Algebraic Computation of Pattern Maximum Likelihood

Jayadev Acharya   Hirakendu Das   Alon Orlitsky   Shengjun Pan

UC San Diego

Abstract—**Pattern maximum likelihood (PML) is a technique for estimating the probability multiset of an unknown distribution. With any random sample, it associates the distribution maximizing the probability of its pattern. The required computation is a maximization of a monomial symmetric polynomial over the monotone simplex. The PML of only very few patterns have been found analytically, and for other patterns, the PML has been approximated by a heuristic algorithm. Taking an algebraic approach, we determine the PML of short patterns by solving a system of multivariate polynomial equations using the method of resultants. Using this approach, we determine the PML of the pattern 1112234, the last length-7 pattern whose PML was unknown. Under two plausible but yet unproved assumptions on the optimal alphabet size and the number of distinct probabilities, we also find the PML distribution of all previously unknown patterns of length up to 14.**

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

I. INTRODUCTION

Pattern Maximum Likelihood (PML) is a recent technique for estimating the probability multiset from an observed sample. The *pattern* of a sequence is the integer sequence obtained by substituting each symbol by its order of appearance, *e.g.,* [1]. For example, the pattern of $ISIT$ is 1213 and the pattern of $abracadabra$ is 12314151231. The pattern reflects the number of times and order in which the symbols appear, while abstracting the actual values.

The *PML distribution* is the distribution maximizing the probability of the observed pattern. Its estimation properties have been studied in the context of universal compression of large-alphabet sources in [1, 2, 3], and it has been shown to better estimate the probability multiset than *Sequence Maximum Likelihood (SML)* which maximizes the sequence probability [4].

For example, Figure 1 shows a uniform distribution over 500 elements, indicated by a solid (blue) line. In a typical collection of 1000 samples from this distribution, 2 elements appeared 7 times, 4 appeared 6 times, and so on, and 79 did not appear at all, as shown in the figure. The SML estimate is the empirical distribution, shown by the dotted (red) line. It not only misses the elements which do not appear, but also misses the uniformity. By contrast, PML distribution is essentially same as the underlying distribution.

Although PML has good estimation properties, its analytical calculation appears difficult. The compu-
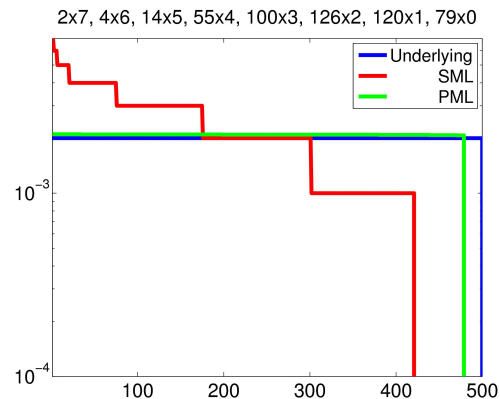
tation involved is that of maximizing a monomial symmetric polynomial over the monotone simplex, for which no general techniques are known. So far, PML distributions have been analytically derived for only special patterns, for example all but one pattern of length ≤ 7, patterns with just two or three distinct symbols, patterns that are close to uniform, and skewed patterns [4, 5, 6, 7].

We consider an algebraic approach for finding the PML. Given an upper bound on the support size of the PML distribution, we maximize the pattern probability by solving the system of multivariate polynomial equations obtained by differentiating the pattern probability with respect to each variable.

In Section III we describe two well known *elimination* techniques, for solving systems of multivariate polynomial equations, Groebner bases and resultants. Given a pattern and bounds on the alphabet size and number of distinct probabilities in its PML distribution, we find it using resultants.

However, severe computational requirements of algebraic elimination methods render this approach time intensive even for small patterns and moderate support sizes. In Sections IV and V, we combine the algebraic approach with analytic upper bounds on the support size and number of distinct probabilities in the PML distribution.

Of all patterns of length ≤ 7, the only one whose

PML was not known, was 1112234. We show that its PML distribution has support size ≤ 17 and ≤ 3 distinct probabilities. We then use the resultants method to show that its PML distribution is the uniform distribution over 5 symbols, $\{\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\}$. While we do not have upper bounds on the alphabet size and number of distinct probabilities of the PML distribution of general patterns, under two such plausible but unproved bounds, we compute the PML of all patterns of length ≤ 14.

