Streaming Algorithms for Estimating Entropy

Nicholas J. A. Harvey
MIT CSAIL
Cambridge, MA
nickh@mit.edu

Jelani Nelson
MIT CSAIL
Cambridge, MA
minilek@mit.edu

Krzysztof Onak
MIT CSAIL
Cambridge, MA
konak@mit.edu

***Abstract*— We give a method for estimating the empirical Shannon entropy of a distribution in the streaming model of computation. Our approach reduces this problem to the well-studied problem of estimating frequency moments. The analysis of our approach is based on new results which establish quantitative bounds on the rate of convergence of Rényi entropy towards Shannon entropy.**

I. Introduction

The Problem. This work considers the following technical problem. There is a vector $A = (A_1, \ldots, A_n) \in \mathbb{Z}^n$, initially zero. We see a sequence of *updates* to this vector, each update of the form "increment A_i by j", where $j = \pm 1$. We see m such updates, one-by-one, in a fixed order. After seeing all updates, we wish to compute the empirical entropy of the distribution $A/\|A\|_1$. Assume for simplicity that $A \geq 0$ at all times.

This is a trivial problem, if we can afford to store the entire vector A or the entire sequence of updates. The problem becomes non-trivial if we allow ourselves significantly less space. Can one estimate the entropy to within ϵ, using only $\mathrm{poly}(1/\epsilon, \log n, \log m)$ bits of space?

The Motivation. Why does this problem deserve study? One application arises in the field of network anomaly detection. Internet service providers' business depends on providing good quality of service to their users; they need to detect a wide range of anomolous conditions quickly, so that corrective action can be taken. The volume of traffic that ISPs process is enormous, too much to be subjected to thorough scrutiny. Instead, lightweight methods are needed to give broad statistical summaries of the general traffic distribution.

Let us consider a concrete example. One form of malicious activity on the internet is *port scanning*, in which attackers probe target machines, trying to find open ports which could be leveraged for further attacks. In contrast, typical internet traffic is directed to a small number of heavily used ports for web traffic, email delivery, etc. Consequently, when a port scanning attack is underway, there is a significant change in the distribution of port numbers in the packets being delivered. Such attacks can be detected by measuring the entropy of the distribution of port numbers. See Lakhina et al. [12] and Xu et al. [17] for further information about such problems and methods for their solution.

The Model. The problem described above is based on a model of computation in which algorithms see a sequence of updates to an object and must estimate properties of that object while using a very small amount of space. This model of computation has become known as the *streaming model* of computation. It has received much attention since the seminal work of Alon, Matias and Szegedy [1]; Muthukrishnan [14] gives a comprehensive survey. Our assumption that $A \geq 0$ at all times is known as the *strict turnstile model*.

The work of Alon et al. focuses on approximating the α^{th} *frequency moment*, defined to be $\|A\|_\alpha^\alpha = \sum_{i=1}^n A_i^\alpha$, for $\alpha \geq 0$. A sequence of papers [2], [10], [11], [16] established a remarkable result: if $\epsilon > 0$ is constant, then a $(1+\epsilon)$-multiplicative approximation can be computed in $\mathrm{poly}(\log n)$ bits of space if $0 \leq \alpha \leq 2$, but $\Omega(n^{1-2/\alpha - o(1)})$ space is necessary for all $\alpha > 2$.

Many other problems have been studied in the streaming model, but estimating frequency moments remains the central, and best-understood problem. It is worth pointing out that the space lower bounds for computing frequency moments [2], [16] are proven using information theoretic techniques. This is a common approach for lower bounds in the streaming model.

The Precursors. Estimating empirical entropy in the streaming model is a problem that has been studied in the recent literature [3], [5], [6], [9], [18]. Two sorts of estimates are typically of interest: multiplicative $(1+\epsilon)$ approximations, and additive ϵ approximations. An attractive feature of additive approximations is that they can be used to give additive approximations of conditional entropy and mutual information. For simplicity, this paper focuses on additive approximations; the full version of this paper extends our techniques to give multiplicative approximations as well.

