Fast Approximations of the Pattern Maximum Likelihood Estimate

Pascal O. Vontobel
Department of Information Engineering
The Chinese University of Hong Kong

Talk at Simons Institute, UC Berkeley, CA, March 20, 2015

Fast Approximations of the Pattern Maximum Likelihood Estimate

Pascal O. Vontobel
Department of Information Engineering
The Chinese University of Hong Kong

Talk at Simons Institute, UC Berkeley, CA, March 20, 2015

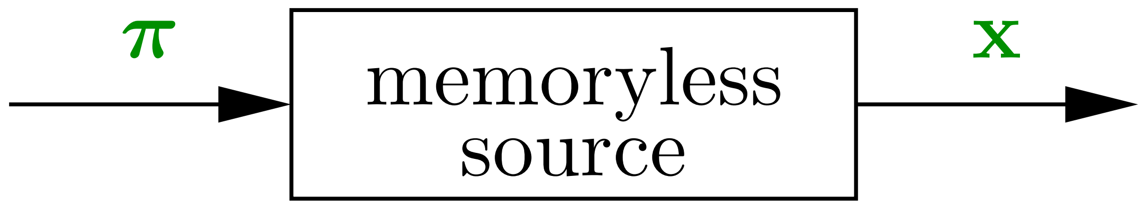
Overview

PML distribution: Pattern Maximum Likelihood distribution

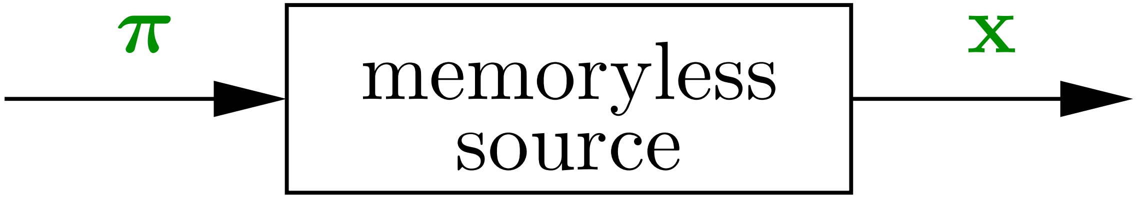
- Definition of the PML distribution
- Gibbs free energy approach to the PML distribution
- Bethe/Sinkhorn approximation to the PML distribution
- Valiant–Valiant estimate of distribution histogram
- Connections
- Conclusions / Outlook

Pattern maximum likelihood estimate

Distribution Estimate



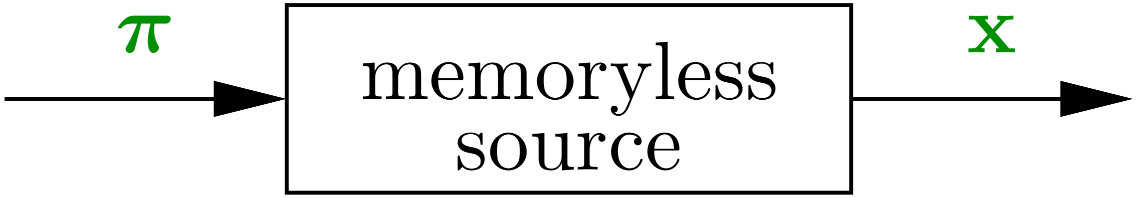
Distribution Estimate



- Memoryless source with finite alphabet $\mathcal{X}$ and distribution $\boldsymbol{\pi}$.

E.g., $\mathcal{X} = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{e}, \mathrm{f}\}.$

Distribution Estimate



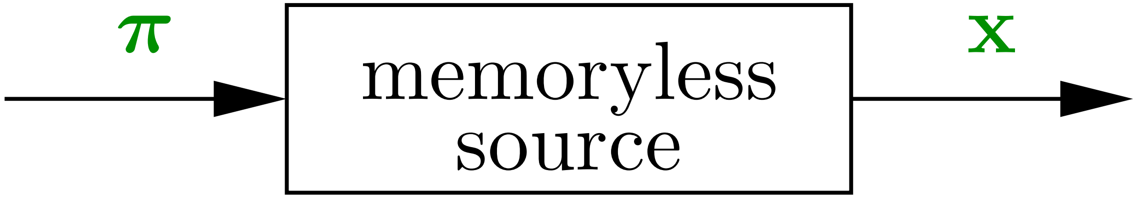
- Memoryless source with finite alphabet $\mathcal{X}$ and distribution $\boldsymbol{\pi}$.

  E.g., $$\mathcal{X} = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{e}, \mathrm{f}\}.$$

- Consider a length-n sequence $\mathbf{x}$ produced by this source.

  E.g., $$\mathbf{x} = \mathrm{c\ c\ a\ d\ c\ d\ d\ e\ a}$$

Distribution Estimate



- Memoryless source with finite alphabet $\mathcal{X}$ and distribution $\boldsymbol{\pi}$.

  E.g., $\mathcal{X} = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{e}, \mathrm{f}\}$.

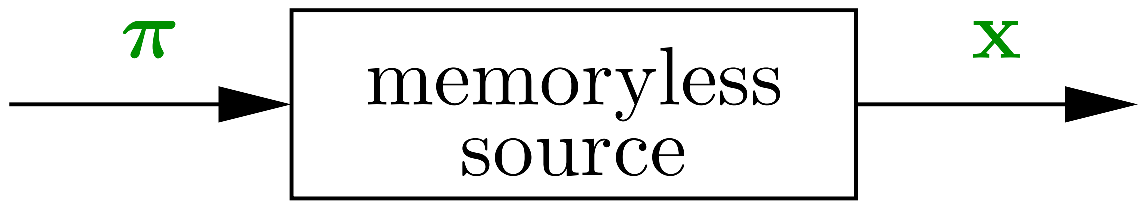
- Consider a length-n sequence $\mathbf{x}$ produced by this source.

  E.g., $\mathbf{x} = \mathrm{c\ c\ a\ d\ c\ d\ d\ e\ a}$

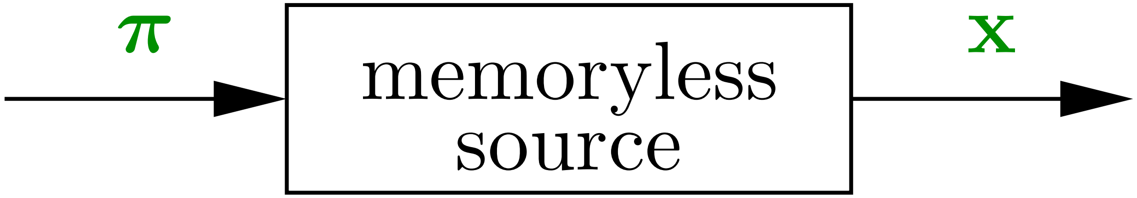
- ML estimate of distribution $\boldsymbol{\pi}$ given sequence $\mathbf{x}$:

$$\hat{\pi}_x \triangleq \frac{\left|\{\ell \,|\, x_\ell = x\}\right|}{n}, \quad x \in \mathcal{X}.$$

Estimates Based on $\boldsymbol{\pi}$



Estimates Based on $\boldsymbol{\pi}$



- Entropy estimate:

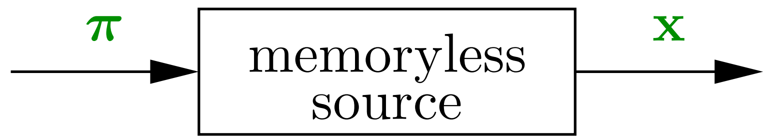
$$\widehat{H(X)} = -\sum_x \hat{\pi}_x \log(\hat{\pi}_x)$$

- Support estimate:

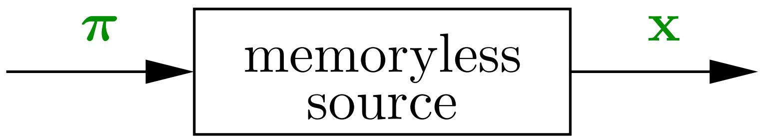
$$|\widehat{\mathrm{supp}(\boldsymbol{\pi})}| = \{x \,:\, \hat{\pi}_x > 0\}$$

- ...

