Recent Results on Pattern Maximum Likelihood

Jayadev Acharya
ECE Department, UCSD
Email: jayadev@ucsd.edu

Alon Orlitsky
ECE & CSE Departments, UCSD
Email: alon@ucsd.edu

Shengjun Pan
CSE Department, UCSD
Email: s1pan@ucsd.edu

Abstract—We derive some general sufficient conditions for the uniformity of the Pattern Maximum Likelihood distribution (PML). We also provide upper bounds on the support size of a class of patterns, and mention some recent results about the PML of 1112234.

I. INTRODUCTION

Estimating the distribution underlying an observed data sample has important applications in a wide range of fields, including statistics, genetics, system design, and compression.

Many of these applications do not require knowing the probability of each element, but just the collection, or *multiset* of probabilities. For example, in evaluating the probability that when a coin is flipped twice both sides will be observed, we don't need to know $p(heads)$ and $p(tails)$, but only the multiset $\{p(heads), p(tails)\}$. Similarly to determine the probability that a collection of resources can satisfy certain requests, we don't need to know the probability of requesting the individual resources, just the multiset of these probabilities, regardless of their association with the individual resources. The same holds whenever just the data "statistics" matters.

One of the simplest solutions for estimating this probability multiset uses *standard maximum likelihood (SML)* to find the distribution maximizing the sample probability, and then ignores the association between the symbols and their probabilities. For example, upon observing the symbols @∧@, SML would estimate their probabilities as $p(@) = 2/3$ and $p(\wedge) = 1/3$, and disassociating symbols from their probabilities, would postulate the probability multiset $\{2/3, 1/3\}$.

SML works well when the number of samples is large relative to the underlying support size. But it falls short when the sample size is relatively small. For example, upon observing a sample of 100 distinct symbols, SML would estimate a uniform multiset over 100 elements. Clearly a distribution over a large, possibly infinite number of elements, would better explain the data. In general, SML errs in never estimating a support size larger than the number of elements observed, and tends to underestimate probabilities of infrequent symbols.

Several methods have been suggested to overcome these problems. One line of work began by Fisher [1], and was followed by Good and Toulmin [2], and Efron and Thisted [3]. Bunge and Fitzpatric [4] provide a comprehensive survey of many of these techniques. A related problem, not considered in this paper estimates the probability of individual symbols for small sample sizes. This problem was considered by Laplace [5], Good and Turing [6], and more recently by McAllester and Schapire [7], Shamir [8], Gemelos and Weissman [9], Jedynak and Khudanpur [10], and Wagner, Viswanath, and Kulkarni [11].

A recent information-theoretically motivated method for the multiset estimation problem was introduced in [12] and a number of results are presented in [13], [14]. It is based on the observation that since we do not care about the association between the elements and their probabilities, we can replace the elements by their order of appearance, called the observation's *pattern*. For example the pattern of @∧@ is 121, and the pattern of *abracadabra* is 12314151231.

Slightly modifying SML, this *pattern maximum likelihood (PML)* method asks for the distribution multiset that maximizes the probability of the observed pattern. For example, the 100 distinct-symbol sample above has pattern 123...100, and this pattern probability is maximized by a distribution over a large, possibly infinite support set, as we would expect.

To evaluate the accuracy of PML we conducted the following experiment. We took a uniform distribution over 500 elements, shown in Figure 1 as the solid (blue) line. We sampled the distribution with replacement 1000 times. In a typical run, of the 500 distribution elements, 6 elements appeared 7 times, 2 appeared 6 times, and so on, and 77 did not appear at all as shown in the figure. The standard ML estimate, which always agrees with empirical frequency, is shown by the dotted (red) line. It underestimates the distribution's support size by over 77 elements and misses the distribution's uniformity. By contrast, the PML distribution, as approximated by the EM algorithm described in [14] and shown by the dashed (green) line, performs significantly better and postulates essentially the correct distribution.