The work of Chakrabarti et al. [5] yields an algorithm using $O(\epsilon^{-2} \log^3 m)$ words[1] of space to give an additive ϵ approximation. However, their algorithm cannot handle deletions: all updates must increment the value of some coordinate A_i. In contrast, the algorithm of Bhuvanagiri and Ganguly [3] can handle deletions, but the space required is roughly $O(\epsilon^{-3} \log^7 m)$ words. Zhao et al. [18] give practical methods for estimating the so-called *entropy norm* of a stream, defined to be $\sum_{i=1}^n A_i \log A_i$, where A is not normalized to be a distribution.

Our Contributions. In this work, we give a clean reduction from the entropy estimation problem to the problem of estimating frequency moments. Our algorithm is very simple, and it gives an additive ϵ approximation of entropy using $\tilde{O}(\epsilon^{-4} \log^4 m)$ words of space. Our algorithm works in the

[1] A *word* is a unit of storage containing $\log(n+m)$ bits of data.

strict turnstile model, or even the more general model in which we only require that $A \geq 0$ at the end of the stream. This is the strongest model in which entropy estimation makes sense. Thus, our algorithm improves on the result of Bhuvanagiri and Ganguly, if ϵ is not too small. The full version of this paper gives a more involved algorithm which further improves the space requirements.

The basic idea of our algorithm is to estimate Rényi α-entropy, for $\alpha \approx 1$, then to use this as an estimate of Shannon entropy. The accuracy of this estimate is guaranteed by lemmas that we prove which analyze the rate of convergence of Rényi entropy towards Shannon entropy.

To our knowledge, no results on this rate of convergence were previously known. Budimir et al. [4] and Dragomir [8] give inequalities bounding the difference between Shannon entropy and Rényi α-entropy, but these bounds diverge as $\alpha \to 1$. Życzkowski [19] also states bounds relating Shannon entropy to Rényi entropies, but some gaps in the proofs were later discovered.

We remark that the algorithm of Zhao et al. [18] also uses frequency moment estimation, although their algorithm is intended for estimating the entropy norm, and seems to only work for certain ranges of parameters.

II. Information Theoretic Results

Let x be a distribution on n elements. For convenience, we will assume that all logarithms are natural (i.e., we measure entropy in nats). For $0 \leq \alpha$ and $\alpha \neq 1$, the Rényi α-entropy is defined

$$H_\alpha(x) \;=\; \frac{1}{1-\alpha} \log\Big(\sum_{i=1}^{n} x_i^\alpha\Big).$$

When x is understood, we will simply write H_α. We also define $H_1 = \lim_{\alpha \to 1} H_\alpha$.

It is known that H_1 equals Shannon entropy, which we denote $H(x) = -\sum_{i=1}^{n} x_i \log x_i$. Our main technical result is an analysis of the rate of convergence of H_α towards H as $\alpha \to 1$.

Theorem 2.1: Let $x \in \mathbb{R}^n$ be a probability distribution whose smallest positive value is at least $1/m$, where $m \geq n$. Let $0 < \epsilon < 1$ be arbitrary. Define $\mu = \epsilon/(4 \log m)$ and $\alpha = 1 + \mu/\big(16 \log(1/\mu)\big)$. Then

$$1 \;\leq\; \frac{H_1}{H_\alpha} \;\leq\; 1 + \epsilon. \tag{1}$$

Define $\nu = \epsilon/(4 \log n \log m)$ and $\alpha = 1 + \nu/\big(16 \log(1/\nu)\big)$. Then

$$0 \;\leq\; H_1 - H_\alpha \;\leq\; \epsilon. \tag{2}$$

The same result holds if we measure entropy in bits rather than nats. The following result shows that the multiplicative bound (Eq. (1)) is nearly tight, so long as ϵ is not too small.

Theorem 2.2: There exists a distribution x whose smallest positive value is $1/m$ such that, for $\frac{10}{\log m} < \epsilon < 1$ and $\alpha = 1 + \frac{\epsilon}{\log m}$, we have

$$\frac{H_1}{H_\alpha} \;\geq\; 1 + \Omega(\epsilon).$$

III. Algorithm

The following pseudocode describes our algorithm for estimating Shannon entropy.

Algorithm 1. Our algorithm for estimating the empirical Shannon entropy of a stream to within an additive error of ϵ. For simplicity, assume that m is known in advance.

