> [there was] an eccentric librarian called Porlock who in the last years of his dusty life had been engaged in examining old books for miraculous misprints such as the substitution of "l" for the second "h" in the word "hither." ... all he sought was the freak itself, the chance that mimics choice, the flaw that looks like a flower.
>
> Vladimir Nabokov, *The Vane Sisters*

1 Polling people

1.1 The model and the question

For this section, we will work with the following model. We have a set of n voters, called $1, 2, \dots, n$, and voter i has some opinion $b_i \in \{0, 1\}$; this opinion can represent a preference for a political party (Democrat or Republican), a position on some issue (Stay or Remain in the EU), or anything else. Everything here can be easily adapted to the case where the opinion is not binary (e.g. more than two political parties), but for simplicity we'll stick to the binary case.

Our goal is to estimate the fraction of the population with each opinion; with the normalization $b_i \in \{0, 1\}$, this fraction is simply

$$x = \frac{1}{n} \sum_{i=1}^{n} b_i.$$

To do this estimation, we will *poll*. For our purposes, we will assume that to poll, we pick some $k \leq n$, then randomly sample k voters, and ask their opinion. If $S \subseteq \{1, \dots, n\}$ is the set of randomly chosen voters, then we will compute

$$X = \frac{1}{k} \sum_{i \in S} b_i,$$

and return X as our estimate of x. Note that X is a random variable, since it depends on the random choice of the set S of cardinality k.

To measure how good our estimate is, there are two parameters, $\delta, \varepsilon \in (0, 1)$, that we're interested in. The condition we want is

$$\Pr\left(|X - x| > \delta\right) < \varepsilon.$$

In other words, we want that with probability at least $1 - \varepsilon$, our estimate X will be within δ of the true value x. The parameters δ and ε are usually called *accuracy* and *precision*, respectively, but perhaps better terminology (reflecting the notation) is *discrepancy* and *error*—a small δ means our estimate deviates little from the truth, while a small ε means that the odds of screwing up are small.

Now, let's suppose that we're running a polling firm, so we pick some values of δ and ε that we want to achieve. Standard choices are $\delta = 0.02$ and $\varepsilon = 0.05$, meaning that with 95% probability, our polled value will be within 2% of the truth. Having fixed these values, let's suppose that we begin varying n, the size of the population. For instance, suppose we wish to estimate the fraction of Democrats in Stanford, in Palo Alto, in Santa Clara county, and in California. For each of these, we conduct a separate poll according to the model above (pick k voters at random), and we want to achieve the same accuracy parameters δ, ε throughout. Then the most basic question one can ask is the following.

Question. *For fixed $\delta, \varepsilon \in (0,1)$, how large does k have to be (as a function of n) to ensure that $\Pr(|X - x| > \delta) < \varepsilon$?*

1.2 A digression on real-world polling

Before answering this question, I want to say a few words about real-world polling. In short, the model described above is wildly unrealistic, for a number of reasons, and this is why real-world polling is so hard. Loosely speaking, the job of a professional pollster is to figure out how to simulate the theoretical model in the real world.

One problem with the model is that we assumed that voters report their true opinion b_i when polled, but people can lie to pollsters. In political polling, this usually isn't a huge problem, though it is a cause for concern sometimes. For instance, the so-called "Bradley effect" says that in elections with one white and one non-white candidate, voters intending to vote for the white candidate will nevertheless lie about this to pollsters, for fear of seeming racist. It's not clear how real this effect is (for instance, some people worried about it before the 2008 election, but in the end Obama's popular vote margin was quite close to what the polls had predicted), but there is some evidence for it in past elections. More recently, in Israel's elections last month, one of the three exit polls was wildly off both from the other exit polls and the truth, and the lead pollster blamed[1] the discrepancy on people lying to her out of spite. However, in non-political contexts, this can be a big problem. Imagine, for instance, the IRS conducting a poll to determine how many people cheat on their taxes; everyone would say no, and the data would be useless. The same problem arises whenever one wishes to determine the frequency of something stigmatized or illegal, such as drug use, abortions, certain illneses, and so on. Many techniques have been developed to deal with this problem, most famously the "randomized response" technique. In the simplest variant, the IRS would ask respondents to secretly flip a coin, and then answer "yes" with probability 1/2, and answer the truth with probability 1/2; in this case, anyone answering "yes" could easily deny having committed tax fraud, while the IRS could easily compute the true percentage of tax cheaters from this data.