II. NOTATION

A. Patterns

The *pattern* $\Psi(\overline{x})$ of a sequence $\overline{x} \stackrel{\text{def}}{=} x_1 x_2 \cdots x_n$ is the integer sequence derived by replacing each symbol in $\overline{x}$ by its order of appearance. For example, in the sequence $ISIT$, the symbol I appears first, the symbol S appears second, and T is the third distinct symbol to appear, hence $\Psi(ISIT) = 1213$. Similarly, $\Psi(abracadabra) = 12314151231$.

We denote the length and the number of distinct symbols in a pattern by n and m respectively. The *multiplicity* μ_ψ of an integer ψ in a pattern $\overline{\psi}$ is the number of times ψ appears in $\overline{\psi}$. For example, the pattern 12314151231, has $n = 11, m = 5, \mu_1 = 5, \mu_2 = \mu_3 = 2$, 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 123323 is $1^3 2^2 3$.

B. Pattern probabilities

The probability of a pattern $\overline{\psi}$ is

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

the probability of observing a sequence with that pattern. Throughout this paper, we consider sequences generated *i.i.d.* according to probability distributions. Let $\overline{\psi} = 1^{\mu_1} 2^{\mu_2} \cdots m^{\mu_m}$ be a length-n canonical pattern and $P = (p_1, p_2, \ldots, p_k)$ be a distribution over the alphabet $[k] \stackrel{\text{def}}{=} \{1, 2, \ldots, k\}$ of size k. Since there are $k^{\underline{m}} \stackrel{\text{def}}{=} k(k-1) \cdots (k-m+1)$ sequences that have the same pattern $\overline{\psi}$, the pattern probability is

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

$$= \sum_{(i_1, i_2, \ldots, i_m) \in [k]^{\underline{[m]}}} p_{i_1}^{\mu_1} p_{i_2}^{\mu_2} \cdots p_{i_m}^{\mu_m}.$$

where $[k]^{\underline{[m]}}$ is the set of all mappings from $[m]$ to $[k]$.

Observe that the pattern probability is a symmetric polynomial in the probabilities. Also, when sequences are generated *i.i.d.*, a pattern and its canonical pattern have the same probability and hence, we only consider the PML of canonical patterns.

For example, if a distribution P assigns probability p_a to an element a, and p_b to element b, and p_c to element c, then the probability of the pattern 121 is the sum of probabilities of the $3^{\underline{2}} = 6$ sequences with pattern 121,

$$P(121) = P(112)$$
$$= P(aab) + P(bba) + P(aac) + P(cca) + P(ccb) + P(bbc)$$
$$= p_a^2 p_b + p_b^2 p_a + p_a^2 p_c + p_c^2 p_a + p_c^2 p_b + p_b^2 p_c.$$

C. PML Distribution

Let $\mathcal{P}_{\mathcal{M}}$ be the set of all monotone distributions,

$$\mathcal{P}_{\mathcal{M}} \stackrel{\text{def}}{=} \{(p_1, p_2, \ldots) : p_i \geq 0, p_i \geq p_{i+1}, \sum_{i=1}^{\infty} p_i = 1\}.$$

The *pattern maximum likelihood* $\widehat{P}_{\overline{\psi}}(\overline{\psi})$ of a pattern $\overline{\psi}$ is its maximum likelihood under all possible distributions, *i.e.*,

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

The distribution $\widehat{P}_{\overline{\psi}} \stackrel{\text{def}}{=} \arg\max_{P \in \mathcal{P}_{\mathcal{M}}} P(\overline{\psi})$ is called the PML distribution of $\overline{\psi}$. Existence and convergence properties of $\widehat{P}_{\overline{\psi}}$ were presented in [4].