Set $\nu = \epsilon/(4 \log n \log m)$, $\alpha = 1 + \nu/\big(16 \log(1/\nu)\big)$, and $\tilde{\epsilon} = \epsilon \cdot (\alpha - 1)$
Process the entire stream:
  Compute $\tilde{F}_\alpha$, a $(1+\tilde{\epsilon})$-approximation of $\|A\|_\alpha^\alpha$
  Compute $\|A\|_1$
Return $\log(\tilde{F}_\alpha / \|A\|_1^\alpha)/(1-\alpha)$

To compute the estimate $\tilde{F}_\alpha$, we use the algorithm of Li [13], which requires only $O(\tilde{\epsilon}^{-2})$ words of space. Computing $\|A\|_1$ is trivial since we assume the strict turnstile model. Since $\tilde{\epsilon} = \tilde{\Omega}\big(\epsilon^2/(\log n \log m)\big)$, the total space required by our algorithm is $\tilde{O}(\epsilon^{-2} \log^2 n \log^2 m)$ words.

The accuracy of our estimates is ensured by the following argument. Let x be the distribution $A/\|A\|_1$. With constant probability, $\tilde{F}_\alpha = (1 \pm \tilde{\epsilon}) \|A\|_\alpha^\alpha$. Then

$$\begin{aligned}
\frac{1}{1-\alpha} \log\left(\frac{\tilde{F}_\alpha}{\|A\|_1^\alpha}\right) \;&=\; \frac{1}{1-\alpha} \log\Big((1 \pm \tilde{\epsilon}) \sum_{i=1}^{n} x_i^\alpha\Big) \\
&=\; H_\alpha(x) \;\pm\; O\Big(\frac{\tilde{\epsilon}}{1-\alpha}\Big) \\
&=\; H(x) \;\pm\; O(\epsilon).
\end{aligned}$$

The last line follows from Theorem 2.1.

IV. Proofs

A. Preliminaries

Claim 4.1: The following inequalities hold.

- Let $y \in \mathbb{R}$. Then $1 - y \leq e^{-y}$.
- Let $0 < y < 1$. Then $1 - y + y^2/3 \leq e^{-y}$.
- Let $y > 0$. Then $e^{-y} < 1 - y + y^2/2$.
- Let $0 < y \leq 1$. Then $e^y < 1 + 2y$.
- Let $y > 0$. Then $1 - y \leq \log(1/y)$.
- Let $0 \leq y \leq 1/2$. Then $1/(1-y) \leq 1 + 2y$.

Claim 4.2: Given any $c \in (0,1)$, we have $\log x \geq (x-1)\log(c)/(c-1)$ for all $x \in [c, 1]$.

Claim 4.3: Let $y > 1$. Then $\log(1 - \frac{1}{y}) \geq -\frac{1}{y} - \frac{1}{y(y-1)}$.

Proof. We have

$$\begin{aligned}
\log(1 - 1/y) \;&=\; \log((y-1)/y) \;\geq\; 1 - \frac{y}{y-1} \\
&=\; -\frac{1}{y-1} \;=\; -\frac{1}{y} - \frac{1}{y(y-1)},
\end{aligned} \tag{3}$$

the inequality via Claim 4.1. ∎

Claim 4.4: Let $y > 1$ and $z > 0$. Then

$$1 - \frac{z}{y} - \frac{z}{y(y-1)} \;\leq\; (1 - 1/y)^z \;\leq\; 1 - \frac{z}{y} + \Big(\frac{z}{y}\Big)^2/2.$$

Proof. By Claim 4.1 and Claim 4.3,

$$(1-1/y)^z \;=\; e^{z\log(1-1/y)} \;\geq\; 1 + z\log(1-1/y) \;\geq\; 1 - \frac{z}{y} - \frac{z}{y(y-1)}.$$

On the other hand, Claim 4.1 also shows

$$(1-1/y)^z \;\leq\; e^{-z/y} \;\leq\; 1 - z/y + (z/y)^2/2,$$

as required. ∎

Claim 4.5: Let $1 \leq a \leq b$ and let $x \in \mathbb{R}^n$. Then $\|x\|_b \leq \|x\|_a \leq n^{1/a-1/b}\|x\|_b$.

Claim 4.6: If $0 \leq \alpha \leq \beta$ then $H_\alpha \geq H_\beta$

Claim 4.7: If $\alpha > 1$ then $\log\big(1/\|x\|_\alpha\big) < (\alpha-1)\cdot H_1$.

