is unfortunate is less than 1. We conclude that there exists a choice of at most $\frac{2^n}{\delta^2}$ columns (i.e., coin-toss sequences) that is fortunate for all the rows.

(b). Consider the set of functions $F=\{f:\{0,1\}^n\rightarrow I\}$, where $I=\{i\cdot 2\epsilon \mid i=1\ldots\lfloor\frac{1}{2\epsilon}\rfloor\}$. For every $f:\{0,1\}^n\rightarrow[0,1]$, let $f'\in F$ be the following **approximation of f in F**: for each $x\in\{0,1\}^n$, the value of $f'(x)$ is the rounding of $f(x)$ to the nearest value in I. Note that $|I|=\frac{1}{2\epsilon}$, and $|\bar{v}_{f'}-\bar{v}_f|\le\epsilon$. Using the technique of part (a), we have that there exists an $(\epsilon,2\delta)$ sampler, S', for functions in F, that tosses only $n+2\log\frac{1}{\delta}+\log\log\frac{1}{2\epsilon}$ coins.

Sampler S'' operates as follows, on any function $f:\{0,1\}^n\rightarrow[0,1]$. Run sampler S' with the following provision: round each sample value to the nearest value in I before giving it to S'. Since sampler S' is an $(\epsilon,2\delta)$-sampler for functions in F, with probability $(1-2\delta)$ the estimate made by sampler S' is less than ϵ away from the average of f', the approximation of f in F. Consequently, this estimate is less than 2ϵ away from the average of f. $\square$

Corollary 2 *Let $n\in\mathbf{N}$. For every $\epsilon>0$ and every $\delta>0$, there exist:*

(a). *An (ϵ,δ)-sampler for boolean functions that uses $\frac{1}{2\epsilon^2}\ln\frac{4}{\delta}$ samples and tosses $n+2\log\frac{2}{\delta}$ coins.*

(b). *An (ϵ,δ)-sampler for functions into $[0,1]$ that uses $\frac{2}{\epsilon^2}\ln\frac{4}{\delta}$ samples and tosses $n+2\log\frac{2}{\delta}+\log\log\frac{1}{\epsilon}$ coins.*

Proof. Consider the sampler S, that chooses t sample points with uniform distribution and outputs the average of the function value at these points. By Hoefding Inequality [Ho], S is an $(\epsilon, 2\mathrm{e}^{-2\epsilon^2 t})$ sampler for every $\epsilon>0$. Thus, for every $\epsilon>0$ and every $\delta>0$, we have that $t=\frac{1}{2\epsilon^2}\ln\frac{2}{\delta}$ samples are sufficient for S to be an (ϵ,δ) sampler. The corollary follows by applying Theorem 3 to sampler S. $\square$

Acknowledgements

We are indebted to Johan Hastad for suggesting a simplification for our proof of Theorem 1.

References

[BGG] M. Bellare, O. Goldreich and S. Goldwasser, "Randomness in Interactive Proofs", *31st FOCS,* 1990, pp.563-571.

[CEG] R. Canetti, Guy Even and O. Goldreich, "Lower Bounds for Sampling Algorithms for Estimating the Average", Technical Report # 789, Department of Computer Science, Technion, Nov. 1993.

[GW] O. Goldreich and A. Wigderson, "Tiny Families of Functions with Random Properties: A Quality-Size Trade-off for Hashing", *26th STOC,* 1994.

[Ho] W. Hoefding, "Probability Inequalities for Sums of Bounded Random Variables", *Journal of the American Statistical Association,* Vol. 58, 1963, pp. 13-30.

[KL] R. M. Karp and M. Luby, "Monte-Carlo Algorithms for Enumeration and Related Problems", *24th FOCS,* 1983, pp.56-64.

[KLM] R. M. Karp, M. Luby and Neal Madras, "Monte-Carlo Approximation Algorithms for Enumeration Problems," *J. of Algorithms*, Vol. 10, No. 3, Sept. 1989, pp. 429-448.

[N] I. Newman, "Private vs. Common Random Bits in Communication Complexity", *Information Processing Letters 39,* 1991, pp. 67-71.

[VT] H. L. Van Trees, *"Detection, Estimation, and Modulation Theory",* Part 1, John Wiley & sons, 1968.