# Exact Calculation of Pattern Probabilities

Jayadev Acharya
ECE, UCSD
jacharya@ucsd.edu

Hirakendu Das
ECE, UCSD
hdas@ucsd.edu

Hosein Mohimani
ECE, UCSD
hoseinm@ucsd.edu

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

Shengjun Pan
CSE, UCSD
s1pan@ucsd.edu

***Abstract*—We describe two algorithms for calculating the probability of $m$-symbol length-$n$ patterns over $k$-element distributions, a partition-based algorithm with complexity roughly $2^{O(m\log m)}$ and a recursive algorithm with complexity roughly $2^{O(m+\log n)}$ with the precise bounds provided in the text. The problem is related to symmetric-polynomial evaluation, and the analysis reveals a connection to the number of connected graphs.**

## I. Introduction

Recent works on estimating distributions over large alphabets have replaced the observed sequence, assuming *i.i.d.*, by its *pattern*, the integer sequence obtained by substituting each symbol by its order of appearance [1, 2, 3, 4] For example the pattern of @∧@ is 121, and the pattern of *abracadabra* is 12314151231. The pattern reflects the number of times and order in which symbols appear, while abstracting their actual values.

It has been shown that typically the maximum likelihood (ML) distribution maximizing the probability of the observed pattern, approximates the underlying distribution better than the ML distribution of the sequence itself.

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, 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 sequence maximum-likelihood (SML) estimate, which always agrees with empirical frequency, is shown by the dotted (red) line. It underestimates that the distribution's support size, and also misses the uniformity. By contrast, PML 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 [5], yet analytical calculation of the PML distribution for specific patterns appears difficult. So far the PML distribution has been analytically derived for only very simple or short (length $\leq 7$) patterns.

Essentially all practical PML's have therefore been evaluated computationally, typically using an Expectation Maximization Algorithm approxmiating the PML [6].

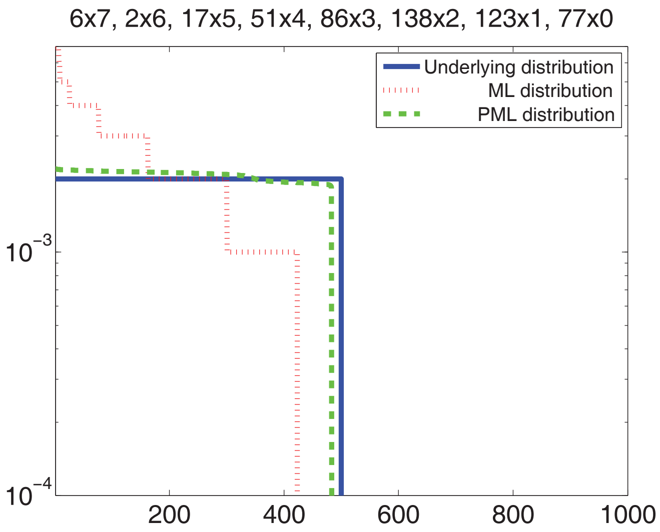

Fig. 1. SML and PML reconstruction of uniform distribution over 500 symbols from 1000 samples

In this paper, we address the more basic problem of precise calculation of pattern probability. The problem is of interest for its own sake, and also as it corresponds to the calculation of symmetric polynomials.

## II. Pattern probability

The probability that a sample has pattern $\overline{\psi}$ is

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

For example, if a distribution $P$ assigns probability $p(a)$ to an element $a$, and $p(b)$ to element $b$, then the probability of the pattern 121 is

$$P(121) = P(aba)+P(bab) = p^2(a)p(b) + p^2(b)p(a).$$

In general, we denote the length of a pattern by $n$ and its number of distinct symbols by $m$. The *multiplicity* of an integer $\psi$ in a pattern $\overline{\psi}$ is the number $\mu_\psi$ of times $\psi$ 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 $\psi$ repeats consecutively $i$ times, we abbreviate it as $\psi^i$. 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, and their probabilities are the same under any distribution. For example, the canonical pattern of 123223 is $1^3 2^2 3$, and $P(123223) = P(1^3 2^2 3)$ for every distribution $P$. We therefore consider only canonical patterns.

