The VC-dimension of quadratic residues in finite fields

Anurag Sahay

University of Rochester

anuragsahay@rochester.edu

18th November, 2022

(joint work with Brian McDonald and Emmett Wyman)

Quadratic residues

Let $\mathbb{F}_q$ be a finite field with $q = p^s$ elements. Our object of study is the set of quadratic residues (i.e. squares) in $\mathbb{F}_q$,

$$\mathcal{S}_q = \{x^2 : x \in \mathbb{F}_q^\times\}.$$

There's some issues surrounding whether it is appropriate to take $0 \in \mathcal{S}_q$, but we will brush this under the rug for the purposes of our discussion.

More generally, our methods apply to the unique subgroup of $\mathbb{F}_q^\times$ having index r,

$$\Gamma^{(r)} = \{x^r : x \in \mathbb{F}_q^\times\},$$

provided that r is fixed.

Vapnik-Chervonenkis dimension in groups

Definition (VC-dimension in groups)

Let $(G, +)$ be an Abelian group, and let $S \subseteq G$ be an arbitrary subset. We say that S *shatters* $Y = \{y_1, \cdots, y_n\} \subseteq G$ if for any $\emptyset \subseteq A \subseteq Y$, there is an $x \in G$ such that

$$y_j \in A \iff y_j \in (S + x)$$

for $1 \leqslant j \leqslant n$.

Further, we define the VC-dimension of S to be the cardinality of the largest set Y that is shattered by S, and we denote it by $\mathrm{VCdim}_G(S)$

Relationship to other notions of VC-dimension

If one defines VC-dimension via set systems, this is the same definition as the VC-dimension relative to the set system of additive translates of S,

$$\mathscr{P}_S = \{S + x : x \in G\}.$$

If one defines it via classifiers, this is the same definitions as the VC-dimension relative to the indicator functions,

$$\mathscr{H}_S = \{\mathbb{1}_{S+x} : x \in G, \}.$$

An upper bound on VC-dimension

A standard argument gives that $\mathrm{VCdim}_G(S) \leqslant \log_2 |G|$.

VC-dimension of quadratic residues

This brings us to our main question, which is to understand how

$$\mathrm{VCdim}_{\mathbb{F}_q}(\mathcal{S}_q)$$

grows as $q \to \infty$ along the prime powers. Our conjecture is the following:

Conjecture (McDonald–S.–Wyman, 2022+)

We have that

$$\mathrm{VCdim}_{\mathbb{F}_q}(\mathcal{S}_q) = (1 + o(1)) \log_2 q.$$

as $q \to \infty$.

Note that this says that asymptotically, the VC-dimension is as large as it can possibly be.

Basis for the conjecture

The main basis for the conjecture comes from sum-product heuristics. The VC-dimension in a group generally measures how much group structure the set has – the lower the VC-dimension, the more structured it is.

In general, we expect additive and multiplicative structure to largely be orthogonal. The set $\mathcal{S}_q$ is a multiplicative subgroup, and so we expect that it would have very little additive structure, and thus, very high VC-dimension.

The conjecture can also be restated in terms of Paley graphs, and in this context can be seen as a manifestation of the pseudorandomness of Paley graphs.

VC-dimension in graphs

Definition (VC-dimension in graphs)

Let G be a directed graph with edge set E. We say that E shatters $Y = \{y_1, \cdots, y_n\} \subseteq G$ if for any $\emptyset \subseteq A \subseteq Y$, there is an $x \in G$ such that

$$y_j \in A \iff x \to y_j \text{ in } G$$

for $1 \leqslant j \leqslant n$. Further, we define the VC-dimension of E to the be cardinality of the largest set Y that is shattered by E, and we denote it by $\mathrm{VCdim}'_G(E)$

This generalizes the definition for groups, since if E is the edge set of the Cayley (di)graph $\mathrm{Cay}(G, S)$, then $\mathrm{VCdim}_G(S) = \mathrm{VCdim}'_G(E)$.

