Estimating Multiple Concurrent Processes

Jayadev Acharya, ECE, UCSD, jayadev@ucsd.edu
Hirakendu Das, ECE, UCSD, hdas@ucsd.edu
Ashkan Jafarpour, ECE, UCSD, ajafarpo@ucsd.edu
Alon Orlitsky, ECE & CSE, UCSD, alon@ucsd.edu
Shengjun Pan, CSE, UCSD, s1pan@ucsd.edu

***Abstract*—We consider two related problems of estimating properties of a collection of point processes: estimating the multiset of parameters of continuous-time Poisson processes based on their activities over a period of time t, and estimating the multiset of activity probabilities of discrete-time Bernoulli processes based on their activities over n time instants.**

For both problems, it is sufficient to consider the observations' profile—the multiset of activity counts, regardless of their process identities. We consider the *profile maximum likelihood (PML)* estimator that finds the parameter multiset maximizing the profile's likelihood, and establish some of its competitive performance guarantees. For Poisson processes, if any estimator approximates the parameter multiset to within distance ϵ with error probability δ, then PML approximates the multiset to within distance 2ϵ with error probability at most $\delta \cdot e^{4\sqrt{t \cdot S}}$, where S is the sum of the Poisson parameters, and the same result holds for Bernoulli processes.

In particular, for the L_1 distance metric, we relate the problems to the long-studied distribution-estimation problem and apply recent results to show that the PML estimator has error probability $e^{-(t \cdot S)^{0.9}}$ for Poisson processes whenever the number of processes is $k = \mathcal{O}(tS\log(tS))$, and show a similar result for Bernoulli processes. We also show experimental results where the EM algorithm is used to compute the PML.

I. Introduction

Service providers like websites or phone companies often want to estimate the usage patterns of their subscribers as a whole. For example, they may want to know how many users are active on an average at any point of time, or the number of heavy users or other *aggregate* usage statistics, based on observations over a period of time. We consider two natural and arguably the simplest models for this problem. In the Poisson model, the activities of different users are independent Poisson processes and in the Bernoulli model, they are independent Bernoulli processes. We want to estimate the multiset of Poisson parameters or success probabilities of the processes in the respective models, given samples from each of them.

A. Notation and problem definition

The Poisson multiset estimation problem is mathematically defined as follows. Let $\Lambda \overset{\text{def}}{=} (\lambda_1, \lambda_2, \ldots, \lambda_k)$ be the parameters of k Poisson processes. Since we are interested in estimating only the multiset of Poisson parameters, without loss of generality we assume that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k \geq 0$. Each of these processes is observed for time t and thus, the sample or multiplicity $\mu(i)$ produced by process i is distributed according to $\text{poi}(\lambda_i \cdot t)$ for $i \in [k]$. Here and throughout this paper, $\text{poi}(\lambda)$ denotes the Poisson distribution with mean λ and $[k] \overset{\text{def}}{=} \{1, 2, \ldots, k\}$ for any positive integer k. We observe

$$\begin{aligned}\overline{\mu} &\overset{\text{def}}{=} (\mu_1, \ldots, \mu_m)\\ &\overset{\text{def}}{=} \big\{\mu(i) : \mu(i) > 0, i \in [k]\big\},\end{aligned}$$

where $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_m > 0$, namely, the multiset of positive multiplicities, or number of times each process that was active, fired till time t. Note that we observe only the $m \leq k$ processes that were active at least once during time t, we are not given any information about k or the $k - m$ that were not active. An estimator Q takes $\overline{\mu}$ as input and outputs a multiset of nonnegative reals $Q_{\overline{\mu}} \overset{\text{def}}{=} (q_1, q_2, \ldots, q_{k'})$ as an estimate of the unknown Λ. To measure the quality of estimation, one may use a suitable distance measure $D(\Lambda, Q)$. One such distance measure is the (sorted) L_1 distance $|\Lambda - Q| \overset{\text{def}}{=} \sum_{i=1}^{\infty} |\lambda_i - q_i|$, where $\lambda_i = 0$ for $i > k$ and $q_i = 0$ for $i > k'$. The following is an example of Poisson multiset estimation.

Example 1. For $\Lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (3.5, 3, 1.2, 0.1)$ and $t = 2$, we have $\mu(1) \sim \text{poi}(7)$, $\mu(2) \sim \text{poi}(6)$, $\mu(3) \sim \text{poi}(2.4)$ and $\mu(4) \sim \text{poi}(0.2)$. Let the samples produced be $\mu(1) = 6, \mu(2) = 8, \mu(3) = 3, \mu(4) = 0$, namely, $\overline{\mu} = (8, 6, 3)$. If $Z \sim \text{poi}(\lambda)$, then the maximum likelihood estimate of λ is Z. Hence, $Q_{\overline{\mu}} = \overline{\mu}/t = (4, 3, 1.5)$ is a reasonable estimate of Λ and $|\Lambda - Q_{\overline{\mu}}| = 0.5 + 0 + 0.3 + 0.1 = 0.9$. □

The Bernoulli multiset estimation problem is defined similarly. Let $B \overset{\text{def}}{=} (\theta_1, \theta_2, \ldots, \theta_k)$, be the success probabilities of k Bernoulli 0-1 distributions, where $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_k \geq 0$. For each $i \in [k]$, let $\overline{X}(i) \overset{\text{def}}{=} X(i,1), X(i,2), \ldots, X(i,n)$ be n samples drawn independently according to Bernoulli(θ_i). The samples $X(i,j)$ take values 1 or 0 depending on whether user i is active or inactive at time instant $j \in [n]$. Let

$$\begin{aligned}\overline{\overline{X}} &\overset{\text{def}}{=} \overline{X}_1, \overline{X}_2, \ldots, \overline{X}_m\\ &\overset{\text{def}}{=} \big\{\overline{X}(i) : \sum_{j=1}^{n} X(i,j) > 0, i \in [k]\big\},\end{aligned}$$

where $m \leq k$, be the multiset of observed sequences, *i.e.,* that have at least one activity. Let μ_i be the number of ones in $\overline{X}_i$ for $i \in [m]$ and let $\overline{\mu} = (\mu_1, \mu_2, \ldots, \mu_m)$ be the multiset of counts of ones in the sequences in $\overline{\overline{X}}$. Without loss of generality, $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_m > 0$. It is sufficient to consider only estimators that depend on $\overline{\overline{X}}$ only through $\overline{\mu}$. This is because we are interested in estimating only the multiset B as a whole and not the success probabilities of specific processes, and furthermore, the sequences are generated *i.i.d.*. See [3, Section 3.1.3] for a simple proof of this fact. Hence, a Bernoulli multiset estimator Q takes $\overline{\overline{X}}$ and outputs $Q_{\overline{\overline{X}}} \overset{\text{def}}{=} Q_{\overline{\mu}} \overset{\text{def}}{=} (q_1, q_2, \ldots, q_{k'})$ as an estimate of B. As earlier, any suitable distance $D(B, Q)$, *e.g.,* the L_1 distance $|B - Q| \overset{\text{def}}{=} \sum_{i=1}^{\infty} |\theta_i - q_i|$, can be used to measure the quality of estimation. The following example illustrates Bernoulli multiset estimation.

