Simple Analysis of Priority Sampling

Majid Daliri[†] Juliana Freire[†] Christopher Musco[†] Aécio Santos[†]
Haoxiang Zhang[†]

Abstract

We prove a tight upper bound on the variance of the priority sampling method (aka sequential Poisson sampling). Our proof is significantly shorter and simpler than the original proof given by Mario Szegedy at STOC 2006, which resolved a conjecture by Duffield, Lund, and Thorup.

1 Background

Suppose we have a list of non-negative numbers $w_1, \ldots, w_n$. A common task in streaming and distributed algorithms is to collect a sample of this list, which can then be used to estimate arbitrary sums of a subset of the numbers. As toy examples, we might hope to estimate $\sum_{i:i \text{ is odd}} w_i$ or $\sum_{i:100\leq i\leq 200} w_i$. Importantly, the condition used to determine the subset will not be known in advance when collecting samples.

It has been observed that to obtain accurate results for subset sum estimation, it is usually important to sample from $w_1, \ldots, w_n$ with "probability proportional to size". I.e., we want to collect a subset of size $k \ll n$ from this list in such a way that larger numbers are sampled with higher probability – ideally proportional to or approximately proportional to their size. Such a subset will typically be more useful in estimating sums than a uniform sample. We will also refer to probability proportional to size as "weighted sampling".

1.1 Threshold Sampling One simple approach for weighted sampling is the so-called *Threshold Sampling* method, also referred to as Poisson sampling [Duffield et al., 2005]. Threshold sampling is used in computer science due to applications in *sample coordination*, a topic beyond the scope of this note [Flajolet, 1990, Cohen and Kaplan, 2013]. For each item w_i, we draw a uniform random variable $u_i \sim \text{Unif}[0, 1]$. Then, we fix a threshold $\tau \geq 0$ and sample all numbers w_i for which $\frac{u_i}{w_i} \leq \tau$. Evidently, w_i gets sampled with probability:

$$p_i = \min(1, w_i\tau).$$

The probabilities $p_1, \ldots, p_n$ are approximately proportional to the weights $w_1, \ldots, w_n$ (in fact, exactly proportional unless $w_i\tau > 1$ for some i). The expected number of items sampled is upper bounded by $\sum_{i=1}^n p_i \leq \sum_{i=1}^n w_i\tau = \tau \cdot W$, where $W = \sum_{i=1}^n w_i$. If $\tau = \frac{k}{W}$, the expected number of items sampled is $\leq k$. To use our samples to estimate the sum of a subset of items $\mathcal{I} \subseteq \{1, \ldots, n\}$, a natural approach is to apply the Horvitz-Thompson estimator:

$$\sum_{i\in\mathcal{I}} \hat{w}_i \approx \sum_{i\in\mathcal{I}} w_i \qquad \text{where for all } i \in 1, \ldots, n, \qquad \hat{w}_i = \mathbb{1}\left[\frac{u_i}{w_i} \leq \tau\right] \cdot \frac{w_i}{p_i}.$$

Here $\mathbb{1}[A]$ denotes the indicator random variable that evaluates to 1 if the event A is true and 0 otherwise. It is not hard to see that $\mathbb{E}[\hat{w}_i] = w_i$, so $\sum_{i\in\mathcal{I}} \hat{w}_i$ is an unbiased estimate for the true subset sum. Since $\hat{w}_1, \ldots, \hat{w}_n$ are independent, the variance of this estimate is $\sum_{i\in\mathcal{I}} \text{Var}[\hat{w}_i]$, a quantity that depends on the unknown set $\mathcal{I}$. So, in lieu of bounding variance, a common goal is to bound the *total variance*, $\sum_{i=1}^n \text{Var}[\hat{w}_i]$.[1] To do so, first note that when $w_i\tau \geq 1$ (i.e., $p_i = 1$), we have $\text{Var}[\hat{w}_i] = 0$. So, we can restrict our attention to terms for which

[†]New York University, Tandon School of Engineering.

