MLE for Gaussian Matching

Jiaxin Hu

March 30, 2022

1 MLE for Gaussian Tensor matching

Recall that we have two random tensors $\mathcal{A}, \mathcal{B}' \in \mathbb{R}^{n^{\otimes m}}$, where $\mathcal{A}(\omega)$ and $\mathcal{B}'(\omega)$ denote the tensor entry indexed by $\omega = (i_1, \ldots, i_m) \in [n]^m$. Suppose $\mathcal{A}$ and $\mathcal{B}'$ are super-symmetric; i.e., $\mathcal{A}(\omega) = \mathcal{A}(f(\omega)), \mathcal{B}(\omega) = \mathcal{B}'(f(\omega))$ for any function f permutes the indices in ω for all $\omega \in [n]^m$. Consider the bivariate generative model for the entries $\{\omega : 1 \leq i_1 \leq \cdots \leq i_m \leq n\}$

$$(\mathcal{A}(\omega), \mathcal{B}'(\omega)) \sim \mathcal{N}\left(0, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right), \quad \text{and} \quad (\mathcal{A}(\omega), \mathcal{B}'(\omega)) \perp (\mathcal{A}(\omega'), \mathcal{B}'(\omega')), \text{ for all } \omega \neq \omega',$$

where the correlation $\rho \in (0, 1)$ and $\perp$ denote the statistical independence.

Suppose we observe the tensor pair $\mathcal{A}$ and $\mathcal{B} \overset{\text{def}}{=} \mathcal{B}' \circ \pi^*$, where $\pi^* : [n] \mapsto [n]$ denotes a permutation on $[d]$, and $\mathcal{B}(i_1, \ldots, i_m) = \mathcal{B}'(\pi(i_1), \ldots, \pi(i_m))$ for all $(i_1, \ldots, i_m) \in [n]^m$. Equivalently, let $\Pi^* \in \{0, 1\}^{n \times n}$ denote the permutation matrix, where

$$\Pi^*_{ij} = \begin{cases} 1 & \pi^*(i) = j \\ 0 & \text{otherwise} \end{cases}.$$

We have

$$\mathcal{B}' = \mathcal{B} \times_1 \Pi^* \times_2 \cdots \times_m \Pi^*.$$

Let $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and $\Sigma^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$. The likelihood function of $\pi : [n] \mapsto [n]$ (or corresponding $\Pi \in \{0, 1\}^{n \times n}$) with given observations $\mathcal{A}, \mathcal{B}$ is

$$\mathcal{L}(\pi, \rho | \mathcal{A}, \mathcal{B}) = \frac{1}{[\det(2\pi\Sigma)]^{n^m}} \exp\left(-\frac{1}{2} \sum_{\omega \in [n]^m} (\mathcal{A}(\omega), \mathcal{B}(\pi \circ \omega))\Sigma^{-1}(\mathcal{A}(\omega), \mathcal{B}(\pi \circ \omega))^T\right)$$

$$= \frac{1}{[\det(2\pi\Sigma)]^{n^m}} \exp\left(-\frac{1}{2(1-\rho^2)} \sum_{\omega \in [n]^m} \mathcal{A}(\omega)^2 + \mathcal{B}(\pi \circ \omega)^2 - 2\rho\mathcal{A}(\omega)\mathcal{B}(\pi \circ \omega)\right).$$

Note that $\sum_{\omega \in [n]^m} \mathcal{B}(\pi \circ \omega)^2 = \sum_{\omega \in [n]^m} \mathcal{B}(\omega)^2$.

If we take ρ as a fixed constant, the only term dependent to π in $\mathcal{L}$ is $\mathcal{A}(\omega)\mathcal{B}(\pi \circ \omega)$. We have the MLE for π^* that satisfies

$$\hat{\pi} = \arg\max_{\pi:[n]\mapsto[n]} \mathcal{L}(\pi|\mathcal{A}, \mathcal{B}) = \arg\max_{\pi:[n]\mapsto[n]} \sum_{\omega\in[n]^m} \rho\mathcal{A}(\omega)\mathcal{B}(\pi \circ \omega) = \arg\max_{\Pi\in\{0,1\}^{n\times n}} \langle \mathcal{A}, \mathcal{B} \times_1 \Pi \times_2 \cdots \times_m \Pi \rangle.$$

Solving $\hat{\pi}$ can be viewed as a special case of *multiway assignment problem* which is a higher-order generalization of *quadratic assignment problem*.

If we consider ρ as an unknown parameter, we consider the log-likelihood ignoring constant terms

$$\ell(\pi, \rho|\mathcal{A}, \mathcal{B}) = -\frac{n^m}{2}\log(1-\rho^2) - \frac{\|\mathcal{A}\|^2 + \|\mathcal{B}\|^2}{2(1-\rho^2)} + \frac{\rho}{1-\rho^2}\sum_{\omega\in[n]^m}\mathcal{A}(\omega)\mathcal{B}(\pi \circ \omega)$$

$$= -\frac{n^m}{2}\log\sigma^2 - \frac{\|\mathcal{A}\|^2 + \|\mathcal{B}\|^2}{2\sigma^2} + \frac{\sqrt{1-\sigma^2}}{\sigma^2}\sum_{\omega\in[n]^m}\mathcal{A}(\omega)\mathcal{B}(\pi \circ \omega),$$

where $\sigma^2 = 1 - \rho^2$. Take the first derivative over σ^2. We have the MLE of σ^2 with given π, denoted $\hat{\sigma}^2(\pi)$, satisfying

$$-\frac{n^m}{2} + \frac{\|\mathcal{A}\|^2 + \|\mathcal{B}\|^2}{2\hat{\sigma}^2(\pi)} - \left(\frac{\sqrt{1-\hat{\sigma}^2(\pi)}}{\hat{\sigma}^2(\pi)} + \frac{1}{\sqrt{1-\hat{\sigma}^2(\pi)}}\right)\sum_{\omega\in[n]^m}\mathcal{A}(\omega)\mathcal{B}(\pi \circ \omega) = 0.$$

Thus, the MLE of ρ is equal to $\hat{\rho}(\pi) = \sqrt{1-\hat{\sigma}^2(\pi)}$. Note that the MLE of ρ is a function of π. We have the MLE of π, denoted $\hat{\pi}$, satisfying

$$\hat{\pi} = \arg\max_{\pi:[n]\mapsto[n]} \mathcal{L}(\pi, \hat{\rho}(\pi)|\mathcal{A}, \mathcal{B}).$$

Therefore, the MLE for (π, ρ) is $(\hat{\pi}, \hat{\rho}(\hat{\pi}))$.

References