U.S. Government work not protected by U.S. copyright

Example 2. Let $B = (\theta(1), \ldots, \theta(5)) = (1, \frac{1}{2}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$. Let $n = 3$ and the sample sequences obtained be $\overline{X}(1) = (1,1,1)$, $\overline{X}(2) = (1,0,1)$, $\overline{X}(3) = (0,0,0)$, $\overline{X}(4) = (1,0,0)$, $\overline{X}(5) = (0,0,0)$. We only observe the sequences $\overline{\overline{X}} = \overline{X}(1), \overline{X}(2), \overline{X}(4)$. The empirical estimator outputs $Q = \overline{\mu}/n = (1, \frac{2}{3}, \frac{1}{3})$. If in addition, we are given that $k = 5$ and that each of the $\theta(i)$ have a uniform prior over $[0,1]$, then one obtains the *Laplace* or *add-one* estimate for each of the processes as $(\frac{3+1}{3+2}, \frac{2+1}{3+2}, \frac{1+1}{3+2}, \frac{0+1}{3+2}, \frac{0+1}{3+2})$ and outputs $Q' = (\frac{4}{5}, \frac{3}{5}, \frac{2}{5}, \frac{1}{5}, \frac{1}{5})$. The multisets $Q'' = (1, \frac{1}{2})$, $Q''' = (1, 1, \frac{1}{3})$. are also allowed estimates, although it is clear that they cannot generate the given sample sequences. Yet, $|B - Q''| = \frac{3}{8} \leq |B - Q| = \frac{5}{8}$. □

B. Empirical estimators

For both Poisson and Bernoulli multiset estimation, it is natural to consider the empirical estimator Q^{emp} that outputs $Q^{\text{emp,p}}_{\overline{\mu}} \stackrel{\text{def}}{=} (\frac{\mu_1}{t}, \ldots, \frac{\mu_m}{t})$ and $Q^{\text{emp,b}}_{\overline{\mu}} \stackrel{\text{def}}{=} (\frac{\mu_1}{n}, \ldots, \frac{\mu_m}{n})$ for the respective problems. The probability that a collection of Poisson processes Λ produce samples $(\mu(1), \ldots, \mu(k))$ is

$$\Lambda(\mu(1), \ldots, \mu(k)) = \prod_{i=1}^{k} \frac{(\lambda_i t)^{\mu(i)} e^{-\lambda_i t}}{\mu(i)!}.$$

Hence $Q^{\text{emp,p}}_{\overline{\mu}} = \max_\Lambda \Lambda(\mu(1), \ldots, \mu(k))$, *i.e.,* the empirical estimator maximizes the likelihood of observing $(\mu(1), \ldots, \mu(k))$ such that $\mu(i) = \mu_i$ for $i \in [m]$ and $\mu(i) = 0$ for $i > m$.

Similarly, the probability that a collection of Bernoulli processes B produce sequences $(\overline{x}(1), \ldots, \overline{x}(k))$, with $(\mu(1), \ldots, \mu(k))$ as the respective counts of ones is

$$B(\overline{x}(1), \ldots, \overline{x}(k)) = \prod_{i=1}^{k} \theta_i^{\mu(i)} (1 - \theta_i)^{n - \mu(i)}.$$

Therefore $Q^{\text{emp,b}}_{\overline{\mu}} = \max_B B(\overline{x}(1), \ldots, \overline{x}(k))$ is the maximum likelihood estimator of $(\overline{x}(1), \ldots, \overline{x}(k))$ such that $\mu(i) = \mu_i$ for $i \in [m]$ and $\mu(i) = 0$ for $i > m$.

For any Λ, let $S_\Lambda \stackrel{\text{def}}{=} \sum_{i=1}^{k} \lambda_i$ and for any B, let $S_B \stackrel{\text{def}}{=} \sum_{i=1}^{k} \theta_i$. It can be shown that $Q^{\text{emp,p}}$ is a good estimate of Λ in terms of L_1 distance when k is small compared to $t \cdot S_\Lambda$ since $\mu(i)$ concentrates around $t \cdot \lambda(i)$ for $i \in [k]$ using Poisson tail bounds of Fact 4. Similarly, $Q^{\text{emp,b}}$ is a good estimate of B in terms of L_1 distance when k is small compared to $n \cdot S_B$ since $\mu(i)$ concentrates around $n \cdot \theta(i)$ for $i \in [k]$ using Chernoff bounds of Fact 7. However, it can also be shown similar to the large alphabet examples in [11], [6] that the empirical estimators may not be good estimates of Λ or B when the number of processes is large and $k = \Omega(tS_\Lambda)$ or $k = \Omega(nS_B)$ respectively. In this paper, we show estimators that have good estimation guarantees even for large alphabets.

