A General Framework for Symmetric Property Estimation

Moses Charikar   Kiran Shiragur   Aaron Sidford

NeurIPS Conference

Symmetric property estimation

Symmetric property estimation: Setup

Setup: Unknown distribution p. Given n i.i.d samples $x^n \sim p$.

Symmetric property estimation: Setup

Setup: Unknown distribution p. Given n i.i.d samples $x^n \sim p$.

Goal: Estimate $f(p)$, where $f(\cdot)$ is a symmetric property.

Symmetric property estimation: Setup

Setup: Unknown distribution p. Given n i.i.d samples $x^n \sim p$.

Goal: Estimate $f(p)$, where $f(\cdot)$ is a symmetric property.

Symmetric Properties: invariant to label permutation.

Symmetric property estimation: Setup

Setup: Unknown distribution p. Given n i.i.d samples $x^n \sim p$.

Goal: Estimate $f(p)$, where $f(\cdot)$ is a symmetric property.

Symmetric Properties: invariant to label permutation.

$f(\cdot)$	**Optimal samples** n	**References**
Support	$\frac{N}{\log N} \log^2 \frac{1}{\varepsilon}$	(**WY15**)
Support coverage	$\frac{m}{\log m} \log \frac{1}{\varepsilon}$	(**OSW16**)
Entropy	$\frac{N}{\log N} \frac{1}{\varepsilon}$	(**VV11a**; **WY16**; **JVHW15**)
Dist. to uniform	$\frac{N}{\log N} \frac{1}{\varepsilon^2}$	(**JHW16**; **VV11b**)

Symmetric property estimation: Setup

Setup: Unknown distribution p. Given n i.i.d samples $x^n \sim p$.

Goal: Estimate $f(p)$, where $f(\cdot)$ is a symmetric property.

Symmetric Properties: invariant to label permutation.

$f(\cdot)$	**Optimal samples** n	**References**
Support	$\frac{N}{\log N} \log^2 \frac{1}{\varepsilon}$	(**WY15**)
Support coverage	$\frac{m}{\log m} \log \frac{1}{\varepsilon}$	(**OSW16**)
Entropy	$\frac{N}{\log N} \frac{1}{\varepsilon}$	(**VV11a**; **WY16**; **JVHW15**)
Dist. to uniform	$\frac{N}{\log N} \frac{1}{\varepsilon^2}$	(**JHW16**; **VV11b**)

Universal estimator: one algorithm that is sample competitive for all symmetric properties?

Profile Maximum Likelihood

Definition of a profile

- **Domain**: $\mathcal{X} = \{a, b, c\}$

Definition of a profile

- **Domain**: $\mathcal{X} = \{a, b, c\}$
- **You receive** $n(= 3)$ **i.i.d symbols**: $x^n = aba \quad (x^n \in \mathcal{X}^n)$

Definition of a profile

- **Domain**: $\mathcal{X} = \{a, b, c\}$
- **You receive** $n (= 3)$ **i.i.d symbols**: $x^n = aba \quad (x^n \in \mathcal{X}^n)$
- **Profile** of x^n: $\phi = \Phi(x^n) = \{2, 1\}$ also: pattern/histogram of histogram/fingerprint

Definition of a profile

- **Domain**: $\mathcal{X} = \{a, b, c\}$
- **You receive** $n(= 3)$ **i.i.d symbols**: $x^n = aba \quad (x^n \in \mathcal{X}^n)$
- **Profile** of x^n: $\phi = \Phi(x^n) = \{2, 1\}$ also: pattern/histogram of histogram/fingerprint
- For any distribution p,

Sequence probability

$\mathbf{P}(p, x^n) = \mathbf{P}(p, aba) = p_a^2 \cdot p_b$

Definition of a profile

- **Domain**: $\mathcal{X} = \{a, b, c\}$
- **You receive** $n(= 3)$ **i.i.d symbols**: $x^n = aba \quad (x^n \in \mathcal{X}^n)$
- **Profile** of x^n: $\phi = \Phi(x^n) = \{2, 1\}$ also: pattern/histogram of histogram/fingerprint
- For any distribution p,