Sorted Distribution

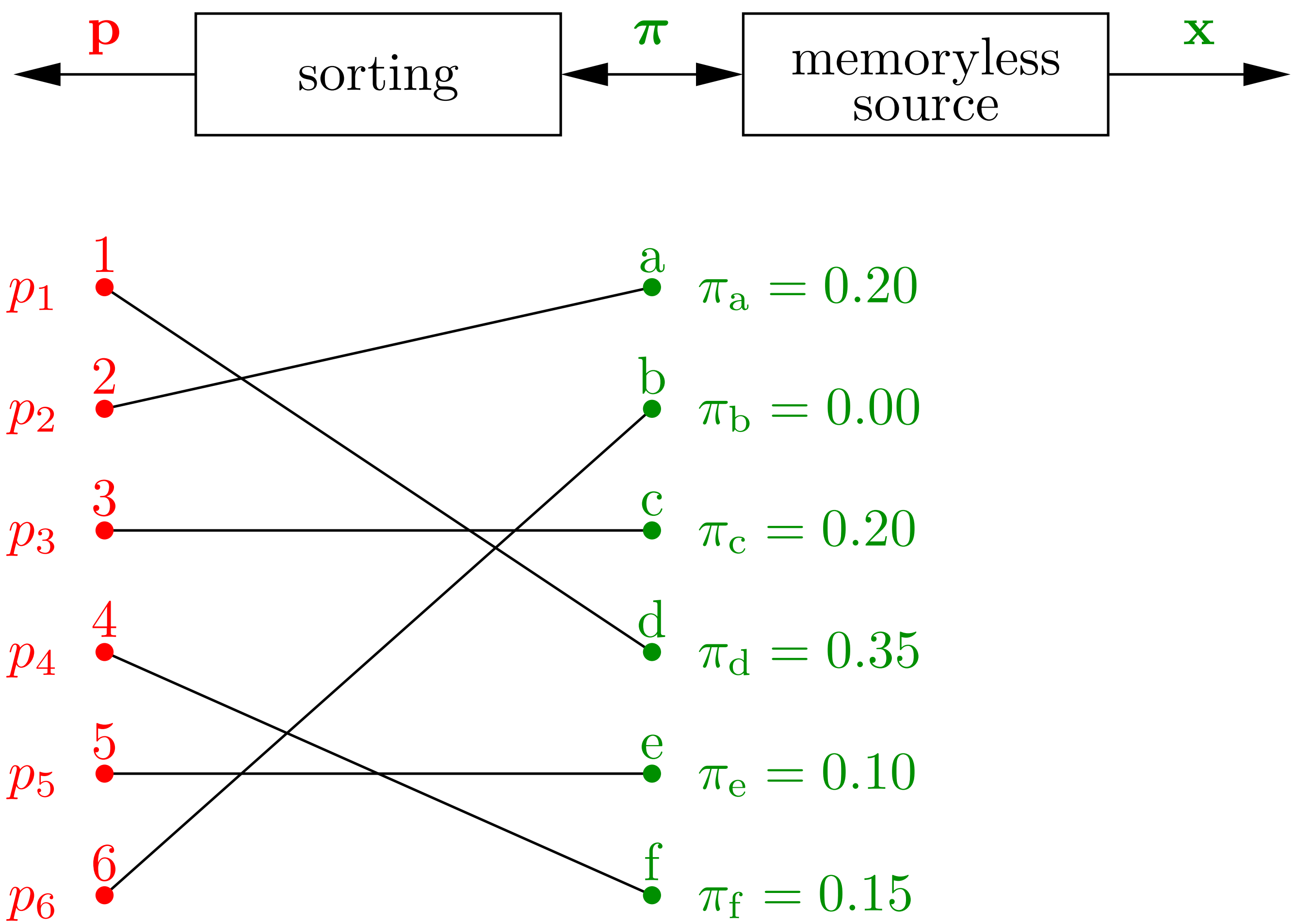


Sorted Distribution



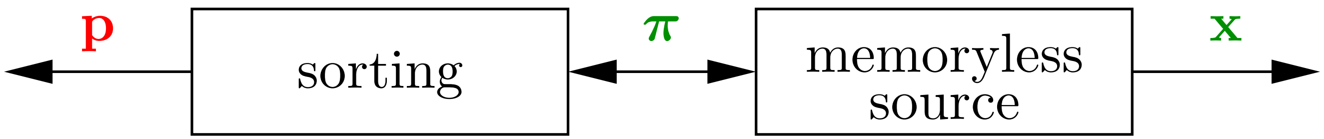
- a $\pi_a = 0.20$
- b $\pi_b = 0.00$
- c $\pi_c = 0.20$
- d $\pi_d = 0.35$
- e $\pi_e = 0.10$
- f $\pi_f = 0.15$

Sorted Distribution

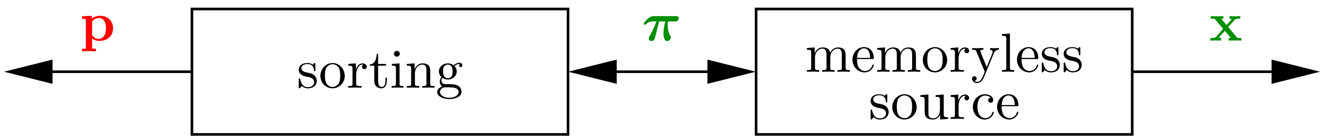


Sorted distribution $\mathbf{p}$: non-increasingly sorted version of $\boldsymbol{\pi}$.

Estimates Based on $\mathbf{p}$



Estimates Based on $\mathbf{p}$



- Entropy estimate:

$$\widehat{H(X)} = -\sum_i \hat{p}_i \log(\hat{p}_i)$$

- Support estimate:

$$|\widehat{\mathrm{supp}(\boldsymbol{\pi})}| = \{i \;:\; \hat{p}_i > 0\}$$

- ...

Pattern of a Sequence

Pattern of a Sequence

- For estimating $\mathbf{p}$ based on $\mathbf{x}$:

The pattern ψ of $\mathbf{x}$ is a sufficient statistic.

Pattern of a Sequence

- For estimating $\mathbf{p}$ based on $\mathbf{x}$:

  The pattern ψ of $\mathbf{x}$ is a sufficient statistic.

- Consider a finite-length sequence $\mathbf{x}$ produced by the source.

  E.g., $\mathbf{x} = \text{c c a d c d d e a}$

Pattern of a Sequence

- For estimating $\mathbf{p}$ based on $\mathbf{x}$:

  The pattern ψ of $\mathbf{x}$ is a sufficient statistic.

- Consider a finite-length sequence $\mathbf{x}$ produced by the source.

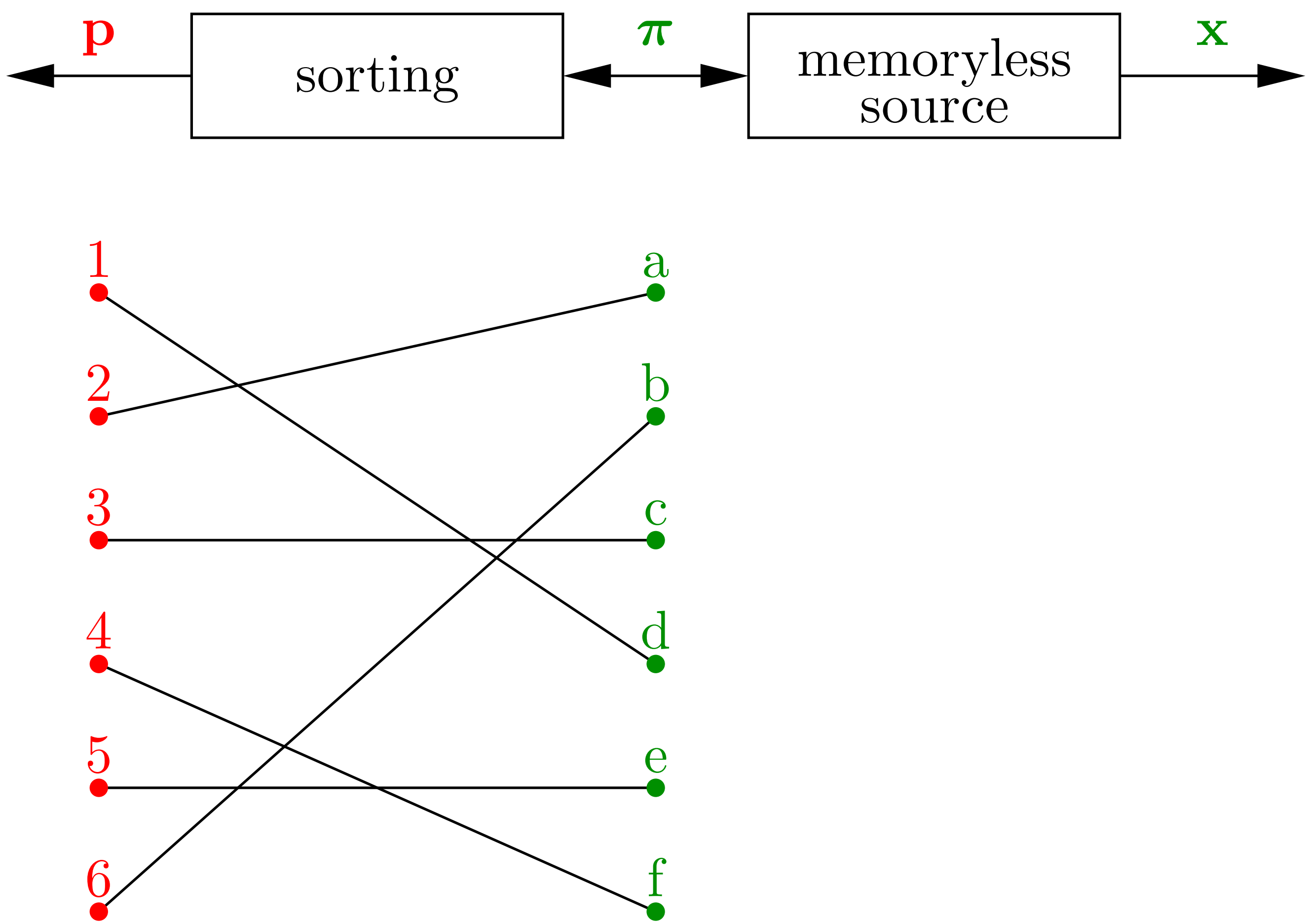
  E.g., $\mathbf{x} = \mathrm{c\ c\ a\ d\ c\ d\ d\ e\ a}$

- Pattern ψ:

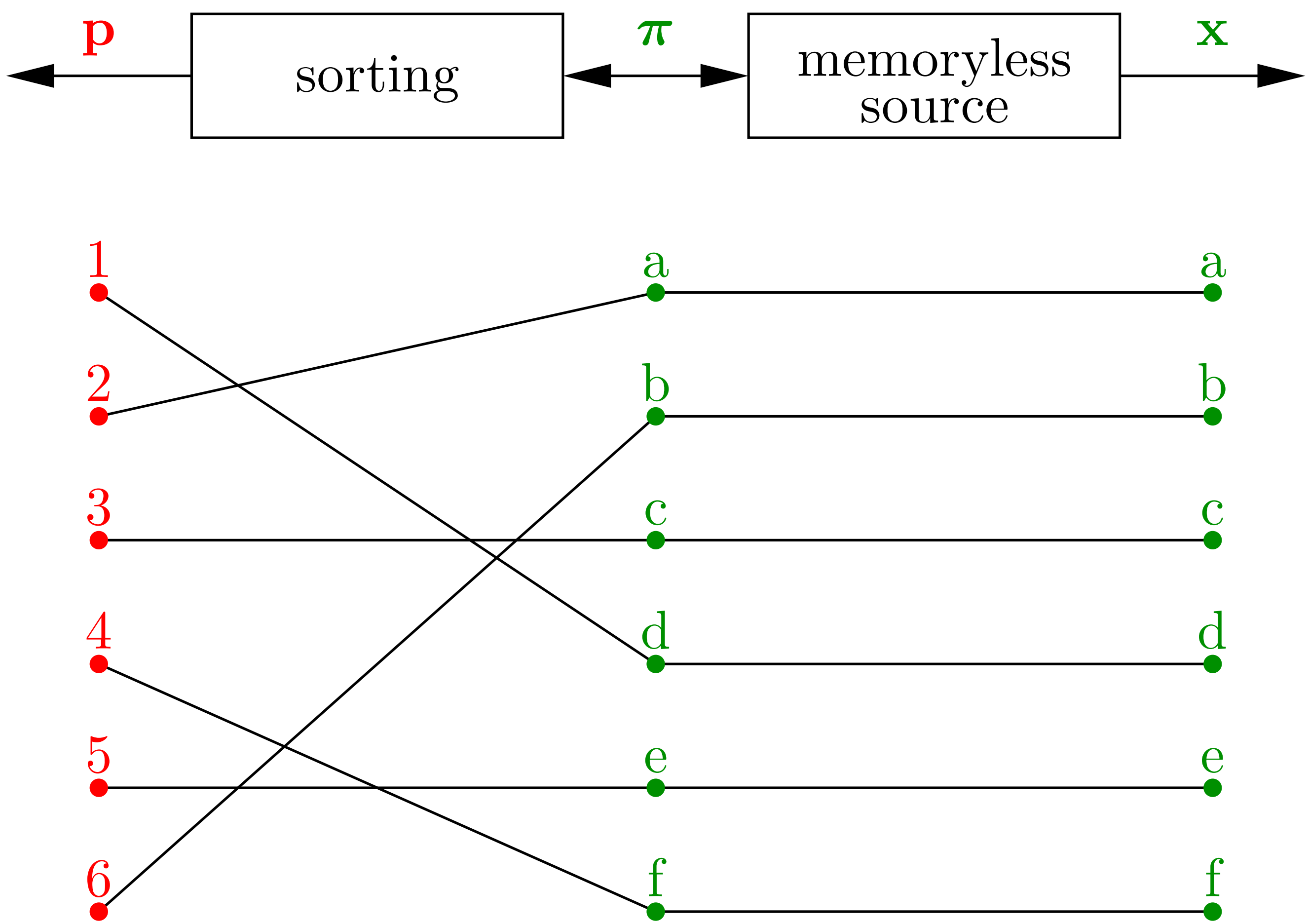
  Replaces the symbols in $\mathbf{x}$ by their order of first appearance.