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

$$\{(p_1, p_2, \dots) : p_1 \geq p_2 \geq \cdots \geq 0, \quad \sum p_i = 1\}.$$

It is easy to see that the pattern probability is a *symmetrized monomial* (also called *monomial symmetric polynomial*)

$$P(1^{\mu_1} 2^{\mu_2} \dots m^{\mu_m}) \\ = \sum_{i_1=1}^{k} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{k} \cdots \sum_{\substack{i_m=1 \\ i_m \neq i_1, i_2, \dots}}^{k} p_{i_1}^{\mu_1} p_{i_2}^{\mu_2} \dots p_{i_m}^{\mu_m}.$$

This expression consists of $k^{\underline{m}}$ terms. In cases of interest, both $k$ and $m$ could be large, rendering straightforward calculation of the pattern probability infeasible.

In this paper we analyze two methods for calculating the pattern probability. One has complexity $m2^m + kn\log n$, and the other $m \cdot 3^m + kn \log n$. While these complexities are large as well, their evaluation will reveal an interesting connection to the number of connected graphs.

## III. Recursive algorithms

For any pattern $\bar{\psi} = 1^{\mu_1} \cdots m^{\mu_m}$ and distribution $P$, $P(\bar{\psi})$ can be written as

$$P(\bar{\psi}) = P(1^{\mu_1}) \cdot P\Big(1^{\mu_2} \cdots (m-1)^{\mu_m}\Big) - \sum_{i=2}^{m} P(\bar{\psi}_i), \quad (1)$$

where $\bar{\psi}_i \overset{\text{def}}{=} 1^{\mu_1+\mu_i} 2^{\mu_2} \cdots (i-1)^{\mu_{i-1}} i^{\mu_{i+1}} \cdots (m-1)^{\mu_m}$ is the pattern obtained from $\bar{\psi}$ by identifying the $i$-th symbol with the first symbol. For example, for $\bar{\psi} = 11223$, $\bar{\psi}_2 = 11112$, $\bar{\psi}_3 = 11122$, and

$$P(11223) = P(11) \cdot P(112) - P(11112) - P(11122).$$

The pattern probability $P(\bar{\psi})$ can be calculated by recursively applying Equation (1) to patterns appearing on its right-hand side.

Observe that all patterns of the form $1^\mu$ appearing in the recursive calculation satisfy $\mu = \sum_{i \in S} \mu_i$ for some $S \subseteq [m]$.

The *prevalence* $\varphi_\mu$ of an integer $\mu$ in a pattern $\bar{\psi}$ is the number of symbols appearing $\mu$ times. The *profile* $\bar{\varphi}$ of a pattern $\bar{\psi}$ is the formal product $\prod_\mu \mu^{\varphi_\mu}$. For example, the prevalences in pattern $1^3 2^3 3^2$ are $\varphi_2 = 1$ and $\varphi_3 = 2$, and the profile is therefore $2^1 3^2$.

Let $\nu_1^{\varphi_1} \nu_2^{\varphi_2} \cdots \nu_d^{\varphi_d}$ be the profile of pattern $\bar{\psi}$, where $d$ is the number of distinct multiplicities. The following theorem bounds the complexity in terms of $k$, $n$, $m$, and $d$. showing in particular that for constant $d$, the complexity is polynomial in the other parameters.

**Theorem 1.** *For any distribution $P$, the probability of pattern $\bar{\psi} = 1^{\mu_1} 2^{\mu_2} \cdots m^{\mu_m}$ can be deterministically calculated in time*

$$O\Bigg(k \min\{n, 2^m\} \log n \\ + \min\Big\{m e^{\pi\sqrt{n}}, nm2^m, m3^m, nm^d, m^{2d}\Big\}\Bigg),$$

*where $d$ is the number of distinct multiplicities.*

*Proof:* Assume that $P(1^{\sum_{i \in S} \mu_i})$ are calculated beforehand for all $S \subseteq [m]$, which can be done in time $O\left(k \min\{n, 2^m\} \log n\right)$.