Proof. $\log\big(1/\|x\|_\alpha\big) = \frac{\alpha-1}{\alpha}H_\alpha(x) < (\alpha-1)\cdot H_\alpha(x) \leq (\alpha-1)\cdot H_1(x)$. ∎

Claim 4.8: Let $y = (y_1,\ldots,y_n)$ and $z = (z_1,\ldots,z_n)$ be probability distributions such that $\|y-z\|_1 \leq 1/2$. Then

$$|H_1(y) - H_1(z)| \;\leq\; \|y-z\|_1 \cdot \log\Big(\frac{n}{\|y-z\|_1}\Big).$$

Proof. See Cover and Thomas [7, 16.3.2]. ∎

B. Proof of Theorem 2.1

Recall that $x \in \mathbb{R}^n$ is a distribution whose smallest positive value is at least $1/m$.

Lemma 4.9: Let $\alpha > 1$, let $\xi = \xi(\alpha)$ denote $4(\alpha-1)H_1(x)$, and let

$$e(\alpha) \;=\; 2\Big(\xi\log n \;+\; \xi\log(1/\xi)\Big).$$

Assume that $\xi(\alpha) < 1/4$. Then $H_\alpha \leq H_1 \leq H_\alpha + e(\alpha)$.

Proof. The first inequality follows from Claim 4.6 so we focus on the second one. Define $f(\alpha) = \log\|x\|_\alpha^\alpha$ and $g(\alpha) = 1-\alpha$, so that $H_\alpha = f(\alpha)/g(\alpha)$. The derivatives are

$$f'(\alpha) \;=\; \frac{\sum_{i=1}^n x_i^\alpha \log x_i}{\|x\|_\alpha^\alpha} \qquad \text{and} \qquad g'(\alpha) \;=\; -1,$$

so $\lim_{\alpha\to1} f'(\alpha)/g'(\alpha)$ exists and equals $H(x)$. Since $\lim_{\alpha\to1} f(\alpha) = \lim_{\alpha\to1} g(\alpha) = 0$, L'Hôpital's rule implies that $\lim_{\alpha\to1} H_\alpha = H(x)$. A stronger version of L'Hôpital's rule is as follows.

Claim 4.10: Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be differentiable functions such that the following limits exist

$$\lim_{\alpha\to1} f(\alpha) = 0, \quad \lim_{\alpha\to1} g(\alpha) = 0, \quad \lim_{\alpha\to1} f'(\alpha)/g'(\alpha) = L.$$

Let ϵ and δ be such that $|\alpha-1| < \delta$ implies that $|f'(\alpha)/g'(\alpha) - L| < \epsilon$. Then $|\alpha-1| < \delta$ also implies that $|f(\alpha)/g(\alpha) - L| < \epsilon$.

Proof. See Rudin [15, p.109]. □

Thus, to prove our lemma, it suffices to show that $|f'(\alpha)/g'(\alpha) - H_1| < e(\alpha)$. (In fact, we actually need $|f'(\beta)/g'(\beta) - H_1| < e(\alpha)$ for all $\beta \in (1,\alpha]$, but this follows by monotonicity of $e(\beta)$ for $\beta \in (1,\alpha]$.)

A key concept in this proof is the "perturbed" probability distribution $x(\alpha)$, defined by $x(\alpha)_i = x_i^\alpha / \|x\|_\alpha^\alpha$. We have the following relationship.

$$\begin{aligned}
\frac{f'(\alpha)}{g'(\alpha)} &= \frac{\sum_{i=1}^n x_i^\alpha \log(1/x_i)}{\|x\|_\alpha^\alpha} \\
&= \frac{\sum_{i=1}^n x_i^\alpha \big(\log(1/x_i) + \log\|x\|_\alpha - \log\|x\|_\alpha\big)}{\|x\|_\alpha^\alpha} \\
&= \frac{\Big(\sum_{i=1}^n x_i^\alpha \log(\|x\|_\alpha / x_i)\Big) \;-\; \Big(\sum_{i=1}^n x_i^\alpha \log\|x\|_\alpha\Big)}{\|x\|_\alpha^\alpha} \\
&= \frac{1}{\alpha}\sum_{i=1}^n \frac{x_i^\alpha}{\|x\|_\alpha^\alpha}\log\left(\frac{\|x\|_\alpha^\alpha}{x_i^\alpha}\right) \;-\; \log\|x\|_\alpha \\
&= \frac{H_1\big(x(\alpha)\big)}{\alpha} \;+\; \log(1/\|x\|_\alpha)
\end{aligned}$$