C. Profile maximum likelihood (PML) estimators

The empirical Poisson multiset estimator considers the likelihood of observing a specific $(\mu(1), \ldots, \mu(k))$ such that the multiset of nonzero multiplicities is $\overline{\mu}$. However, for estimating the multiset Λ, it is natural to consider the overall likelihood of $\overline{\mu}$ *i.e.,* the probability of observing any $(\mu(1), \ldots, \mu(k))$ whose multiplicity multiset is $\overline{\mu}$. The information in $\overline{\mu}$ is equivalently conveyed by the profile $\overline{\varphi} \stackrel{\text{def}}{=} \varphi(\overline{\mu}) \stackrel{\text{def}}{=} (\varphi_1, \varphi_2, \ldots)$ where $\varphi_\mu \stackrel{\text{def}}{=} |\{\mu_i : \mu_i = \mu, i \in [m]\}|$, called the *prevalence* of μ, is the number of multiplicities in $\overline{\mu}$ that are equal to μ. We henceforth use $\overline{\mu}$ and its profile $\overline{\varphi}$ with the same meaning. The likelihood of a $\overline{\mu}$ or its profile $\overline{\varphi}$ under a collection Λ is

$$\begin{aligned} \Lambda(\overline{\varphi}) \stackrel{\text{def}}{=} \Lambda(\overline{\mu}) &\stackrel{\text{def}}{=} \Pr\left((\mu(1), \ldots, \mu(k)) \in \mathcal{S}_{\overline{\varphi}}\right) \\ &= \sum_{(\mu(1), \ldots, \mu(k)) \in \mathcal{S}_{\overline{\varphi}}} \Lambda(\mu(1), \ldots, \mu(k)), \end{aligned}$$

where $\mathcal{S}_{\overline{\varphi}} \stackrel{\text{def}}{=} \mathcal{S}_{\overline{\mu}}$ is the collection of all $(\mu(1), \ldots, \mu(k))$ whose multiplicity multiset is $\overline{\mu}$. Similarly, in the Bernoulli multiset estimation problem, the probability of a $\overline{\mu}$ and its profile $\overline{\varphi}$ under a collection B is

$$\begin{aligned} B(\overline{\varphi}) \stackrel{\text{def}}{=} B(\overline{\mu}) &\stackrel{\text{def}}{=} \Pr\left((\overline{X}(1), \ldots, \overline{X}(k)) \in \mathcal{S}_{\overline{\varphi}}\right) \\ &= \sum_{(\overline{x}(1), \ldots, \overline{x}(k)) \in \mathcal{S}_{\overline{\varphi}}} B(\overline{x}(1), \ldots, \overline{x}(k)), \end{aligned}$$

where $\mathcal{S}_{\overline{\varphi}}$ is the collection of all $(\overline{x}(1), \ldots, \overline{x}(k))$ whose multiplicity multiset is $\overline{\mu}$.

For both problems we consider the profile maximum likelihood (PML) estimator that maximizes the likelihood of observing the profile of the given observations. For Poisson multiset estimation, the PML distribution is $Q^{\text{PML,p}}_{\overline{\varphi}} \stackrel{\text{def}}{=} \hat{\Lambda}_{\overline{\varphi}} \stackrel{\text{def}}{=} \arg\max_\Lambda \Lambda(\overline{\varphi})$ and the PML is $\hat{\Lambda}(\overline{\varphi}) \stackrel{\text{def}}{=} \max_\Lambda \Lambda(\overline{\varphi}) = \hat{\Lambda}_{\overline{\varphi}}(\overline{\varphi})$. When the maximization is limited to a class of Poisson multisets $\mathcal{L}$, we use the notations $Q^{\text{PML,p}}_{\mathcal{L},\overline{\varphi}}$ and $\hat{\Lambda}_{\mathcal{L},\overline{\varphi}}$. Likewise, for Bernoulli multiset estimation, we have $Q^{\text{PML,b}}_{\overline{\varphi}} \stackrel{\text{def}}{=} \hat{B}_{\overline{\varphi}} \stackrel{\text{def}}{=} \arg\max_B B(\overline{\varphi})$ and the PML is $\hat{B}(\overline{\varphi}) \stackrel{\text{def}}{=} \max_B B(\overline{\varphi}) = \hat{B}_{\overline{\varphi}}(\overline{\varphi})$. When the maximization is limited to a class of Bernoulli multisets $\mathcal{B}$, we use the notations $Q^{\text{PML,b}}_{\mathcal{B},\overline{\varphi}}$ and $\hat{B}_{\mathcal{B},\overline{\varphi}}$.

In general, Q^{PML} is different from Q^{emp}. The PML technique, also known as *pattern maximum likelihood*, has been used earlier in [10], [11], [9] for estimating the probability multiset of distributions in the context of universal compression of large alphabet data sources. As we observe later, computation of PML Poisson multiset is almost identical to computing PML distribution, but PML Bernoulli multiset has a much different structure. We also develop useful connections between distribution estimation and Poisson and Bernoulli multiset estimation. For other recent works that efficiently exploit profiles for problems related to distribution multiset estimation and property testing, see [13] and references therein.

D. Summary of main results

We show competitive estimation guarantees for the PML estimators that can be described informally as follows. In the Poisson model, if there is an estimator Q that estimates any Λ to within ϵ in some distance D with high probability δ, *i.e.,* $\Pr(D(\Lambda, Q_{\overline{\varphi}}) \geq \epsilon) \leq \delta$, then the PML estimator provides a similar guarantee that $\Pr\left(D(\Lambda, \hat{\Lambda}_{\overline{\varphi}}) \geq 2\epsilon\right) \leq \delta \cdot e^{4\sqrt{t \cdot S_\Lambda}} + e^{-t \cdot S_\Lambda / 3}$. In the Bernoulli model, if an estimator Q is a good estimate of any B in some distance measure D such that $\Pr(D(B, Q_{\overline{\varphi}}) \geq \epsilon) \leq \delta$, then the PML estimator is such that $\Pr\left(D(B, \hat{B}_{\overline{\varphi}}) \geq 2\epsilon\right) \leq \delta \cdot e^{4\sqrt{n \cdot S_B}} + e^{-n \cdot S_B / 3}$. The results are shown using a similar competitive property of maximum likelihood estimators in general. Similar arguments have been used previously in [9] to show consistency properties of PML for distribution estimation.

For these results to be useful, we need to show estimators whose error probability δ is smaller than $e^{-4\sqrt{t \cdot S_\Lambda}}$ in the