  Here, $\psi = 1\ 1\ 2\ 3\ 1\ 3\ 3\ 4\ 2$

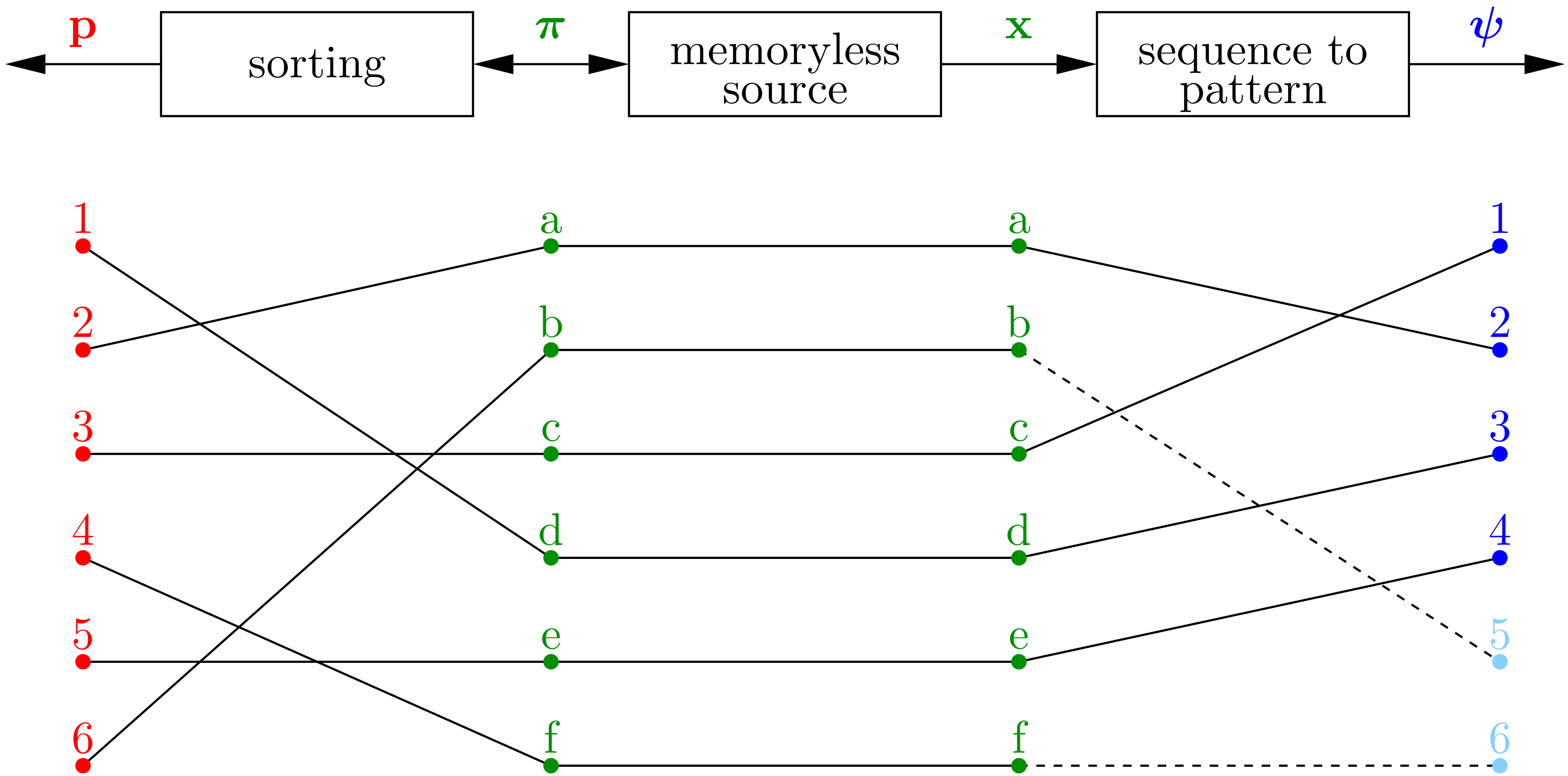
Pattern of a Sequence



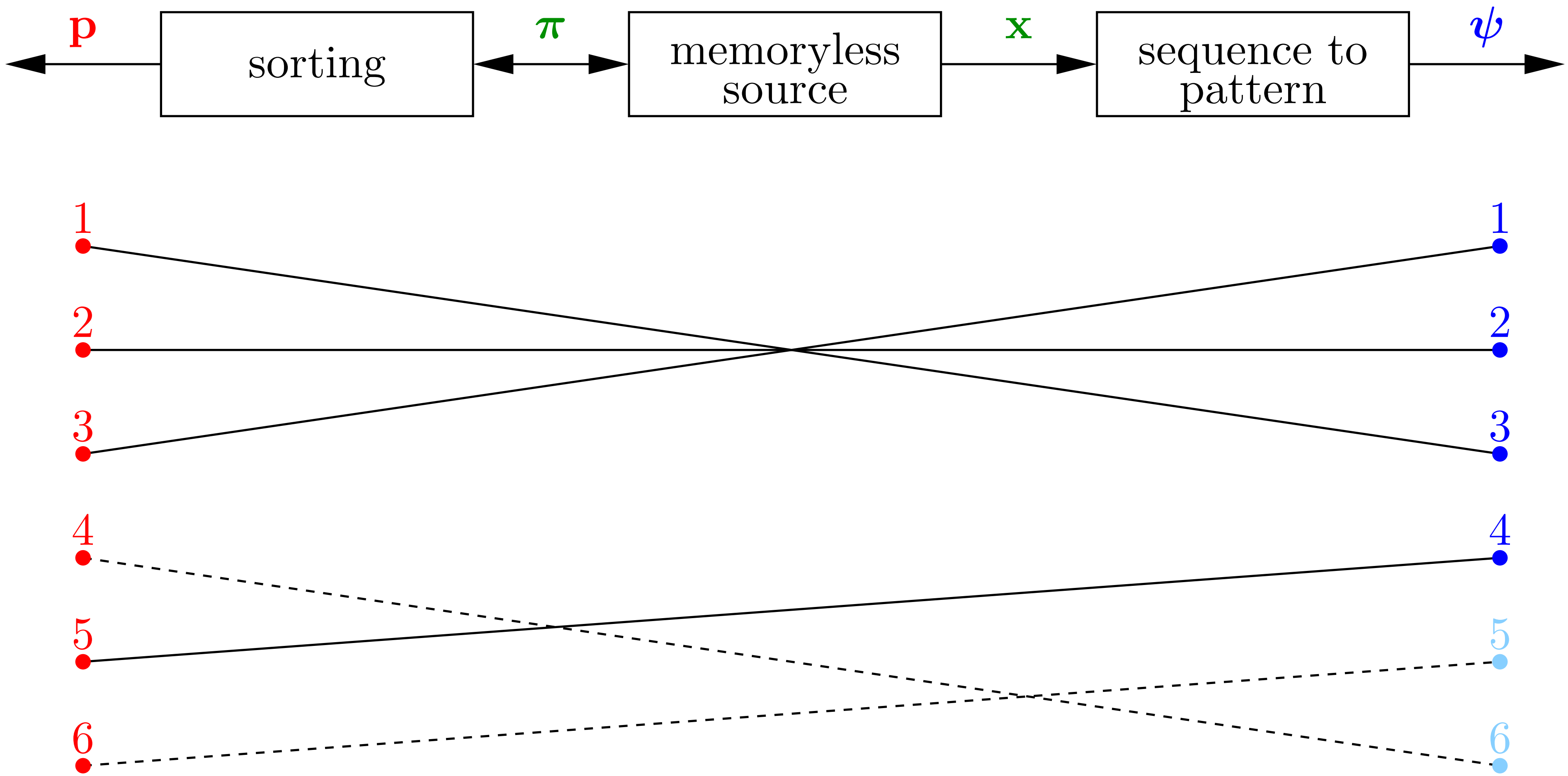
Pattern of a Sequence



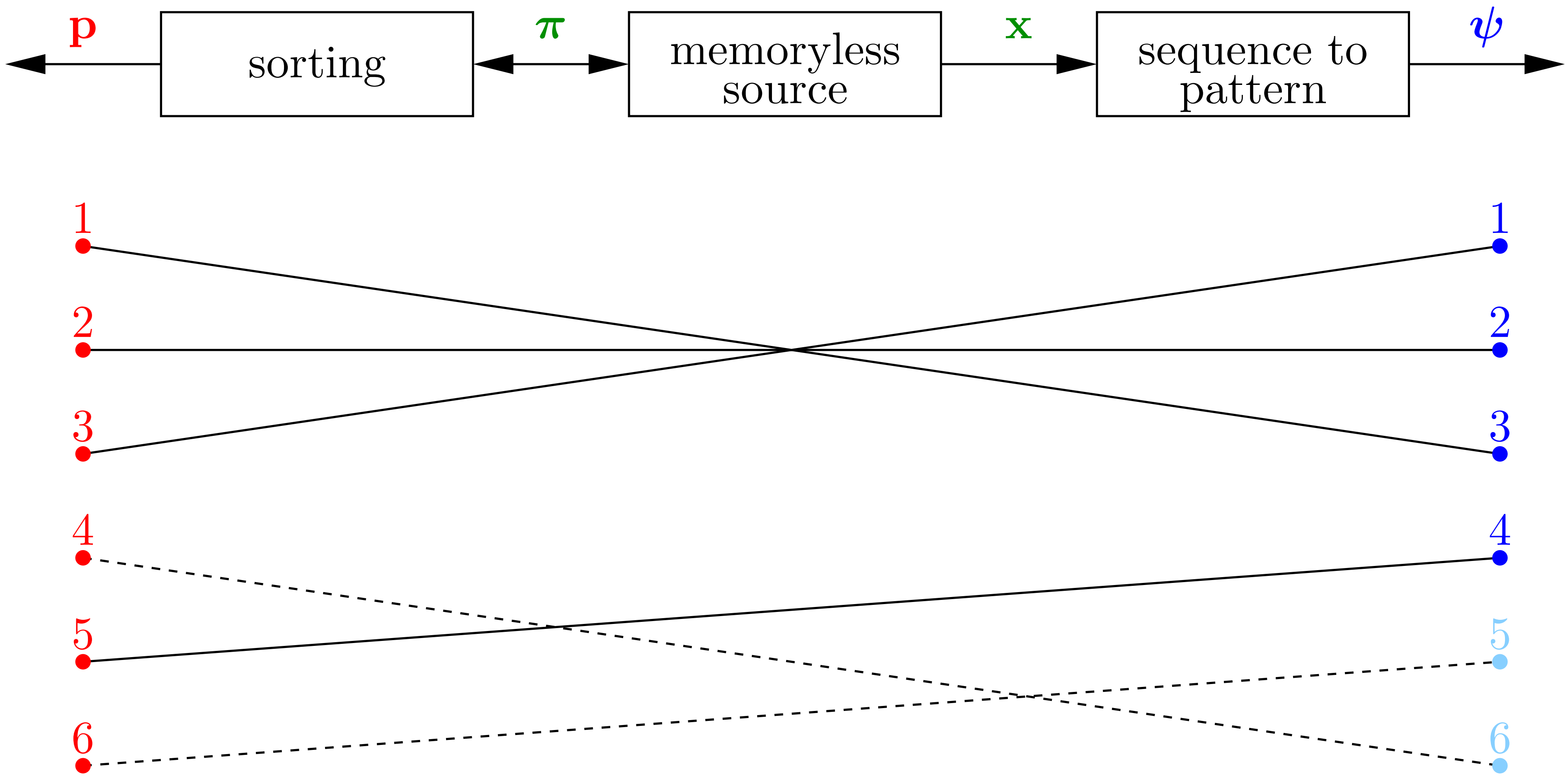
Pattern of a Sequence



Pattern of a Sequence

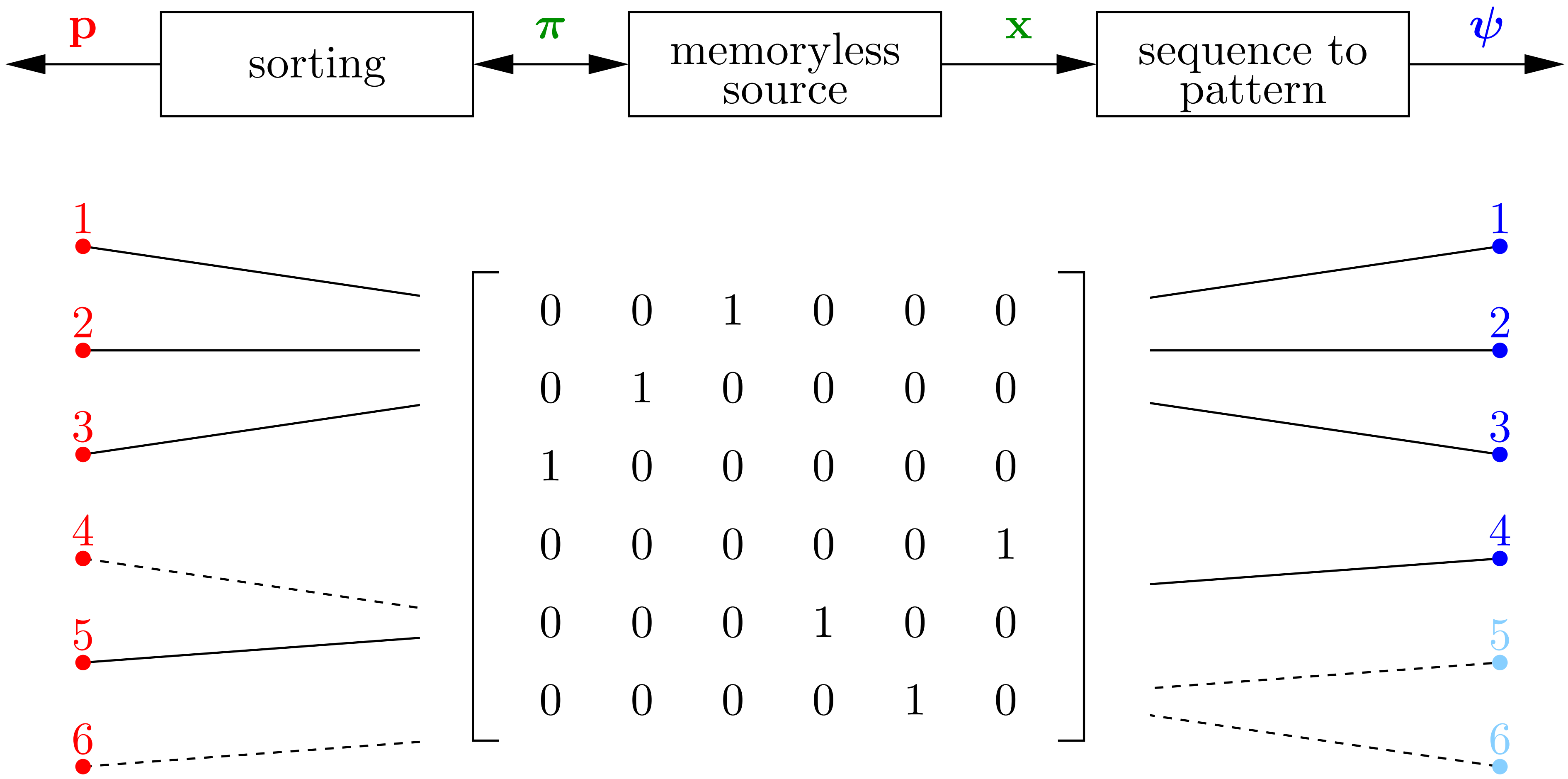


Pattern of a Sequence



Permutation σ : $\{1, \ldots, 6\} \to \{1, \ldots, 6\}$.

Pattern of a Sequence



Permutation matrix M_σ.

Pattern ML Distribution

Pattern maximum likelihood (PML) distribution (Orlitsky *et al.*):

Given ψ, what is the most likely $\mathbf{p}$?

Pattern ML Distribution

Pattern maximum likelihood (PML) distribution (Orlitsky *et al.*):

$$\mathbf{p}^{\mathrm{PML}}(\boldsymbol{\psi}) \triangleq \arg\max_{\mathbf{p}} \ P(\boldsymbol{\psi} \mid \mathbf{p}).$$

Pattern ML Distribution

- The above probability can be expressed as follows:

$$P(\boldsymbol{\psi} \mid \mathbf{p}) = \sum_{\sigma} {p_1}^{\mu_{\sigma(1)}} \, {p_2}^{\mu_{\sigma(2)}} \, \cdots \, {p_k}^{\mu_{\sigma(k)}}.$$

Pattern ML Distribution

- The above probability can be expressed as follows:

$$P(\boldsymbol{\psi} \mid \mathbf{p}) = \sum_{\sigma} {p_1}^{\mu_{\sigma(1)}} \, {p_2}^{\mu_{\sigma(2)}} \ \cdots \ {p_k}^{\mu_{\sigma(k)}}.$$

- This probability can be expressed as follows:

$$P(\boldsymbol{\psi} \mid \mathbf{p}) \ \propto \ \operatorname{perm}\big(\boldsymbol{\theta}(\mathbf{p}, \boldsymbol{\psi})\big),$$

with

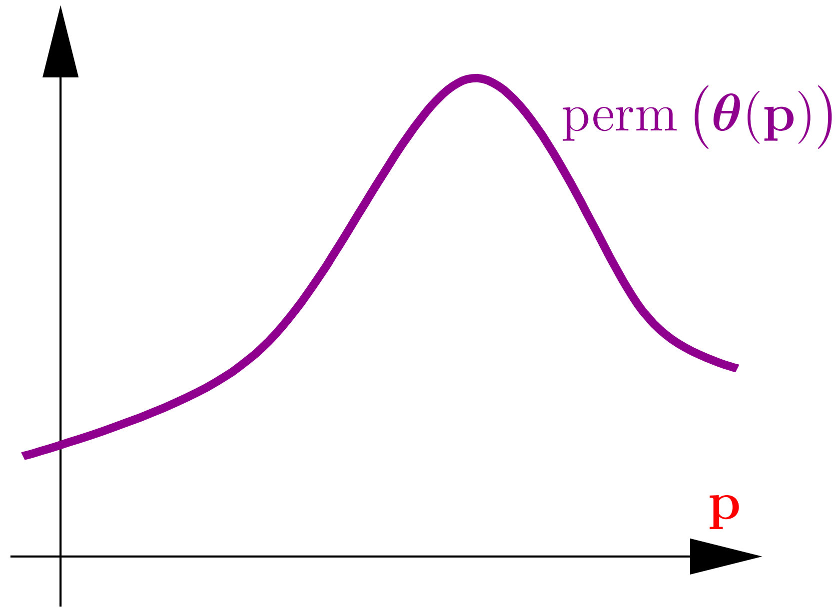
$$\boldsymbol{\theta}(\mathbf{p}, \boldsymbol{\psi}) \triangleq \begin{pmatrix} p_1^{\mu_1} & p_1^{\mu_2} & \cdots & p_1^{\mu_k} \\ p_2^{\mu_1} & p_2^{\mu_2} & \cdots & p_2^{\mu_k} \\ \vdots & \vdots & & \vdots \\ p_k^{\mu_1} & p_k^{\mu_2} & \cdots & p_k^{\mu_k} \end{pmatrix},$$

where $\boldsymbol{\mu} \triangleq \boldsymbol{\mu}(\boldsymbol{\psi})$ are the multiplicities of the integers in the pattern.

