HAL open science

Optimal approximations made easy

Mónika Csikós, Nabil H Mustafa

▶ **To cite this version:**

Mónika Csikós, Nabil H Mustafa. Optimal approximations made easy. Information Processing Letters, 2022, 176, pp.106250. 10.1016/j.ipl.2022.106250 . hal-04501852

HAL Id: hal-04501852

https://hal.science/hal-04501852

Submitted on 12 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimal Approximations Made Easy

Mónika Csikós Nabil H. Mustafa

Université Gustave Eiffel, LIGM, Equipe A3SI, ESIEE Paris,
Cité Descartes, 2 boulevard Blaise Pascal, 93162 Noisy-le-Grand Cedex, France.
Emails: monika.csikos@esiee.fr, mustafan@esiee.fr

Abstract

The fundamental result of Li, Long, and Srinivasan [LLS01] on approximations of set systems has become a key tool across several communities such as learning theory, algorithms, computational geometry, combinatorics and data analysis.

The goal of this paper is to give a modular, self-contained, intuitive proof of this result for finite set systems. The only ingredient we assume is the standard Chernoff's concentration bound. This makes the proof accessible to a wider audience, readers not familiar with techniques from statistical learning theory, and makes it possible to be covered in a single self-contained lecture in a geometry, algorithms or combinatorics course.

Keywords: relative approximations, VC theory, chaining.

Funding: The work of the authors has been supported by the grants ANR ADDS (ANR-19-CE48-0005) and ANR SAGA (JCJC-14-CE25-0016-01).

1 Introduction

Given a finite set system $(X, \mathcal{F})$, our goal is to construct a small set $A \subseteq X$ such that each set of $\mathcal{F}$ is 'well-approximated' by A. Research on such approximations started in the 1950s, with random sampling being the key tool for showing their existence. A breakthrough in the study of approximations dates back to 1971 when Vapnik and Chervonenkis studied set systems with finite VC-dimension [VC71]. The *VC-dimension* of $(X, \mathcal{F})$, denoted by VC-dim$(X, \mathcal{F})$, is the size of the largest $Y \subseteq X$ for which $\mathcal{F}|_Y = 2^Y$, where $\mathcal{F}|_Y = \{Y \cap S \colon S \in \mathcal{F}\}$. Since then, the notion of approximations has become a fundamental structure across several communities—learning theory, statistics, combinatorics and algorithms.

Relative (ϵ, δ)-approximations. Given a set system $(X, \mathcal{F})$ with $n = |X|$ and parameters $0 < \epsilon, \delta < 1$, a set A of size t is a *relative (ϵ, δ)-approximation* for $(X, \mathcal{F})$ if for all $S \in \mathcal{F}$,

$$\left| \frac{|S|}{n} - \frac{|A \cap S|}{t} \right| \leq \delta \cdot \max\left\{ \frac{|S|}{n}, \epsilon \right\}, \quad \text{or equivalently,} \quad |A \cap S| = \frac{|S|\, t}{n} \pm \delta t \max\left\{ \frac{|S|}{n}, \epsilon \right\}.$$

A basic upper-bound on sizes of relative (ϵ, δ)-approximations follows immediately from Chernoff's bound, which we first recall (see [AS12]).

Theorem A (Chernoff's bound)**.** *Let X be a set of n elements and A be a uniform random sample of X of size t. Then for any $S \subseteq X$ and $\eta > 0$,*

$$\Pr\left[|A \cap S| \notin \left(\frac{|S|t}{n} - \eta,\ \frac{|S|t}{n} + \eta\right)\right] \leq 2\ \exp\left(-\frac{\eta^2 n}{2|S|t + \eta n}\right).$$

In particular, setting $\eta = \delta t \max\left\{\frac{|S|}{n}, \epsilon\right\}$, a uniform random sample A of size t fails to be a relative (ϵ, δ)-approximation for a fixed $S \in \mathcal{F}$ with probability at most $2\exp\left(-\frac{\epsilon\delta^2 t}{3}\right)$.