[1]Other proxy performance measures besides total variance have also been studied, like the average variance for random subsets of a fixed size [Szegedy and Thorup, 2007].

$w_i\tau < 1$ (i.e., $p_i = w_i\tau$). Specifically, letting $\mathcal{K}$ be a set containing all i for which $w_i\tau < 1$, and setting $\tau = \frac{k}{W}$ (so we take at most k samples in expectation), the total variance can bounded by:

$$\sum_{i=1}^{n} \mathrm{Var}[\hat{w}_i] = \sum_{i\in\mathcal{K}} \mathrm{Var}[\hat{w}_i] = \sum_{i\in\mathcal{K}} \frac{w_i^2}{p_i^2} \mathrm{Var}\left[\mathbb{1}\left[\frac{u_i}{w_i} \leq \tau\right]\right] = \sum_{i\in\mathcal{K}} \frac{w_i^2}{p_i^2} p_i(1-p_i) \leq \sum_{i\in\mathcal{K}} \frac{w_i^2}{p_i} = \sum_{i\in\mathcal{K}} \frac{w_i}{\tau} \leq \frac{W}{\tau} = \frac{W^2}{k}.$$

This upper bound of $\frac{W^2}{k}$ for threshold sampling is known to be optimal in the sense that any sampling scheme generating a sequence of random variables $\hat{w}_1, \ldots, \hat{w}_n$ such that $\mathbb{E}[\hat{w}_i] = w_i$ cannot have a lower total variance if the expected number of non-zero variables is $\leq k$ [Duffield et al., 2007].

1.2 Priority Sampling While variance optimal, a disadvantage of threshold sampling is that it only guarantees that k samples are taken *in expectation*. Ideally, we want a scheme that samples *exactly* k items, while still sampling with probabilities (approximately) proportional to the weights $w_1, \ldots, w_n$. Many such schemes exist, including pivotal sampling, reservoir sampling methods, and conditional Poisson sampling [Tillé, 2023]. In computer science, one method of particular interest is *Priority Sampling*, which was introduced to the field by [Duffield et al., 2004], but had been previously studied in statistics under the name “Sequential Poisson Sampling” [Ohlsson, 1998]. Similar to threshold sampling, priority sampling is often preferred in computer science over methods like pivotal sampling due to applications in coordinated random sampling.

In fact, priority sampling is almost identical to threshold sampling. The one (major) difference is that the threshold τ is chosen adaptively to equal the $(k+1)^{\mathrm{st}}$ smallest item in the list $\{\frac{u_1}{w_1}, \ldots, \frac{u_n}{w_n}\}$. Let $\mathcal{S}$ contain all values of i such that $\frac{u_i}{w_i} < \tau$ (i.e., the indices of the k smallest items in the list).[2] We define

$$\hat{w}_i = \begin{cases} \frac{w_i}{\min(1, w_i\tau)} & i \in S \\ 0 & i \notin S \end{cases}. \tag{1.1}$$

As before, to estimate the sum of a subset $\mathcal{I} \subseteq \{1, \ldots, n\}$, we return $\sum_{i\in\mathcal{I}} \hat{w}_i$. Analyzing this estimator is trickier than threshold sampling because τ is now a random number that depends on $u_1, \ldots, u_n$. As a result, $\hat{w}_1, \ldots, \hat{w}_n$ are no longer independent random variables. However, the following (surprising) fact is well known (see [Duffield et al., 2007] or our proof in Appendix A).