Sequence probability

$\mathbf{P}(p, x^n) = \mathbf{P}(p, aba) = p_a^2 \cdot p_b$

Profile probability

$\mathbf{P}(p, \phi) = \sum_{\{y^n \in \mathcal{X}^n : \Phi(y^n) = \phi\}} \mathbf{P}(p, y^n)$

Definition of a profile

- **Domain**: $\mathcal{X} = \{a, b, c\}$
- **You receive $n(=3)$ i.i.d symbols**: $x^n = aba \quad (x^n \in \mathcal{X}^n)$
- **Profile** of x^n: $\phi = \Phi(x^n) = \{2, 1\}$ also: pattern/histogram of histogram/fingerprint
- For any distribution p,

Sequence probability

$\mathbf{P}(p, x^n) = \mathbf{P}(p, aba) = p_a^2 \cdot p_b$

Profile probability

$\mathbf{P}(p, \phi) = \sum_{\{y^n \in \mathcal{X}^n : \Phi(y^n) = \phi\}} \mathbf{P}(p, y^n)$

- Sequences with profile $\{2, 1\}$: $\{aab, aac, bba, bbc, cca, ccb\}$
  Permutations of aab: $\{aab, aba, baa\}$

Definition of a profile

- **Domain**: $\mathcal{X} = \{a, b, c\}$
- **You receive $n(=3)$ i.i.d symbols**: $x^n = aba \quad (x^n \in \mathcal{X}^n)$
- **Profile** of x^n: $\phi = \Phi(x^n) = \{2, 1\}$ also: pattern/histogram of histogram/fingerprint
- For any distribution p,

Sequence probability

$\mathbf{P}(p, x^n) = \mathbf{P}(p, aba) = p_a^2 \cdot p_b$

Profile probability

$\mathbf{P}(p, \phi) = \sum_{\{y^n \in \mathcal{X}^n : \Phi(y^n) = \phi\}} \mathbf{P}(p, y^n)$

- Sequences with profile $\{2, 1\}$: $\{aab, aac, bba, bbc, cca, ccb\}$
  Permutations of aab: $\{aab, aba, baa\}$
- $\mathbf{P}(p, \{2, 1\}) = 3 \times \left(p_a^2 p_b + p_a^2 p_c + p_b^2 p_a + p_b^2 p_c + p_c^2 p_a + p_c^2 p_b\right)$

Sequence v/s Profile Maximum Likelihood

Given any sequence $x^n \in \mathcal{X}^n$, $\phi = \Phi(x^n)$ be its profile.

Sequence v/s Profile Maximum Likelihood

Given any sequence $x^n \in \mathcal{X}^n$, $\phi = \Phi(x^n)$ be its profile.

Sequence maximum likelihood (SML)

$p_{sml} = \max_p \mathbf{P}(p, x^n)$

Sequence v/s Profile Maximum Likelihood

Given any sequence $x^n \in \mathcal{X}^n$, $\phi = \Phi(x^n)$ be its profile.

Sequence maximum likelihood (SML)

$p_{sml} = \max_p \mathbf{P}(p, x^n)$

Profile maximum likelihood (PML)

$p_{pml} = \max_p \mathbf{P}(p, \phi)$

Sequence v/s Profile Maximum Likelihood

Given any sequence $x^n \in \mathcal{X}^n$, $\phi = \Phi(x^n)$ be its profile.

Sequence maximum likelihood (SML)

$p_{sml} = \max_p \mathbf{P}(p, x^n)$

Profile maximum likelihood (PML)

$p_{pml} = \max_p \mathbf{P}(p, \phi)$

Introduced by (**OSVZ04**)

Sequence v/s Profile Maximum Likelihood

Given any sequence $x^n \in \mathcal{X}^n$, $\phi = \Phi(x^n)$ be its profile.