However, by far the biggest problem with our model is the assumption that we can get a uniformly random sample of k voters. In reality, getting a truly random sample is essentially impossible, for a number of reasons. Generally speaking, in America, the people who respond to polls tend to be older, wealthier, and whiter than the average citizen,

[1] See `https://www.globes.co.il/news/article.aspx?did=1001281723` (in Hebrew).

since those are the sorts of demographics that correlate with people who have the time and are willing to talk to a stranger about their political opinions. Moreover, many pollsters exacerbate this problem by doing things like only calling landlines (which means they're less likely to find younger voters) and only conducting their polls in English (leading to fewer Latino respondents). If these demographic traits were uncorrelated with political beliefs, this bias in the polling sample wouldn't matter, but in fact there's often a strong correlation (e.g. younger voters are less likely to own a landline and more likely to be Democrats). Because of this problem, pollsters employ a technique called *weighting*, wherein they pick a set of demographic traits, compute the average of responses within each demographic slice, and then compute a weighted average of these according to the relative populations. So, for instance, if a community is 70% black and 30% white, the pollster would multiply the average among black respondents by .7 and add it to .3 times the average among white respondents to get their final estimate. Unfortunately, weighting also has a host of problems associated with it; the biggest one is that if you partition a population into many small demographic slices and then conduct a poll, some of the slices might have a dangerously small sample size. One notorious example was in the 2016 election, when a U.S.C/LA Times poll consistently gave Trump a few more points than other polls. There were several reasons for this, but it seems that one of them is that they divided their panel of 3000 respondents into extremely fine demographic categories. Moreover, one of their panelists was a 19-year-old black man from Illinois who supported Trump. Since they had very few young black men in their panel, they weighted his opinion extremely highly (roughly 30 times more than the average respondent), and concluded that a large fraction of young black men in America were Trump supporters, which was not the case[2]. For what it's worth, I think that this poll should be commended rather than mocked: they picked a non-standard methodology, stuck with it even when it gave problematic results, and made public all their data so that independent analyses could be conducted. Many pollsters hide behind a layer of opacity and tweak their data to agree with the consensus (a process called "herding"), which can lead to bad polling misses.

Finally, let me mention that these questions about non-representative samples aren't purely academic, but in fact very pressing political issues. The 2020 census is coming up soon, and the Trump administration is trying to add a citizenship question to it. The Supreme Court is currently debating the legality of this question, but opponents argue that it would discourage certain minority communities from responding to the census; according to one government estimate, 6.5 million people might not be counted as a result of this question. This has real consequences, since census data is used to determine the allocation of some federal funds, as well as seats in the House of Representatives and the Electoral College.

[2]This analysis comes from `https://www.nytimes.com/2016/10/13/upshot/how-one-19-year-old-illinois-man-is-distorting-national-polling-averages.html`; see also `https://www.latimes.com/politics/la-na-pol-daybreak-poll-questions-20161013-snap-story.html` for a rebuttal.

1.3 The answer

Recall that we fixed some $\delta, \varepsilon \in (0,1)$, and were interested in how k depends on n in order to guarantee that $\Pr(|X - x| > \delta) < \varepsilon$. Astonishingly, the answer is that it *doesn't*.

Theorem 1. *For every $\delta, \varepsilon \in (0,1)$, there exists some $k \in \mathbb{N}$ such that for every integer $n > k$ and every $b_1, \dots, b_n \in \{0,1\}$, we have*

$$\Pr(|X - x| > \delta) < \varepsilon,$$

where

$$x = \frac{1}{n}\sum_{i=1}^{n} b_i \text{ and } X = \frac{1}{k}\sum_{i \in S} b_i,$$

where $S \subseteq \{1, \dots, n\}$ is a randomly chosen subset of cardinality k.