As shown in the above and other experiments, PML's empirical performance seems promising. In addition, several results have proved its convergence to the underlying distribution [13]. As mentioned in [12], [15] the analytical calculation of PML appears to be a difficult problem. Thus, analytic methods for finding the PML distribution are of interest. In [12], [15] certain patterns were shown to have uniform PML distribution. We prove PML uniformity for a larger set of patterns. Finding a good upper bound on the number of elements or the number of distinct elements in the PML multiset is an important issue, especially for the practical implementation of the algorithms that estimate the PML of a given pattern. We present some upper bounds on the maximum number of discrete elements in the PML distribution of some specific patterns. We also mention some recent progress that we have made in finding the PML of 1112234, the only length
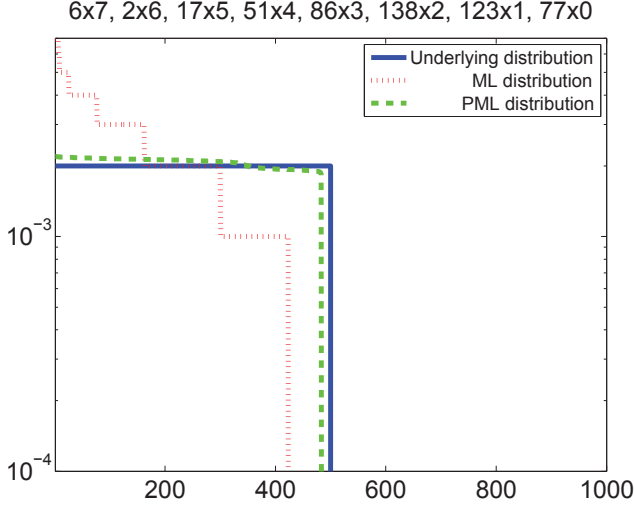
6x7, 2x6, 17x5, 51x4, 86x3, 138x2, 123x1, 77x0

Fig. 1.   SML and PML reconstruction of uniform distribution over 500 symbols from 1000 samples

7 pattern whose PML is not known analytically.

II. NOTATION

The *pattern* $\psi(\overline{x})$ of a sequence $\overline{x} \stackrel{\text{def}}{=} x_1^n$ is the integer sequence obtained by replacing each symbol x in $\overline{x}$ by the number of distinct symbols up to (and including) x's first appearance. For example, $\psi(abracadabra) = 12314151231$.

We denote the length of a pattern by n and its number of distinct symbols by m. The *multiplicity* of an integer ψ in a pattern $\overline{\psi}$ is the number μ_ψ of times ψ appears in $\overline{\psi}$. For example, for 12314151231, $n = 11$, $m = 5$, $\mu_1 = 5$, $\mu_2 = \mu_3 = 2$, and $\mu_4 = \mu_5 = 1$.

For simplicity, if a number ψ repeats consecutively i times, we abbreviate it as ψ^i. For example, we may write the pattern 11222111 as $1^2 2^3 1^3$. A pattern of the form $1^{\mu_1} 2^{\mu_2} \cdots m^{\mu_m}$ with $\mu_1 \geq \cdots \geq \mu_m$ is *canonical*. Clearly every pattern has a canonical pattern with the same multiplicities. For example, the canonical pattern of 123223 is $1^3 2^2 3$.

We now define pattern probabilities. To be most general, we consider *mixed* distributions that assign probabilities to discrete elements and a continuous interval. For example, a distribution P may assign probability $p(a)$ to an element a, $p(b)$ to an element b, and $1 - p(a) - p(b)$ to the interval $[0, 1]$.