Patterns of the same canonical form have the same probability. The *computational graph* for calculating $P(\bar{\psi})$ is a directed graph, consisting of canonical forms of all patterns ever appearing in the recursive calculation, and the (outgoing) neighbors of pattern $\bar{\psi}$ are patterns appearing on the right-hand side of Equation (1), namely $1^{\mu_1}, 1^{\mu_2} \cdots (m-1)^{\mu_m}$, and $\bar{\psi}_i$'s.

All patterns in the computational graph have length at most $n$. Since a canonical pattern of length $n$ can be uniquely represented as the partition of its length into its multiplicities: $n = \sum_{i=1}^m \mu_i$, the number of patterns in the computational graph is at most the number of partitions of integers up to $n$, which is approximately $O(e^{\pi\sqrt{n}})$. Thus the computational time for the recursive steps is $O(m \cdot e^{\pi\sqrt{n}})$.

Furthermore, patterns of the same profile have the same probability. Any pattern in the computational graph can be obtained by first removing a subset of symbols from $\bar{\psi}$ , and then identifying another subset of the remaining symbols. Any such pattern can be represented by its *profile form* $(\nu_0; \phi_1, \phi_2, \dots, \phi_d)$ : one *pivot* symbol appearing $\nu_0$ times, and $\phi_i$ $(\leq \varphi_i)$ symbols appearing $\nu_i$ times, where $\nu_0$ is the sum of multiplicities from the removed subset.

For example, $\bar{\psi} = 1^3 2^2 3^2 45$ has multiplicities $\{3, 2, 2, 1, 1\}$, where the distinct ones are $\nu_1 = 3$, $\nu_2 = 2$, and $\nu_3 = 1$. Pattern $1^5 2^2 3$ appears in the computational graph, corresponding to removing the subset $\{1\}$ from $\{3,2,1,1\}$ and adding elements in the subset $\{3,2\}$, and hence it can written as $(5;0,1,1)$.

Then Equation (1) for a pattern with profile form $(\nu_0;\phi_1,\phi_2,\dots,\phi_d)$ can be rewritten as

$$\begin{aligned}P(\nu_0;\,&\phi_1,\phi_2,\dots,\phi_d)\\ &= P(1^{\nu_0})\cdot P(\nu_{i^*};\,\phi_1,\dots,\phi_{i^*}-1,\dots,\phi_d)\\ &\quad -\sum_{i=1}^{d}\phi_i P(\nu_0+\nu_i;\,\phi_1,\dots,\phi_i-1,\dots,\phi_d),\end{aligned} \tag{2}$$

where $i^* = \min\{i : \phi_i \neq 0\}$ is the smallest index $i$ such that $\phi_i$ is not zero. We have rewritten the pattern $(0;\phi_1,\phi_2,\dots,\phi_i,\dots,\phi_d)$ as $(\nu_{i^*};\,\phi_1,\dots,\phi_{i^*}-1,\dots,\phi_d)$ by taking one symbol that appears $\nu_{i^*}$ times to be the pivot symbol.

The *profile computational graph* for calculating $P(\bar{\psi})$ is a directed graph, consisting of profile forms $(\nu_0;\phi_1,\phi_2,\dots,\phi_d)$ representing patterns appearing in the recursive calculation, and the (outgoing) neighbors of any profile form are those appearing on the right-hand side of Equation (2).

For all patterns $(\nu_0;\,\phi_1,\phi_2,\dots,\phi_d)$ in the profile computational graph, the multiplicity $\nu_0$ is the sum of a subset of multiplicities of $\bar{\psi}$. Thus it can be written as $\nu_0=\sum_{i=1}^d c_i\nu_i$, where $c_i$ is the number of multiplicities $\nu_i$ added to $\nu_0$. Note that $c_i+\phi_i\le\varphi_i$. Thus a profile form corresponds to writing each $\varphi_i$ as a sum of three nonnegative integers: $\varphi_i=c_i+\varphi_i+\ell_i$. It follows that the number of profile forms is at most $\prod_{i=1}^d\binom{\varphi_i+2}{2}$.