In order to slightly simplify the computations, we will assume that, rather than picking a random subset of k voters, we will instead sample a random voter k times. The difference is that we'll be drawing "without replacement", meaning that a voter can be sampled multiple times. Equivalently, we allow S to be a multiset, rather than an ordinary set. The difference is minute since, for $k \ll n$, we will likely pick k distinct voters anyway; moreover, everything we'll say below can be made to work for the previous model, but the computations are a tiny bit messier.

Since this result is so important and surprising, we'll present three proofs.

Proof 1. Suppose we sample k random voters one by one, each chosen uniformly at random among the n voters. Let $X_1, \dots, X_k \in \{0,1\}$ be the opinions of these voters. Then note that the X_i are independent random variables, with

$$X_i = \begin{cases} 0 & \text{with probability } 1 - x \\ 1 & \text{with probability } x \end{cases}$$

and $X = \frac{1}{k}\sum_{i=1}^{k} X_i$. Thus, X is the average of k independent random variables, each of mean x. The weak law of large numbers says that for any $\delta > 0$,

$$\lim_{k \to \infty} \Pr\left(\left|\frac{1}{k}\sum_{i=1}^{k} X_i - x\right| > \delta\right) = 0.$$

In particular, for some k sufficiently large, this quantity will be less than ε. Moreover, this "sufficiently large" only depends on δ and ε, and *not* on n. In fact, n is completely irrelevant in this argument, since all that matters is that we're adding up k independent random variables of mean x. □

One disadvantage of this high-level argument is that it tells us nothing about how large k has to be; it only asserts that some such k exists. We can actually get some quantitative information by avoiding the law of large numbers.

Proof 2. Let X_i be as above. By linearity of expectation, we know that

$$\mathbb{E}[X] = \frac{1}{k}\sum_{i=1}^{k} \mathbb{E}[X_i] = x.$$

Moreover, since the X_i are independent, we can also compute

$$\mathrm{Var}(X) = \frac{1}{k^2}\sum_{i=1}^{k} \mathrm{Var}(X_i) = \frac{1}{k}x(1-x) \leq \frac{1}{4k},$$

where the final inequality uses the fact that $x(1-x) \leq 1/4$ for $x \in [0,1]$. Therefore, if we let $\sigma = \sqrt{\mathrm{Var}(X)}$ denote the standard deviation of X, we find that $\sigma \leq 1/2\sqrt{k}$. Chebyshev's inequality says that for any $\lambda > 0$,

$$\Pr(|X - x| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}.$$

Plugging in $\lambda = \delta/\sigma$, we find that

$$\Pr(|X - x| \geq \delta) \leq \frac{\sigma^2}{\delta^2} \leq \frac{1}{4k\delta^2}.$$

Therefore, if we pick $k \geq (4\varepsilon\delta^2)^{-1}$, we find that

$$\Pr(|X - x| \geq \delta) \leq \varepsilon,$$

as desired. Thus, we even find that the dependence of k on δ and ε isn't so bad; it's linear in one and quadratic in the other. However, crucially, again, k does not depend on n. □

As it turns out, Chebyshev's inequality is very general, and is rather weak when applied to sums of bounded independent variables. We can slightly improve the dependence of k on ε by using a stronger bound that holds in this case.

Proof 3. The Chernoff bound (also called, in this context, Hoeffding's inequality) gives exponential tail bounds for sums of independent $\{0,1\}$-valued random variables. For X_i and X as above, it says that

$$\Pr(|X - x| > \delta) < 2e^{-2\delta^2 k}.$$

Therefore, if we take $k \geq \log(2/\varepsilon)/2\delta^2$, then we obtain the desired result. Note that this improves the dependence of k on ε from linear to logarithmic, and in fact this dependence (as well as the quadratic dependence on δ) is correct; if we poll fewer than $O(\delta^{-2}\log(\varepsilon^{-1}))$ voters, then we will not be able to obtain the accuracy guarantees δ and ε. □

All three proofs above used some black-box result (though some are harder than others; for instance, Chebyshev's inequality is quite simple to prove). However, it is also possible to prove Theorem 1 directly, by bounding a bunch of binomial coefficients. The computation is messy and fairly uninteresting, but it's important to realize that Theorem 1 is not some deep fact about hard probability theory: instead, it's a straightforward fact about objects that one can work with directly.