Fact 1. *Let $\hat{w}_1, \ldots, \hat{w}_n$ be as defined in* (1.1). *For all i, $\mathbb{E}[\hat{w}_i] = w_i$ and for all $i \neq j$, $\mathbb{E}[\hat{w}_i\hat{w}_j] = w_iw_j$. In other words, the random variables are equal to $w_1, \ldots, w_n$ in expectation, and are pairwise uncorrelated.*

It follows that for any subset $\mathcal{I}$, $\mathbb{E}\left[\sum_{i\in\mathcal{I}} \hat{w}_i\right] = \sum_{i\in\mathcal{I}} w_i$. So, samples collected via priority sampling can be used to obtain an unbiased estimate for subset sums. Additionally, since $\hat{w}_1, \ldots, \hat{w}_n$ are pairwise uncorrelated, we have that $\mathrm{Var}[\sum_{i\in\mathcal{I}} \hat{w}_i] = \sum_{i\in\mathcal{I}} \mathrm{Var}[\hat{w}_i]$, as was the case for threshold sampling. So, a natural goal is still to bound the total variance $\sum_{i=1}^{n} \mathrm{Var}[\hat{w}_i]$. It was shown in [Alon et al., 2005] that $\sum_{i=1}^{n} \mathrm{Var}[\hat{w}_i] = O\left(\frac{W^2}{k}\right)$, where $W = \sum_{i=1}^{n} w_i$. This matches the $\frac{W^2}{k}$ bound for threshold sampling up to a constant factor. However, it was conjectured in that work that the bound could be improved to $\frac{W^2}{k-1}$, which is only just worse than the optimal $\frac{W^2}{k}$. This conjecture was resolved in a 2006 paper by Szegedy:

Theorem 2 ([Szegedy, 2006])**.** *Let $\hat{w}_1, \ldots, \hat{w}_n$ be as defined in* (1.1)*, let $\hat{W} = \sum_{i=1}^{n} \hat{w}_i$, and let $W = \sum_{i=1}^{n} w_i$.*

$$\mathrm{Var}[\hat{W}] = \sum_{i=1}^{n} \mathrm{Var}[\hat{w}_i] \leq \frac{W^2}{k-1}.$$

We note that such a bound is also known to hold for other related sampling methods amenable to sample coordination, like the successive weighted sampling without replacement (PPSWOR) method [Cohen, 2015].

Szegedy's proof of Theorem 2 is quite involved, as it is based on an explicit integral formula for the total variance, and several pages of detailed calculations. We provide a simple alternative proof below.

[2] We note that some papers, including the early work in [Duffield et al., 2004], define as “priorities” $\frac{w_1}{u_1}, \ldots, \frac{w_n}{u_n}$ and select the k indices with the *largest* priority. This is of course equivalent to selecting the indices with the smallest values of $\frac{u_i}{w_i}$.

2 Main Analysis

As in prior work (e.g. [Duffield et al., 2007]) we introduce a new random variable τ_i for each item i. τ_i is equal to the k^{th} smallest value of $\frac{u_j}{w_j}$ for $j \in \{1, \ldots, n\} \setminus \{i\}$. Note that τ_i is independent from u_i, and the probability that i is included in our set of k samples $\mathcal{S}$ is exactly equal to $\Pr[\frac{u_i}{w_i} \le \tau_i] = \min(1, \tau_i w_i)$. Moreover, conditioned on the event that $i \in S$, we have that $\tau = \tau_i$. Accordingly, for $i \in S$, $\frac{w_i}{\min(1,\tau w_i)} = \frac{w_i}{\min(1,\tau_i w_i)}$, and thus $\hat{w}_i$ can equivalently be written as:

$$\hat{w}_i = \begin{cases} \frac{w_i}{\min(1,\tau_i w_i)} & i \in S \\ 0 & i \notin S \end{cases}$$

With this definition in place, we prove some intermediate claims.