On the other hand, since $\nu_0\le n$ and $0\le\phi_i\le\varphi_i$, the number of profile forms can also be bounded by $n\prod_{i=1}^d(\varphi_i+1)$. Thus the number of profile forms is at most $\min\left\{n\prod_{i=1}^d(\varphi_i+1),\prod_{i=1}^d\binom{\varphi_i+2}{2}\right\}$.

Since each recursive step requires at most $O(d)$ calculations, the overall computational time is $O\left(d\min\left\{n\prod_{i=1}^d(\varphi_i+1),\prod_{i=1}^d\binom{\varphi_i+2}{2}\right\}\right)$, which can be further bounded as follows. Since $\sum_{i=1}^d\varphi_i=m$,

$$\prod_{i=1}^d\binom{\varphi_i+2}{2}\le\left(\frac{m}{2d}+1\right)^d\le 1.5^m,$$
$$\prod_{i=1}^d(\varphi_i+1)\le\left(\frac{m}{d}+1\right)^d\le 2^m.$$

Hence

$$dn\prod_{i=1}^d(\varphi_i+1)\le dn\left(\frac{m}{d}+1\right)^d=O(nm^d),$$
$$d\prod_{i=1}^d\binom{\varphi_i+2}{2}=d\prod_{i=1}^d(\varphi_i+1)\prod_{i=1}^d\left(\frac{\varphi_i+2}{2}\right)=O(m^{2d}),$$

and also

$$dn\prod_{i=1}^d(\varphi_i+1)\le nm2^m,\text{ and } d\prod_{i=1}^d\binom{\varphi_i+2}{2}\le m3^m. \blacksquare$$

**Example:** For any distribution $P$ with fixed support size the probability of pattern $1^1 2^2\cdots m^m$ can be calculated in time $O\left(m3^m\right)=O\left(3^{m(1+o(1))}\right)$.

## IV. Formulation in Power sums

The recursive algorithms using Equation (1) achieve efficiency in time at the cost of memory storage. Given pattern $\bar{\psi}=1^{\mu_1}2^{\mu_2}\cdots m^{\mu_m}$ and distribution $P=(p_1,p_2,\dots,p_k)$, a direct calculation of $P(\bar{\psi})$ as sum of sequence probabilities requires constant memory space. However, the computational time is in the order of the size of $\bar{\psi}$, i.e., $k^{\underline{m}}$.

We show that the pattern probabilities can be calculated in time $O(k\min\{n,2^m\}\log n+m^{m+1}(\log m)^4)$ by writing $P(\bar{\psi})$ as a polynomial in power sums $P(1^t)=\sum_{i=1}^k p_i^t$, $2\le t\le n$. For example, $P(112)=P(11)-P(111)$.

### A. Expansion over graphs

Let $M\overset{\text{def}}{=}\{\mu_1,\mu_2,\dots,\mu_m\}$ be the multiset of multiplicities of pattern $\bar{\psi}$. The *probability of graph* $G$ over $M$ is $P(G)\overset{\text{def}}{=}\prod_{i=1}^t P(1^{\sum_{\mu\in V_i}\mu})$, where $V_1,V_2,\dots,V_t$ are the vertex sets of components of $G$.

For example, suppose $\bar{\psi}=1^{\mu_1}2^{\mu_2}3^{\mu_3}4^{\mu_4}5^{\mu_5}$, and $G$ has edges $\{\mu_1,\mu_2\}$, $\{\mu_2,\mu_3\}$, and $\{\mu_4,\mu_5\}$. Then $G$ has two components with vertex sets $V_1=\{\mu_1,\mu_2,\mu_3\}$, and $V_2=\{\mu_4,\mu_5\}$, and $P(G)=P(1^{\mu_1+\mu_2+\mu_3})P(1^{\mu_4+\mu_5})$.