2 Property testing

Above, our notion of polling us wanted to approximate the true average x to within a discrepancy of δ with probability at least $1-\varepsilon$. A task simpler than approximation, which we will deal with from now on, is that of *testing:* we are promised that the value of x is either 0 or at least δ, and we only wish to determine which scenario we are in, rather than approximating the true value of x. In the language of polling above, this would be like asking us to determine whether there are any Republicans in some voting population, given the promise that there are either none or at least, say 2%. As before, we want to succeed with probability at least $1-\varepsilon$. Note that if we can accurately approximate, then we can also accurately test, simply by estimating x to within an accuracy of $\delta/2$ with high probability, and then checking if our approximate value is below or above $\delta/2$. In fact, for polling as above, testing and approximating are more or less equivalent problems, but we will soon consider a more general setup where testing is more natural than approximating.

In our previous scenario, we had some finite set of "voters", and we had a unary relation on them (each element in the set was assigned a $\{0,1\}$ value). A natural generalization is to suppose that we have some *binary* relation, wherein every (unordered) pair of voters is assigned a 0 or a 1. In other words, we are given a graph on the vertex set $\{1,\dots,n\}$. We will be interested in determining whether this graph has certain properties (e.g. whether it's connected, or planar, or 17-colorable). Of course, if we know the graph, then we can just determine the answer to this question, but as with the simple polling above, we're interested in approximately determining the answer when we aren't allowed to see the whole graph, but are instead allowed only to sample pairs and ask them if they form edges. As before, since we're only allowed a randomized procedure like this, we cannot ask for certainty guarantees, but instead only want to be approximately correct most of the time. More formally, we make the following definitions.

Definition 1. Given two graphs G, H with vertex sets $\{1,\dots,n\}$, and a number $\delta \in (0,1)$, we say that G is δ-*close* to H if

$$|E(G) \triangle E(H)| \leq \delta \binom{n}{2}.$$

Equivalently, we can get from G to H by adding or deleting at most $\delta\binom{n}{2}$ edges.

Definition 2. Given a graph property P (e.g. connectedness), a graph G on n vertices, and a number $\delta \in (0,1)$, we say that G is δ-*close* to P if G is δ-close to some graph H satisfying the property P. In particular, if G already satisfies property P, then it is δ-close to P for all $\delta > 0$. If G is not δ-close to P, we say it's δ-*far* from P.

Now, fix some graph property P and some parameters $\delta, \varepsilon \in (0,1)$. Suppose we are given a graph G on n vertices, and we are promised that either G satisfies property P or it is δ-far from P. We would like to conduct a poll to figure out which of the two scenarios we're in. More precisely, we pick some $k \in \mathbb{N}$, and are then allowed to query up to k pairs of vertices to determine whether or not they form edges. Then, we say either that G has property

P or that it's δ-far from having property P, and we'd like to be right with probability at least $1-\varepsilon$. This task is called *graph property testing*, and it was introduced by Goldreich, Goldwasser, and Ron in 1996.

As with the simpler task of polling, the most basic question one can ask is the following.

Question. *Fix P, δ, ε. How large does k have to be as a function of n so that we can test the property P with accuracy parameters δ, ε?*

Before answering this question, it's worth observing that, just as with polling, this isn't a purely academic question. Enormous graphs are extremely common in today's world (e.g. the Facebook friends graph, the internet graph, the graph of neuron connections in the brain), and we'd often like to understand what properties they satisfy; for instance, Milgram's famous "six degrees of separation" result more or less says[3] that the graph of human acquaintance has diameter at most 6. However, many of these graphs are simply too large to be able to analyze directly—any algorithm that tries to look at the whole graph will not terminate in a reasonable amount of time, not to mention the fact that some of these graphs are too large to even be stored on a single computer. Therefore, a polling approach like the one described above is very natural and very useful in practice, and finding good property testing algorithms for natural properties is actually useful for many applications.