Sequence maximum likelihood (SML)

$p_{sml} = \max_p \mathbf{P}(p, x^n)$

Profile maximum likelihood (PML)

$p_{pml} = \max_p \mathbf{P}(p, \phi)$

Introduced by (**OSVZ04**)

Example 1

For a sequence: aba, profile: $\{2, 1\}$

Sequence v/s Profile Maximum Likelihood

Given any sequence $x^n \in \mathcal{X}^n$, $\phi = \Phi(x^n)$ be its profile.

Sequence maximum likelihood (SML)

$p_{sml} = \max_p \mathbf{P}(p, x^n)$

Profile maximum likelihood (PML)

$p_{pml} = \max_p \mathbf{P}(p, \phi)$

Introduced by (**OSVZ04**)

Example 1

For a sequence: aba, profile: $\{2, 1\}$

$p_{sml} = (2/3, 1/3)$

Sequence v/s Profile Maximum Likelihood

Given any sequence $x^n \in \mathcal{X}^n$, $\phi = \Phi(x^n)$ be its profile.

Sequence maximum likelihood (SML)

$p_{sml} = \max_p \mathbf{P}(p, x^n)$

Profile maximum likelihood (PML)

$p_{pml} = \max_p \mathbf{P}(p, \phi)$

Introduced by (**OSVZ04**)

Example 1

For a sequence: aba, profile: $\{2, 1\}$

$p_{sml} = (2/3, 1/3)$ $p_{pml} = (1/2, 1/2)$

Sequence v/s Profile Maximum Likelihood

Given any sequence $x^n \in \mathcal{X}^n$, $\phi = \Phi(x^n)$ be its profile.

Sequence maximum likelihood (SML)

$p_{sml} = \max_p \mathbf{P}(p, x^n)$

Profile maximum likelihood (PML)

$p_{pml} = \max_p \mathbf{P}(p, \phi)$

Introduced by (**OSVZ04**)

Example 1

For a sequence: aba, profile: $\{2, 1\}$

$p_{sml} = (2/3, 1/3)$ $p_{pml} = (1/2, 1/2)$

Example 2

For a sequence: $abca$, profile: $\{2, 1, 1\}$

Sequence v/s Profile Maximum Likelihood

Given any sequence $x^n \in \mathcal{X}^n$, $\phi = \Phi(x^n)$ be its profile.

Sequence maximum likelihood (SML)

$p_{sml} = \max_p \mathbf{P}(p, x^n)$

Profile maximum likelihood (PML)

$p_{pml} = \max_p \mathbf{P}(p, \phi)$

Introduced by (**OSVZ04**)

Example 1

For a sequence: aba, profile: $\{2, 1\}$

$p_{sml} = (2/3, 1/3)$ $p_{pml} = (1/2, 1/2)$

Example 2

For a sequence: $abca$, profile: $\{2, 1, 1\}$

$p_{sml} = (1/2, 1/4, 1/4)$

Sequence v/s Profile Maximum Likelihood

Given any sequence $x^n \in \mathcal{X}^n$, $\phi = \Phi(x^n)$ be its profile.

Sequence maximum likelihood (SML)

$p_{sml} = \max_p \mathbf{P}(p, x^n)$

Profile maximum likelihood (PML)

$p_{pml} = \max_p \mathbf{P}(p, \phi)$

Introduced by (**OSVZ04**)

Example 1

For a sequence: aba, profile: $\{2, 1\}$

$p_{sml} = (2/3, 1/3)$ $p_{pml} = (1/2, 1/2)$

Example 2

For a sequence: $abca$, profile: $\{2, 1, 1\}$

$p_{sml} = (1/2, 1/4, 1/4)$ $p_{pml} = (1/5, 1/5, 1/5, 1/5, 1/5)$

Probability of a profile computation

Probability of a profile computation

- For any distribution p, and profile $\phi = \Phi(x^n)$:

Probability of a profile computation

- For any distribution p, and profile $\phi = \Phi(x^n)$:

$$\mathbf{P}(p, \phi) \propto \mathrm{perm} Q_{p,\phi}$$

  where $Q_{p,\phi}$ (depends on p, ϕ) is a generalized Vandermonde matrix.