Let $\mathscr{G}_M$ be the set of all graphs over $M$. Let $\operatorname{sign}(G)\overset{\text{def}}{=}(-1)^{|E(G)|}$, i.e., $\operatorname{sign}(G)$ is 1 if $G$ has even number of edges; otherwise $\operatorname{sign}(G)=-1$.

**Theorem 2.** *For any pattern $\bar{\psi}$ and distribution $P$,*

$$P(\bar{\psi})=\sum_{G\in\mathscr{G}_M}\mathit{sign}(G)P(G). \tag{3}$$

*Proof:* Let $\bar{\psi}=1^{\mu_1}2^{\mu_2}\cdots m^{\mu_m}$, and define

$$U\overset{\text{def}}{=}\{x_1^{\mu_1}x_2^{\mu_2}\cdots x_m^{\mu_m}: x_i\in A\text{ for all }i\in[m]\},$$

the set of sequences consisting of runs of lengths $\mu_1,\mu_2,\dots,\mu_m$, where $A$ is the alphabet set of distribution $P$, and $x_i$'s are not necessarily distinct.

Let $\binom{[m]}{2}$ be the set of all 2-element subsets of $[m]$. For any pair $\{i,j\}\in\binom{[m]}{2}$, let

$$S_{i,j}\overset{\text{def}}{=}\{x_1^{\mu_1}x_2^{\mu_2}\cdots x_m^{\mu_m}\in U: x_i=x_j\}.$$

Then

$$\bar{\psi}=\bigcap_{\{i,j\}\in\binom{[m]}{2}}(U\setminus S_{i,j})=U\setminus\left(\bigcup_{\{i,j\}\in\binom{[m]}{2}}S_{i,j}\right).$$

Thus $P(\bar{\psi}) = P(U) - P\left(\cup_{\{i,j\}\in\binom{[m]}{2}} S_{i,j}\right)$. Using the inclusion exclusion principle, we get

$$P(\bar{\psi}) = P(U) + \sum_{I\subseteq\binom{[m]}{2}, I\neq\emptyset} (-1)^{|I|} P\left(\underset{\{i,j\}\in I}{\cap} S_{i,j}\right). \quad (4)$$

Observe that any subset $I \subseteq \binom{[m]}{2}$ uniquely determines a graph $G_I$ over $M$ with edge set $I$. For $I \neq \emptyset$,

$$P\left(\underset{\{i,j\}\in I}{\cap} S_{i,j}\right) = P(G_I), \text{ and } P(U) = P(G_{\emptyset}).$$

It follows from Equation (4) that $P(\bar{\psi})$ can be written as $\sum_{G\in\mathscr{G}_M} \text{sign}(G)P(G)$. ■

**Example**: Consider pattern $1^{\mu_1}2^{\mu_2}3^{\mu_3}$. There are 8 graphs over the multiset $M = \{\mu_1, \mu_2, \mu_3\}$: the empty graph, three graphs with one edge, three graphs with two edges, and the complete graph. Then

$$\begin{aligned}
&P(1^{\mu_1}2^{\mu_2}3^{\mu_3})\\
&= P(1^{\mu_1})P(1^{\mu_2})P(1^{\mu_3}) - P(1^{\mu_2+\mu_3})P(1^{\mu_1})\\
&\quad - P(1^{\mu_1+\mu_3})P(1^{\mu_2}) - P(1^{\mu_1+\mu_2})P(1^{\mu_3})\\
&\quad + 3P(1^{\mu_1+\mu_2+\mu_3}) - P(1^{\mu_1+\mu_2+\mu_3}).
\end{aligned}$$

A direct application of Equation (3) for calculating $P(\bar{\psi})$ requires time

$$O\left(k \min\{n, 2^m\} \log n + m \cdot 2^{\binom{m}{2}}\right).$$

For example, under any distribution with fixed support size the probability of pattern $1^1 2^2 \cdots m^m$ can be calculated in time $O\left(m2^{m(m-1)/2}\right) = O\left(\sqrt{2}^{\,m^2(1+o(1))}\right)$.