Claim 3. $\mathrm{Var}\left[\hat{w}_i\right] \le w_i \cdot \mathbb{E}\left[\frac{1}{\tau_i}\right]$

Proof. We begin by analyzing $\mathbb{E}\left[\hat{w}_i^2\right]$. Conditioning on τ_i, we have:

$$\mathbb{E}\left[\hat{w}_i^2 \mid \tau_i\right] = \frac{w_i^2}{\min(1,\tau_i w_i)^2} \cdot \Pr[i \in S] = \frac{w_i^2}{\min(1,\tau_i w_i)^2} \cdot \min(1,\tau_i w_i) = \frac{w_i^2}{\min(1,\tau_i w_i)} = w_i^2 \cdot \max\left(1, \frac{1}{\tau_i w_i}\right).$$

From the law of total expectation, we thus have that $\mathbb{E}\left[\hat{w}_i^2\right] = w_i^2 \cdot \mathbb{E}\left[\max\left(1, \frac{1}{\tau_i w_i}\right)\right]$. Combined with the fact that $\mathbb{E}\left[\hat{w}_i\right]^2 = w_i^2$ (from Fact 1) we have that $\mathrm{Var}\left[\hat{w}_i\right] = \mathbb{E}\left[\hat{w}_i^2\right] - \mathbb{E}[w_i]^2$ equals:

$$\mathrm{Var}\left[\hat{w}_i\right] = w_i^2 \cdot \left(\mathbb{E}\left[\max\left(1, \frac{1}{\tau_i w_i}\right)\right] - 1\right) = w_i^2 \cdot \mathbb{E}\left[\max\left(0, \frac{1}{\tau_i w_i} - 1\right)\right] = w_i^2 \cdot \mathbb{E}\left[\max\left(0, \frac{1}{\tau_i w_i}\right)\right].$$

And since $\tau_i \cdot w_i$ is a positive value, we obtain:

$$\mathrm{Var}\left[\hat{w}_i\right] \le w_i^2 \cdot \mathbb{E}\left[\frac{1}{\tau_i w_i}\right] = w_i \cdot \mathbb{E}\left[\frac{1}{\tau_i}\right]. \qquad \square$$

Claim 4. $\mathbb{E}\left[\frac{1}{\tau}\right] \le \frac{W}{k}$.

Proof. Consider the random variable $\hat{W} = \sum_{i=1}^{n} \hat{w}_i$. Note that $\hat{W}$ can be rewritten as:

$$\hat{W} = \sum_{i \in S} \hat{w}_i = \sum_{i \in S} \frac{w_i}{\min\left(1, \tau w_i\right)} = \sum_{i \in S} \max\left(w_i, \frac{1}{\tau}\right).$$

Hence, $\hat{W} \ge \sum_{i \in S} \frac{1}{\tau} = \frac{k}{\tau}$. We also know from Fact 1 that $\mathbb{E}[\hat{W}] = W$. If the random variable $\hat{W}$ is always larger than the random variable $\frac{k}{\tau}$, it holds that: $\mathbb{E}\left[\frac{k}{\tau}\right] \le \mathbb{E}\left[\hat{W}\right] = W$. Dividing by k proves the result. $\square$

Claim 5. $\mathbb{E}\left[\frac{1}{\tau_i}\right] \le \frac{W}{k-1}$

Proof. Simply apply Claim 4 to the setting where we collect $k-1$ priority samples from the set of weights $\{w_j : j \in \{1, \ldots, n\} \setminus \{i\}\}$. Note that $W' = \sum_{j=1, j \ne i}^{n} w_j$ is no larger than W, so $\frac{W'}{k-1} \le \frac{W}{k-1}$. $\square$

We are now ready to prove the main result of [Szegedy, 2006].