In summary, we have shown that

$$\left|\frac{f'(\alpha)}{g'(\alpha)} - \frac{H_1\big(x(\alpha)\big)}{\alpha}\right| \;\leq\; \log(1/\|x\|_\alpha) \;\leq\; (\alpha-1)\cdot H_1(x), \tag{4}$$

the last inequality following from Claim 4.7. To use this bound, we observe that:

$$\begin{aligned}
&\left|\frac{f'(\alpha)}{g'(\alpha)} - H_1\big(x(\alpha)\big)\right| \\
&\quad= \left|\frac{f'(\alpha)}{g'(\alpha)} - \frac{H_1\big(x(\alpha)\big)}{\alpha} + \left(\frac{1}{\alpha}-1\right)H_1\big(x(\alpha)\big)\right| \\
&\quad\leq \left|\frac{f'(\alpha)}{g'(\alpha)} - \frac{H_1\big(x(\alpha)\big)}{\alpha}\right| \;+\; |1/\alpha - 1|\cdot H_1\big(x(\alpha)\big)
\end{aligned}$$

We substitute Eq. (4) into this expression, and use $|1/\alpha - 1| \leq \alpha - 1$ (valid since $\alpha \geq 1$). This yields:

$$\left|\frac{f'(\alpha)}{g'(\alpha)} - H_1\big(x(\alpha)\big)\right| \leq (\alpha-1)\cdot H_1(x) + (\alpha-1)\cdot H_1\big(x(\alpha)\big) \tag{5}$$

Recall that our goal is to analyze $|f'(\alpha)/g'(\alpha) - H_1(x)|$. We do this by showing that $H_1\big(x(\alpha)\big) \approx H_1(x)$, and that the right-hand side of Eq. (5) is at most $e(\alpha)$. This is done using Claim 4.8; the key step is bounding $\|x - x(\alpha)\|_1$.

Claim 4.11: Suppose that $1 < \alpha \leq 1 + 1/(2\log n)$. Then $1/\|x\|_\alpha^\alpha < 1 + 3(\alpha-1)H_1(x)$.

Proof. From Claim 4.5 and $\|x\|_1 = 1$, we obtain $1/\|x\|_\alpha \leq n^{1-1/\alpha} < n^{\alpha-1}$. Our hypothesis on α implies that

$$\alpha\cdot\log(1/\|x\|_\alpha) \;<\; \alpha\cdot(\alpha-1)\log n \;<\; 2\cdot(\alpha-1)\log n \;\leq\; 1. \tag{6}$$

Thus

$$\begin{aligned}
\frac{1}{\|x\|_\alpha^\alpha} \;&=\; e^{\alpha\log(1/\|x\|_\alpha)} \;<\; 1 + 2\cdot\alpha\log(1/\|x\|_\alpha) \\
&<\; 1 + 3(\alpha-1)H_1(x).
\end{aligned}$$

The first inequality is from Claim 4.1 and Eq. (6), and the second from Claim 4.7. □

Recall that $\xi = 4(\alpha-1)H_1(x)$.

Claim 4.12: $\|x - x(\alpha)\|_1 \le \xi$.

Proof. To avoid the absolute values, we shall split the sum defining $\|x - x(\alpha)\|_1$ into two cases. For that purpose, let $S = \{\, i \,:\, x(\alpha)_i \ge x_i \,\}$. Then

$$\begin{aligned}
&\|x - x(\alpha)\|_1 \\
&\quad = \sum_{i\in S} \big(x(\alpha)_i - x_i\big) \;+\; \sum_{i\notin S} \big(x_i - x(\alpha)_i\big) \\
&\quad = \sum_{i\in S} x_i \cdot \left(\frac{x_i^{\alpha-1}}{\|x\|_\alpha^\alpha} - 1\right) + \sum_{i\notin S} x_i \cdot \left(1 - \frac{x_i^{\alpha-1}}{\|x\|_\alpha^\alpha}\right)
\end{aligned}$$