Poisson model or $e^{-4\sqrt{n\cdot S_B}}$ in the Bernoulli model. We show the existence of such estimators when L_1 is used as distance by relating these problems to that of distribution multiset estimation. It additionally shows that distribution multiset estimators can be easily adapted for Poisson and Bernoulli multiset estimation. In the distribution estimation problem, one wants to estimate the probability multiset $\{p_1, p_2, \ldots, p_k\}$ of an unknown *distribution* P, *i.e.,* $\sum_{i=1}^{k} p_i = 1$, given ℓ samples generated *i.i.d.* according to P. This problem has been studied for a long time, *e.g.,* see [11], [9], [5], [14], [13] and references therein for an overview of past and current developments. In particular, estimators are shown in [13] that can approximate distributions to within an L_1 distance of ϵ with high probability, whenever their support size is $k = \mathcal{O}(\epsilon^{2.1}\ell\log(\ell))$. We show that in the Poisson model, estimating a Λ to within an L_1 distance of ϵS_Λ is equivalent to estimating the distribution $P = \frac{\Lambda}{S_\Lambda}$ to within an L_1 distance of ϵ using $\ell = \mathcal{O}(t \cdot S_\Lambda)$ samples. Furthermore, an error probability of $\delta(\ell)$ for distribution estimation translates to $\delta(t \cdot S_\Lambda)$ in the Poisson model. Likewise, for a Bernoulli multiset B, a distribution estimator $\widetilde{Q}$ such that $\Pr(|\frac{B}{S_B} - \widetilde{Q}_{\overline{\varphi}}| \le \epsilon) \le \delta(\ell)$ implies a Bernoulli multiset estimator Q such that $\Pr(|B - Q_{\overline{\varphi}}| \le \epsilon) \le \delta(n \cdot S_B)$. where $\ell = \mathcal{O}(n \cdot S_B)$. We use the results for distribution estimation in [13] along with the competitive estimation guarantees for the PML estimator to show that for any Λ such that $k = \mathcal{O}(\epsilon^{2.1} t S_\Lambda \log(tS_\Lambda))$, $\Pr(|\Lambda - Q^{\mathrm{PML}}_{\overline{\varphi}}| \ge \epsilon) \le e^{-(tS_\Lambda)^{0.9}}$ and for any B such that $k = \mathcal{O}(\epsilon^{2.1} n S_B \log(nS_B))$, $\Pr(|B - Q^{\mathrm{PML}}_{\overline{\varphi}}| \ge \epsilon) \le e^{-(nS_B)^{0.9}}$.

Computational aspects of finding the PML multiset by leveraging existing techniques for computing PML distribution are discussed towards the end of the paper.

II. Competitive estimation guarantees for the PML estimator

We first show a general competitive estimation guarantee for maximum likelihood (ML) estimators. Let $\mathcal{Z}$ be a discrete alphabet of size $|\mathcal{Z}|$ and $\mathcal{P}$ be a collection of probability distributions on $\mathcal{Z}$. Given a sample Z generated according to an unknown distribution $P \in \mathcal{P}$, we want to estimate P. An estimator $Q : \mathcal{Z} \to \mathcal{P}$, outputs a distribution $Q_z \in \mathcal{P}$ when given input $z \in \mathcal{Z}$. The ML estimator outputs a distribution $\hat{P}_z \stackrel{\mathrm{def}}{=} \arg\max_{P\in\mathcal{P}} P(z)$. Let $D(\cdot,\cdot)$ be a distance measure defined on distributions in $\mathcal{P}$. The next lemma shows that $\hat{P}_Z$ is almost as good as any other estimator.

Lemma 3. *For some $\epsilon \ge 0$ and $\delta \in [0,1]$, let Q be an estimator such that for all $P \in \mathcal{P}$, when $Z \sim P$, $\Pr\big(D(P, Q_Z) \ge \epsilon\big) \le \delta$. Then, $\Pr\big(D(P, \hat{P}_Z) \ge 2\epsilon\big) \le \delta \cdot |\mathcal{Z}|$.*

Proof Sketch. If $Z = z$ is such that $P(z) > \delta$, then $D(P, \hat{P}_z) \le 2\epsilon$. To see this, note that $D(P, Q_z) \le \epsilon$, otherwise if $D(P, Q_z) \ge \epsilon$, then

$$\begin{aligned}\Pr\big(D(P,Q_Z) \ge \epsilon\big) &= \sum_{z' \in \mathcal{Z}: D(P,Q_{z'}) \ge 2\epsilon} P(z') \\ &\ge P(z) > \delta,\end{aligned}$$

contradicting that Q has error probability at most δ. By a similar reasoning, $D(\hat{P}_z, Q_z) \le \epsilon$, since Q is a good estimator of $\hat{P}_z \in \mathcal{P}$ as well and $\hat{P}_z(z) \ge P(z) > \delta$. Hence, $D(P, \hat{P}_z) \le D(P, Q_z) + D(Q_z, \hat{P}_z) \le 2\epsilon$ and

$$\begin{aligned}&\Pr\big(D(P,\hat{P}_Z) \ge 2\epsilon\big) \\ &\quad= \Pr\big((D(P,\hat{P}_Z) \ge 2\epsilon) \wedge (P(Z) > \delta)\big) \\ &\qquad\quad + \Pr\big((D(P,\hat{P}_Z) \ge 2\epsilon) \wedge (P(Z) \le \delta)\big) \\ &\quad\le 0 + \Pr(P(Z) \le \delta) \le \delta \cdot |\mathcal{Z}|. \qquad \square\end{aligned}$$

For showing such an estimation guarantee for PML in the Poisson model, we use two more facts. In the multiset estimation problem, $\mathcal{Z}$ is the set of all profiles or all $\overline{\mu}$, which is infinite. However, we show that the profiles generated by any Λ concentrate over a much smaller subset. For any $\overline{\varphi}$ and its corresponding $\overline{\mu}$ let $S_{\overline{\varphi}} = S_{\overline{\mu}} = \sum_{i=1}^{m} \mu_i$ be the sum of multiplicities. If $\overline{\varphi} \sim \Lambda$, then $S_{\overline{\varphi}} = \sum_{i=1}^{k} \mu(i)$ is distributed according to $\mathrm{poi}(tS_\Lambda)$ by the well known property of sum of Poisson random variables. Hence, $S_{\overline{\varphi}}$ concentrates around S_Λ by the Poisson tail bounds given below.

Fact 4. *(Also [12, Corollary 32].) For all $\epsilon \in (0,1)$ and sufficiently large λ, if $X \sim \mathrm{poi}(\lambda)$, then $\Pr\big(|X - \lambda| \ge \epsilon\lambda\big) \le 2\exp(-\epsilon^2\lambda/3)$. For $\alpha \ge 2$, $\Pr\big(X \ge \alpha\lambda\big) \le \exp(-\alpha\lambda/6)$.* □

Secondly, we use the fact that the set $\Phi^S \stackrel{\mathrm{def}}{=} \{\overline{\varphi} : S_{\overline{\varphi}} = \sum_i \mu_i = S\}$ is in 1-1 correspondence with the integer partitions of S [10] and can therefore be bounded by this well known fact about *partition number* [4].

Fact 5. *For all S, $|\Phi^S| = p(S) \le e^{\pi\sqrt{\frac{2}{3}}\sqrt{S}} < e^{3\sqrt{S}}$.* □

Using the general lemma and these facts, we have the following results.