Paley (di)graphs

The Paley (di)graph $\mathcal{P}_q$ on q vertices is defined as $\mathrm{Cay}(\mathbb{F}_q, \mathcal{S}_q)$. That is, the underlying edge set of $\mathcal{P}_q$ is $\mathbb{F}_q$, and

$$u \to v \text{ in } \mathcal{P}_q \iff (v - u) \in \mathcal{S}_q.$$

$\mathcal{P}_q$ is well-known to be a pseudorandom graph (a deterministic model for a random graph); this follows, for example, by the seminal work of Chung–Graham–Wilson. This might also suggest that the VC-dimension should be as large as possible.

The Main Theorem

Theorem (McDonald–S.–Wyman, 2022+)

For any $\epsilon > 0$, we have that for all $q \gg_\epsilon 1$, and for every subset $Y \subseteq \mathbb{F}_q$ with $|Y| \leqslant (\frac{1}{2} - \epsilon) \log_2 q$, the set of squares $\mathcal{S}_q$ shatters Y.

This immediately implies as a corollary that

$$\mathrm{VCdim}_{\mathbb{F}_q}(\mathcal{S}_q) \geqslant (\tfrac{1}{2} + o(1)) \log_2 q.$$

The Main Lemma

The key fact is the following lemma:

Lemma

Let $r \geqslant 2$, $n \geqslant 1$, $Y = \{y_1, \cdots, y_n\} \subseteq \mathbb{F}_q$ *and* $\mathbf{t} = (t_1, \cdots, t_n) \in (\mathbb{F}_q^\times)^n$. *Then,*

$$\Pr_{x \in \mathbb{F}_q}\left[y_j - x \in t_j \Gamma^{(r)}, 1 \leqslant j \leqslant n\right] = \frac{1}{r^n} + O\Big(\frac{n}{q^{1/2}}\Big),$$

where the implicit constant can be chosen to be 1 *and hence is uniform in all parameters, and* x *is sampled from the uniform distribution.*

Recall that $\Gamma^{(r)}$ is the unique multiplicative subgroup of index r, and here $t_j\Gamma^{(r)}$ is a coset of that subgroup.

Pseudorandomness of bounded index subgroups

Note that for a fixed j, y_j, t_j,

$$\Pr_{x \in \mathbb{F}_q}\left[y_j - x \in t_j \Gamma^{(r)}\right] = \frac{1}{r},$$

while the lemma says

$$\Pr_{x \in \mathbb{F}_q}\left[y_j - x \in t_j \Gamma^{(r)}, 1 \leqslant j \leqslant n\right] = \frac{1}{r^n} + O\Big(\frac{n}{q^{1/2}}\Big),$$

Thus, the lemma can be viewed as saying that for a fixed choice of n and r, the values of $x - y_j$ all equidistribute in the cosets of $\Gamma^{(r)}$ independently of each other as $q \to \infty$. This may be seen as a pseudorandomness phenomenon for bounded index subgroups of $\mathbb{F}_q^\times$.

Lemma $\implies$ Theorem

Let $\epsilon > 0$, and $n = |Y| \leqslant (\frac{1}{2} - \epsilon) \log_2 q$. This implies that

$$\Pr_{x \in \mathbb{F}_q} \left[y_j - x \in t_j \mathcal{S}_q, 1 \leqslant j \leqslant n \right] > 0.$$

uniformly in $Y = \{y_1, \cdots, y_n\}$ and $\mathbf{t} \in (\mathbb{F}_q^\times)^n$, for q large enough since by the main lemma, the probability above is at least

$$\frac{1}{2^n} - \frac{n}{q^{1/2}} \geqslant \frac{1}{q^{1/2-\epsilon}} - \frac{\log_2 q}{q^{1/2}}.$$

Since $\log_2 q = o(q^\epsilon)$ as $q \to \infty$,

$$\Pr_{x \in \mathbb{F}_q} \left[y_j - x \in t_j \Gamma^{(r)}, 1 \leqslant j \leqslant n \right] \gg q^{-1/2+\epsilon} > 0.$$

as claimed.