If P is sampled independently with replacement then

$$P(\overline{\psi}) \stackrel{\text{def}}{=} P(\{\overline{x} : \psi(\overline{x}) = \overline{\psi}\})$$

is the probability that the sample has pattern $\overline{\psi}$. For example, the distribution P above assigns to the pattern 121 probability

$$P(121) = P(aba) + P(bab) + P(\{xyx : x \in \{a, b\}, y \in [0, 1]\})$$
$$= p^2(a)(1 - p(a)) + p^2(b)(1 - p(b)).$$

Note that the pattern probability is determined by just the multiset of discrete probabilities, hence P can be identified with a vector in the monotone simplex

$$\mathcal{P} \stackrel{\text{def}}{=} \{(p_1, p_2, \ldots) : p_1 \geq p_2 \geq \cdots \geq 0, \quad \sum p_i \leq 1\}.$$

We call $q \stackrel{\text{def}}{=} 1 - \sum p_i$, the *continuous part* of P. The *maximum-likelihood (PML) probability* of a pattern $\overline{\psi}$ is

$$\widehat{P}_{\overline{\psi}}(\overline{\psi}) \stackrel{\text{def}}{=} \max_{P \in \mathcal{P}} P(\overline{\psi}),$$

the highest probability assigned to $\overline{\psi}$ by any distribution, and its *maximum-likelihood (PML) distribution* $\widehat{P}_{\overline{\psi}}$ is the distribution achieving this highest probability. We let $\widehat{k} = \widehat{k}_{\overline{\psi}}$ denote the discrete support size of $\widehat{P}_{\overline{\psi}}$. Observe that every distribution assigns the same probability to a pattern as it does to its canonical form. Hence the two have the same PML distribution. From now on we therefore consider without loss of generality only canonical patterns.

III. SOME KNOWN RESULTS

A pattern is binary if, like 11122, it has $m = 2$. Theorem 11 in [12] shows that all binary patterns have $\widehat{k} = 2$, and the PML distribution can then be determined.

A pattern is *uniform* if, as in 121323, all multiplicities μ_i are equal. A pattern is *1-uniform* if any two multiplicities differ by at most one. A pattern is *quasi-uniform* if the square of the difference between any two multiplicities is at most their sum, namely for all i, j, $(\mu_i - \mu_j)^2 \leq \mu_i + \mu_j$. For example, the pattern 111223 is quasi-uniform. Note that a binary pattern is quasi-uniform if $(\mu_1 - \mu_2)^2 \leq n$.

Theorem 11 in [12] shows also that all quasi-uniform binary patterns have PML $(\frac{1}{2}, \frac{1}{2})$. The following lemma from [15] extends this result to non-binary patterns when the underlying distribution is limited to support size m.

Lemma 1: If an m-symbol pattern is quasi-uniform then among all discrete distributions with support size m, its probability is maximized by the uniform distribution.   ∎

For example, the lemma implies that among all distributions over three elements, $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ maximizes the probability of 111223.

Using Theorem 6 in [12] which states that

$$\widehat{k} \leq m + \frac{m - 1}{2^{\mu_m} - 2} \tag{1}$$

with Lemma 1,

Corollary 2: The PML distribution of a quasi-uniform pattern with $\mu_m > \log_2(m + 1)$ is uniform over m symbols. ∎

An important application of PML is to estimate the underlying distribution's support size $\widehat{k}$. Inequality (1) bounds the support size when the lowest multiplicity, μ_m, is at least 2. The next theorem [15] upper bounds $\widehat{k}$ when $\mu_m = 1$ and all other multiplicities are at least 2, namely exactly one element appears once, for example as in the pattern 11122334. We call such patterns *unique-singleton*. We will later use this result to establish the PML distribution of ternary patterns.

Theorem 3: For unique-singleton patterns,

$$\widehat{k} \leq 2(m - 1).$$ ∎

A pattern ψ is *1-uniform* if $\mu_i - \mu_j \leq 1$ for all i, j, namely all multiplicities are within one from each other as in 1112233.

252

As shown in [13], all 1-uniform patterns have a uniform PML distribution and can thus be evaluated.