Lemma 6. *Let $\mathcal{L}$ be a class of Poisson multisets Λ such that $S_\Lambda \ge 2$ and $D(\cdot,\cdot)$ be a distance measure on $\mathcal{L}$. Suppose an estimator Q is such that for some $\epsilon, \delta > 0$, when $\overline{\varphi} \sim \Lambda \in \mathcal{L}$, $\Pr\big(D(\Lambda, Q_{\overline{\varphi}}) \ge \epsilon\big) \le \delta$. Then, $\Pr\big(D(\Lambda, \hat{\Lambda}_{\mathcal{L},\overline{\varphi}}) \ge 2\epsilon\big) \le \delta e^{4\sqrt{tS_\Lambda}} + e^{-tS_\Lambda/3}$.*

Proof Sketch. If $\overline{\varphi} \sim \Lambda \in \mathcal{L}$, then

$$\begin{aligned}\Pr\big(D(\Lambda, \hat{\Lambda}_{\mathcal{L},\overline{\varphi}}) \ge 2\epsilon\big) &\le \Pr\big(S_{\overline{\varphi}} > 2tS_\Lambda\big) \\ &\quad + \Pr\Big(\big(D(\Lambda, \hat{\Lambda}_{\mathcal{L},\overline{\varphi}}) \ge 2\epsilon\big) \wedge \big(S_{\overline{\varphi}} \le 2tS_\Lambda\big)\Big) \\ &\le e^{-tS_\Lambda/3} + \delta e^{4\sqrt{tS_\Lambda}}.\end{aligned}$$

In the last inequality, the bound on the first term follows from $S_{\overline{\varphi}} \sim \mathrm{poi}(tS_\Lambda)$ and Fact 4. For the second term, we use Lemma 3, along with Fact 5 which implies that $|\mathcal{Z}| = |\{\overline{\varphi} : S_{\overline{\varphi}} \le 2S_\Lambda\}| \le 2tS_\Lambda \cdot |\Phi^{2tS_\Lambda}| \le e^{4\sqrt{tS_\Lambda}}$. □

A similar estimation guarantee can be shown for PML Bernoulli multiset estimator. This time, we use the fact that if $\overline{\varphi} \sim B$, then $S_{\overline{\varphi}} = \sum_{i=1}^{k}\sum_{j=1}^{n} X(i,j)$ is a sum of independent 0-1 random variables and concentrates around its mean nS_B using the Chernoff bounds below.

Fact 7. *(Chernoff bounds.) Let $X = \sum_{i=1}^{n} Y_i$ be a sum of independent 0-1 random variables $Y_1, \ldots, Y_n$ such that $\Pr(Y_i = 1) = p_i$. Let $\mu = E[X] = \sum_i p_i$. For $\epsilon \in [0,1]$, $\Pr(|X - \mu| \ge \epsilon\mu) \le 2e^{-\mu\epsilon^2/3}$. For $\epsilon \ge 1$, $\Pr(X \ge (1+\epsilon)\mu) \le e^{-\mu\epsilon/3}$.* □

Lemma 8. *Let $\mathcal{B}$ be a class of Bernoulli multisets B and $D(\cdot,\cdot)$ be a distance measure on $\mathcal{B}$. For large n,* *let Q be an estimator such that for some $\epsilon, \delta > 0$ when $\overline{\overline{X}} \sim B \in \mathcal{B}$, $\Pr\left(D(B, Q_{\varphi(\overline{\overline{X}})}) \geq \epsilon\right) \leq \delta$. Then,* $\Pr\left(D(B, \hat{B}_{\mathcal{B},\varphi(\overline{\overline{X}})}) \geq 2\epsilon\right) \leq \delta e^{4\sqrt{nS_B}} + e^{-nS_B/3}$. □

To make use of these results, we show Poisson and Bernoulli multiset estimators whose error probability is less than $e^{-5\sqrt{tS_\Lambda}}$ and $e^{-5\sqrt{nS_B}}$ respectively by relating these problems to distribution multiset estimation.

III. Relationship between Poisson and Bernoulli multiset estimation and distribution estimation

A distribution multiset estimator $\widetilde{Q}$ takes as input a sequence $\overline{Y} = Y_1, \ldots, Y_\ell$ of ℓ samples drawn *i.i.d.* according to an unknown distribution $P = (p_1, \ldots, p_k)$ and outputs $\widetilde{Q}_{\overline{Y}} = (q_1, \ldots, q_{k'})$ as an estimate of P. We assume the probabilities in P and Q are arranged in decreasing order. We use $\mu(i) \stackrel{\text{def}}{=} \mu_{\overline{Y}}(i)$ to denote the number of appearances of symbol i (whose probability is p_i) in $\overline{Y}$. As earlier, $\overline{\mu} \stackrel{\text{def}}{=} (\mu_1, \ldots, \mu_m) \stackrel{\text{def}}{=} \{\mu(i); \mu(i) > 0, i \in [k]\}$ is the multiset of nonzero multiplicities and $\overline{\varphi} \stackrel{\text{def}}{=} \varphi(\overline{\mu}) \stackrel{\text{def}}{=} \varphi(\overline{Y})$ denotes the corresponding profile. Without loss of generality, we assume $\widetilde{Q}$ depends on $\overline{Y}$ only through its profile, *i.e.*, $\widetilde{Q}_{\overline{Y}} = \widetilde{Q}_{\overline{\varphi}}$.

To relate the various estimation problems, it is useful to consider the well known useful technique of *Poissonization* which is summarized below.

Fact 9. *Let $\overline{Y}'$ be a sequence of $\ell' \sim \mathrm{poi}(\ell)$ samples drawn* i.i.d. $\sim P$*. Then, for all $i \in [k]$, $\mu_{\overline{Y}'}(i) \sim \mathrm{poi}(\ell p_i)$ and is independent of $\mu(i')$ for all other $i' \in [k]$.* □

In the next definition and lemma, we show that good distribution multiset estimators can be used to construct good Poisson multiset estimators, both under L_1 distance guarantees.