Proof of Theorem 2. Applying Claim 3 and Claim 5, we have that:

$$\sum_{i=1}^{n} \mathrm{Var}\left[\hat{w}_i\right] \le \sum_{i=1}^{n} w_i \cdot \mathbb{E}\left[\frac{1}{\tau_i}\right] \le \sum_{i=1}^{n} w_i \frac{W}{k-1} = \frac{W^2}{k-1}. \qquad \square$$

3 Discussion and Pedagogical Perspective

It is natural to ask how the proof above avoids the complexity of [Szegedy, 2006]. In fact, it is even simpler than the proof that establishes a looser $O(W^2/k)$ bound from [Alon et al., 2005], which invokes a bucketing argument combined with Chernoff bounds. Where's the magic? We do not have a fully satisfying answer, except to point out that a key step in our proof is to reduce the problem to bounding $\mathbb{E}[1/\tau]$. At first glance, this does not seem productive: as the k^{th} smallest value of n *scaled* uniform random variables, τ is a complicated random variable. In particular, its distribution depends on each of $w_1, \ldots, w_n$ in an involved way. However, as we show in Claim 4, a simple comparison argument can be used to upper bound $\mathbb{E}[1/\tau]$ without even writing down the probability density function (PDF) of τ.

This analysis might be interesting from a pedagogical perspective even when all weights are uniform. In this case, τ is the $(k+1)^{\text{st}}$ smallest out of n uniform draws. Such random variables appear frequently in course material on randomized algorithms, for example in analyzing the elegant distinct elements algorithm from [Bar-Yossef et al., 2002] or when studying the k-minimum values (KMV) sketch. In these applications, it is necessary to compute the expected value and variance of $1/\tau$, which typically involves an explicit expression for the PDF of τ (which is beta distributed), combined with involved calculations [Beyer et al., 2007]. Our approach, on the other hand, gives a simple argument from first principles, which we outline below.

Corollary 6. *Let τ be the $(k+1)^{st}$ smallest out of n uniform random variables $u_1, \ldots, u_n$.*

$$\mathbb{E}\left[\frac{1}{\tau}\right] = \frac{k}{n} \qquad \textit{and} \qquad \mathrm{Var}\left[\frac{1}{\tau}\right] = \frac{n^2 - nk}{k^2(k-1)}.$$

Proof. Let $\mathcal{S}$ denote the set of k indices i for which $u_i < \tau$. Additionally, let τ_1 equal the k^{th} smallest out of $\{u_1, \ldots, u_n\} \setminus \{u_1\}$. There is nothing special about the choice of 1: τ_1 is simply an auxilary variable used in our analysis. We could have instead chosen the k^{th} smallest out of $\{u_1, \ldots, u_n\} \setminus \{u_i\}$ for any i. Consider the random variable X defined equivalently (using the same argument as in the previous section) as:

$$X = \begin{cases} \frac{1}{\tau} & \text{if } 1 \in \mathcal{S} \\ 0 & \text{otherwise} \end{cases} = \begin{cases} \frac{1}{\tau_1} & \text{if } 1 \in \mathcal{S} \\ 0 & \text{otherwise.} \end{cases}$$

From the second definition, we observe that:

$$\mathbb{E}[X] = \mathbb{E}\left[\mathbb{E}\left[X \mid \tau_1\right]\right] = \mathbb{E}\left[\Pr[1 \in \mathcal{S} \mid \tau_1] \cdot \frac{1}{\tau_1}\right] = 1.$$

Alternatively, consider the first definition. The value of τ is independent from the event that $1 \in \mathcal{S}$, and by symmetry, $\Pr[1 \in \mathcal{S}] = \frac{k}{n}$. So,

$$\mathbb{E}[X] = \mathbb{E}\left[\frac{1}{\tau}\right] \cdot \Pr[1 \in \mathcal{S}] = \mathbb{E}\left[\frac{1}{\tau}\right] \cdot \frac{k}{n}.$$

We conclude that in order for $\mathbb{E}\left[\frac{1}{\tau}\right] \cdot \frac{k}{n}$ to equal 1, it must be that

$$\mathbb{E}\left[\frac{1}{\tau}\right] = \frac{n}{k}. \tag{3.2}$$

We can then compute the variance of $1/\tau$ using a similar argument (and applying (3.2) in the last step):

$$\mathbb{E}[X^2] = \mathbb{E}\left[\mathbb{E}[X^2 \mid \tau_1]\right] = \mathbb{E}\left[\Pr[1 \in \mathcal{S} \mid \tau_1] \cdot \frac{1}{\tau_1^2}\right] = E\left[\tau_1 \cdot \frac{1}{\tau_1^2}\right] = \mathbb{E}\left[\frac{1}{\tau_1}\right] = \frac{n-1}{k-1}.$$