As mentioned earlier, the simplest patterns are binary, and their PML distribution was derived in [12], showing in particular that all of them have $\widehat{k} = 2$. The next simplest patterns are ternary, and have $m = 3$. Three types of ternary patterns can be addressed by existing results.

1) Uniform ($1^r 2^r 3^r$). Of these, 123 has $\widehat{P} = ()$, and all others have $\widehat{P} = (1/3, 1/3, 1/3)$ [12].
2) 1-uniform ($1^r 2^r 3^{r-1}$ or $1^r 2^{r-1} 3^{r-1}$). Of these, 1123 has $\widehat{P} = (1/5, 1/5, 1/5, 1/5, 1/5)$, and all others have $\widehat{P} = (1/3, 1/3, 1/3)$ [13].
3) Skewed ($1^r 23$). Of these, 1123 is 1-uniform and addressed above, and all others have $\widehat{P} = (\frac{r}{r+2})$. This result is proved in [16].

It is easy to see that all ternary patterns not covered by these cases have at most one symbol appearing once, for example 111223 and 111122233. For all those, it was shown in [15] that the PML distribution has support size 3.

Theorem 4: All ternary patterns with at most one symbol appearing once have $\widehat{k} = 3$.

The theorem allows us to compute the PML distribution of all ternary patterns by combining the theorem with the Kuhn-Tucker conditions.

Corollary 5: For all ternary patterns with at most one symbol appearing once,

$$\widehat{P}_{1^{\mu_1} 2^{\mu_2} 3^{\mu_3}} = (p_1, p_2, p_3),$$

where p_1, p_2, p_3 are solutions to the following three polynomial equations,

$$p_1 + p_2 + p_3 = 1,$$

$$\sum \mu_{j_1} p_1^{\mu_{j_1}-1} p_2^{\mu_{j_2}} p_3^{\mu_{j_3}} = \sum \mu_{j_1} p_2^{\mu_{j_1}-1} p_3^{\mu_{j_2}} p_1^{\mu_{j_3}},$$

$$\sum \mu_{j_1} p_1^{\mu_{j_1}-1} p_2^{\mu_{j_2}} p_3^{\mu_{j_3}} = \sum \mu_{j_1} p_3^{\mu_{j_1}-1} p_1^{\mu_{j_2}} p_2^{\mu_{j_3}}.$$

where the summation is over all six permutations (j_1, j_2, j_3) of $(1, 2, 3)$.

Using these results, in [15], the PML distribution of all patterns of length up to seven, with the exception of one was found out. These include 111223, 1112223, 1111223, whose PML can be found in the table. The only exception is 1112234, which we conjecture to have PML $(1/5, 1/5, \ldots, 1/5)$ but have not been able to prove yet. The PML distributions of these patterns are shown in Table I along with references to where they were shown.

IV. NEW RESULTS

In the next three subsections we present the recent results, which were mentioned briefly in the Introduction.

A. Patterns with uniform PML distribution

It was shown in [13] that all 1-uniform patterns have a uniform PML distribution. Lemma 1 showed that under certain support size constraints, the probability of quasi-uniform patterns is maximized by uniform distributions. Slightly tightening the quasi-uniformity condition, we remove the support-size