Maximizing the Permanent

Finding the PML distribution means finding the pmf $\mathbf{p}$ that maximizes

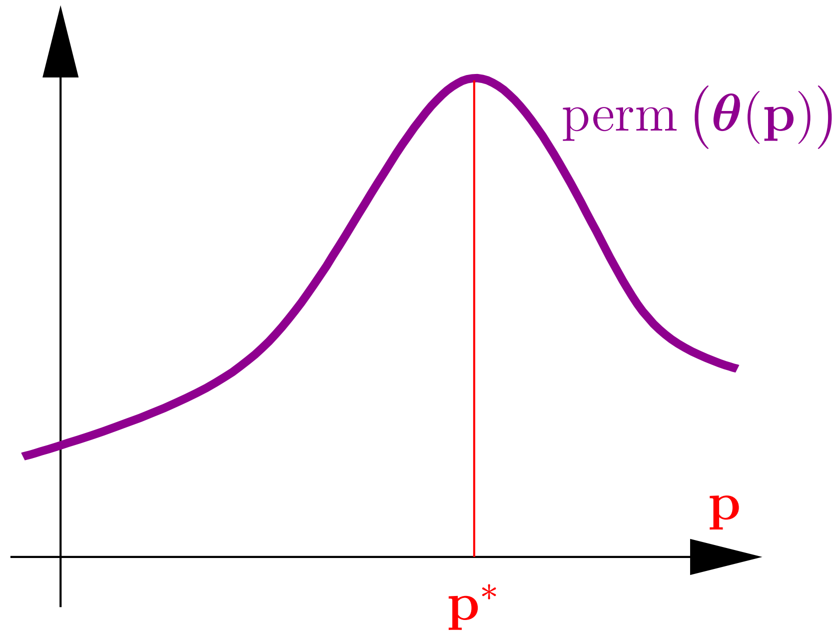
$$\mathbf{p}^* = \arg\max_{\mathbf{p}} \ \mathrm{perm}\left(\boldsymbol{\theta}(\mathbf{p})\right).$$



Maximizing the Permanent

Finding the PML distribution means finding the pmf $\mathbf{p}$ that maximizes

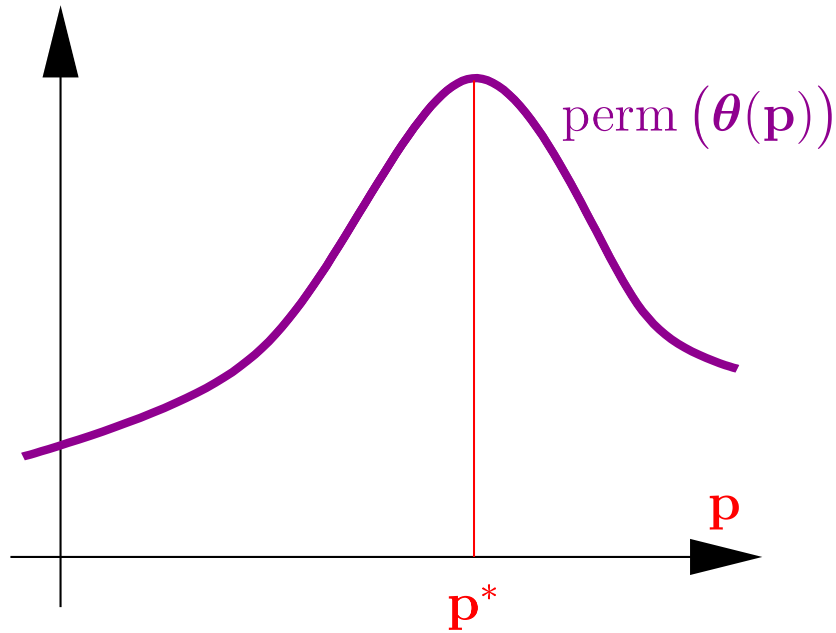
$$\mathbf{p}^* = \arg\max_{\mathbf{p}} \operatorname{perm}\big(\boldsymbol{\theta}(\mathbf{p})\big).$$



Maximizing the Permanent

Finding the PML distribution means finding the pmf $\mathbf{p}$ that maximizes

$$\mathbf{p}^* = \arg\max_{\mathbf{p}} \ \mathrm{perm}\left(\boldsymbol{\theta}(\mathbf{p})\right).$$

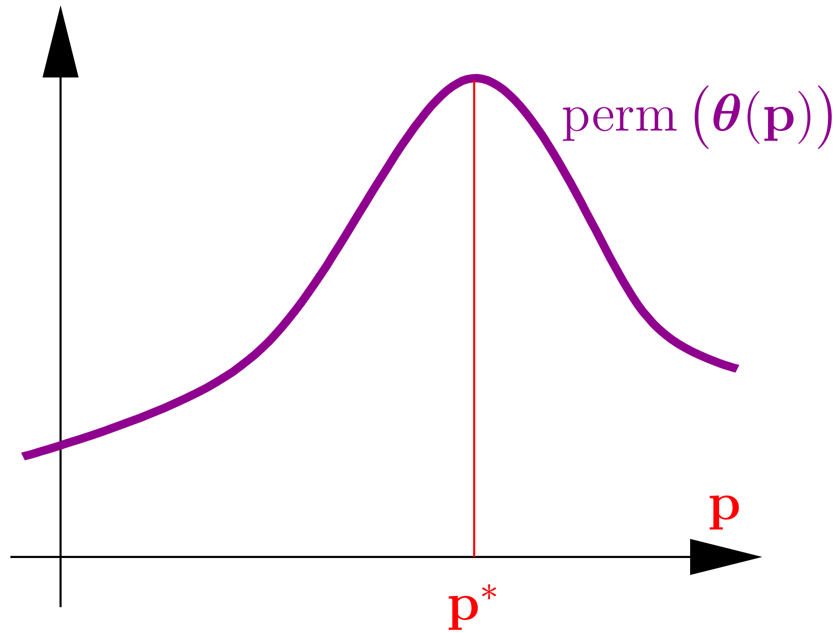


$\Rightarrow$ This problem appears intractable for practically relevant problem sizes.

Maximizing the Permanent

Finding the PML distribution means finding the pmf $\mathbf{p}$ that maximizes

$$\mathbf{p}^* = \arg\max_{\mathbf{p}} \ \mathrm{perm}\left(\boldsymbol{\theta}(\mathbf{p})\right).$$

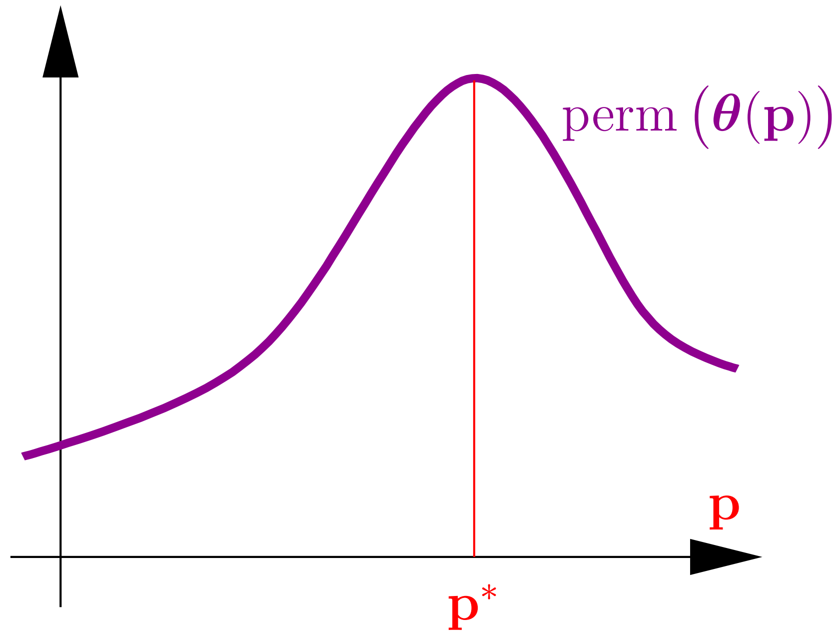


$\Rightarrow$ One needs to come up with approximate optimization algorithms:

Maximizing the Permanent

Finding the PML distribution means finding the pmf $\mathbf{p}$ that maximizes

$$\mathbf{p}^* = \arg\max_{\mathbf{p}} \ \mathrm{perm}\left(\boldsymbol{\theta}(\mathbf{p})\right).$$



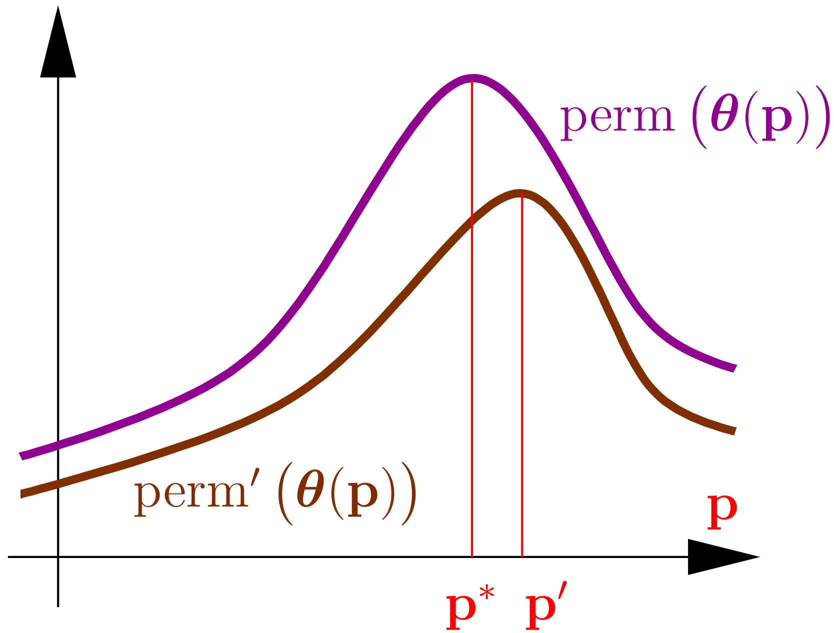
$\Rightarrow$ One needs to come up with approximate optimization algorithms:

- Monte Carlo Markov chain (MCMC) based approaches.