### B. Expansion over partitions

Note that, for any graph $G \in \mathscr{G}_M$, the vertex sets of components form a partition of $M$, which is sufficient for calculating $P(G)$. The *probability of a partition* $\mathbb{P}$ of $M$, denoted $P(\mathbb{P})$, is the probability of any graph whose vertex sets of components is $\mathbb{P}$, namely $\prod_{C\in\mathbb{P}} P(1^{\sum_{\mu\in C}\mu})$.

Thus in Equation (3) of Theorem 2, we may combine terms over graphs whose vertex sets of components form the same partition. Let $\mathscr{P}_M$ be the set of all partitions of $M$.

**Theorem 3.**

$$P(\bar{\psi}) = \sum_{\mathbb{P}\in\mathscr{P}_M} (-1)^{m-|\mathbb{P}|} P(\mathbb{P}) \prod_{C_i\in\mathbb{P}} (|C_i| - 1)!, \quad (5)$$

To prove Theorem 3, we first show the following lemma, which is of its own interest, shows that the number of connected $n$-graphs with an even number of edges, minus the number of connected $n$-graphs with an odd number of edges is $(-1)^n(n-1)!$.

Let $C_n$ denote the number of connected graphs on $n$ vertices, and let $C_n^{\text{e}}$ and $C_n^{\text{o}}$ denote the number of connected graphs on $n$ vertices with even and odd number of edges respectively. While a closed-form formula for

$$C_n = C_n^{\text{e}} + c_n^{\text{o}}$$

is unknown, the next lemma shows that

$$C_n^{\text{e}} - c_n^{\text{o}} = (-1)^{n-1}(n-1)!.$$

In a recent private communication, Philippe Flajolet showed that this result can also be derived using generating functions related to ones used to enumerate trees with any given number of inversions. The lemma therefore also provides a combinatorial proof for the number of trees with zero inversions [7].

Let $\mathscr{G}_n$ be the set of all connected graphs over vertex set $[n]$. Then

**Lemma 4.**

$$\sum_{G\in\mathscr{G}_n} \textit{sign}(G) = (-1)^{n-1}(n-1)!.$$

*Proof of Lemma 4:* Let $f(n) \overset{\text{def}}{=} \sum_{G\in\mathscr{G}_n} \text{sign}(G)$. For any $G \in \mathscr{G}_{n+1}$, a connected graph of $n+1$ vertices, let $G_1, G_2, \ldots, G_m$ be the connected subgraphs after removing vertex $n+1$ from $G$. Let $\varphi_\mu$ be the number of $G_i$'s of size $\mu$. Given $\varphi_\mu$'s, there are

$$\frac{n!}{\prod_\mu (\mu!)^{\varphi_\mu}\varphi_\mu!}$$

partitions of $[n]$ having $\varphi_\mu$ parts of size $\mu$ for all $\mu$. Furthermore, note that, for all connected graphs of $n+1$ vertices having the same $G_i$'s, their signs cancel, except for those with only one edge from vertex $n+1$ to each $G_i$. Thus

$$f(n+1) = \sum_{\varphi:|\varphi|=n} \frac{n!}{\prod_\mu (\mu!)^{\varphi_\mu}\varphi_\mu!} \prod_\mu (-1)^{\varphi_\mu} \prod_\mu f(\mu)^{\varphi_\mu},$$

where the negative sign comes from a single edge from $n+1$ to each $G_i$. Note that $\prod_\mu (-1)^{\varphi_\mu + (\mu-1)\varphi_\mu} = (-1)^n$. Using induction, it's sufficient to show that

$$\sum_{\varphi:|\varphi|=n} \frac{n!}{\prod_\mu (\mu!)^{\varphi_\mu}\varphi_\mu!} \cdot \prod_\mu [(\mu-1)!]^{\varphi_\mu} = n!,$$

which is true since each term in the left-hand side is the number of permutations having $\varphi_\mu$ cycles of size $\mu$ for all $\mu$. ■

We use Lemma 4 to prove Theorem 3.