The first sum is upper-bounded using $x_i^{\alpha-1} \le 1$ and $\sum_{i\in S} x_i \le 1$. The second sum is upper-bounded using $\|x\|_\alpha^\alpha \le 1$ and $1 - x_i^{\alpha-1} \le \log\big(1/x_i^{\alpha-1}\big)$ (see Claim 4.1).

$$\begin{aligned}
&\le \left(\frac{1}{\|x\|_\alpha^\alpha} - 1\right) \;+\; (\alpha-1)\sum_{i\notin S} x_i \log(1/x_i) \\
&\le 3(\alpha-1)H_1(x) \;+\; (\alpha-1)H_1(x),
\end{aligned}$$

using Claim 4.11. This completes the proof. □

Thus, by our assumption that $\xi(\alpha) < 1/4$, by Claim 4.8, by Claim 4.12, and by the fact that $x \mapsto x\log(1/x)$ is monotonically increasing for $x \in (0, 1/4)$, we obtain that

$$|H_1(x) - H_1(x(\alpha))| \;\le\; \xi\log n + \xi\log(1/\xi).$$

Now we assemble the error bounds. Our result from Eq. (5) yields

$$\begin{aligned}
&\left|\frac{f'(\alpha)}{g'(\alpha)} - H_1(x)\right| \\
&\quad\le \left|\frac{f'(\alpha)}{g'(\alpha)} - H_1(x(\alpha))\right| \;+\; |H_1(x) - H_1(x(\alpha))| \\
&\quad\le \Big((\alpha-1)H_1(x) + (\alpha-1)H_1(x(\alpha))\Big) \\
&\qquad\quad +\; |H_1(x) - H_1(x(\alpha))| \\
&\quad\le 2(\alpha-1)H_1(x) \;+\; \alpha\cdot|H_1(x) - H_1(x(\alpha))| \\
&\quad\le 2\Big(\xi\log n + \xi\log(1/\xi)\Big)
\end{aligned}$$

This completes the proof. ■

We now use Lemma 4.9 to show that $H_\alpha \approx H_1$, if α is sufficiently small.

Proof (of Theorem 2.1). First we focus on Eq. (1). The lower bound is immediate from Claim 4.6, so we show the upper-bound. For an arbitrary $\mu \in (0,1)$, we have

$$\mu^2 \;<\; \frac{\mu}{2\log(1/\mu)} \;<\; \mu;$$

this follows since $\mu\log(1/\mu) < 1/2$ for all μ. Let $\tilde\mu = \mu/\big(2\log(1/\mu)\big)$. Then

$$\tilde\mu\log(1/\tilde\mu) \;<\; \mu.$$

This follows since $\mu^2 < \tilde\mu \implies 1/\tilde\mu < 1/\mu^2 \implies \log(1/\tilde\mu) < 2\log(1/\mu)$.

The hypotheses of Theorem 2.1 give $\alpha = 1 + \tilde\mu/8$. Hence,

$$\begin{aligned}
e(\alpha) \;&=\; 8(\alpha-1)H_1\Big[\log n \;+\; \log\Big(1/\big(4(\alpha-1)H_1\big)\Big)\Big] \\
&\le\; \tilde\mu H_1\Big[\log n \;+\; \log\big(2/(\tilde\mu H_1)\big)\Big]
\end{aligned}$$

Since $H_1 \ge (\log m)/m$ for any distribution satisfying our hypotheses, this is at most

$$\begin{aligned}
&\le\; \tilde\mu H_1\Big(\log n \;+\; \log(1/\tilde\mu) \;+\; \log m\Big) \\
&\le\; (\log m)\mu H_1 \;<\; (\epsilon/2)H_1,
\end{aligned}$$

since our hypotheses give $\mu = \epsilon/(4\log m)$. Applying Lemma 4.9, we obtain that

$$\begin{aligned}
H_1 - H_\alpha \;&\le\; (\epsilon/2)H_1 \\
\implies\quad (1-\epsilon/2)H_1 \;&\le\; H_\alpha \\
\implies\quad \frac{H_1}{H_\alpha} \;&\le\; \frac{1}{1-\epsilon/2} \;\le\; 1+\epsilon,
\end{aligned}$$

the last inequality following from Claim 4.1. This establishes Eq. (1).