Lemma $\implies$ Theorem (contd.)

To prove that $Y = \{y_1, \cdots, y_n\}$ is shattered by $\mathcal{S}_q$, we need to find, for every $\emptyset \subseteq A \subseteq Y$, an element $x \in \mathbb{F}_q$ with the property that

$$A = Y \cap (\mathcal{S}_q + x).$$

We do this as follows. Define $\mathbf{t} = (t_1, \cdots, t_n)$ by

$$t_j = \begin{cases} 1 & \text{if } y_j \in A, \\ -1 & \text{if } y_j \notin A. \end{cases}$$

If $q \neq 3 \bmod 4$, then we replace -1 here by some non-square in $\mathbb{F}_q$.

By the previous slide, there must be an $x \in \mathbb{F}_q$ with the property that $y_j - x \in t_j\mathcal{S}_q$ for each $1 \leqslant j \leqslant n$. This means that $y_j \in (\mathcal{S}_q + x)$ if and only $t_j = 1$ which, by construction, happens if and only if $y_j \in A$. Since this is true for any choice of A, we have shown that Y is shattered by $\mathcal{S}_q$, proving the theorem.

Fourier analysis on finite Abelian groups

Recall that a character is a group homomorphism $\chi : G \to S^1 \subseteq \mathbb{C}^\times$. The character theory of finite Abelian groups tells us that there are $|G|$ characters on G, and further that if $H \subseteq G$, then

$$\frac{1}{|G/H|} \sum_{\chi \in \widehat{G/H}} \chi(x)\overline{\chi(t)} = \begin{cases} 1 & \text{if } x \in tH, \\ 0 & \text{otherwise.} \end{cases}$$

Here the sum runs over characters of G/H.

Specializing to $G = \mathbb{F}_q^\times$ and $H = \Gamma^{(r)}$ gives

$$\mathbb{1}_{t\Gamma^{(r)}}(x) = \frac{1}{r} \sum_{\chi \in (\mathbb{F}_q^\times/\Gamma^{(r)})^\wedge} \chi(x)\overline{\chi(t)}.$$

Since $\mathbb{F}_q^\times$ is cyclic, so is $(\mathbb{F}_q^\times/\Gamma^{(r)})^\wedge$, and hence the above can be written as

$$\mathbb{1}_{t\Gamma^{(r)}}(x) = \frac{1}{r}\left(1 + \sum_{k=1}^{r-1} \chi_r^k(x)\overline{\chi_r^k(t)}\right),$$

where χ_r is a generator of $(\mathbb{F}_q^\times/\Gamma^{(r)})^\wedge$.

The Weil bound

Our main tool is a deep result from arithmetic geometry known as the Weil bound for multiplicative character sums, which we will use as a black box.

Lemma (Weil)

For $r \geqslant 2$, suppose that $f \in \mathbb{F}_q[x]$ has n distinct roots and that f is not an rth power. Then, we have

$$\left| \sum_{x \in \mathbb{F}_q} \chi_r(f(x)) \right| \leqslant (n-1)\sqrt{q}.$$

This is equivalent to the Riemann hypothesis for curves over finite fields, and was proved by André Weil. An account of this result which assumes minimal knowledge of algebraic geometry can be found in Chapter 11 of Iwaniec & Kowalski.

Proof Sketch

For conciseness, let $E \subseteq \mathbb{F}_q$ be the event in the lemma (namely, that $y_j - x \in t_j\Gamma^{(r)}$ for every $1 \leqslant j \leqslant n$). Further, for $1 \leqslant j \leqslant n$, define $E_j \subseteq \mathbb{F}_q$ to be the event that $y_j - x \in t_j\Gamma^{(r)}$. Then,

$$\mathbb{1}_E(x) = \prod_{j=1}^{n} \mathbb{1}_{E_j}(x) = \prod_{j=1}^{n} \mathbb{1}_{t_j\Gamma^{(r)}}(y_j - x).$$