We denote the alphabet size of $\widehat{P}_{\overline{\psi}}$ by $\widehat{k} \stackrel{\text{def}}{=} \widehat{k}(\overline{\psi})$ and the number of distinct probabilities in $\widehat{P}_{\overline{\psi}}$ by $\widehat{\Delta} \stackrel{\text{def}}{=} \widehat{\Delta}(\overline{\psi})$. Similar to $\widehat{P}_{\overline{\psi}}$, we define $\widehat{P}_{\overline{\psi}}^k$ as the distribution maximizing $P(\overline{\psi})$ among all distributions with at most k nonzero probabilities, and $\widehat{P}_{\overline{\psi}}^{k,\Delta}$ as the maximizing distribution when further restricting to distributions with at most Δ distinct nonzero probabilities.

III. ALGEBRAIC APPROACH FOR FINDING THE PML

Consider the set

$$\mathcal{P}^k \stackrel{\text{def}}{=} \{(p_1, p_2, \ldots, p_k) : p_i \geq 0, \sum_{i=1}^{k} p_i = 1\}$$

of distributions whose support size is at most k. The Kuhn-Tucker conditions imply that if a distribution $P \in \mathcal{P}^k$ is a local maximum of $P(\overline{\psi})$, then for all $i, j \in [k]$ such that $p_i, p_j > 0$,

$$\frac{\partial P(\overline{\psi})}{\partial p_i} = \frac{\partial P(\overline{\psi})}{\partial p_j}.$$

To find $\widehat{P}_{\overline{\psi}}^k$, we can solve the system of equations

$$\frac{\partial P(\overline{\psi})}{\partial p_i} = \frac{\partial P(\overline{\psi})}{\partial p_{i+1}}, \quad i = 1, 2, \ldots, \kappa - 1 \tag{2a}$$

$$p_1 + p_2 + \cdots + p_\kappa = 1, \tag{2b}$$

for each $\kappa \in \{m, m+1, \ldots, k\}$, and among all solutions, find the one maximizing $P(\overline{\psi})$. Note that since the Kuhn-Tucker equalities hold only for the nonzero probabilities, it is not sufficient to consider solutions to system of equations (1a) and (1b) for just $\kappa = k$.

As an example of this basic method and the need to solve the equations for all $\kappa \leq k$, consider $\widehat{P}_{112}^3$. For $\kappa = 2$, the equations yield

$$p_1^2 = p_2^2,$$
$$p_1 + p_2 = 1,$$

whose unique solution is $P = (p_1, p_2) = (\frac{1}{2}, \frac{1}{2})$ with $P(112) = \frac{1}{4}$. For $\kappa = 3$, Equations (1a) and (1b), yield

$$p_1(2 - 3p_1) = p_2(2 - 3p_2) = p_3(2 - 3p_3),$$
$$p_1 + p_2 + p_3 = 1,$$

whose only solution is $P' = (p_1, p_2, p_3) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Furthermore, $P'(112) = \frac{2}{9} < \frac{1}{4} = P(112)$, hence $\widehat{P}_{112}^3 = (\frac{1}{2}, \frac{1}{2})$.

While such simple manipulations work for small patterns, for longer patterns, we need a systematic approach for solving the set of polynomial equations obtained. The natural approach for solving a system of polynomial equations is to generalize the Gaussian-elimination method for linear equations to address polynomial equations. There are two well-known approaches for doing that.

The first uses Buchberger's algorithm and its variations that yield a *Groebner basis* for the original polynomials. However, the degrees of the resulting polynomials, and hence also the computation time of these algorithms may in general be doubly exponential in the number of variables. For more information, see, *e.g.,* [8].

The second approach uses resultants, *e.g.,* [9, 8]. While it too may require doubly-exponential time, in our experiments it has performed more efficiently, and we describe it here.

The *resultant* of a degree-u polynomial $f = f_0 x^u + f_1 x^{u-1} + \cdots + f_u$ and a degree-v polynomial $g = g_0 x^v + g_1 x^{v-1} + \cdots + g_v$ is the determinant of a corresponding $(u+v) \times (u+v)$ *Sylvester matrix*,

$$\text{Res}(f, g, x) \stackrel{\text{def}}{=} \begin{vmatrix} f_0 & & & & g_0 & & & \\ f_1 & f_0 & & & g_1 & g_0 & & \\ \vdots & f_1 & \ddots & f_0 & \vdots & g_1 & \ddots & g_0 \\ f_u & \vdots & \ddots & f_1 & g_v & \vdots & \ddots & g_1 \\ & f_u & & \vdots & & g_v & & \vdots \\ & & \ddots & f_u & & & \ddots & g_v \end{vmatrix},$$

where v columns correspond to f, u columns to g, and blank spaces are zeros.

