Critical Threshold for Arithmetic Progressions in Random Subsets

Deepseek

26.01.2025

Problem Statement

Let $\xi_n, n \in [N]$ be independent, identically distributed random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{P}[\xi_n = 1] = N^{-\delta}$ and $\mathbb{P}[\xi_n = 0] = 1 - N^{-\delta}$ for $0 < \delta < 1$. The random subset $\{n\xi_n\}_{n\in[N]}$ is formed by including each n independently with probability $N^{-\delta}$. For which δ does the probability that the subset contains at least one k-term arithmetic progression (AP) approach 1 as $N \to \infty$?

Solution

Step 1: Expected Number of k-term APs

Let X denote the number of k-term APs in the subset. The total number of k-term APs in $[N]$ is approximately $\Theta(N^2)$. The probability that a specific k-term AP is included is $N^{-k\delta}$. Thus:

$$\mathbb{E}[X] \asymp N^2 \cdot N^{-k\delta} = N^{2-k\delta}.$$

For $\mathbb{E}[X] \to \infty$, we require:

$$2 - k\delta > 0 \quad \Longrightarrow \quad \delta < \frac{2}{k}.$$

Step 2: Variance Analysis

To show $\mathrm{Var}(X)/\mathbb{E}[X]^2 \to 0$, expand X as a sum over all APs:

$$X = \sum_A I_A, \quad \text{where } I_A \text{ is the indicator for AP } A.$$

The variance is:

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \sum_A \mathbb{E}[I_A] + \sum_{A \neq B} \mathrm{Cov}(I_A, I_B).$$

Covariance for Overlapping APs

For two APs A and B sharing $m \geq 1$ terms:

$$\mathrm{Cov}(I_A, I_B) = N^{-2k\delta}(N^{m\delta} - 1) \asymp N^{-2k\delta+m\delta}.$$

The dominant contribution comes from $m = 1$ (most common overlap):

$$\mathrm{Cov}(I_A, I_B) \asymp N^{-2k\delta+\delta}.$$

Bounding the Variance

The number of pairs of APs sharing m terms is $\Theta(N^{4-m})$. Summing over m:

$$\mathrm{Var}(X) \lesssim N^{2-k\delta} + \sum_{m=1}^{k-1} N^{4-m} \cdot N^{-2k\delta+\delta m}.$$

The dominant term is for $m = 1$:

$$\mathrm{Var}(X) \lesssim N^{2-k\delta} + N^{3-2k\delta+\delta}.$$

Step 3: Variance-to-Mean Ratio

Using $\mathbb{E}[X]^2 \asymp N^{4-2k\delta}$:

$$\frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} \lesssim \frac{N^{2-k\delta} + N^{3-2k\delta+\delta}}{N^{4-2k\delta}} = N^{-2+k\delta} + N^{-1-k\delta+\delta}.$$

For $\delta < \frac{2}{k}$:

- $N^{-2+k\delta} \to 0$ (since $k\delta < 2$),
- $N^{-1-k\delta+\delta} \to 0$ (since $\delta(k-1) > 0$).

Conclusion

By the second moment method, if $\mathrm{Var}(X)/\mathbb{E}[X]^2 \to 0$, then:

$$\mathbb{P}[X \geq 1] \geq 1 - \frac{\mathrm{Var}(X)}{\mathbb{E}[X]^2} \to 1.$$

Thus, the critical threshold is:

$$\boxed{\delta < \frac{2}{k}}$$