Applying orthogonality and expanding out the product,

$$\frac{1}{r^n}\prod_{j=1}^{n}\left(1 + \sum_{k=1}^{r-1} \chi_r^k(y_j - x)\overline{\chi_r^k(t_j)}\right) = \frac{1}{r^n}\left(1 + \sum_{\substack{0 \leqslant \mathbf{k} \leqslant r-1 \\ \mathbf{k} \neq \mathbf{0}}} b(\mathbf{k}, \mathbf{t})\chi_r(f_{\mathbf{k}}(x))\right),$$

where

$$b(\mathbf{k}, \mathbf{t}) = \overline{\chi_r}\left(\prod_{j=1}^{n} t_j^{k_j}\right), \qquad f_{\mathbf{k}}(x) = \prod_{j=1}^{n}(y_j - x)^{k_j}.$$

Proof Sketch (contd.)

By linearity of expectation

$$\Pr_{x \in \mathbb{F}_q}[E] = \mathbb{E}_x[\mathbb{1}_E(x)] = \frac{1}{r^n} + \frac{1}{r^n} \sum_{\substack{0 \leqslant \mathbf{k} \leqslant r-1 \\ \mathbf{k} \neq \mathbf{0}}} b(\mathbf{k}, \mathbf{t}) \mathbb{E}_x[\chi_r(f_{\mathbf{k}}(x))].$$

Here the expectation is over $x \in \mathbb{F}_q$. Now, $|b(\mathbf{k}, \mathbf{t})| \leqslant 1$. Thus, by the triangle inequality, $\Pr[E] - r^{-n}$ is bounded by

$$\frac{1}{r^n} \sum_{\substack{0 \leqslant \mathbf{k} \leqslant r-1 \\ \mathbf{k} \neq \mathbf{0}}} \left| \mathbb{E}_x[\chi_r(f_{\mathbf{k}}(x))] \right| \leqslant \max_{\substack{0 \leqslant \mathbf{k} \leqslant r-1 \\ \mathbf{k} \neq \mathbf{0}}} \left| \mathbb{E}_x[\chi_r(f_{\mathbf{k}}(x))] \right|.$$

However, this last quantity is bounded by $\frac{n}{q^{1/2}}$ using the Weil bound, proving the lemma.

Results about $\Gamma^{(r)}$

The general conjecture is the following:

Conjecture (McDonald–S.–Wyman, 2022+)

We have that

$$\mathrm{VCdim}_{\mathbb{F}_q}(\Gamma^{(r)}) = (1 + o_r(1)) \log_2 q.$$

as $q \to \infty$.

Our partial progress is:

Theorem (McDonald–S.–Wyman, 2022+)

For any $\epsilon > 0$, $r \geqslant 2$ *we have that for all* $q \gg_{r,\epsilon} 1$, *and for every subset* $Y \subseteq \mathbb{F}_q$ *with* $|Y| \leqslant (\frac{1}{2} - \epsilon) \log_r q$, *the set of* r*th powers* $\Gamma^{(r)}$ *shatters* Y.

This implies immediately that $\mathrm{VCdim}_{\mathbb{F}_q}(\mathcal{S}_q) \geqslant (\frac{1}{2} + o_r(1)) \log_r q$.

Thank You!

Appendix: Possible future questions

We posed some open questions connected to this work, including:

1. For any $\delta > 0$, is it always the case that for q large enough, there is a set $Y = \{y_1, \cdots, y_n\}$ with $n = |Y| \geqslant (\frac{1}{2} + \delta) \log_2 q$ such that $\mathcal{S}_q$ does not shatter?
2. What does the VC-dimension of an Erdős-Rényi random graph behave like? This problem has been studied in the sparse setting as $n \to \infty$ with $p = p(n) = o(1)$ by Anthony, Brightwell, and Cooper. However, there appear to be no results in the dense regime where p is constant as $n \to \infty$.
3. A similar question can be posed for random Cayley graphs, or other random graph models.