Maximizing the Permanent

Finding the PML distribution means finding the pmf $\mathbf{p}$ that maximizes

$$\mathbf{p}^* = \arg\max_{\mathbf{p}} \ \text{perm}\left(\boldsymbol{\theta}(\mathbf{p})\right).$$



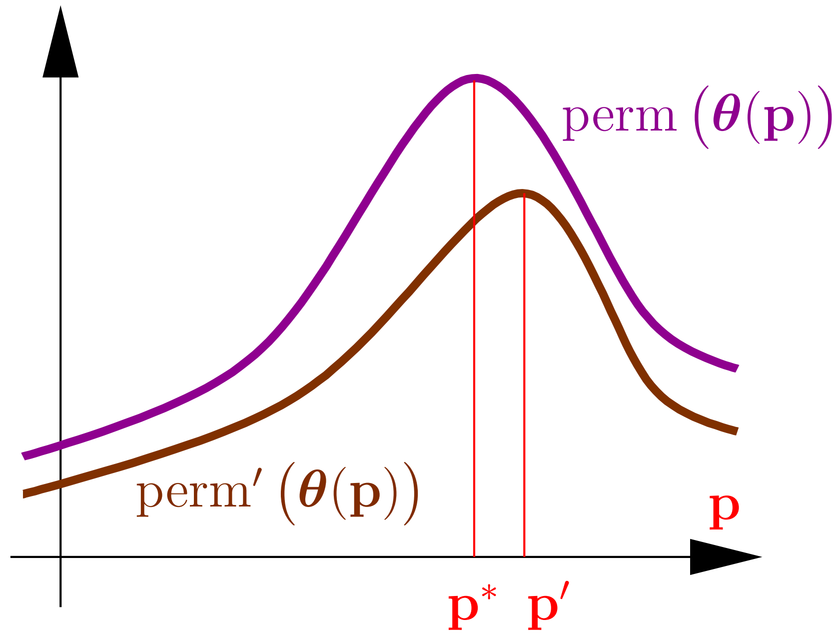
$\Rightarrow$ One needs to come up with approximate optimization algorithms:

- Monte Carlo Markov chain (MCMC) based approaches.
- Surrogate function based approaches.

Maximizing the Permanent

Finding the PML distribution means finding the pmf $\mathbf{p}$ that maximizes

$$\mathbf{p}^* = \arg\max_{\mathbf{p}} \ \mathrm{perm}\left(\boldsymbol{\theta}(\mathbf{p})\right).$$

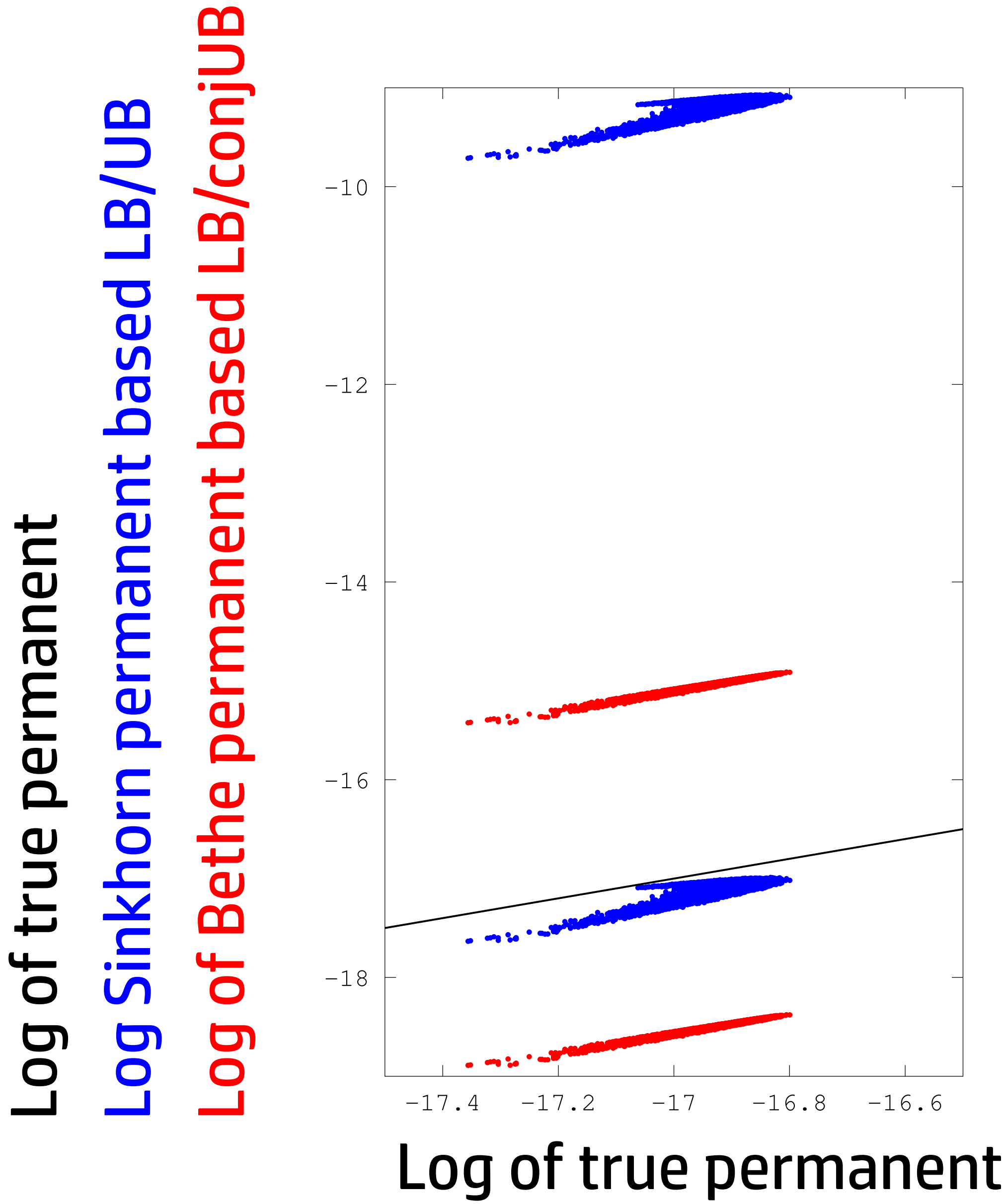


$\Rightarrow$ One needs to come up with approximate optimization algorithms:

- Monte Carlo Markov chain (MCMC) based approaches.
- Surrogate function based approaches.
- ...

Estimating the Permanent of a Matrix

10000 experiments with matrices of size 10×10 and structure $\theta(\mathbf{p})$.



Estimating the Permanent of a Matrix

- Sinkhorn permanent based LB/UB give

  a deterministic polynomial-time algorithm to approximate the permanent of a non-negative matrix up to a multiplicative factor of e^n.

  [Linial, Samorodnitsky, Wigderson, 2000]

- Bethe permanent based LB/UB give

  a deterministic polynomial-time algorithm to approximate the permanent of a non-negative matrix up to a multiplicative factor of 2^n (conjecture: $\sqrt{2}^n$).

  [Gurvits, Samorodnitsky, 2014]

Gibbs free energy approach to PML distribution

Pattern ML Distribution

$$\mathbf{p}^{\mathrm{PML}} = \arg\max_{\mathbf{p}} \ \mathrm{perm}\left(\boldsymbol{\theta}(\mathbf{p})\right)$$

Pattern ML Distribution

We replace $\mathrm{perm}\,\big(\boldsymbol{\theta}(\mathbf{p})\big)$ by the solution of an optimization problem:

$$\mathbf{p}^{\mathrm{PML}} = \arg\max_{\mathbf{p}}\ \mathrm{perm}\,\big(\boldsymbol{\theta}(\mathbf{p})\big)$$

$$\uparrow$$

$$\mathrm{perm}(\boldsymbol{\theta}) = \max_{\gamma}\ \exp\big(-F_{\mathrm{Gibbs}}(\gamma;\boldsymbol{\theta})\big).$$

Pattern ML Distribution

We replace $\mathrm{perm}\,\big(\boldsymbol{\theta}(\mathbf{p})\big)$ by the solution of an optimization problem:

$$\mathbf{p}^{\mathrm{PML}} = \arg\max_{\mathbf{p}}\ \mathrm{perm}\,\big(\boldsymbol{\theta}(\mathbf{p})\big)$$

$$\uparrow$$

$$\mathrm{perm}(\boldsymbol{\theta}) = \max_{\gamma}\ \exp\big(-F_{\mathrm{Gibbs}}(\boldsymbol{\gamma};\boldsymbol{\theta})\big).$$

Gibbs free energy

Pattern ML Distribution

We replace $\mathrm{perm}\,\big(\boldsymbol{\theta}(\mathbf{p})\big)$ by the solution of an optimization problem:

$$\mathbf{p}^{\mathrm{PML}} = \arg\max_{\mathbf{p}}\ \mathrm{perm}\,\big(\boldsymbol{\theta}(\mathbf{p})\big)$$