Alternatively, again using that τ is independent from $\mathbb{1}\left[1 \in \mathcal{S}\right]$, we have that $\mathbb{E}[X^2] = \mathbb{E}\left[\frac{1}{\tau^2}\right] \cdot \frac{k}{n}$ So, we conclude that $\mathbb{E}\left[\frac{1}{\tau^2}\right] = \frac{n(n-1)}{k(k-1)}$. Finally, we have the bound:

$$\mathrm{Var}\left[\frac{1}{\tau}\right] = \mathbb{E}\left[\frac{1}{\tau^2}\right] - \mathbb{E}\left[\frac{1}{\tau}\right]^2 = \frac{n^2 - nk}{k^2(k-1)}. \qquad \square$$

This matches the formula given e.g., in [Beyer et al., 2007], and establishes that $\mathrm{Var}\left[\frac{1}{\tau}\right] \leq \epsilon^2 \mathbb{E}\left[\frac{1}{\tau}\right]^2$ when $k = O(1/\epsilon^2)$, which is the bound needed to prove that the [Bar-Yossef et al., 2002] distinct elements method gives a $(1 \pm \epsilon)$ relative error approximation with $k = O(1/\epsilon^2)$ space

Acknowledgements

This work was supported by NSF Award CCF-204623. We thank Xiaoou Cheng for carefully reviewing our proof, and also the anonymous reviewers, whose comments helped improve the presentation of this paper.

A Proof of Fact 1

For completeness, we prove the following important and well-known fact about priority sampling, following the approach of existing proofs [Duffield et al., 2007].

Fact 1. *Let $\hat{w}_1, \ldots, \hat{w}_n$ be as defined in* (1.1)*. For all i, $\mathbb{E}[\hat{w}_i] = w_i$ and for all $i \neq j$, $\mathbb{E}[\hat{w}_i \hat{w}_j] = w_i w_j$. In other words, the random variables are equal to $w_1, \ldots, w_n$ in expectation, and are pairwise uncorrelated.*

Proof. Let $\tau_1, \ldots, \tau_n$ be as defined in Section 2. Recall that $\hat{w}_i$ can equivalently be written as:

$$\hat{w}_i = \begin{cases} \frac{w_i}{\min(1, \tau_i w_i)} & i \in S \\ 0 & i \notin S \end{cases}$$

Then, we can compute its expectation:

$$\mathbb{E}[w_i] = \mathbb{E}\left[\mathbb{E}\left[w_i \mid \tau_i\right]\right] = \mathbb{E}\left[\frac{w_i}{\min(1, \tau_i w_i)} \Pr[i \in \mathcal{S} \mid \tau_i]\right] = w_i.$$

In the last step we use that $\Pr[i \in \mathcal{S} \mid \tau_i] = \min(1, \tau_i w_i)$, which follows from noting that, for i to be in $\mathcal{S}$, it must be that $\frac{u_i}{w_i}$ is less than τ_i.

Next we show that $\mathbb{E}\left[\hat{w}_i \hat{w}_j\right] = w_i w_j = \mathbb{E}\left[\hat{w}_i\right] \mathbb{E}\left[\hat{w}_j\right]$. Let $\tau_{i,j}$ denote the $(k-1)^{\text{st}}$ smallest value of $\frac{u_r}{w_r}$ for $r \in \{1, \ldots, n\} \setminus \{i, j\}$. If either i or j is absent from the S, then either $\hat{w}_i$ or $\hat{w}_j$ is 0. So we have:

$$\mathbb{E}[\hat{w}_i \hat{w}_j \mid \tau_{i,j}] = \frac{w_i}{\min(1, \tau_{i,j} w_i)} \frac{w_j}{\min(1, \tau_{i,j} w_j)} \Pr[i, j \in S \mid \tau_{i,j}]$$