Theorem A together with the union bound gives the following trivial upper-bound on relative (ϵ, δ)-approximation sizes for *any* finite set system.

Theorem 1. *Let $(X, \mathcal{F})$ be a finite set system and $0 < \epsilon, \delta, \gamma < 1$ be given parameters. Then a uniform random sample $A \subseteq X$ of size at least $\frac{3}{\epsilon\delta^2} \ln \frac{2|\mathcal{F}|}{\gamma}$ is a relative (ϵ, δ)-approximation for $\mathcal{F}$ with probability at least $1 - \gamma$.*

This note addresses the following influential result of Li, Long, and Srinivasan [LLS01], described as 'the pinnacle of a long sequence of papers' in [HP11, Section 7.4].[1]

Theorem 2 ([LLS01])**.** *There exists an absolute constant $c \geq 1$ such that the following holds. Let $(X, \mathcal{F})$ be a set system such that $|\mathcal{F}|_Y| \leq (e|Y|/d)^d$ for all $Y \subseteq X$ with $|Y| \geq d$, and let $0 < \delta, \epsilon, \gamma < 1$ be given parameters. Then a uniform random sample $A \subseteq X$ of size*

$$\frac{c}{\epsilon\delta^2} \cdot \left(d \ln \frac{1}{\epsilon} + \ln \frac{1}{\gamma}\right)$$

is a relative (ϵ, δ)-approximation for $(X, \mathcal{F})$ with probability at least $1 - \gamma$.

Remark. As VC-dim $(X, \mathcal{F}) \leq d$ implies that $|\mathcal{F}|_Y| \leq (e|Y|/d)^d$ for any $Y \subseteq X$ [Sau72, She72], Theorem 2 also applies to set systems with VC-dim $(X, \mathcal{F}) \leq d$. This bound is asymptotically tight, and immediately implies many other approximation bounds such as ϵ-approximations (Vapnik and Chervonenkis [VC71], Talagrand [Tal94]), sensitive ϵ-approximations (Brönnimann et al. [BCM93]), and ϵ-nets (Haussler and Welzl [HW87], Komlós et al. [KPW92]).

The original proof of Theorem 2 uses two techniques:

Symmetrization. To prove that a random sample A satisfies the required properties, one takes another random sample G, sometimes called a 'ghost sample'. Properties of A are then proven by comparing it with G. Note that G is not used in the algorithm or its construction—it is solely a method of analysis, a 'thought experiment' of sorts.

Chaining. The idea is to analyse the interaction of the sets in $\mathcal{F}$ with a random sample by partitioning each $S \in \mathcal{F}$ into a logarithmic number of smaller sets, each belonging to a distinct 'level'. The number of sets increase with increasing level while the size of each set decreases. The overall sum turns out to be a geometric series, which then gives the optimal bounds [KT59, Tal16].

What makes the proof of Theorem 2 in [LLS01] difficult is that it combines chaining and symmetrization intricately. All the tail bounds are stated in their 'symmetrized' forms and symmetrization is carried through the entire proof. It is not an easy proof to explain to undergraduate or even graduate students in computer science, as it is difficult to see what is really going on in terms of the significance and intuition of these two ideas. In fact, even the proofs of simpler statements involving just symmetrization, as given in textbooks[2]—e.g., see [KV94, DGL96, Mat99, Cha00, Mat02, AB09, HP11, AS12]—often come with the caveat that the idea is ingenious but difficult to understand intuitively (e.g., "one might be tempted to believe that it works by some magic" [Mat02, Section 10.2]).

[1] The original result was stated using the notion of (ϵ, δ)-samples, but they are asymptotically equivalent: an (ϵ, δ)-sample is a relative $(\epsilon, 4\delta)$-approximation and a relative (ϵ, δ)-approximation is an (ϵ, δ)-sample; see [HS11].

[2] Also used in teaching; to pick two arbitrary examples, see here for an example from the perspective of statistics/learning and here from the algorithmic side.

Our Results.