Appendix: Numerical evidence

We found three pieces of computational evidence:

1. A direct computation of the VC-dimension of $\mathcal{S}_q$ for prime values of q going up to 300.
2. A computation of the size of the largest shattered arithmetic progression in $\mathbb{F}_q$ for prime values of q going up to 200000.
3. A set of plots of the likelihoods that a random set whose size is a given proportion of the theoretical maximum VC-dimension, $\log_2 q$, is shattered for prime values

$$58 \approx 2^{5/0.85} \leqslant q \leqslant 2^{12/0.7} \approx 144716.$$

Appendix: Some Pictures I

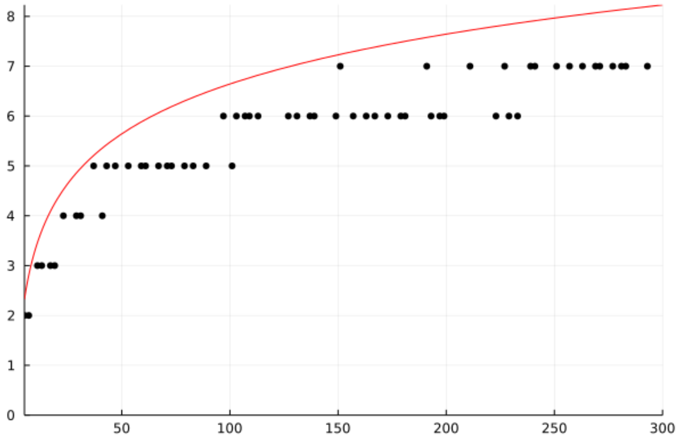

Figure: On the horizontal axis: primes q from 5 to 300. On the vertical axis: the size of the largest shattered subset found by the first experiment for this q. The red curve is the graph of $\log_2 q$, for reference.

Appendix: Some Pictures II

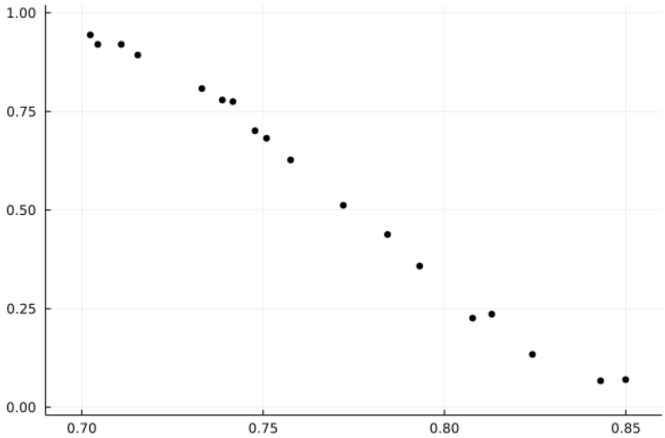

Figure: Along the vertical axis is probability. Along the horizontal axis is the proportion $n/\log_2(q)$ for $n = 5$. The plot is generated for randomly selected q in the range $0.7 \leqslant n/\log_2 q \leqslant 0.85$.

Appendix: Some Pictures III

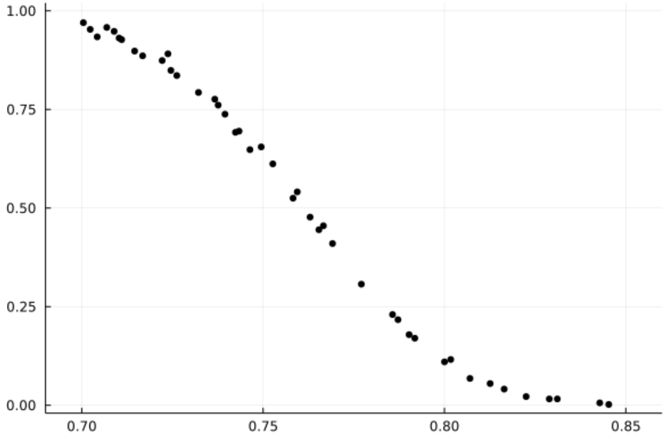

Figure: Along the vertical axis is probability. Along the horizontal axis is the proportion $n/\log_2(q)$ for $n = 6$. The plot is generated for randomly selected q in the range $0.7 \leqslant n/\log_2 q \leqslant 0.85$.