Let us now consider the above argument, replacing μ with $\nu = \epsilon/(4\log n\log m)$. We obtain

$$e(\alpha) \;\le\; (\log m)\nu H_1 \;\le\; \epsilon/4,$$

since $H_1 \le \log n$. Thus, Eq. (2) follows directly. ■

C. Proof of Theorem 2.2

We consider the very simple distribution with $x_1 := 1 - 1/m$ and $x_2 := 1/m$. The Shannon entropy is trivial to analyze.

$$H_1(x) \;=\; \big(\tfrac{m-1}{m}\big)\log\big(\tfrac{m}{m-1}\big) + \tfrac{\log m}{m} \;>\; \tfrac{\log m}{m} \qquad (7)$$

Now recall that $\frac{10}{\log m} < \epsilon < 1$ and $\alpha = 1 + \frac{\epsilon}{\log m}$. We have

$$H_\alpha \;=\; -\left(\tfrac{\log m}{\epsilon}\right)\log\big(x_1^\alpha + x_2^\alpha\big).$$

To prove the theorem, the main task is to prove a lower bound on $\log(x_1^\alpha + x_2^\alpha)$. First, Claim 4.4 directly shows that

$$1 - \frac{\alpha}{m} - \frac{\alpha}{m(m-1)} \;\le\; x_1^\alpha \;\le\; 1 - \frac{\alpha}{m} + \frac{\alpha^2}{2m^2}. \qquad (8)$$

Next, by Claim 4.1, we have

$$x_2^\alpha \;=\; \frac{1}{m^{1+\epsilon/\log m}} \;=\; \frac{e^{-\epsilon}}{m} \;\ge\; \frac{1 - \epsilon + \epsilon^2/3}{m}. \qquad (9)$$

Claim 4.13: $\log(x_1^\alpha)/(x_1^\alpha - 1) \le 1 + O(1/m)$.

Proof. Claim 4.3 implies that

$$\log(x_1^\alpha) \;=\; \alpha\,\log(1 - 1/m) \;\ge\; -\frac{\alpha}{m} - \frac{\alpha}{m(m-1)}.$$

On the other hand, Eq. (8) shows that

$$x_1^\alpha - 1 \;\le\; -\frac{\alpha}{m} + \frac{\alpha^2}{2m^2}.$$

We now upper bound $\log(x_1^\alpha)/(x_1^\alpha - 1)$. (Note both numerator and denominator are negative).

$$\begin{aligned}\frac{\log(x_1^\alpha)}{x_1^\alpha - 1} &\leq \frac{-\alpha/m - \alpha/\big(m(m-1)\big)}{-\alpha/m + \alpha^2/(2m^2)} = \frac{1 + 1/(m-1)}{1 - \alpha/(2m)} \\ &\leq \big(1 + 1/(m-1)\big)\big(1 + \alpha/m\big),\end{aligned}$$

as required. □

Thus, by Eq. (8) and Eq. (9),

$$\begin{aligned}&\log(x_1^\alpha + x_2^\alpha) \\ &\quad\geq \log\left(\Big(1 - \frac{\alpha}{m} - \frac{\alpha}{m(m-1)}\Big) + \Big(\frac{1 - \epsilon + \epsilon^2/3}{m}\Big)\right) \\ &\quad\geq \log\left(1 - \frac{\epsilon - \epsilon^2/3}{m} - \frac{\epsilon}{m\log m} - O(1/m^2)\right)\end{aligned}$$

We lower bound this using Claim 4.2, taking $c = x_1^\alpha$ and using Claim 4.13 to lower bound $\log(c)/(c-1)$.

$$\begin{aligned}&\geq \left(-\frac{\epsilon - \epsilon^2/3}{m} - \frac{\epsilon}{m\log m} - O(1/m^2)\right)\big(1 + O(1/m)\big) \\ &\geq \frac{-\epsilon}{m}\Big(1 - \epsilon/3 + \frac{1}{\log m} + O\big(1/(\epsilon m)\big)\Big)\big(1 + O(1/m)\big)\end{aligned}$$