This work is an attempt to improve this state of affairs. We show that in fact one can separate the roles of chaining and symmetrization, giving two separate statements which together immediately imply Theorem 2. The role of symmetrization is to get a bound on relative (ϵ, δ)-approximations that is independent of $|\mathcal{F}|$:

Theorem 3. *There exists an absolute constant c_1 such that the following holds. Let $(X, \mathcal{F})$ be a set system such that $|\mathcal{F}|_Y| \leq (e|Y|/d)^d$ for all $Y \subseteq X$, $|Y| \geq d$, and let $0 < \delta, \epsilon, \gamma < 1$ be given parameters. Then a uniform random sample $A \subseteq X$ of size at least*

$$\frac{c_1}{\epsilon\delta^2} \cdot \left(d \ln \frac{1}{\epsilon\delta} + \ln \frac{1}{\gamma} \right)$$

is a relative (ϵ, δ)-approximation for $(X, \mathcal{F})$ with probability at least $1 - \gamma$.

Remark: Theorem 3 is well-known; for completeness we give a proof in the appendix. In fact, symmetrization is not really necessary for finite set systems[3] and can be replaced by a more intuitive argument that makes it obvious, pedagogically, why the bound is independent of $|\mathcal{F}|$.

On the other hand, the role of chaining is to get rid of logarithmic factors that arise when applying union bound, by more carefully analyzing the failure probability for a collection of events. The key observation is that Chernoff's bound for a set $S \in \mathcal{F}$ *improves* as the size of S decreases. One can take advantage of this by partitioning each $S \in \mathcal{F}$ into a logarithmic number of smaller sets, each belonging to a distinct level, such that the levels strike a proper balance—the number of sets (arising from partitioning every $S \in \mathcal{F}$) increase each level, but their size across levels decreases geometrically. This way one gets an improved bound by applying the union bound separately to sets of different levels.

The resulting bound is captured in the next statement:

Theorem 4. *There exists an absolute constant c_2 such that the following holds. Let $(X, \mathcal{F})$ be a set system such that $|\mathcal{F}|_Y| \leq (e|Y|/d)^d$ for all $Y \subseteq X$, $|Y| \geq d$, and let $0 < \delta, \epsilon, \gamma < 1$ be given parameters. Then a uniform random sample $A \subseteq X$ of size at least*

$$c_2 \ \max \left\{ \frac{1}{\epsilon\delta} \ln \frac{|\mathcal{F}|}{\gamma}, \quad \frac{1}{\epsilon\delta^2} \ln \left(\frac{1}{\epsilon^d \gamma} \right) \right\}$$

is a relative (ϵ, δ)-approximation for $(X, \mathcal{F})$ with probability at least $1 - \gamma$.

The proof of this is given in Section 2.

The above two statements immediately imply a proof of Theorem 2: given $(X, \mathcal{F})$, apply Theorem 3 to get a relative $(\epsilon, \frac{\delta}{3})$-approximation $A_1 \subseteq X$ of $\mathcal{F}$, of size $O\left(\frac{1}{\epsilon\delta^2} \ln \frac{1}{\epsilon^d \delta^d \gamma}\right)$. Now apply Theorem 4 to $\mathcal{F}|_{A_1}$ to get a relative $(\epsilon, \frac{\delta}{3})$-approximation $A_2 \subseteq A_1$ of $\mathcal{F}|_{A_1}$, of size

$$O\left(\max \left\{ \frac{1}{\epsilon\delta} \ln \frac{\left(\frac{e}{d\epsilon\delta^2} \ln \frac{1}{\epsilon^d \delta^d \gamma}\right)^d}{\gamma}, \quad \frac{1}{\epsilon\delta^2} \ln \left(\frac{1}{\epsilon^d \gamma} \right) \right\} \right) = O\left(\frac{1}{\epsilon\delta^2} \cdot \left(d \ln \frac{1}{\epsilon} + \ln \frac{1}{\gamma} \right) \right).$$

Thus A_2 is a relative (ϵ, δ)-approximation of $\mathcal{F}$ of the required size.

