

Problem Sheet for Math-448 Sheet 9 2025

1. We are studying customers $\{1, \dots, n\}$ who buy products $\{1, \dots, m\}$. An edge variable takes the value one if customers buy a particular product, this forming the data matrix X which is $n \times m$.
 - (a) Assume that the expectation of X_{ij} is a product of a term capturing the customer activity α_i and a term capturing product popularity β_j . Write down $\mathbb{E}X_{ij} = P_{ij}$.
 - (b) Calculate the in and out degrees of the data matrix X . Determine its distribution, assuming every entry of the data matrix is independent.
 - (c) Describe how to estimate α and β using the degrees.
 - (d) Calculate the mean and variance of the estimated parameters.
 - (e) Describe how to test that $\alpha_i = \alpha$ a constant.
2. Referring back to question 1, the proposed model captures an effect due to each individual, combined with an effect due to each product. What is there are clear clusters in the data? Instead we assume that there are rows r_i belonging to cluster i and columns c_i that belong to the same cluster. Describe how we would estimate the cluster expected value θ_i . Is it more important to estimate cluster membership or cluster probabilities correctly?