Thus

$$\begin{aligned}H_\alpha(x) &= \frac{1}{1-\alpha}\log(x_1^\alpha + x_2^\alpha) \\ &= -\frac{\log m}{\epsilon}\cdot\log(x_1^\alpha + x_2^\alpha) \\ &\leq -\frac{\log m}{\epsilon}\cdot\left(\frac{-\epsilon}{m}\Big(1 - \frac{\epsilon}{3} + \frac{1}{\log m} + O\big(\tfrac{1}{\epsilon m}\big)\Big)\right. \\ &\qquad\left.\cdot\big(1 + O(\tfrac{1}{m})\big)\right) \\ &= \frac{\log m}{m}\Big(1 - \epsilon/3 + \frac{1}{\log m} + O\big(1/(\epsilon m)\big)\Big) \\ &\qquad\cdot\big(1 + O(1/m)\big) \\ &= \frac{\log m}{m}\Big(1 - \epsilon/3 + \frac{1}{\log m} + O\big(1/(\epsilon m)\big)\Big)\end{aligned}$$

Comparing this with Eq. (7) shows $H_\alpha(x) \leq H_1\big(1 - \Omega(\epsilon)\big)$, since $\epsilon > 10/\log(m)$.

Acknowledgements

The authors thank Piotr Indyk and Ping Li for several helpful discussions on this topic.

N. Harvey is supported by a Natural Sciences and Engineering Research Council of Canada PGS Scholarship, by NSF contract CCF-0515221 and by ONR grant N00014-05-1-0148. J. Nelson is supported by a National Defense Science and Engineering Graduate (NDSEG) Fellowship. K. Onak is supported in part by NSF contract 0514771.

References

[1] N. Alon, Y. Matias, and M. Szegedy. The Space Complexity of Approximating the Frequency Moments. *Journal of Computer and System Sciences*, 58(1):137–147, 1999.

[2] Z. Bar-Yossef, T. S. Jayram, R. Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004.

[3] L. Bhuvanagiri and S. Ganguly. Estimating entropy over data streams. In *Proceedings of the 14th Annual European Symposium on Algorithms*, pages 148–159, 2006.

[4] I. Budimir, S. Dragomir, and J. Pečarić. Further Reverse Results for Jensen's Discrete Inequality and Applications in Information Theory. *Journal of Inequalities in Pure and Applied Mathematics*, 2(1), 2001.

[5] A. Chakrabarti, G. Cormode, and A. McGregor. A near-optimal algorithm for computing the entropy of a stream. In *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 328–335, 2007.

[6] A. Chakrabarti, K. Do Ba, and S. Muthukrishnan. A near-optimal algorithm for computing the entropy of a stream. In *Proceedings of the 23rd Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 196–205, 2006.

[7] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley Interscience, 1991.

[8] S. Dragomir. On some inequalities for the Renyi α-entropy. *Revista de Matematica e Estatistica*, 19:349–362, 2001.

[9] S. Guha, A. McGregor, and S. Venkatasubramanian. Streaming and sublinear approximation of entropy and information distances. In *Proceedings of the 17th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 733–742, 2006.

[10] P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *Journal of the ACM*, 53(3):307–323, 2006.

[11] P. Indyk and D. P. Woodruff. Optimal approximations of the frequency moments of data streams. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC)*, pages 202–208, 2005.

[12] A. Lakhina, M. Crovella, and C. Diot. Mining anomalies using traffic feature distributions. In *Proceedings of the ACM SIGCOMM Conference*, pages 217–228, 2005.

[13] P. Li. Estimators and Tail Bounds for Dimension Reduction in L_α $(0 < \alpha \leq 2)$ Using Stable Random Projections. In *Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2008.

[14] S. Muthukrishnan. Data Streams: Algorithms and Applications. *Foundations and Trends in Theoretical Computer Science*, 1(2):117–236, 2005.

[15] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, third edition, 1976.

[16] M. E. Saks and X. Sun. Space lower bounds for distance approximation in the data stream model. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing (STOC)*, pages 360–369, 2002.

[17] K. Xu, Z.-L. Zhang, and S. Bhattacharyya. Profiling internet backbone traffic: behavior models and applications. In *Proceedings of the ACM SIGCOMM Conference*, pages 169–180, 2005.

[18] H. Zhao, A. Lall, M. Ogihara, O. Spatscheck, J. Wang, and J. Xu. A Data Streaming Algorithm for Estimating Entropies of OD Flows. In *Proceedings of the Internet Measurement Conference (IMC)*, 2007.

[19] K. Życzkowski. Rényi Extrapolation of Shannon Entropy. *Open Systems & Information Dynamics*, 10(3):297–310, 2003.