*Proof of Theorem 3:* For any graph $G \in \mathscr{G}_M$, let $\mathbb{P}_G$ be the collection of vertex sets of components. it's easy to verify that, for any partition $\mathbb{P} \in \mathscr{P}_M$,

$$\sum_{G\in\mathscr{G}_M:\mathbb{P}_G=\mathbb{P}} \text{sign}(G) = \prod_{C_i\in\mathbb{P}} \sum_{G\in\mathscr{G}_{m_i}} \text{sign}(G),$$

where $m_i \overset{\text{def}}{=} |C_i|$. By Theorem 2,

$$P(\bar{\psi}) = \sum_{\mathbb{P} \in \mathscr{P}_M} P(\mathbb{P}) \prod_{C_i \in \mathbb{P}} \sum_{G \in \mathscr{G}_{m_i}} \text{sign}(G).$$

By Lemma 4, $\sum_{G \in \mathscr{G}_{m_i}} \text{sign}(G) = (-1)^{m_i - 1}(m_i - 1)!$. Then

$$\begin{aligned} P(\bar{\psi}) &= \sum_{\mathbb{P} \in \mathscr{P}_M} P(\mathbb{P}) \prod_{C_i \in \mathbb{P}} (-1)^{m_i - 1}(m_i - 1)! \\ &= \sum_{\mathbb{P} \in \mathscr{P}_M} (-1)^{m - |\mathbb{P}|} P(\mathbb{P}) \prod_{C_i \in \mathbb{P}} (m_i - 1)!. \quad \blacksquare \end{aligned}$$

To evaluate $P(\bar{\psi})$ using Equation (5), note that $P(\mathbb{P}) = \prod_{C \in \mathbb{P}} P(1^{\sum_{\mu \in C} \mu})$. Evaluating $P(1^{\sum_{\mu \in C} \mu})$ for all $C \subseteq M$ can be done in time $O(k \min\{n, 2^m\} \log n)$. Since there are at most $m^m$ partitions in $\mathscr{P}_M$, the sum in Equation (5) has at most $m^m$ terms.

We can show that, using the Schönhage-Strassen algorithm [8], the product of $m$ integers can be computed in time $O((N + m)(\log N)^2 \log m)$, where $N$ is the number of digits in the product. Since $\prod_{C_i \in \mathbb{P}} (m_i - 1)!$ has at most $m \log m$ digits, it can be computed in time $O(m(\log m)^4)$. Hence $P(\bar{\psi})$ can be computed in time

$$O\big(k \min\{n, 2^m\} \log n + m^{m+1}(\log m)^4\big).$$

For example, under any distribution with fixed support size the probability of pattern $1^1 2^2 \cdots m^m$ can be calculated in time $O(m^{m+1}(\log m)^4) = O(m^{m(1+o(1))})$.

### C. Expansion over multi-profiles

Further improvement can be obtained by considering profiles. Note that any partition $\mathbb{P} \in \mathscr{P}_M$ *induces* a partition of the pattern $\bar{\psi}$ into shorter patterns. Let the *multi-profile* of a partition $\mathbb{P} \in \mathscr{P}_M$ be the multiset of profiles of the shorter patterns induced by $\mathbb{P}$.

For example, for $\bar{\psi} = 1^5 2^3 3^3 4^3$, $\mu_1 = 5$ and $\mu_2 = \mu_3 = \mu_4 = 3$. The partition $\mathbb{P} = \{\{\mu_1, \mu_2\}, \{\mu_3, \mu_4\}\}$ induces a partition of $\bar{\psi}$ into two shorter patterns $1^{\mu_1} 2^{\mu_2} = 1^5 2^3$ and $1^{\mu_3} 2^{\mu_4} = 1^3 2^3$ with profiles $3^1 5^1$ and $3^2$. Then the multi-profile of $\mathbb{P}$ is the multiset $\{3^1 5^1, 3^2\}$, namely one part has one 5 and one 3 and the other part has two 3's.