[3] This is typically the case in its use in algorithms, computational geometry, combinatorics. The infinite case can usually be reduced to the finite case by a sufficiently fine grid, see [MWW93].

2 Proof of Theorem 4

Let $n = |X|$ and $t = |A|$. We use the following consequence of Theorem 3 (though better bounds exist [Hau95]):

Lemma 5. *There is an absolute constant c_3 such that the following holds. Let $\alpha \geq 2$ and let $\mathcal{P} \subseteq \mathcal{F}$ be an α-packing of $\mathcal{F}$; that is, for any $S, S' \in \mathcal{P}$, the symmetric difference of S and S', denoted by $\Delta(S, S')$, has size at least α. Then $|\mathcal{P}| \leq \left(\frac{c_3 n}{\alpha}\right)^{2d}$.*

Proof. Let $\mathcal{G} = \{\Delta(S, S') : S, S' \in \mathcal{P}\}$. By Theorem 3 there exists a relative $(\frac{\alpha}{n}, \frac{1}{2})$-approximation A' for $\mathcal{G}$ of size $\frac{4c_1 dn^2}{\alpha^2}$. Then $A' \cap S \neq A' \cap S'$ for any $S, S' \in \mathcal{P}$ since

$$|\Delta(S, S') \cap A'| \geq \frac{|\Delta(S, S')|\,|A'|}{n} - \frac{|A'|}{2} \cdot \max\left\{\frac{|\Delta(S, S')|}{n}, \frac{\alpha}{n}\right\} = \frac{1}{2} \cdot \frac{|\Delta(S, S')|\,|A'|}{n} > 0,$$

and so $|\mathcal{P}| = |\mathcal{P}|_{A'}| \leq \left(\frac{4ec_1 n^2}{\alpha^2}\right)^d \leq \left(\frac{4\sqrt{c_1} n}{\alpha}\right)^{2d}$. □

Set $k = \lceil \log \frac{1}{\delta} \rceil$ and for $i \in [0, k]$, let $\mathcal{P}_i$ be a *maximal* $\frac{\epsilon n}{2^i}$-packing of $\mathcal{F}$ and set $\mathcal{P}_{k+1} = \mathcal{F}$. For any $S \in \mathcal{P}_{i+1} \setminus \mathcal{P}_i$ there exists a set $F_S \in \mathcal{P}_i$ such that $|\Delta(S, F_S)| < \frac{\epsilon n}{2^i}$. Define

$$\mathcal{A}_i = \{S \setminus F_S \colon S \in \mathcal{P}_{i+1} \setminus \mathcal{P}_i\} \qquad \text{and} \qquad \mathcal{B}_i = \{F_S \setminus S \colon S \in \mathcal{P}_{i+1} \setminus \mathcal{P}_i\}.$$

Lemma 5 implies that

$$|\mathcal{A}_i|, |\mathcal{B}_i| \leq |\mathcal{P}_{i+1}| \leq \left(\frac{c_3 \cdot 2^i}{\epsilon}\right)^{2d}.$$

Claim 6. *Let $\epsilon_i = \sqrt{(i+1)/2^i}\,\epsilon$. With probability $1 - \gamma$, A is simultaneously (i) a relative (ϵ, δ)-approximation for $\mathcal{A}_k \cup \mathcal{B}_k$, and (ii) a relative (ϵ_i, δ)-approximation for $\mathcal{A}_i \cup \mathcal{B}_i$ for all $i \in [0, k-1]$, and (iii) a relative (ϵ, δ)-approximation for $\mathcal{P}_0$.*

Proof. (i) Each set in $\mathcal{A}_k \cup \mathcal{B}_k$ has size less than $\frac{\epsilon n}{2^k} \leq \epsilon n \delta$ and so Theorem A with $\eta = \delta t \epsilon$ implies that, for a large-enough value of c_2, this fails with probability at most

$$|\mathcal{F}| \cdot 2\exp\left(-\frac{\delta^2 t^2 \epsilon^2 \cdot n}{2\epsilon n \delta \cdot t + \delta t \epsilon \cdot n}\right) = |\mathcal{F}| \cdot 2\exp\left(-\frac{\delta \epsilon t}{3}\right) \leq \frac{\gamma}{3}.$$

