Submatrix Detection

Parker Knight

September 5, 2022

1 Introduction

This set of notes concerns the submatrix detection problem, following [1].

Our setup is as follows: we observe a high dimensional random matrix, and want to test if it contains a submatrix (of much smaller size) with large entries. In particular, let

$$Y_{ij} = s_{ij} + \varepsilon_{ij}$$

for $i \in [N], j \in [M]$, and the ε_{ij} are iid $N(0,1)$. We test the following hypothesis:

$$H_0 : s_{ij} = 0 \quad \forall i, j$$

versus

$$H_1 : \exists C = A \times B \subset [N] \times [M] \quad \text{s.t.} \quad s_{ij} \geq a \quad \forall i, j \in C$$

where $|A| = n$ and $|B| = m$. We assume that the size of the matrix (i.e., the values of n and m) is known. Let $\mathcal{S}_{nm,a}$ denote the set of all $N \times M$ matrices satisfying the alternative hypothesis, and let $\mathcal{C}_{nm}$ denote the set of all possible $n \times m$ submatrices of an $N \times M$ matrix[1]. Let $\mathbb{P}_0 \{E\}$ denote the probability of event E under the null hypothesis, and let $\mathbb{P}_S \{E\}$ denote the probability of E under some $S \in \mathcal{S}_{nm,a}$.

A test ψ is any measurable function of the observations $\{Y_{ij}\}$ taking values in $\{0, 1\}$. We introduce the following notation:

[1] In other words, $\mathcal{C}_{nm} = \{A \times B : A \subset [N], B \subset [M], |A| = n, |M| = m\}$

$$
\begin{aligned}
&\text{Type I Error:} \quad \alpha(\psi) = \mathbb{P}_0 \{\psi = 1\} \\
&\text{Type II Error:} \quad \beta(\psi, S) = \mathbb{P}_S \{\psi = 0\} \\
&\text{Maximal Type II Error:} \quad \beta_{nm,a}(\psi) = \sup_{S \in \mathcal{S}_{nm,a}} \mathbb{P}_S \{\psi = 0\} \\
&\text{Risk:} \quad \gamma_{nm,a}(\psi) = \alpha(\psi) + \beta_{nm,a}(\psi)
\end{aligned}
$$

We consider the following high dimensional sparse asymptotic regime: $N, M, n, m \to \infty$ with $p = n/N, q = m/M \to 0$.

References

[1] C. Butucea and Y. I. Ingster, "Detection of a sparse submatrix of a high-dimensional noisy matrix," vol. 19, no. 5.