Appendix: Some Pictures IV

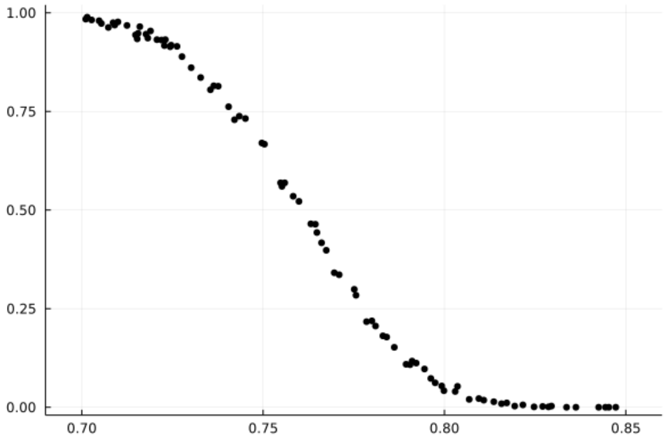

Figure: Along the vertical axis is probability. Along the horizontal axis is the proportion $n/\log_2(q)$ for $n = 7$. The plot is generated for randomly selected q in the range $0.7 \leqslant n/\log_2 q \leqslant 0.85$.

Appendix: Some Pictures V

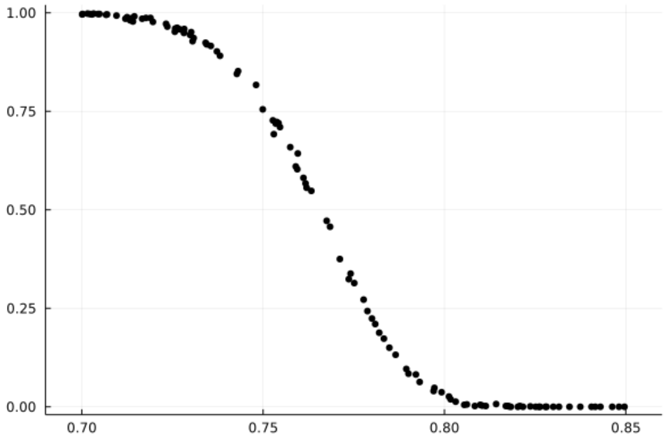

Figure: Along the vertical axis is probability. Along the horizontal axis is the proportion $n/\log_2(q)$ for $n = 8$. The plot is generated for randomly selected q in the range $0.7 \leqslant n/\log_2 q \leqslant 0.85$.

Appendix: Some Pictures VI

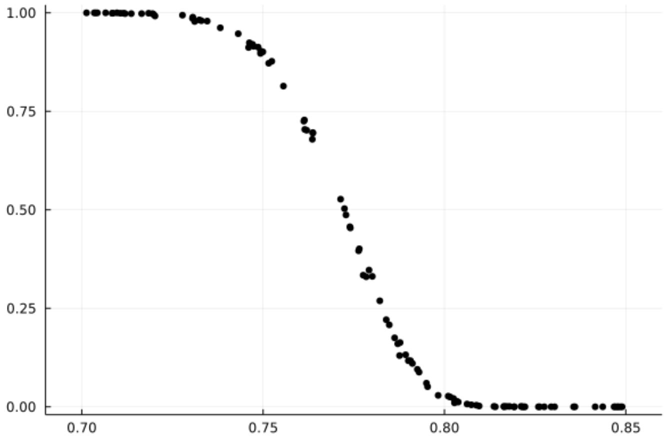

Figure: Along the vertical axis is probability. Along the horizontal axis is the proportion $n/\log_2(q)$ for $n = 9$. The plot is generated for randomly selected q in the range $0.7 \leqslant n/\log_2 q \leqslant 0.85$.