Definition 10. Let $\mathcal{L}$ be a class of Poisson multisets and let $\mathcal{P} \stackrel{\text{def}}{=} \{\frac{\Lambda}{S_\Lambda} = (\frac{\lambda_1}{S_\Lambda}, \ldots, \frac{\lambda_k}{S_\Lambda}) : \Lambda \in \mathcal{L}\}$ be the corresponding class of normalized distributions. Let $\widetilde{Q}$ be a distribution multiset estimator for $\mathcal{P}$. Then, the corresponding Poisson multiset estimator Q^{poi} outputs $Q^{\text{poi}}_{\overline{\varphi}} \stackrel{\text{def}}{=} \frac{S_{\overline{\varphi}}}{t} \cdot \widetilde{Q}_{\overline{\varphi}}$. □

Lemma 11. *For $\epsilon \in (0, 1)$, let $\mathcal{L}$ be a class of Poisson multisets such that S_Λ is sufficiently large for all $\Lambda \in \mathcal{L}$. Let $\mathcal{P} \stackrel{\text{def}}{=} \{\Lambda/S_\Lambda : \Lambda \in \mathcal{L}\}$. Let $\widetilde{Q}$ be a distribution estimator such that when $\ell \geq \min_{\Lambda \in \mathcal{L}} \frac{S_\Lambda}{2}$ and given $\overline{Y} \sim P^\ell$, $\Pr\left(|P - \widetilde{Q}_{\varphi(\overline{Y})}| > \epsilon\right) \leq \delta(\ell)$, where δ decreases monotonically in ℓ. Then the estimator Q^{poi} corresponding to $\widetilde{Q}$ is such that when $\overline{\varphi} \sim \Lambda \in \mathcal{L}$,*

$$\Pr\left(|\Lambda - Q^{\text{poi}}_{\overline{\varphi}}| > 2\epsilon S_\Lambda\right) \leq \delta\big(\frac{tS_\Lambda}{2}\big) + e^{-tS_\Lambda/12} + 2e^{-\epsilon^2 tS_\Lambda/3}.$$

Proof Sketch. Let $\overline{\mu} \sim \Lambda \in \mathcal{L}$. Then,

$$\begin{aligned}\Pr\big(|\frac{\Lambda}{S_\Lambda} - \widetilde{Q}_{\overline{\varphi}}| \leq \epsilon\big) &\leq \Pr\big(S_{\overline{\varphi}} \leq \frac{tS_\Lambda}{2}\big) \\ &\quad + \Pr\Big(\big(|\frac{\Lambda}{S_\Lambda} - \widetilde{Q}_{\overline{\varphi}}| \leq \epsilon\big) \wedge \big(S_{\overline{\varphi}} \geq \frac{tS_\Lambda}{2}\big)\Big) \\ &\leq e^{-tS_\Lambda/12} + \delta\big(\frac{tS_\Lambda}{2}\big).\end{aligned}$$

Here, the first term is due to the Poisson tail bounds of Fact 4. For the second term, we use Fact 9, which implies that $\varphi(\overline{\mu})$ has the same distribution as $\varphi(\overline{Y}')$, where $\overline{Y}'$ is an *i.i.d.* sequence of length $S_{\overline{\mu}} \sim \mathrm{poi}(tS_\Lambda)$ drawn according to the distribution $\frac{\Lambda}{S_\Lambda}$. Since the number of samples in the input $\overline{Y}'$ to $\widetilde{Q}$ is $S_{\overline{\mu}} \geq \frac{S_\Lambda}{2}$ and δ is monotonically decreasing, the error probability is at most $\delta\big(\frac{tS_\Lambda}{2}\big)$.

Again by Fact 4, since $S_{\overline{\varphi}} \sim \mathrm{poi}(tS_\Lambda)$,

$$\Pr\left(|S_{\overline{\varphi}} - tS_\Lambda| \geq \epsilon tS_\Lambda\right) \leq 2e^{-\epsilon^2 tS_\Lambda/3}.$$

Lastly, if $|\frac{\Lambda}{S_\Lambda} - \widetilde{Q}_{\overline{\varphi}}| \leq \epsilon$ and $|S_{\overline{\varphi}} - tS_\Lambda| \leq \epsilon tS_\Lambda$, then

$$\begin{aligned}|\Lambda - Q^{\text{poi}}_{\overline{\varphi}}| &= \big|\Lambda - \frac{S_{\overline{\varphi}}}{t} \cdot \widetilde{Q}_{\overline{\varphi}}\big| \\ &= \frac{1}{t}\big|tS_\Lambda \cdot \frac{\Lambda}{S_\Lambda} - tS_\Lambda \cdot \widetilde{Q}_{\overline{\varphi}} + tS_\Lambda \cdot \widetilde{Q}_{\overline{\varphi}} - S_{\overline{\varphi}} \cdot \widetilde{Q}_{\overline{\varphi}}\big| \\ &\leq S_\Lambda\big|\frac{\Lambda}{S_\Lambda} - \widetilde{Q}_{\overline{\varphi}}\big| + \frac{1}{t}|tS_\Lambda - S_{\overline{\varphi}}| \leq 2\epsilon S_\Lambda.\end{aligned}$$

Combining the above observations, by union bound,

$$\begin{aligned}\Pr\left(|\Lambda - Q^{\text{poi}}_{\overline{\varphi}}| > 2\epsilon S_\Lambda\right) &\leq \Pr\big(|\frac{\Lambda}{S_\Lambda} - \widetilde{Q}_{\overline{\varphi}}| \leq \epsilon\big) \\ &\quad + \Pr\left(|S_\Lambda - S_{\overline{\varphi}}| \geq \epsilon S_\Lambda\right) \\ &\leq \delta\big(tS_\Lambda/2\big) + e^{-tS_\Lambda/12} + 2e^{-\epsilon^2 tS_\Lambda/3}. \quad \square\end{aligned}$$

We use similar arguments for converting good distribution multiset estimators into good Bernoulli multiset estimators under L_1 distance guarantees.

Definition 12. Let $\mathcal{B}$ be a class of Bernoulli multisets and $\mathcal{P} \stackrel{\text{def}}{=} \{\frac{B}{S_B} = (\frac{\theta_1}{S_B}, \ldots, \frac{\theta_k}{S_B}) : B \in \mathcal{B}\}$ be the corresponding class of normalized distributions. Let $\widetilde{Q}$ be a distribution multiset estimator for $\mathcal{P}$. We define as follows a corresponding Bernoulli multiset estimator Q^{bern} that takes input $\overline{\overline{X}} \sim B \in \mathcal{B}$. For $i = 1, \ldots, m$, generate independent $n_i \sim \mathrm{poi}(n/2)$. If some $n_i > n$, terminate the estimation process and output error. Otherwise, for each of $i = 1, \ldots, m$, let $\overline{Y}_i$ consist of first n_i samples of $\overline{X}_i$, *i.e.*, $\overline{Y}_i = X_{i,1}, X_{i,2}, \ldots, X_{i,n_i}$.