If f and g are multivariate polynomials with x as one of the variables, then viewing f and g as polynomials in x whose coefficients are polynomials in the other variables, $\text{Res}(f, g, x)$ is a polynomial in the remaining variables. The important property of resultants that makes them useful for elimination is that $\text{Res}(f, g, x) = a \cdot f + b \cdot g$ for two polynomials a and b. Hence solving the equations $f = g = 0$ is equivalent to solving the system $f = \text{Res}(f, g) = 0$.

To eliminate variables $p_2, \ldots, p_{\kappa-1}$ from Equations (1a), we use resultants to eliminate p_2 from the $\kappa - 1$ equations in (1a) to obtain $\kappa - 2$ equations in $p_1, p_3, \ldots, p_\kappa$ from which we eliminate p_3 and obtain $\kappa - 3$ equations. We proceed similarly to eliminate $p_4, \ldots, p_{\kappa-1}$, until we are left with a single homogenous equation in p_1 and p_κ. We use this to solve for $\frac{p_\kappa}{p_1}$ and by backsubstution, obtain $(\frac{p_{\kappa-1}}{p_1}, \frac{p_{\kappa-2}}{p_1}, \ldots, \frac{p_2}{p_1})$. Finally, using Equation (1b), we obtain all the probabilities $(p_1, p_2, \ldots, p_\kappa)$.

While the resultant of two polynomials can be obtained by explicitly computing the determinant of their Sylvester matrix or using other well known determinantal formulae for resultants, a brute-force computation is challenging. Other computationally efficient methods for computing resultants that use interpolation techniques are discussed in [10, 9]. Nevertheless, it is easy to see that after eliminating $p_2, \ldots, p_{\kappa-1}$, the degree of the final polynomial in p_1 and p_κ can be $O(n^{2^\kappa})$. This makes the resultant calculations intensive for even small values of n and κ.

The number of calculations is smaller when considering distributions with at most Δ distinct probabilities. For such distributions, it suffices to consider for each $d \in 1, 2, \ldots, \Delta$, partitions of $\{p_1, p_2, \ldots, p_\kappa\}$ into d parts where within each part all probabilities are equal, and perform the elimination with d variables.

When evaluating resultants of two polynomials, we remove their common factors, otherwise the resultant is zero. While we do not discuss mixed distributions that have discrete probabilities as well as a continuous part, the method can be easily extended to this case by adding an additional variable q for the probability of the continuous part. A summary of the complete method is listed in Algorithm 1, which computes the PML distribution $\widehat{P}_{\overline{\psi}}^{k,\Delta}$, given as input a pattern $\overline{\psi}$ and bounds k and Δ.

We implemented Algorithm 1 in MATHEMATICA. Due to computational limitations, the program can be used to compute $\widehat{P}_{\overline{\psi}}^{k,\Delta}$ for patterns of length ≤ 14 with $k \leq 17$ and and $\Delta \leq 4$. In the next two sections, we consider bounds on $\widehat{k}$ and $\widehat{\Delta}$ which are used with the program to compute the PML of 1112234 and other patterns.

IV. PML OF 1112234

We show a sufficient condition for a pattern to have no continuous part in its PML distribution. Following

Algorithm 1 Computation of $\widehat{P}_{\overline{\psi}}^{k,\Delta}$ using resultants