Canonical $\bar{\psi}$	$\widehat{P}_{\bar{\psi}}$	Reference
1	any distribution	Trivial
11, 111, 111, …	(1)	Trivial
12, 123, 1234, …	()	Trivial
112, 1122, 1112, 11122, 111122	$(1/2, 1/2)$	[12]
11223, 112233, 1112233	$(1/3, 1/3, 1/3)$	[13]
111223, 1112223,	$(1/3, 1/3, 1/3)$	[15]
1123, 1122334	$(1/5, 1/5, \ldots, 1/5)$	[12]
11234	$(1/8, 1/8, \ldots, 1/8)$	[13]
11123	$(3/5)$	[16]
11112	$(0.7887.., 0.2113..)$	[12]
111112	$(0.8322.., 0.1678..)$	[12]
111123	$(2/3)$	[16]
111234	$(1/2)$	[16]
112234	$(1/6, 1/6, \ldots, 1/6)$	[13]
112345	$(1/13, \ldots, 1/13)$	[13]
1111112	$(0.857.., 0.143..)$	[12]
1111122	$(2/3, 1/3)$	[12]
1112345	$(3/7)$	[16]
1111234	$(4/7)$	[16]
1111123	$(5/7)$	[16]
1111223	$\left(\frac{1}{\sqrt{7}}, \frac{\sqrt{7}-1}{2\sqrt{7}}, \frac{\sqrt{7}-1}{2\sqrt{7}}\right)$	[15]
1123456	$(1/19, \ldots, 1/19)$	[13]
1112234	$(1/5, 1/5, \ldots, 1/5)?$	Conjectured

TABLE I
PML DISTRIBUTIONS OF ALL PATTERNS OF LENGTH ≤ 7

constraints, thereby greatly extending the lemma's applicability.

Theorem 6: Let $\bar{\psi} \overset{\text{def}}{=} 1^{\mu_1} 2^{\mu_2} \cdots m^{\mu_m}$. Suppose that $(\mu_t - \mu_{t'})^2 \leq \mu_t + \mu_{t'} - 2$ for all $t, t' \in [m]$. Then the PML of $\bar{\psi}$ can be achieved at a uniform distribution.

Proof of Theorem 6: The proof is along the similar lines of the proof of 1-uniform and quasi-uniform patterns. Details are being omitted due to lack of space. ∎

If $\widehat{P}_{1^{\mu_1} 2^{\mu_2} \cdots m^{\mu_m}} = (\widehat{p}_1, \widehat{p}_2, \cdots, \widehat{p}_k)$, the two lemmas below bound the ratio $\widehat{p}_1 / \widehat{p}_k$, which we use to generalize Theorem 6.

Lemma 7: For all patterns,

$$\frac{\widehat{p}_1}{\widehat{p}_k} \leq \frac{\mu_1 - 1}{\mu_m - 1}.$$

For $t \neq t' \in [m]$, define $d_{t,t'} = (\mu_t - \mu_{t'})^2 - (\mu_t + \mu_{t'} - 2)$. Then,

Lemma 8: If $\sum_{i=1}^k \widehat{p}_i = 1$ and $\widehat{p}_1 \neq \widehat{p}_k$,

$$\left(\frac{\widehat{p}_1}{\widehat{p}_k}\right)^{\mu_1 - \mu_m} > \frac{\sum_{d_{t,t'}<0} |d_{t,t'}|}{\sum_{d_{t,t'}>0} d_{t,t'}},$$

where the summations are over unordered pairs $\{t, t'\}$.

The following theorem is the most general sufficiency condition we have for a pattern to have a uniform PML.

Theorem 9: Let $\bar{\psi} \overset{\text{def}}{=} 1^{\mu_1} 2^{\mu_2} \cdots m^{\mu_m}$. If

$$\frac{\sum_{d_{t,t'}<0} |d_{t,t'}|}{\sum_{d_{t,t'}>0} d_{t,t'}} \geq \left(\frac{\mu_1 - 1}{\mu_m - 1}\right)^{\mu_1 - \mu_m},$$

then the PML of $\bar{\psi}$ can be achieved at a uniform distribution.

Proof of Theorem 9: The condition of the theorem violates Lemmas 7 and 8 unless $\widehat{p}_1 = \widehat{p}_k$. This implies that the discrete

part of PML is uniform. We can also prove that for such patterns we can increase the pattern probability by adding two probabilities p_i and p_j. This can be used to show that PML can be achieved at a uniform distribution. ∎

Note that we can obtain the PML distribution for 1-uniform patterns as well as for patterns in Theorem 6 as a special case for this theorem, as we have $d_{t,t'} \leq 0$ for all t, t'.