(ii) For a fixed $S \in \mathcal{A}_i \cup \mathcal{B}_i$, as $|S| \leq \frac{\epsilon n}{2^i} \leq \epsilon_i n$, applying Theorem A with $\eta = \delta t \epsilon_i$ implies that the failure probability for S is at most

$$2\exp\left(-\frac{\delta^2 t^2 \epsilon_i^2 n}{2|S| t + \delta \epsilon_i t n}\right) \leq 2\exp\left(-\frac{\delta^2 t \epsilon^2 (i+1)/2^i}{2\epsilon/2^i + \delta\epsilon\sqrt{(i+1)/2^i}}\right) \leq 2\exp\left(-\frac{\epsilon \delta^2 t (i+1)}{3}\right).$$

Thus the overall probability failure, for a large-enough value of c_2, is at most

$$\sum_{i=0}^{k-1} |\mathcal{A}_i \cup \mathcal{B}_i| \cdot 2\exp\left(-\frac{\epsilon \delta^2 t (i+1)}{3}\right) \leq \sum_{i=0}^{k-1} 2\left(\frac{c_3 \cdot 2^i}{\epsilon}\right)^{2d} 2\left(\epsilon^d \gamma\right)^{c_2 (i+1)/3} \leq \gamma \sum_{i=1}^{\infty} \frac{1}{5^i} \leq \frac{\gamma}{3}.$$

(iii) Theorem 1 implies that this failure probability is at most $\frac{\gamma}{3}$ if $t \geq \frac{3}{\epsilon\delta^2} \ln \frac{2\left(\frac{c_3}{\epsilon}\right)^{2d}}{(\gamma/3)}$. □

Observe that for any set $S \in \mathcal{F}$, there exists a set $S_k \in \mathcal{P}_k$, with $A_k = S \setminus S_k \in \mathcal{A}_k$ and $B_k = S_k \setminus S \in \mathcal{B}_k$, such that $S = (S_k \setminus B_k) \cup A_k$. Similarly, one can express S_k in terms of $S_{k-1} \in \mathcal{P}_{k-1}$, $A_{k-1} \in \mathcal{A}_{k-1}$,

$B_{k-1} \in \mathcal{B}_{k-1}$ and so on until we reach $S_0 \in \mathcal{P}_0$. Thus using Claim 6, with probability at least $1-\gamma$,

$$\begin{aligned}
&\left|\frac{|S|}{n} - \frac{|A\cap S|}{t}\right| = \left|\frac{|S_k|}{n} - \frac{|B_k|}{n} + \frac{|A_k|}{n} - \left(\frac{|A\cap S_k|}{t} - \frac{|A\cap B_k|}{t} + \frac{|A\cap A_k|}{t}\right)\right| \\
&\overset{(i)}{\leq} \left|\frac{|S_k|}{n} - \frac{|A\cap S_k|}{t}\right| + \delta\max\left\{\epsilon, \frac{|A_k|}{n}\right\} + \delta\max\left\{\epsilon, \frac{|B_k|}{n}\right\} = \left|\frac{|S_k|}{n} - \frac{|A\cap S_k|}{t}\right| + 2\delta\epsilon \;\leq\; \cdots \\
&\overset{(ii)}{\leq} \left|\frac{|S_0|}{n} - \frac{|A\cap S_0|}{t}\right| + 2\delta\sum_{j=0}^{k-1}\epsilon_j + 2\delta\epsilon \\
&\overset{(iii)}{\leq} \delta\max\left\{\epsilon, \frac{|S_0|}{n}\right\} + 14\delta\epsilon \leq \delta\frac{|S|}{n} + 16\delta\epsilon \leq 2\delta\max\left\{\frac{|S|}{n}, 16\epsilon\right\},
\end{aligned}$$

where the second-last step uses the fact that $|S_0| \leq |S| + \sum_{j=0}^{k}|B_i| \leq |S| + \sum_{j=0}^{\infty}\frac{\epsilon n}{2^j} \leq |S| + 2\epsilon n$.