Probability of a profile computation

- For any distribution p, and profile $\phi = \Phi(x^n)$:

$$\mathbf{P}(p, \phi) \propto \mathrm{perm} Q_{p,\phi}$$

  where $Q_{p,\phi}$ (depends on p, ϕ) is a generalized Vandermonde matrix.

- For instance: domain $\mathbf{X} = \{a, b, c\}$ and profile $\phi = \{2, 1\}$, then:

$$Q_{p,\phi} = \begin{bmatrix} \mathbf{p}_a^2 & \mathbf{p}_a & \mathbf{p}_a^0 \\ \mathbf{p}_b^2 & \mathbf{p}_b & \mathbf{p}_b^0 \\ \mathbf{p}_c^2 & \mathbf{p}_c & \mathbf{p}_c^0 \end{bmatrix}$$

Probability of a profile computation

- For any distribution p, and profile $\phi = \Phi(x^n)$:

$$\mathbf{P}(p, \phi) \propto \mathrm{perm} Q_{p,\phi}$$

  where $Q_{p,\phi}$ (depends on p, ϕ) is a generalized Vandermonde matrix.
- For instance: domain $\mathbf{X} = \{a, b, c\}$ and profile $\phi = \{2, 1\}$, then:

$$Q_{p,\phi} = \begin{bmatrix} \mathbf{p}_a^2 & \mathbf{p}_a & \mathbf{p}_a^0 \\ \mathbf{p}_b^2 & \mathbf{p}_b & \mathbf{p}_b^0 \\ \mathbf{p}_c^2 & \mathbf{p}_c & \mathbf{p}_c^0 \end{bmatrix}$$

- (**Val79**) #-P hard in general.

Probability of a profile computation

- For any distribution p, and profile $\phi = \Phi(x^n)$:

$$\mathbf{P}(p, \phi) \propto \mathrm{perm} Q_{p,\phi}$$

  where $Q_{p,\phi}$ (depends on p, ϕ) is a generalized Vandermonde matrix.

- For instance: domain $\mathbf{X} = \{a, b, c\}$ and profile $\phi = \{2, 1\}$, then:

$$Q_{p,\phi} = \begin{bmatrix} \mathbf{p}_a^2 & \mathbf{p}_a & \mathbf{p}_a^0 \\ \mathbf{p}_b^2 & \mathbf{p}_b & \mathbf{p}_b^0 \\ \mathbf{p}_c^2 & \mathbf{p}_c & \mathbf{p}_c^0 \end{bmatrix}$$

- (**Val79**) #-P hard in general.
- (**JSV04**) Gives an efficient randomized algorithm to approximate permanent of non-negative matrices within $(1 + \epsilon)$ accuracy.

Approximate profile maximum likelihood computation

Recall, given a profile ϕ, $\max_p \operatorname{perm} Q_{p,\phi}$.

$$\max_{p_a+p_b+p_c=1} \operatorname{perm} \begin{bmatrix} \mathbf{p}_a^2 & \mathbf{p}_a & \mathbf{p}_a^0 \\ \mathbf{p}_b^2 & \mathbf{p}_b & \mathbf{p}_b^0 \\ \mathbf{p}_c^2 & \mathbf{p}_c & \mathbf{p}_c^0 \end{bmatrix}$$

Approximate profile maximum likelihood computation

Recall, given a profile ϕ, $\max_p \operatorname{perm} Q_{p,\phi}$.