Let $\mu'_i \stackrel{\text{def}}{=} \mu(\overline{Y}_i)$ be the number of 1's in $\overline{Y}_i$ for $i \in [m]$. And let $\overline{\varphi}' = (\varphi'_1, \varphi'_2, \ldots)$ be the profile corresponding to $\overline{\mu}' \stackrel{\text{def}}{=} \{\mu'_i : \mu'_i > 0, i \in [m]\}$. The output of Q^{bern} is $Q^{\text{bern}}_{\overline{\varphi}} \stackrel{\text{def}}{=} \frac{S_{\overline{\varphi}}}{n} \cdot \widetilde{Q}_{\overline{\varphi}'}$. □

Lemma 13. *Let $\mathcal{B}$ be a class of distribution multisets and let $\mathcal{P} \stackrel{\text{def}}{=} \{B/S_B : B \in \mathcal{B}\}$. Let $\widetilde{Q}$ be a distribution estimator such that for large n and $\ell \geq n \cdot \min_{B \in \mathcal{B}} \frac{S_B}{4}$, when $\overline{Y} \sim P^\ell$, $\Pr\left(|P - \widetilde{Q}_{\varphi(\overline{Y})}| > \epsilon\right) \leq \delta(\ell)$, where δ decreases monotonically in ℓ. Then, the corresponding Q^{bern} is such that when $\overline{\overline{X}} \sim B \in \mathcal{B}$,*

$$\begin{aligned}\Pr\left(\big|B - Q^{\text{bern}}_{\varphi(\overline{\overline{X}})}\big| > 2\epsilon S_B\right) &\leq \delta\big(\frac{nS_B}{4}\big) + 2e^{-\epsilon^2 nS_B/3} \\ &\quad + e^{-nS_B/12} + ke^{-n/6}.\end{aligned}$$

Proof. Let $\overline{\overline{X}} \sim B \in \mathcal{B}$ and $\overline{\varphi} = \varphi(\overline{\overline{X}})$. We analyze the error probability in each of the intermediate steps of Definition 12. Using the Poisson tail bounds in Fact 4 and union bound, probability that $n_i > n$ for some $i \in \{1, \ldots, m\}$ is at most $me^{-n/6} \leq ke^{-n/6}$.

If all $n_i < n$, by Fact 9 on Poissonization, all $\mu'_i \sim \mathrm{poi}(n\theta_i/2) = \mathrm{poi}\big((nS_B/2) \cdot (\theta_i/S_B)\big)$. Again by Fact 9, $\overline{\varphi}'$ has the same distribution as the profile of a sequence $\overline{Y}'$ consisting of $n' \sim \mathrm{poi}(nS_B/2)$ samples drawn *i.i.d.* from the distribution $\frac{B}{S_B}$. Hence $\overline{\varphi}'$ has length $S_{\overline{\varphi}'} = \frac{nS_B}{4}$ with probability $\geq 1 - e^{-nS_B/12}$ by Poisson tail bounds. In that case, the estimation guarantee for $\widetilde{Q}$ implies $\Pr\left(|\frac{B}{S_B} - \widetilde{Q}_{\overline{\varphi}'}| \ge \epsilon\right) \le \delta(S_{\overline{\varphi}'}) \le \delta\big(\frac{nS_B}{4}\big)$.

Using Chernoff bounds in Fact 7 and that $S_{\overline{\varphi}} = \sum_{i=1}^{n}\sum_{j=1} X(i,j)$ is a sum of independent 0-1 random variables and has mean $\mathrm{E}[S_{\overline{\varphi}}] = nS_B$, we have $\Pr\left(|S_{\overline{\varphi}} - nS_B| \ge \epsilon nS_B\right) \le 2e^{-\epsilon^2 nS_B/3}$. Similar to the proof of Lemma 11, if $|\frac{B}{S_B} - \widetilde{Q}_{\overline{\varphi}'}| \le \epsilon$ and $|S_{\overline{\varphi}} - nS_B| \le \epsilon nS_B$, then

$$|B - Q^{\mathrm{bern}}_{\overline{\varphi}}| = \left|S_B \cdot \frac{B}{S_B} - \frac{S_{\overline{\varphi}}}{n} \cdot \widetilde{Q}_{\overline{\varphi}'}\right| \le 2\epsilon S_B.$$

Putting these observations together, and using union bound for bounding the overall error probability,

$$\begin{aligned}
&\Pr\left(|B - \widetilde{Q}_{\varphi(\overline{\overline{X}})}| > 2\epsilon S_B\right)\\
&\quad\le \Pr\left(n_i > n \text{ for some } i \in [m]\right)\\
&\qquad + \Pr\left(S_{\overline{\varphi}'} < \frac{nS_B}{4}\right)\\
&\qquad + \Pr\left(\left(|\frac{B}{S_B} - \widetilde{Q}_{\overline{\varphi}'}| \ge \epsilon\right) \wedge \left(S_{\overline{\varphi}'} \ge \frac{nS_B}{4}\right)\right)\\
&\qquad + \Pr\left(|S_{\overline{\varphi}} - nS_B| \ge \epsilon nS_B\right)\\
&\quad\le ke^{-n/6} + e^{-nS_B/12} + \delta\big(\frac{nS_B}{4}\big) + 2e^{-\epsilon^2 nS_B/3}. \qquad \square
\end{aligned}$$

We note that in both Lemma 11 and Lemma 13, an L_1 distance guarantee of $2\epsilon S_\Lambda$ and $2\epsilon S_B$ is reasonable since the maximum L_1 distance between any two multisets, each of whose sum of parameters is S, is at most $2S$. As an application of these lemmas, we state and use the main result in [12], [13] that shows an estimator which can approximate distributions to within a small *relative earthmover distance*, and hence small L_1 distance, even when the support size k is superlinear in the number of samples ℓ. While the error probability shown in [12] is $e^{-\ell^{0.03}}$, it can be improved to arbitrarily close to exponential, say $e^{-\ell^{0.9}}$, by minor modifications to the various constant parameters of their estimator.