$$\uparrow$$

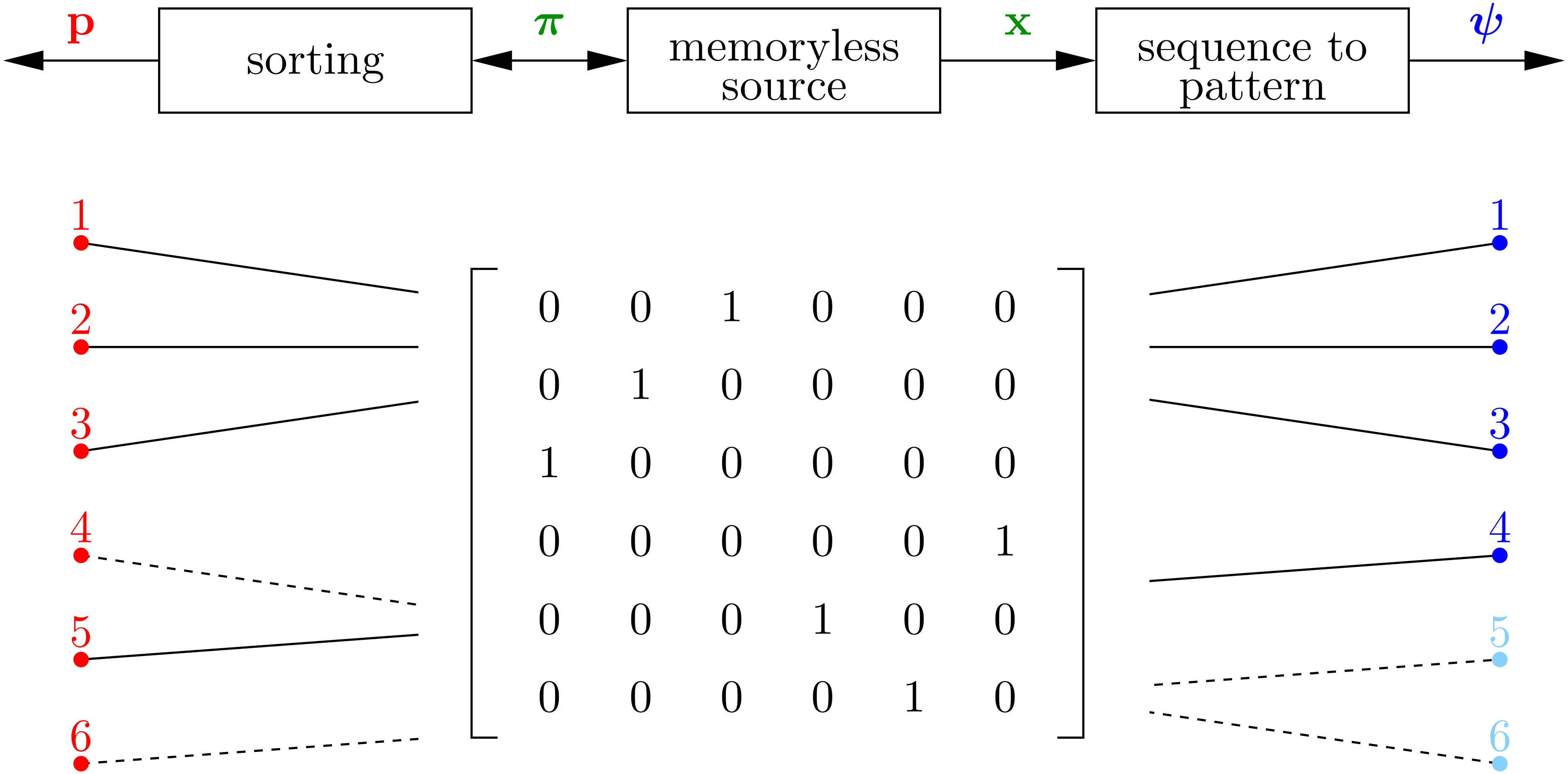
$$\mathrm{perm}(\boldsymbol{\theta}) = \max_{\gamma}\ \exp\big(-F_{\mathrm{Gibbs}}(\gamma;\boldsymbol{\theta})\big).$$

Gibbs free energy

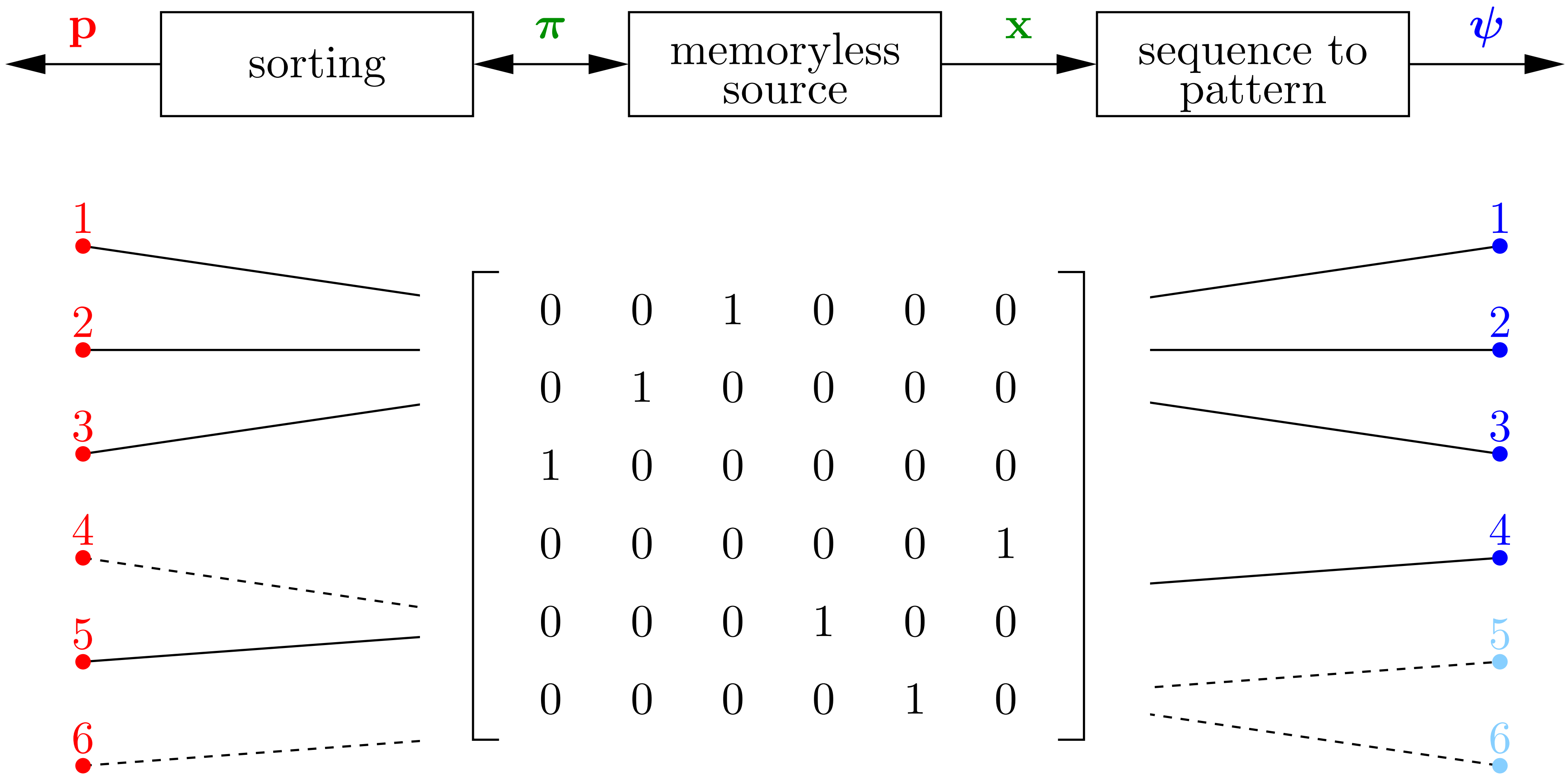
Combined:

$$\mathbf{p}^{\mathrm{PML}} = \arg\max_{\mathbf{p}}\ \max_{\gamma}\ \exp\Big(-F_{\mathrm{Gibbs}}\big(\gamma;\boldsymbol{\theta}(\mathbf{p})\big)\Big).$$

Pattern ML Distribution



Pattern ML Distribution



$$\gamma^{*}_{\text{Gibbs}} = \sum_{\sigma} P(\sigma|\mathbf{p}, \boldsymbol{\psi}) \cdot \mathbf{M}_{\sigma}$$

Pattern ML Distribution

We replace $\mathrm{perm}\big(\boldsymbol{\theta}(\mathbf{p})\big)$ by the solution of an optimization problem:

$$\mathbf{p}^{\mathrm{PML}} = \arg\max_{\mathbf{p}} \ \mathrm{perm}\big(\boldsymbol{\theta}(\mathbf{p})\big)$$

$$\uparrow$$

$$\mathrm{perm}(\boldsymbol{\theta}) = \max_{\gamma} \ \exp\big(-F_{\mathrm{Gibbs}}(\gamma;\boldsymbol{\theta})\big).$$

Gibbs free energy

Combined:

$$\mathbf{p}^{\mathrm{PML}} = \arg\max_{\mathbf{p}} \ \max_{\gamma} \ \exp\Big(-F_{\mathrm{Gibbs}}\big(\gamma;\boldsymbol{\theta}(\mathbf{p})\big)\Big).$$

Pattern ML Distribution

This suggests the following alternating maximization algorithm:

- Fix some $\mathbf{p}^{(0)}$.
- For $t = 1, 2, \ldots$ do:
  - First half:

$$\boldsymbol{\gamma}^{(t)} = \arg\max_{\boldsymbol{\gamma}} \exp\Big(- F_{\text{Gibbs}}\big(\boldsymbol{\gamma}; \boldsymbol{\theta}(\mathbf{p}^{(t-1)})\big)\Big)$$

  - Second half:

$$\mathbf{p}^{(t)} = \arg\max_{\mathbf{p}} \exp\Big(- F_{\text{Gibbs}}\big(\boldsymbol{\gamma}^{(t)}; \boldsymbol{\theta}(\mathbf{p})\big)\Big)$$

Pattern ML Distribution

This suggests the following alternating maximization algorithm:

- Fix some $\mathbf{p}^{(0)}$.
- For $t = 1, 2, \ldots$ do:
  - First half:

$$\gamma^{(t)} = \arg\max_{\gamma} \exp\Big(- F_{\text{Gibbs}}\big(\gamma; \boldsymbol{\theta}(\mathbf{p}^{(t-1)})\big)\Big)$$

  - Second half:

$$\mathbf{p}^{(t)} = \arg\max_{\mathbf{p}} \exp\Big(- F_{\text{Gibbs}}\big(\gamma^{(t)}; \boldsymbol{\theta}(\mathbf{p})\big)\Big)$$

This algorithm is equivalent to an expectation maximization (EM) algorithm proposed by Orlitsky *et al.*

Pattern ML Distribution

This suggests the following alternating maximization algorithm:

- Fix some $\mathbf{p}^{(0)}$.
- For $t = 1, 2, \ldots$ do:
  - First half:

$$\gamma^{(t)} = \arg\max_{\gamma} \exp\Big(- F_{\text{Gibbs}}\big(\gamma; \boldsymbol{\theta}(\mathbf{p}^{(t-1)})\big)\Big)$$

  - Second half:

$$\mathbf{p}^{(t)} = \arg\max_{\mathbf{p}} \exp\Big(- F_{\text{Gibbs}}\big(\gamma^{(t)}; \boldsymbol{\theta}(\mathbf{p})\big)\Big)$$

Can be approximated with the help of MCMC based techniques.