---

Initialize solution set $\mathcal{S} = \{\}$
for $\kappa = m$ **to** k and $d = 1$ **to** $\max\{\Delta, \kappa\}$ **do**
  for unordered partitions $(\mathcal{K}_1, \ldots, \mathcal{K}_d)$ of $[\kappa]$ **do**
    for $i = 1$ **to** $d - 1$ **do**
      $g_{1,i} := \frac{\partial P(\overline{\psi})}{\partial p_\imath} - \frac{\partial P(\overline{\psi})}{\partial p_\jmath}$ for some $\imath \in \mathcal{K}_i$ and $\jmath \in \mathcal{K}_{i+1}$
      Set $p_\imath = z_j$ for all $\imath \in \mathcal{K}_j$ for all j
    end for
    $\mathcal{G}_1 := \{g_{1,1}, g_{1,2}, \ldots, g_{1,d-1}\}$
    // Solve the system of equations $\mathcal{G}_1 = 0$
    for $j = 2$ **to** $d - 1$ **do**
      for $i = 1$ **to** $d - j$ **do**
        // Eliminate z_j by taking resultants
        $g_{j,i} = \mathrm{Res}(g_{j-1,i}, g_{j-1,i+1}, z_j)$
        Remove any monomial factors of $g_{j,i}$ and trivial factors of form $(z_\imath - z_\jmath)$
        Return error for current k and d if $g_{j,i}$ is a constant polynomial
      end for
    end for
    Set $z_1 = 1$ and solve the triangular system of equations $\{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_{d-1}\}$ for variables $z_2, z_3, \ldots, z_d$ by backsubstitution
    For solutions $(z_1, z_2, \ldots, z_d)$ such that all z_i are real and positive, find $Z = \sum_{i=1}^{d} z_i |\mathcal{K}_i|$, find $(p_1, p_2, \ldots, p_\kappa)$ such that $p_\imath = z_j / Z$ where $\imath \in \mathcal{K}_j$ and $S = S \cup \{sort(p_1, p_2, \ldots, p_\kappa)\}$
  end for
end for
Output $\arg\max_{P \in \mathcal{S}} P(\overline{\psi})$

---

the notation in [1], let $\varphi_\mu(\overline{\psi}) = \varphi_\mu$ denote the number of symbols that have appeared μ times in a pattern $\overline{\psi}$.

Lemma 1. *The PML distribution of patterns with $\varphi_2 \geq \frac{n-1}{2(m-1)}\binom{\varphi_1}{2}$ has no continuous part.*

Proof Outline: Let P be the PML distribution of $\overline{\psi}$. Suppose P has a positive continuous probability $q > 0$. Let Q^ϵ be the distribution obtained from $\widehat{P}_{\overline{\psi}}$ by replacing q with $q - \epsilon$ and a new discrete probability ϵ. The main idea is to show that if $\varphi_2 < \frac{n-1}{2(m-1)}\binom{\varphi_1}{2}$, then there exists $0 < \epsilon \leq q$ such that $Q^\epsilon(\overline{\psi})$ is strictly greater than $P(\overline{\psi})$, a contradiction. The details are omitted for brevity. ∎

For the pattern 1112234, $\varphi_2 = 1 = \frac{7-1}{2(4-1)}\binom{2}{2} = \frac{n-1}{2(m-1)}\binom{\varphi_1}{2}$. Hence Lemma 1 implies that its PML distribution has no continuous part.

The next lemma upper bounds the number of distinct probabilities of the PML distribution. It follows from the main result of [11].

Lemma 2. *The PML distribution of any length-n pattern has $\leq \lfloor \frac{n}{2} \rfloor$ distinct discrete probabilities.* ∎

The following theorem upper bounds $\widehat{k}(1112234)$, and the techniques used can potentially be extended to other patterns.

Theorem 3. $\widehat{k}(1112234) \leq 17$.

Proof Outline: Let $P = (p_1, p_2, \ldots, p_k)$ be the PML distribution of pattern 1112234, where $p_1 \geq p_2 \geq \cdots \geq p_k > 0$. Let $\tilde{p}_1 > \tilde{p}_2 > \cdots > \tilde{p}_d > 0$ be the distinct probabilities in P and k_i be the number of occurrences of $\tilde{p}_i$ in P, for $i = 1, \ldots, d$. By Lemma 2 $d \leq 3$.

Given any $i \in [d]$, fix $k_i \tilde{p}_i$ as well as k_j and $\tilde{p}_j$ for all $j \neq i$. Then $P(1112234)$ can be viewed as a function of k_i. This function can be shown to decrease in k_i for $k_i \geq 4$ and $\tilde{p}_i \leq 0.04327$. The details are omitted. Thus to maximize $P(1112234)$ we must have, for any $i \in [d]$,

$$k_i \leq 3, \text{ or } \tilde{p}_i > 0.04327.$$

On the other hand we also obtain the following bounds which are not proved here for lack of space.

$$p_1 > 0.1990, \text{ and } p_1 + p_2 + p_3 > 0.5139.$$

From this it follows that $k_1 \leq 5$. Hence if $k_1 \geq 4$, then $k_1 \tilde{p}_1 \geq 4 \cdot 0.1990 = 0.794$ and $k \leq 5 + 3 \cdot \max\left\{3, \left\lfloor \frac{1-0.794}{0.04327} \right\rfloor\right\} = 17$.

