Sampling Lemma for Unbounded Kernels

Panna Tímea Fekete, Dávid Kunszenti-Kovács
Eötvös Loránd University, Budapest, Hungary
and Alfréd Rényi Institute of Mathematics, Budapest, Hungary

Graph sequences with limits that can be represented by an unbounded, L^p kernel for some $p < \infty$ arise naturally from various random graph models, among others the dense Preferential Attachment Graph random multigraph model. Whilst it was shown that even for unbounded kernels W one has probability 1 convergence in density of the W-random graph sequence $\mathbb{G}(W, n)$ to W itself, this does not imply the same convergence in cut distance. Here we first present high probability bounds on the closeness of the cut norms of the kernel and its samples, i.e., a First Sampling lemma for unbounded kernels.

For some fixed $k \in \mathbb{Z}_+$ let X be a random ordered k-subset of $[0, 1]$. Let $U[X]$ denote the step-function on $[0, 1]^2$ with uniform steps of length $1/k$ in each variable, and values given by $(U(X_i, X_j))_{i \neq j \in [k]}$, and 0 on the main diagonal.

Given an $n \times n$ matrix A and a function $U \in L^1([0, 1]^2)$, their cut norm is:

$$\|A\|_\square := \frac{1}{n^2} \max_{S,T \subseteq [n]} \left| \sum_{i \in S, j \in T} A_{ij} \right| \quad \text{and} \quad \|U\|_\square := \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} U(x, y) dx dy \right|.$$

Our goal is to provide a high probability bound on the difference of the norm $\|U[X]\|_\square$ of the random sample and the original norm $\|U\|_\square$. The typical application would be to check with very high probability that $\|U\|_\square$ is small, via looking at the samples.

Our main result is a generalization of the graphon sampling lemma ([1], [2]):

Theorem 1. *(First Sampling Lemma for Unbounded Kernels). Let $\gamma > 0$ and $m > 4$ be constants such that $m\gamma > 2$. Then for any function $U \in L^m_{sym}([0, 1]^2)$, there exist constants $C, C_0, C_1 > 0$ only depending on γ, m and $\|U\|_m$ such that with probability at least $1 - C_1 k^{2-\gamma m}$,*

$$-C_0 k^{-1/2+\gamma} (\ln k)^{1/2} \leq \|U[X]\|_\square - \|U\|_\square \leq C \cdot k^{-1/4+\gamma/2} (\ln k)^{1/4}.$$

Thus we got that under mild conditions the bound is arbitrarily close in order to the bound for L^∞ graphons, but at the cost of a much larger, only polynomially small (as opposed to exponentially small) exceptional event set. We note, however, that such a polynomial sized exceptional set is still adequate for most applications.

This research was prepared with the professional support of the Doctoral Student Scholarship Program of the Co-operative Doctoral Program of the Ministry of Innovation and Technology financed from the National Research, Development and Innovation Fund and supported by ERC Synergy Grant No. 40089.

[1] N. Alon, W. Fernandez de la Vega, R. Kannan and M. Karpinski, Random sampling and approximation of MAX-CSPs, *J. Comput. System Sci.* **67** (2003), 212–243.

[2] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós and K. Vesztergombi, Convergent Graph Sequences I: Subgraph frequencies, metric properties, and testing, *Advances in Math.* **219** (2008), 1801–1851.