Bethe approximation to PML distribution

Pattern ML Distribution

Recall:

$$\mathbf{p}^{\mathrm{PML}} = \arg\max_{\mathbf{p}} \ \mathrm{perm}\big(\boldsymbol{\theta}(\mathbf{p})\big)$$

$$\uparrow$$

$$\mathrm{perm}(\boldsymbol{\theta}) = \max_{\gamma} \ \exp\big(-F_{\mathrm{Gibbs}}(\gamma; \boldsymbol{\theta})\big).$$

Gibbs free energy

Combined:

$$\mathbf{p}^{\mathrm{PML}} = \arg\max_{\mathbf{p}} \ \max_{\gamma} \ \exp\Big(-F_{\mathrm{Gibbs}}\big(\gamma; \boldsymbol{\theta}(\mathbf{p})\big)\Big).$$

Bethe Approximation to the Pattern ML Distribution

Now:

$$\mathbf{p}^{\mathrm{PML}} = \arg\max_{\mathbf{p}} \ \mathrm{perm}\big(\boldsymbol{\theta}(\mathbf{p})\big)$$

$$\uparrow$$

$$\mathrm{perm}(\boldsymbol{\theta}) \approx \max_{\gamma} \ \exp\big(-F_{\mathrm{Bethe}}(\gamma;\boldsymbol{\theta})\big).$$

Bethe free energy

Combined:

$$\mathbf{p}^{\mathrm{PML}} \approx \mathbf{p}^{\mathrm{BPML}} \triangleq \arg\max_{\mathbf{p}} \ \max_{\gamma} \ \exp\Big(-F_{\mathrm{Bethe}}\big(\gamma;\boldsymbol{\theta}(\mathbf{p})\big)\Big).$$

Bethe Approximation to the Pattern ML Distribution

We can rewrite this as an alternating minimization algorithm:

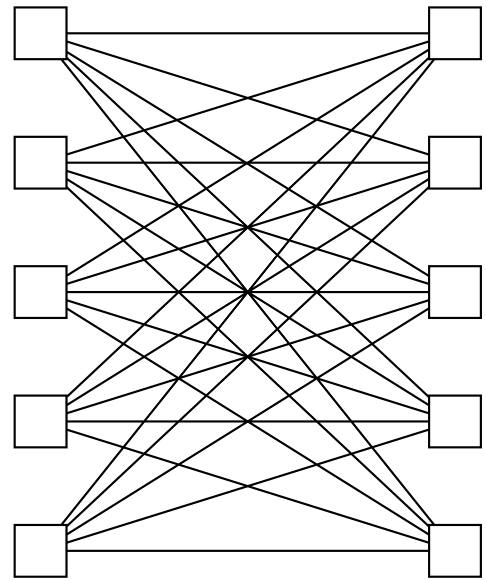
- Fix some $\mathbf{p}^{(0)}$.
- For $t = 1, 2, \ldots$ do:
  - First half:

$$\gamma^{(t)} = \arg\max_{\gamma} \exp\Big(- F_{\mathrm{Bethe}}\big(\gamma; \boldsymbol{\theta}(\mathbf{p}^{(t-1)})\big)\Big)$$

  - Second half:

$$\mathbf{p}^{(t)} = \arg\max_{\mathbf{p}} \exp\Big(- F_{\mathrm{Bethe}}\big(\gamma^{(t)}; \boldsymbol{\theta}(\mathbf{p})\big)\Big)$$

Bethe Approximation to the Pattern ML Distribution



We can rewrite this as an alternating minimization algorithm:

- Fix some $\mathbf{p}^{(0)}$.
- For $t = 1, 2, \ldots$ do:
  - First half:

$$\boldsymbol{\gamma}^{(t)} = \arg\max_{\boldsymbol{\gamma}} \exp\Big(- F_{\mathrm{Bethe}}\big(\boldsymbol{\gamma}; \boldsymbol{\theta}(\mathbf{p}^{(t-1)})\big)\Big)$$

  - Second half:

$$\mathbf{p}^{(t)} = \arg\max_{\mathbf{p}} \exp\Big(- F_{\mathrm{Bethe}}\big(\boldsymbol{\gamma}^{(t)}; \boldsymbol{\theta}(\mathbf{p})\big)\Big)$$

Sinkhorn approximation to PML distribution

Pattern ML Distribution

Recall:

$$\mathbf{p}^{\mathrm{PML}} = \arg\max_{\mathbf{p}} \ \mathrm{perm}\big(\boldsymbol{\theta}(\mathbf{p})\big)$$

$$\uparrow$$

$$\mathrm{perm}(\boldsymbol{\theta}) = \max_{\gamma} \ \exp\big(- F_{\mathrm{Gibbs}}(\gamma; \boldsymbol{\theta})\big).$$

Gibbs free energy

Combined:

$$\mathbf{p}^{\mathrm{PML}} = \arg\max_{\mathbf{p}} \ \max_{\gamma} \ \exp\Big(- F_{\mathrm{Gibbs}}(\gamma; \boldsymbol{\theta}(\mathbf{p}))\Big).$$

Sinkhorn Approximation to the Pattern ML Distribution

Now:

$$\mathbf{p}^{\mathrm{PML}} = \arg\max_{\mathbf{p}} \ \mathrm{perm}\big(\boldsymbol{\theta}(\mathbf{p})\big)$$

$$\uparrow$$

$$\mathrm{perm}(\boldsymbol{\theta}) \approx \max_{\gamma} \ \exp\big(-F_{\mathrm{Sinkhorn}}(\gamma;\boldsymbol{\theta})\big).$$

Sinkhorn free energy

Combined:

$$\mathbf{p}^{\mathrm{PML}} \approx \mathbf{p}^{\mathrm{BPML}} \triangleq \arg\max_{\mathbf{p}} \ \max_{\gamma} \ \exp\Big(-F_{\mathrm{Sinkhorn}}\big(\gamma;\boldsymbol{\theta}(\mathbf{p})\big)\Big).$$

Sinkhorn Approximation to the Pattern ML Distribution

We can rewrite this as an alternating minimization algorithm:

- Fix some $\mathbf{p}^{(0)}$.
- For $t = 1, 2, \ldots$ do:
  - First half:

$$\boldsymbol{\gamma}^{(t)} = \arg\max_{\boldsymbol{\gamma}} \exp\Big(- F_{\text{Sinkhorn}}\big(\boldsymbol{\gamma}; \boldsymbol{\theta}(\mathbf{p}^{(t-1)})\big)\Big)$$

  - Second half:

$$\mathbf{p}^{(t)} = \arg\max_{\mathbf{p}} \exp\Big(- F_{\text{Sinkhorn}}\big(\boldsymbol{\gamma}^{(t)}; \boldsymbol{\theta}(\mathbf{p})\big)\Big)$$

Valiant–Valiant estimate of the distribution histogram

Valiant–Valiant estimate of the distribution histogram

Key ingredients:

Valiant–Valiant estimate of the distribution histogram

Key ingredients:

- Poissonization trick:

  instead of sequences of length n

  consider sequences of length n' where $n' \sim \mathrm{Poisson}(n)$

Valiant–Valiant estimate of the distribution histogram

Key ingredients:

- Poissonization trick:

  instead of sequences of length n

  consider sequences of length n' where $n' \sim \mathrm{Poisson}(n)$

  $\Rightarrow$ multiplicities of pattern symbols are independent!

Valiant–Valiant estimate of the distribution histogram

Key ingredients:

- Poissonization trick:

  instead of sequences of length n

  consider sequences of length n' where $n' \sim \mathrm{Poisson}(n)$

  $\Rightarrow$ multiplicities of pattern symbols are independent!

- A source symbol with probablity p will yield a pattern symbol with multiplicity μ where

$$\mu \sim \mathrm{Poisson}(n \cdot p).$$

Valiant–Valiant estimate of the distribution histogram

Key ingredients:

- Set up a linear program that looks for a sorted distribution

$$\mathbf{p} = \big(\underbrace{p^{(1)}, \ldots, p^{(1)}}_{\text{length } k^{(1)}}, \underbrace{p^{(2)}, \ldots, p^{(2)}}_{\text{length } k^{(2)}}, \ \ldots, \ \underbrace{p^{(L)}, \ldots, p^{(L)}}_{\text{length } k^{(L)}}\big)$$

  - such that the expected multiplicity histogram "matches" the observed mulitiplicty vector $\boldsymbol{\mu}$
  - such that $p^{(\ell)} \in \mathcal{Q}$ for some finite set $\mathcal{Q}$
  - and such that $k^{(1)} + k^{(2)} + \cdots + k^{(L)} = k$.

Valiant–Valiant estimate of the distribution histogram

Key ingredients:

- Set up a linear program that looks for a sorted distribution

$$\mathbf{p} = \Big(\underbrace{p^{(1)}, \ldots, p^{(1)}}_{\text{length } k^{(1)}}, \underbrace{p^{(2)}, \ldots, p^{(2)}}_{\text{length } k^{(2)}}, \ \ldots, \ \underbrace{p^{(L)}, \ldots, p^{(L)}}_{\text{length } k^{(L)}}\Big)$$

  - such that the expected multiplicity histogram "matches" the observed mulitiplicty vector μ
  - such that $p^{(\ell)} \in \mathcal{Q}$ for some finite set $\mathcal{Q}$
  - and such that $k^{(1)} + k^{(2)} + \cdots + k^{(L)} = k$.

Note: there is a bijection between

sorted distributions and distribution histograms.

Connections

Based on the sorted distribution $\mathbf{p}^*$ found by the above LP, one can define a $k \times k$ matrix $\boldsymbol{\gamma}^*$ with entries

$$\gamma^*_{i,j} \triangleq \mathrm{e}^{-np^*_i} \cdot \frac{(np^*_i)^{\mu_j}}{\mu_j! \cdot \varphi_{\mu_j}}, \qquad (i,j) \in [k]^2.$$

such that

- The matrix $\boldsymbol{\gamma}^*$ is approximately doubly stochastic.
  By this we mean
  - that all entries are non-negative and
  - that the row and column sums are approximately 1.
- The vector-matrix pair $(\mathbf{p}^*, \boldsymbol{\gamma}^*)$ is close to being a stationary point of $F_{\mathrm{Sinkhorn}}\big(\boldsymbol{\gamma}^{(t)}; \boldsymbol{\theta}(\mathbf{p})\big)$.

Conclusions / Outlook

Conclusions / Outlook

Conclusions / Outlook

- We have defined the PML estimate and various approximations.

Conclusions / Outlook

- We have defined the PML estimate and various approximations.
- We have defined the Valiant–Valiant estimate of the distribution histogram.

Conclusions / Outlook

- We have defined the PML estimate and various approximations.
- We have defined the Valiant–Valiant estimate of the distribution histogram.
- We have discussed connections between these estimates.

Conclusions / Outlook

- We have defined the PML estimate and various approximations.
- We have defined the Valiant–Valiant estimate of the distribution histogram.
- We have discussed connections between these estimates.
- The key object for establishing these connections and for establishing properties of these estimates is the matrix γ and its approximations.

Conclusions / Outlook

- We have defined the PML estimate and various approximations.
- We have defined the Valiant–Valiant estimate of the distribution histogram.
- We have discussed connections between these estimates.
- The key object for establishing these connections and for establishing properties of these estimates is the matrix γ and its approximations.
- Use insights to speed up Bethe PML and Sinkhorn PML algorithms.


Thank you!