But, if $k_1 \leq 3$, there are at most 2 distinct probabilities in $\{p_4, p_5, \ldots, p_k\}$. Since $p_1 + p_2 + p_3 \geq 0.5139$, we have $k \leq 3 + 2 \cdot \max\left\{3, \left\lfloor \frac{1-0.5139}{0.04327} \right\rfloor\right\} = 17$. ∎

The above bounds on $\widehat{k}$ and $\widehat{\Delta}$ of 1112234 along with Algorithm 1 yields that $\widehat{P}_{1112234} = \widehat{P}_{1112234}^{17,3} = \{\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\}$, *i.e.*, the uniform distribution over 5 symbols.

V. PML OF SHORT PATTERNS UNDER TWO ASSUMPTIONS

While we do not have good general upper bounds on $\widehat{k}$ and $\widehat{\Delta}$ two plausible assumptions are:

A1 $\widehat{\Delta}$ is at most the number of distinct multiplicities in the pattern.

A2 $\widehat{P}_{\overline{\psi}}^k(\overline{\psi})$ is strictly increasing for $m \leq k \leq \widehat{k}$.

One possible justification for A1 is that each of the probabilities may correspond to an observed symbol and then symbols whose multiplicities are equal would be assigned equal probability estimates. For A2, it may be plausible that if P_1 and P_2 are two distributions whose alphabet sizes are k and $k + 2$, and $P_1(\overline{\psi}) > P_2(\overline{\psi})$, then there may exist a distribution P_3 whose support size is $k + 1$ such that $P_1(\overline{\psi}) < P_3(\overline{\psi}) \leq P_2(\overline{\psi})$.

Under these assumptions, given a pattern $\overline{\psi}$, we use Algorithm 1 with Δ as the number of distinct multiplicities in $\overline{\psi}$ for $k = m, m+1, \ldots$ until $\widehat{P}_{\overline{\psi}}^{k,\Delta}(\overline{\psi}) = \widehat{P}_{\overline{\psi}}^{k+1,\Delta}(\overline{\psi})$.

n	$\overline{\psi}$	$\widehat{P}_{\overline{\psi}}$	$\widehat{P}_{\overline{\psi}}(\overline{\psi})$
8	$1^4 2^2 34$	$\{0.462.., 0.134..^4\}$	4.08×10^{-4}
	$1^3 2^3 34$	$\{0.25^4\}$	3.66×10^{-4}
	$1^3 2^2 345$	$\{0.15625^6, 0.0625\}$	4.38×10^{-4}
9	$1^5 2^2 34$	$\{0.553.., 0.0893..^5\}$	1.99×10^{-4}
	$1^4 2^3 34$	$\{0.389..^2\}, 0.222..$	1.33×10^{-4}
	$1^4 2^2 3^2 4$	$\{0.25^4\}$	9.16×10^{-5}
	$1^4 2^2 345$	$\{0.433.., 0.0708..^8\}$	1.10×10^{-4}
	$1^3 2^3 3^2 4$	$\{0.25^4\}$	9.16×10^{-5}
	$1^3 2^3 345$	$\{0.333..^2\}, 0.333..$	1.02×10^{-4}
	$1^3 2^2 3456$	$\{0.138..^6\}, 0.171..$	1.57×10^{-4}
10	$1^6 2^2 34$	$\{0.599.., 0.0800..^5\}$	1.15×10^{-4}
	$1^5 2^3 34$	$\{0.4^2\}, 0.2$	5.24×10^{-5}
	$1^5 2^2 3^2 4$	$\{0.461.., 0.180..^3\}$	2.67×10^{-5}
	$1^5 2^2 345$	$\{0.499.., 0.0626..^5\}$	5.02×10^{-5}
	$1^4 2^4 34$	$\{0.4^2\}, 0.2$	5.24×10^{-5}
	$1^4 2^3 3^2 4$	$\{0.25^4\}$	2.29×10^{-5}
	$1^4 2^3 345$	$\{0.35^2\}, 0.3$	3.47×10^{-5}
	$1^4 2^2 3^2 45$	$\{0.304.., 0.139..^5\}$	1.23×10^{-5}
	$1^4 2^2 3456$	$\{0.3^2\}, 0.4$	3.73×10^{-5}
	$1^3 2^3 3^3 4$	$\{0.25^4\}$	2.29×10^{-5}
	$1^3 2^3 3^2 45$	$\{0.2^5\}$	1.23×10^{-5}
	$1^3 2^3 3456$	$\{0.3^2\}, 0.4$	3.73×10^{-5}
	$1^3 2^2 34567$	$\{0.157..^4\}, 0.371..$	7.16×10^{-5}