B. Upper bounds on support size of Patterns

As seen in the previous works, good upper bounds on the support size, or on the number of distinct multiplicities of the PML, is of practical value when estimating the PML. With good support-size bounds, we can apply algorithms such as EM more efficiently, and can be more certain about their convergence. We can also use such bounds to find the exact PML of some patterns [15]. In Theorem 4 of [12] the number of distinct probabilities that occur in PML of any pattern was shown to be at most $\min\{n-1, 2^m\}$. This means that if we can bound the number of times a certain value can repeat in the PML multiset, we can bound the support size. If no multiplicity is equal to 2 we obtain the following result.

Theorem 10: Let $\bar{\psi} = 1^{\mu_1} 2^{\mu_2} \cdots m^{\mu_m}$. If for all $i \in [m]$, $\mu_i \neq 2$, $\widehat{k} \leq 3m^2(n-1)$.

Proof of Theorem 10: For a probability distribution $P = (p_1, p_2, \ldots)$, let $P_i = (p_1, \ldots, p_{i-1}, 0, p_{i+1}, \ldots)$ be the *sub-distribution* agreeing with P on all probabilities, except p_i, which is set to 0. Note that the probabilities in P_i, including q, sum to $1 - p_i$, hence if $p_i > 0$ then P_i is not a distribution but a point inside the probability simplex $\mathcal{P}$.

Similarly for a pattern $\overline{\psi} = 1^{\mu_1} 2^{\mu_2} \cdots m^{\mu_m}$, let $\overline{\psi}_i \overset{\text{def}}{=} 1^{\mu_1} \cdots (i-1)^{\mu_{i-1}} i^{\mu_{i+1}} \cdots (m-1)^{\mu_m}$ be the pattern obtained by deleting all appearances of the ith symbol. For a set $S \subseteq [m]$ denote by $\bar{\psi}_{\bar{S}}$ the pattern which misses the elements in S.

As in [12], the pattern probability is a polynomial of the underlying distribution multiset, and its derivative is a degree $n-1$ polynomial whose roots contain all values of the PML distribution. We then assume that p_i appears k_i times, with $\alpha_i \overset{\text{def}}{=} k_i p_i$. Using the notation defined the following holds for a fixed value of α_i

$$P(\psi) = \sum_{S \subseteq [m]} f_S\left(\frac{1}{k_i}\right) \cdot \alpha_i^{n_S} P_i(\bar{\psi}_{\bar{S}}), \qquad (2)$$

where $m_S = |S|$, $n_S = \sum_{i \in S} \mu_i$ and

$$f_S(x) \overset{\text{def}}{=} x^{n_S - m_S} \prod_{t=1}^{m_S - 1} (1 - tx).$$

Under the conditions assumed in the theorem it can be shown that for

$$x < \frac{1}{3m^2},$$

$f_S''(x) > 0$ for all $S \subseteq [m]$. Thus, when the value of x is less than the value stated above, we know that the function f_S is convex, and thus can have a maxima only at the boundaries. One of the boundaries is $x = 0$, which converts α_i into a

continuous part, and thus contributes nothing to the discrete support size. The other boundary obtained from the bound on x yields $k_i < 3m^2$. Since the number of distinct probabilities is at most $n-1$ the result follows. ∎

C. Recent work on the pattern 1112234

Following are some results on the PML of the pattern 1112234. Claim 1 shows that without a continuous part, the number of distinct discrete probabilities is at most 4 , and with a continuous part, that number is at most 3. Note that this improves the above $n - 1 = 6$ bound.

Claim 1: The number of distinct probabilities in PML is at most 4 if there is no continuous probability, else it is at most 3.