We can thus compute the overall expectation as:

$$\begin{aligned}\mathbb{E}[\hat{w}_i \hat{w}_j] = \mathbb{E}\left[\mathbb{E}\left[\hat{w}_i \hat{w}_j \mid \tau_{i,j}\right]\right] &= \mathbb{E}\left[\frac{w_i}{\min(1, \tau_{i,j} w_i)} \cdot \frac{w_j}{\min(1, \tau_{i,j} w_j)} \Pr[i, j \in \mathcal{S} \mid \tau_{i,j}\right] \\ &= \mathbb{E}\left[\frac{w_i}{\min(1, \tau_{i,j} w_i)} \cdot \frac{w_j}{\min(1, \tau_{i,j} w_j)} \min(1, \tau_{i,j} w_i) \cdot \min(1, \tau_{i,j} w_j)\right] = w_i w_j.\end{aligned}$$

Above we have used the fact that, for i and j to both be in s, it must be that both $\frac{u_i}{w_i}$ and $\frac{u_j}{w_j}$ are less than $\tau_{i,j}$, which happens with probability $\min(1, \tau_{i,j} w_i) \cdot \min(1, \tau_{i,j} w_j)$. □

References

[Alon et al., 2005] Alon, N., Duffield, N., Lund, C., and Thorup, M. (2005). Estimating arbitrary subset sums with few probes. In *Proceedings of the 24th Symposium on Principles of Database Systems (PODS)*, pages 317–325.

[Bar-Yossef et al., 2002] Bar-Yossef, Z., Jayram, T. S., Kumar, R., Sivakumar, D., and Trevisan, L. (2002). Counting distinct elements in a data stream. In *Proceedings of the 6th International Workshop on Randomization and Computation (RANDOM)*.

[Beyer et al., 2007] Beyer, K., Haas, P. J., Reinwald, B., Sismanis, Y., and Gemulla, R. (2007). On synopses for distinct-value estimation under multiset operations. In *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, pages 199–210.

[Cohen, 2015] Cohen, E. (2015). Stream sampling for frequency cap statistics. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 159–168.

[Cohen and Kaplan, 2013] Cohen, E. and Kaplan, H. (2013). What you can do with coordinated samples. In *Proceedings of the 16th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems (APPROX)*, pages 452–467.

[Duffield et al., 2004] Duffield, N., Lund, C., and Thorup, M. (2004). Flow sampling under hard resource constraints. In *Proceedings of the Joint International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS 2004)*, pages 85–96.

[Duffield et al., 2005] Duffield, N., Lund, C., and Thorup, M. (2005). Learn more, sample less: control of volume and variance in network measurement. *IEEE Transactions on Information Theory*, 51(5):1756–1775.

[Duffield et al., 2007] Duffield, N., Lund, C., and Thorup, M. (2007). Priority sampling for estimation of arbitrary subset sums. *Journal of the ACM*, 54(6).

[Flajolet, 1990] Flajolet, P. (1990). On adaptive sampling. *Computing*, 43(4):391–400.

[Ohlsson, 1998] Ohlsson, E. (1998). Sequential poisson sampling. *Journal of Official Statistics*, 14(2):149.

[Szegedy, 2006] Szegedy, M. (2006). The DLT priority sampling is essentially optimal. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, pages 150–158.

[Szegedy and Thorup, 2007] Szegedy, M. and Thorup, M. (2007). On the variance of subset sum estimation. In *Proceedings of the European Symposium on Algorithms (ESA)*, pages 75–86. Springer Berlin Heidelberg.

[Tillé, 2023] Tillé, Y. (2023). Remarks on some misconceptions about unequal probability sampling without replacement. *Computer Science Review*, 47.