Approximate profile maximum likelihood (APML)

A distribution p^{β}_{pml} is a β-APML if,

$$\mathbf{P}(p^{\beta}_{pml}, \phi) \geq \beta \max_{p} \mathbf{P}(p, \phi)$$

Approximate profile maximum likelihood computation

Recall, given a profile ϕ, $\max_p \mathrm{perm}\, Q_{p,\phi}$.

Approximate profile maximum likelihood (APML)

A distribution p^{β}_{pml} is a β-APML if,

$$\mathbf{P}(p^{\beta}_{pml}, \phi) \geq \beta \max_{p} \mathbf{P}(p, \phi)$$

For any profile ϕ, uniform distribution is a $\exp(-\tilde{O}(n))$-APML distribution.

Approximate profile maximum likelihood computation

Recall, given a profile ϕ, $\max_p \mathrm{perm} Q_{p,\phi}$.

Approximate profile maximum likelihood (APML)

A distribution p^{β}_{pml} is a β-APML if,

$$\mathbf{P}(p^{\beta}_{pml}, \phi) \geq \beta \max_p \mathbf{P}(p, \phi)$$

For any profile ϕ, uniform distribution is a $\exp(-\tilde{O}(n))$-APML distribution.

- **Heuristic algorithms:**
  - Expectation maximization (EM) (**OSVZ04**)
  - Bethe free energy approximation of permanent (**Von12**)
  - Dynamic programming heuristic (**PJW17**)

Approximate profile maximum likelihood computation

Recall, given a profile ϕ, $\max_p \mathrm{perm} Q_{p,\phi}$.

Approximate profile maximum likelihood (APML)

A distribution p_{pml}^{β} is a β-APML if,

$$\mathbf{P}(p_{pml}^{\beta}, \phi) \geq \beta \max_p \mathbf{P}(p, \phi)$$

For any profile ϕ, uniform distribution is a $\exp(-\tilde{O}(n))$-APML distribution.

- **Heuristic algorithms:**
  - Expectation maximization (EM) (**OSVZ04**)
  - Bethe free energy approximation of permanent (**Von12**)
  - Dynamic programming heuristic (**PJW17**)
- **No non-trivial guarantees!**

Connection between PML and symmetric property estimation

PML is a universal symmetric property estimator

Profile is a **sufficient statistic** for estimating symmetric properties.

PML is a universal symmetric property estimator

Profile is a **sufficient statistic** for estimating symmetric properties.

Theorem (ADOS16)

Let n be the optimal sample complexity for estimating symmetric property $f(\cdot)$ within accuracy ϵ, then

PML is a universal symmetric property estimator

Profile is a **sufficient statistic** for estimating symmetric properties.

Theorem (ADOS16)

Let n be the optimal sample complexity for estimating symmetric property $f(\cdot)$ within accuracy ϵ, then

$$f(p_{pml}) \overset{2\epsilon}{\approx} f(p)$$

PML is a universal symmetric property estimator

Profile is a **sufficient statistic** for estimating symmetric properties.

Theorem (ADOS16)

Let n be the optimal sample complexity for estimating symmetric property $f(\cdot)$ within accuracy ϵ, then

$$f(p_{pml}) \overset{2\epsilon}{\approx} f(p)$$

Holds when $\epsilon^2 > \frac{1}{n^{0.499}}$.

PML is a universal symmetric property estimator

Profile is a **sufficient statistic** for estimating symmetric properties.

Theorem (ADOS16)

Let n be the optimal sample complexity for estimating symmetric property $f(\cdot)$ within accuracy ϵ, then

$$f(p_{pml}) \overset{2\epsilon}{\approx} f(p)$$

Holds when $\epsilon^2 > \frac{1}{n^{0.499}}$.

Approximation suffices*:*

Any β-APML where $\beta > e^{-n^{1-\delta}}$ for some constant $\delta > 0$, we have

PML is a universal symmetric property estimator

Profile is a **sufficient statistic** for estimating symmetric properties.