Proof of Claim 1: We first find the polynomial to which the probabilities in PML are roots of. It is given by

$$-42x^6 + 24x^5 + 5x^4(2\widehat{P}(11) - \widehat{P}(12)) + 8x^3(\widehat{P}(111) - \widehat{P}(112))$$
$$+ 3x^2(\widehat{P}(1123) - 2\widehat{P}(1112)) + 2x(\widehat{P}(11123) - \widehat{P}(11122))$$
$$+ 2\widehat{P}(111223) - 7\widehat{P}(1112234) = 0.$$

Using the fact that the PML is majorized by the SML [12], we can check that $\widehat{P}(11) - \widehat{P}(12)$, and $\widehat{P}(111) - \widehat{P}(112)$ are both negative. Also the constant term can be seen to be zero when there is a continuous part, and non-positive otherwise. Using these we can bound the number of sign changes and thus the number of positive roots. ∎

Corollary 11: $\widehat{k}(1112234) \leq 19$.

Proof of Corollary 11: We first prove that either a given probability value will occur at most 4 times or the value itself is larger than 0.044. Combining this with the bound on the number of distinct probabilities we can get the result. ∎

We are currently working towards obtaining more conditions on the PML of 1112234, as well as reducing the support size to a reasonable size when we can run a a program that can find the PML distribution.

REFERENCES

[1] R. Fisher, A. Corbet, and C. Williams, "The relation between the number of species and the number of individuals in a random sample of an animal population," *Journal of Animal Ecology*, vol. 12, pp. 42–48, 1943.

[2] I. Good and G. Toulmin, "The number of new species and the increase in population coverage when the sample is increased," *Biometrika*, vol. 43, no. 1, pp. 45–63, 1956.

[3] B. Efron and R. Thisted, "Estimating the number of unseen species: How many words did Shakespeare know," *Biometrika*, vol. 63, pp. 435–447, 1976.

[4] J. Bunge and M. Fitzpatrick, "Estimating the number of species: a review," *Journal of the American Statistical Association*, vol. 88, pp. 364–373, 1993.

[5] P. Laplace, *Philosophical essays on probabilities*, Translated by A. Dale from the 5th (1825) ed. Springer Verlag, New York, 1995.

[6] I. Good, "The population frequencies of species and the estimation of population parameters," *Biometrika*, vol. 40, no. 3/4, pp. 237–264, December 1953.

[7] D. McAllester and R. Schapire, "On the convergence rate of Good Turing estimators," in *Proceedings of the Thirteenth Annual Conference on Computational Learning Theory*, 2000.

[8] G. Shamir, "Universal lossless compression with unknown alphabets—the average case," *IEEE Transactions on Information Theory*, vol. 52, no. 11, pp. 4915–4944, November 2006.

[9] G. M. Gemelos and T. Weissman, "On the entropy rate of pattern processes," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 3994–4007, 2006.

[10] B. M. Jedynak and S. Khudanpur, "Maximum likelihood set for estimating a probability mass function," *Neural Computation*, vol. 17, pp. 1–23, 2005.

[11] A. B. Wagner, P. Viswanath, and S. R. Kulkarni, "A better good-turing estimator for sequence probabilities," *CoRR*, vol. abs/0704.1455, 2007.

[12] A. Orlitsky, N. Santhanam, K. Viswanathan, and J. Zhang, "On modeling profiles instead of values," in *Proceedings of the 20th conference on Uncertainty in artificial intelligence*, 2004.

[13] ——, "Pattern maximum likelihood: existence and properties," In preparation, 2009.

[14] A. Orlitsky, Sajama, N. Santhanam, K. Viswanathan, and J. Zhang, "Pattern maximum likelihood: computation and experiments," In preparation, 2009.

[15] J. Acharya, A. Orlitsky, and S. Pan, "The maximum likelihood probability of unique-singleton, ternary, and length-7 patterns," Accepted at IEEE International Symposium on Information Theory, 2009.

[16] A. Orlitsky and S. Pan, "The maximum likelihood probability of skewed patterns," Accepted at IEEE International Symposium on Information Theory, 2009.