Appendix: Some Pictures VII

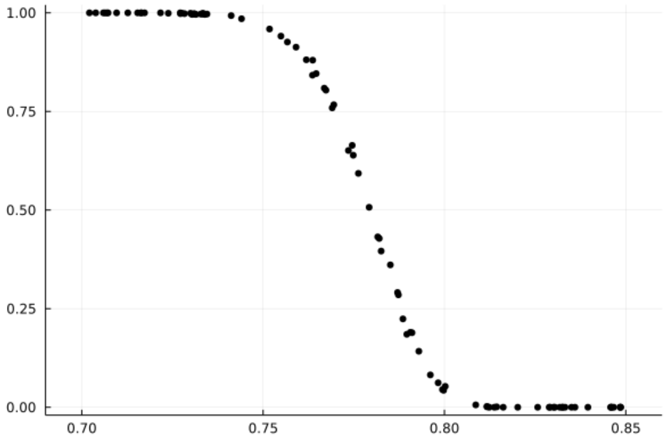

Figure: Along the vertical axis is probability. Along the horizontal axis is the proportion $n/\log_2(q)$ for $n = 10$. The plot is generated for randomly selected q in the range $0.7 \leqslant n/\log_2 q \leqslant 0.85$.

Appendix: Some Pictures VIII

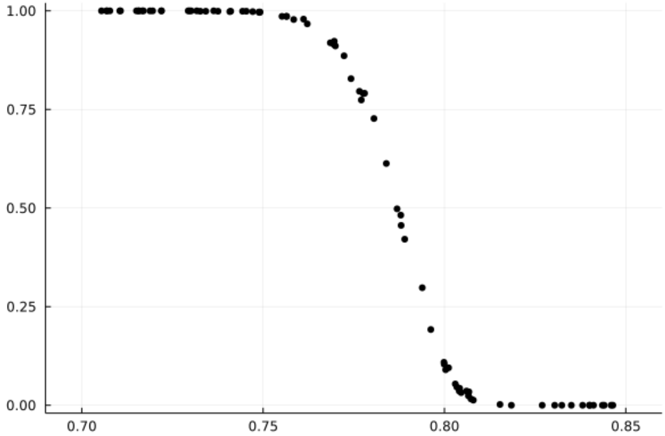

Figure: Along the vertical axis is probability. Along the horizontal axis is the proportion $n/\log_2(q)$ for $n = 11$. The plot is generated for randomly selected q in the range $0.7 \leqslant n/\log_2 q \leqslant 0.85$.

Appendix: Some Pictures IX

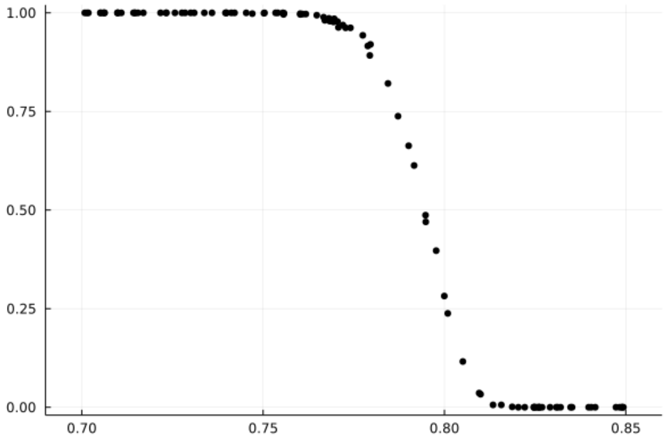

Figure: Along the vertical axis is probability. Along the horizontal axis is the proportion $n/\log_2(q)$ for $n = 12$. The plot is generated for randomly selected q in the range $0.7 \leqslant n/\log_2 q \leqslant 0.85$.

Thank You (x2)!