For example, for the canonical pattern $\overline{\psi} = 1^5 2^2 3^2 45$ of *abracadabra*, the multiplicities of the symbols are $(5, 2, 2, 1, 1)$ and there are 3 distinct multiplicities 5, 2 and 1. Assumption A1 implies that the number of distinct probabilities is at most 3. Since $m = 5$, we run Algorithm 1 with $\Delta = 3$ and $k = 5, 6, 7, \ldots$, and observe that $\widehat{P}^{5,3}_{1^5 2^2 3^2 45} = 3.241.. \times 10^{-6}$, $\widehat{P}^{6,3}_{1^5 2^2 3^2 45} = \widehat{P}^{7,3}_{1^5 2^2 3^2 45} = 4.073.. \times 10^{-6}$. Hence we stop and output $\widehat{P}^{6,3}_{1^5 2^2 3^2 45} = \left\{ \frac{\alpha}{5+\alpha}, \left(\frac{1}{5+\alpha}\right)^5 \right\} = \{0.4429.., 0.1114..^5\}$, where $\alpha = 3.976..$ is a root of $6x^4 - 19x^3 - 19x^2 - x - 1 = 0$.

Under these assumptions, we computed the PML of all patterns of length ≤ 14. For space considerations, Table I shows the PML only of patterns of length ≤ 10. Furthermore, the PML of all binary, ternary, skewed, and quasi-uniform patterns have been determined before, and the table shows the remaining patterns. The PML distribution $\widehat{P}_{\overline{\psi}}$ is represented in the form $\{\tilde{p}_1^{k_1}, \ldots, \tilde{p}_d^{k_d}\}, q$ indicating that for $i = 1, \ldots, d$ it consists of k_i symbols whose probability is $\tilde{p}_i$, and that the continuous part is $q = 1 - \sum_{i=1}^{d} k_i \tilde{p}_i$, shown only when nonzero. Note that all numbers are algebraic, and are truncated to a few significant digits.

REFERENCES

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

[2] A. Orlitsky, N. Santhanam, K. Viswanathan, and J. Zhang, "Limit results on pattern entropy," *itt*, vol. 52, pp. 2954–2964, 2006.

[3] G. I. Shamir, "Universal lossless compression with unknown alphabets - the average case," *CoRR*, vol. abs/cs/0603068, 2006.

[4] 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.

[5] J. Acharya, A. Orlitsky, and S. Pan, "Recent results on pattern maximum likelihood," in *IEEE Information Theory Workshop*, 2009.

[6] ——, "Maximum likelihood probability of unique-singleton, ternary, and length-7 patterns," in *ISIT*, 2009.

[7] A. Orlitsky and S. Pan, "Maximum likelihood probability of skewed patterns," in *ISIT*, 2009.

[8] D. A. Cox, J. Little, and D. O'Shea, *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*. Springer-Verlag New York, 2007.

[9] ——, *Using Algebraic Geometry*. Springer, 2004.

[10] D. Manocha and J. F. Canny, "Multipolynomial resultant algorithms," *Journal of Symbolic Computation*, vol. 15, pp. 99–122, 1993.

[11] V. Timofte, "On the positivity of symmetric polynomial functions. Part I: General results," *Journal of Mathematical Analysis and Applications*, vol. 284, no. 1, pp. 174 – 190, 2003.