Theorem 14. *(Also [12, Theorem 3].) For $\epsilon > 0$ and sufficiently large ℓ, there is an estimator $\widetilde{Q}$ such that for all P whose support size is $k = \mathcal{O}(\epsilon^{2.1}\ell\log(\ell))$, when $\overline{Y} \sim P^\ell$, $\Pr\left(|P - \widetilde{Q}_{\varphi(\overline{Y})}| > \epsilon\right) \le e^{-\ell^{0.9}}$.* $\square$

Corollary 15. *For $\epsilon > 0$, and for all t and Λ such that $t \cdot S_\Lambda$ is sufficiently large and $k = \mathcal{O}(\epsilon^{2.1}tS_\Lambda\log(tS_\Lambda))$, when $\overline{\varphi} \sim \Lambda$, $\Pr\left(|\Lambda - \hat{\Lambda}_{\overline{\varphi}}| \ge 2\epsilon S_\Lambda\right) \le e^{-(tS_\Lambda)^{0.8}}$.*

For all n and B such that $n \cdot S_B$ is sufficiently large and $k = \mathcal{O}(\epsilon^{2.1}nS_B\log(nS_B))$, when $\overline{\varphi} \sim B$, $\Pr\left(|B - \hat{\Lambda}_{\overline{\varphi}}| \ge 2\epsilon S_B\right) \le e^{-(nS_B)^{0.8}}$. In both cases, the maximum likelihood is calculated over multisets of the respective support size bounds. $\square$

IV. Calculation of PML and Experimental Results

We consider some of the computational aspects of PML in this brief final section. Due to space constraints, without going into details, we state that for any $\overline{\varphi}$, due to the similarity between the expressions for likelihoods $\Lambda(\overline{\varphi})$ and $P(\overline{\varphi})$, $\hat{\Lambda}_{\overline{\varphi}} = \frac{S_{\overline{\varphi}}}{t} \cdot \hat{P}_{\overline{\varphi}}$. Thus, computation of PML Poisson multiset is equivalent to computing the PML distribution multiset of a given profile. However, computation of PML Bernoulli multiset $\hat{B}_{\overline{\varphi}}$ is somewhat different from that of $\hat{P}_{\overline{\varphi}}$. Nonetheless, the computation of both $\hat{P}_{\overline{\varphi}}$ and $\hat{B}_{\overline{\varphi}}$ seem to be difficult in general, *e.g.,* see [11] and subsequent works [8], [2], [1] along these lines. We conclude with an example of an experimental result for Bernoulli multiset estimation, as shown in Figure 1. The underlying multiset B is taken to be $\theta_i = 0.05$ for $i \in \{1, 2, \ldots, 500\}$, *i.e.,* $k = 500$. The empirical estimate, referred to as SML or *sequence maximum likelihood* in the figure is clearly not a good estimate of B, both in terms of support size and shape. Note that we do not get to even observe 154 out of the 500 processes. However, the PML multiset is seen to be a very good estimate of B. It is computed approximately using an EM-MCMC algorithm similar to that used in [11], [7] for distribution estimation.

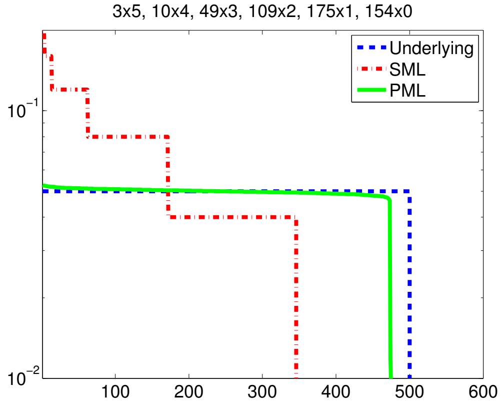

Fig. 1. Bernoulli multiset estimation using empirical (SML) and PML estimators

References

[1] J. Acharya, H. Das, A. Orlitsky, and S. Pan. Algebraic computation of pattern maximum likelihood. In *Proceedings of IEEE Symposium on Information Theory (ISIT)*, pages 400–404, 2011.

[2] J. Acharya, A. Orlitsky, and S. Pan. The maximum likelihood probability of unique-singleton, ternary, and length-7 patterns. In *Proceedings of IEEE Symposium on Information Theory (ISIT)*, pages 1135 –1139, 2009.

[3] Tugkan Batu. *Testing properties of distributions.* PhD thesis, Cornell University, 2001.

[4] G.H. Hardy and S. Ramanujan. Asymptotic formulae in combinatory analysis. *Proceedings of London Mathematics Society*, 17(2):75–115, 1918.

[5] Bruno M. Jedynak and Sanjeev Khudanpur. Maximum likelihood set for estimating a probability mass function. *Neural Computation*, 17:1–23, 2005.

[6] B. Kelly, T. Tularak, A. B. Wagner, and P. Viswanath. Universal hypothesis testing in the learning-limited regime. In *Proceedings of IEEE Symposium on Information Theory (ISIT)*, pages 1478–1482, 2010.

[7] A. Orlitsky, S. Pan, Sajama, N.P. Santhanam, and K. Viswanathan. Pattern maximum likelihood: computaiton and experiments. *In preparation*, 2012.

[8] A. Orlitsky and Shengjun Pan. The maximum likelihood probability of skewed patterns. In *Proceedings of IEEE Symposium on Information Theory (ISIT)*, pages 1130–1134, 2009.

[9] A. Orlitsky, N.P. Santhanam, K. Viswanathan, and J. Zhang. Pattern maximum likelihood: existence and properties. *In Preparation*, 2012.

[10] A. Orlitsky, N.P. Santhanam, and J. Zhang. Universal compression of memoryless sources over unknown alphabets. *IEEE Transactions on Information Theory*, 50:1469–1481, 2004.

[11] Alon Orlitsky, Narayana P. Santhanam, Krishnamurthy Viswanathan, and Junan Zhang. On modeling profiles instead of values. In *UAI '04*, pages 426–435, 2004.

[12] Gregory Valiant and Paul Valiant. Estimating the unseen: A sublinear-sample canonical estimator of distributions. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:180, 2010.

[13] Gregory Valiant and Paul Valiant. Estimating the unseen: An $n/\log(n)$-sample estimator for entropy, support size, and other distribution properties, with a proof of optimality via two new central limit theorems. In *STOC '11: Proceedings of the 42nd annual ACM symposium on Theory of computing*, 2011.

[14] Aaron B. Wagner, Pramod Viswanath, and Sanjeev R. Kulkarni. Probability estimation in the rare-events regime. *IEEE Transactions on Information Theory*, 57(6):3207–3229, 2011.