It's easy to see that two partitions with the same multi-profile have the same probability. In the previous example, let $\mathbb{P}' = \{\{\mu_1, \mu_3\}, \{\mu_2, \mu_4\}\}$. Then $\mathbb{P}$ and $\mathbb{P}'$ have the same multi-profile, and hence they have the same probability $P(1^{5+3})P(1^{3+3})$.

Let $\mathbb{P}_{\boldsymbol{\varphi}}$ be the set of all partitions of $M$ having multi-profile $\boldsymbol{\varphi}$. The number of partitions with the same multi-profile $\boldsymbol{\varphi}$, namely $|\mathbb{P}_{\boldsymbol{\varphi}}|$, can be calculated as follows. Let $\varphi^1, \varphi^2, \ldots$ be the distinct profiles in $\boldsymbol{\varphi}$, let $d_i$ be the number of $\varphi^i$'s in $\boldsymbol{\varphi}$, and let $\varphi^i_\mu$ be the prevalence of $\mu$ in $\varphi^i$. Then $|\mathbb{P}_{\boldsymbol{\varphi}}| = \prod_\mu \binom{\varphi_\mu}{\varphi^1_\mu, \varphi^2_\mu, \ldots} / \prod_i d_i!$.

The *probability of multi-profile* $\boldsymbol{\varphi}$, denoted $P(\boldsymbol{\varphi})$, is the probability of any partition having multi-profile $\boldsymbol{\varphi}$. Equation (5) in Theorem 3 can be rewritten by grouping terms over partitions having the same multi-profile.

Let $\Phi$ be the set of distinct multi-profiles of partitions in $\mathscr{P}_M$, and for any profile $\varphi$ let $|\varphi| \overset{\text{def}}{=} \{\mu : \varphi_\mu > 0\}$.

**Corollary 5.**

$$\begin{aligned} P(\bar{\psi}) = \sum_{\boldsymbol{\varphi} \in \Phi} &(-1)^{m - |\boldsymbol{\varphi}|} P(\boldsymbol{\varphi}) \\ &\cdot \prod_{\varphi \in \boldsymbol{\varphi}} (|\varphi| - 1) \cdot \frac{\prod_\mu \binom{\varphi_\mu}{\varphi^1_\mu, \varphi^2_\mu, \ldots}}{\prod_i d_i!}. \end{aligned} \tag{6}$$

Using Corollary 5 $P(\bar{\psi})$ can be calculated in time

$$O\left(k \min\{n, 2^m\} \log n + m(\log m)^4 |\Phi|\right).$$

For patterns with all distinct multiplicities, Equation (6) is the same as Equation (5). For patterns with few distinct multiplicities, Equation (6) achieves better complexity. For example, for uniform patterns, $|\Phi|$ is the same as the number of partitions of the number $m$, which is asymptotically bounded by $e^{\pi\sqrt{m}}$. Under any distribution with fixed support size, the probability of uniform patterns can be computed in time $O(m(\log m)^4 e^{\pi\sqrt{m}}) = O(e^{\pi\sqrt{m}(1+o(1))})$.

## References

[1] A. Orlitsky, N. Santhanam, and J. Zhang, "Universal compression of memoryless sources over unknown alphabets," *IEEE Transactions on Information Theory*, vol. 50, no. 7, pp. 1469– 1481, July 2004.

[2] 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.

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

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

[5] A. Orlitsky, N. Santhanam, K. Viswanathan, and J. Zhang, "Pattern maximum likelihood: existence and properties," In preparation, 2009.

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

[7] C. L. Mallows and J. Riordan, "The inversion enumerator for labeled trees," *Bull. Amer. Math. Soc.*, vol. 74, no. 1, 1968.

[8] A. Schönhage and V. Strassen, "Schnelle multiplikation gro$\beta$er zahlen," *Computing*, vol. 7, pp. 281–292, 1971.