Theorem (ADOS16)

Let n be the optimal sample complexity for estimating symmetric property $f(\cdot)$ within accuracy ϵ, then

$$f(p_{pml}) \overset{2\epsilon}{\approx} f(p)$$

Holds when $\epsilon^2 > \frac{1}{n^{0.499}}$.

Approximation suffices:

Any β-APML where $\beta > e^{-n^{1-\delta}}$ for some constant $\delta > 0$, we have

$$f(p_{pml}^{\beta}) \overset{2\epsilon}{\approx} f(p)$$

Holds when $\epsilon^2 > \max\left(\frac{1}{n^{0.499}}, \frac{n^{1-\delta}}{n}\right)$.

PML is a universal symmetric property estimator

Profile is a **sufficient statistic** for estimating symmetric properties.

Theorem (ADOS16)

Let n be the optimal sample complexity for estimating symmetric property $f(\cdot)$ within accuracy ϵ, then

$$f(p_{pml}) \overset{2\epsilon}{\approx} f(p)$$

Holds when $\epsilon^2 > \frac{1}{n^{0.499}}$.

Approximation suffices*:*

Any β-APML where $\beta > e^{-n^{1-\delta}}$ for some constant $\delta > 0$, we have

$$f(p^{\beta}_{pml}) \overset{2\epsilon}{\approx} f(p)$$

Holds when $\epsilon^2 > \max\left(\frac{1}{n^{0.499}}, \frac{n^{1-\delta}}{n}\right)$. Exact PML is no better than $e^{-\sqrt{n}}$-APML.

Open Question

Open Question

Can we get rid of the condition on ϵ?

Our Results

Open Question

Can we get rid of the condition on ϵ? **Psuedo PML**

Universal estimator approaches

- **Linear Programming (VV11a)**
  - Suboptimal dependence on ε.

Universal estimator approaches

- **Linear Programming (VV11a)**
  - Suboptimal dependence on ε.
- **Local Moment Matching (HJW18)**
  - Recovers the distribution in sorted order.

Universal estimator approaches

- **Linear Programming (VV11a)**
  - Suboptimal dependence on ε.
- **Local Moment Matching (HJW18)**
  - Recovers the distribution in sorted order.
- **Profile Maximum Likelihood**
  - (**ADOS16**) Broad applicability, testing (**HO19**).

Open Question: Can we get rid of the condition on ϵ?

Psuedo PML

Property specific estimator for estimating f

Empirical estimate + Sophisticated Tools

Psuedo PML

Property specific estimator for estimating f

Empirical estimate + Sophisticated Tools

Psuedo PML for estimating f

Empirical estimate + PML estimate

Psuedo PML

Property specific estimator for estimating f

Empirical estimate + Sophisticated Tools

Psuedo PML for estimating f

Empirical estimate + PML estimate

- Weakly depends on the property.

Psuedo PML

Property specific estimator for estimating f

Empirical estimate + Sophisticated Tools

Psuedo PML for estimating f

Empirical estimate + PML estimate

- Weakly depends on the property.
- Speed up in the running times (Next slide).

Psuedo PML

Property specific estimator for estimating f

Empirical estimate + Sophisticated Tools

Psuedo PML for estimating f

Empirical estimate + PML estimate

- Weakly depends on the property.
- Speed up in the running times (Next slide).
- **Independent work: Truncated PML (HO19)**.

Samples size	10^3	$5 * 10^3$	10^4	$5 * 10^4$	10^5	$5 * 10^5$	10^6	$5 * 10^6$	10^7
EmpFrac	0.18382	0.31654	0.37150	0.50457	0.56239	0.69533	0.75245	0.88554	0.94282
Speedup	0.824	1.205	1.669	3.561	4.852	9.552	13.337	12.196	10.204