Therefore A is a relative $(16\epsilon, 2\delta)$-approximation of $\mathcal{F}$ with probability at least $1-\gamma$. Repeating the same arguments with $\delta' = \delta/2$ and $\epsilon' = \epsilon/16$ we get a relative (ϵ, δ)-approximation of $\mathcal{F}$, as required. $\square$

References

[AB09] M. Anthony and P. L. Bartlett. *Neural network learning: Theoretical foundations.* Cambridge University Press, 2009.

[AS12] N. Alon and J. Spencer. *The Probabilistic Method.* John Wiley, 2012.

[BCM93] H. Brönnimann, B. Chazelle, and J. Matoušek. Product range spaces, sensitive sampling, and derandomization. *Proc. Symposium on Foundations of Computer Science*, pages 400–409, 1993.

[Cha00] B. Chazelle. *The Discrepancy Method: Randomness and Complexity.* Cambridge University Press, New York, NY, USA, 2000.

[DGL96] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition.* Springer, Berlin, 1996.

[Hau95] D. Haussler. Sphere Packing Numbers for Subsets of the Boolean n-Cube with Bounded Vapnik-Chervonenkis Dimension. *J. Comb. Theory, Ser. A*, 69(2):217–232, 1995.

[HP11] S. Har-Peled. *Geometric Approximation Algorithms.* American Mathematical Society, Boston, MA, USA, 2011.

[HS11] S. Har-Peled and M. Sharir. Relative (p, ϵ)-Approximations in Geometry. *Discrete & Computational Geometry*, 45(3):462–496, 2011.

[HW87] D. Haussler and E. Welzl. ε-nets and simplex range queries. *Discrete & Computational Geometry*, 2:127–151, 1987.

[KPW92] J. Komlós, J. Pach, and G. Woeginger. Almost tight bounds for ε-nets. *Discrete & Computational Geometry*, 7:163–173, 1992.

[KT59] AN Kolmogorov and VM Tikhomirov. ϵ-entropy and ϵ-capacity. *Uspekhi Mat. Nauk*, 14:3–86, 1959.

[KV94] M. J. Kearns and U. V. Vazirani. *An Introduction to Computational Learning Theory.* MIT Press, 1994.

[LLS01] Y. Li, P. M. Long, and A. Srinivasan. Improved Bounds on the Sample Complexity of Learning. *J. Comput. Syst. Sci.*, 62(3):516–527, 2001.

[Mat99] J. Matoušek. *Geometric Discrepancy: An Illustrated Guide.* Springer, 1999.

[Mat02] J. Matoušek. *Lectures in Discrete Geometry.* Springer-Verlag, New York, NY, 2002.

[MWW93] J. Matoušek, E. Welzl, and L. Wernisch. Discrepancy and approximations for bounded VC-dimension. *Combinatorica*, 13(4):455–466, 1993.

[Sau72] N. Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13:145–147, 1972.

[She72] S. Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41:247–261, 1972.

[Tal94] M. Talagrand. Sharper bounds for Gaussian and empirical processes. *Annals of Probability*, 22:28–76, 1994.

[Tal16] M. Talagrand. *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems.* Springer Berlin Heidelberg, 2016.

[VC71] V. Vapnik and A. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16(2):264–280, 1971.

Proof of Theorem 3

The folklore argument is similar to the discrepancy-based argument in [MWW93] used for ϵ-approximations, though it is somewhat simpler as it does not need discrepancy, and it applies to the more general notion of a relative (ϵ, δ)-approximation.

To see the intuition, observe that since $|\mathcal{F}| \leq (e|X|/d)^d$, the bound of Theorem 1 depends only on $|X|$—in particular that a random sample $A_1 \subseteq X$ of size $O\left(\frac{1}{\epsilon\delta^2} \ln |X|^d\right) = O\left(\frac{d}{\epsilon\delta^2} \ln |X|\right)$ is a relative (ϵ, δ)-approximation. The size of A_1 is much smaller than that of X and so applying Theorem 1 again to $\mathcal{F}|_{A_1}$ gives a relative (ϵ, δ)-approximation $A_2 \subseteq A_1$ for $\mathcal{F}|_{A_1}$, with

$$|A_2| = O\left(\frac{1}{\epsilon\delta^2} \ln |A_1|^d\right) = O\left(\frac{d}{\epsilon\delta^2} \ln\left(\frac{d}{\epsilon\delta^2} \ln |X|\right)\right) = O\left(\frac{d}{\epsilon\delta^2} \ln \frac{d}{\epsilon\delta} + \frac{d}{\epsilon\delta^2} \ln \ln |X|\right).$$

The size of A_2 is again much smaller than that of A_1. Furthermore, it follows immediately from the definition of relative (ϵ, δ)-approximations that A_2 is a relative $(\epsilon, 3\delta)$-approximation for $\mathcal{F}$. With each successive application of Theorem 1, the size of the set decreases rapidly, while the error of approximation increases only linearly, giving the required bound that is independent of $|\mathcal{F}|$.

Now we present the formal proof. Let $T(\epsilon, \delta, \gamma)$ be the smallest integer such that a uniform random sample of size at least $T(\epsilon, \delta, \gamma)$ from X is a relative (ϵ, δ)-approximation for $\mathcal{F}$ with probability at least $1 - \gamma$. When $\delta \leq \frac{1}{\sqrt{|X|}}$, we have $|X| \leq \frac{1}{\delta^2}$ and thus $T(\epsilon, \delta, \gamma)$ is upper-bounded as required.

Otherwise, a random sample $A' \subseteq X$ of size $T\left(\epsilon, \frac{\delta}{3}, \frac{\gamma}{2}\right)$ is a relative $\left(\epsilon, \frac{\delta}{3}\right)$-approximation for $\mathcal{F}$ with probability at least $1 - \frac{\gamma}{2}$. By Theorem 1 let A be a random sample of A' that is a relative $\left(\epsilon, \frac{\delta}{3}\right)$-approximation for $\mathcal{F}|_{A'}$ with probability $1 - \frac{\gamma}{2}$. Thus A is a uniform random sample of X that is a relative (ϵ, δ)-approximation for $\mathcal{F}$ with probability at least $1 - \gamma$, implying the recurrence

$$T(\epsilon, \delta, \gamma) \leq |A| = \frac{3}{\epsilon(\delta/3)^2} \ln \frac{2\,|\mathcal{F}|_{A'}|}{(\gamma/2)} \leq \frac{27}{\epsilon\delta^2} \ln\left(\frac{4}{\gamma}\left(\frac{e\,T\left(\epsilon, \frac{\delta}{3}, \frac{\gamma}{2}\right)}{d}\right)^d\right).$$

The required bound on $T(\epsilon, \delta, \gamma)$ now follows by induction. As $\left(1 + \frac{1}{d} \ln \frac{2}{\gamma}\right)^d \leq \frac{2}{\gamma}$,

$$\frac{27}{\epsilon\delta^2} \ln\left(\frac{4}{\gamma}\left(\frac{e\,\frac{9\,c_1}{\epsilon\delta^2}\left(d \ln \frac{3}{\epsilon\delta} + \ln \frac{2}{\gamma}\right)}{d}\right)^d\right) \leq \frac{27}{\epsilon\delta^2} \ln\left(\frac{4}{\gamma}\left(\frac{e\,27\,c_1}{\epsilon^2\delta^3}\right)^d\left(1 + \frac{1}{d} \ln \frac{2}{\gamma}\right)^d\right) \leq \frac{c_1}{\epsilon\delta^2} \ln \frac{1}{(\epsilon\delta)^d\,\gamma},$$

for a large-enough c_1. $\square$