Based on what we saw for polling, it's natural to guess that the answer to this question is again that k can be taken to be a constant, independent of n. However, it turns out that for general properties, this is not true. In fact, there exist properties P for which k must be quadratic in n; in other words, to effectively test these properties, one needs to examine a constant fraction of all pairs of vertices, thus essentially achieving nothing by using a sampling algorithm rather than simply examining the whole graph. Nevertheless, for many natural properties, it turns out that k *is* independent of n. Let's start with a fairly simple-looking example.

Theorem 2. *Let P be the property of being triangle-free. Then for every $\delta, \varepsilon \in (0,1)$, there is some $k \in \mathbb{N}$ for which we can test P. In other words, there is a randomized algorithm that queries k vertex pairs from a graph G, outputs either that G is triangle-free or that it's δ-far from being triangle-free, and is correct with probability at least $1-\varepsilon$.*

Proof. Based on the polling ideas we've already seen, there's an obvious guess for the testing algorithm: pick some random subset of the vertices, query all edges inside, and see if there's a triangle. If there is, then the graph is certainly not triangle-free, so output that it's δ-far. If we see no triangle, output that the graph is triangle-free. Note that this is directly analogous to our polling algorithm from before: we pick a random subset of the population, ask its opinions, and guess that this is representative of the entire population.

However, unlike in the polling case, it's not clear how to analyze this algorithm. In particular, it's not clear that the size of the subset we pick can really be taken to be a constant (i.e. independent of n). To do this, we need to know that if a graph is δ-far from

[3] In fact, Kleinberg observed that Milgram's experiment implies a lot more about the structure of this graph; see `https://doi.org/10.1145/335305.335325`.

being triangle-free, then a random subset will contain a triangle with good probability, and this is not obviously true. In fact, the truth of this statement is a famous theorem, called the *triangle removal lemma*.

Theorem 3 (Ruzsa–Szemerédi, 1976)**.** *For every $\delta > 0$, there exists some $\alpha > 0$ such that the following holds. If G is an n-vertex graph that's δ-far from being triangle-free, then it contains at least $\alpha\binom{n}{3}$ triangles.*

This theorem is often stated in the contrapositive (hence its name): if G has fewer than $\alpha\binom{n}{3}$ triangles, then they can all be removed by deleting at most $\delta\binom{n}{2}$ edges. Though this theorem seems like something one could prove in a first course in graph theory, no simple proof is known; every known proof uses ideas related to Szemerédi's regularity lemma, which is perhaps the most powerful tool ever developed in graph theory.

As such, we'll omit the proof of the triangle removal lemma. However, once we know it, proving that the tester works becomes fairly straightforward. Indeed, if G is triangle-free, then our algorithm will output the right answer with probability 1 (since every subset will also be triangle-free, so we'll always output that). On the other hand, if G is δ-far from being triangle-free, then it contains at least $\alpha\binom{n}{3}$ triangles, for α as in Theorem 3. This means that a randomly chosen triple of vertices has probability at least α of forming a triangle in G. Now, let's partition the random subset of vertices our algorithm chooses into a bunch of disjoint triples; each such triple spans a triangle with probability at least α, and all these events are independent[4]. So the probability that our algorithm gets the wrong answer is at most the probability that all these triples are not triangles, which is at most $(1-\alpha)^{\Omega(k)} \leq e^{-\Omega(\alpha k)}$. So picking $k = \Omega(\log(1/\varepsilon)/\alpha)$ gives the desired result. In particular, the bound on k does not depend on n, only on ε and α, and thus ε and δ. □

This proves Theorem 2. As we did previously, now that we know that k is independent of n, we might still be curious about how k depends on ε and δ. The dependence on ε is logarithmic, as before. The dependence on α is linear, which looks good, so we need only understand how α depends on δ in Theorem 3. However, this is where things get bad. Ruzsa and Szemerédi's proof gave a bound of the form

$$\frac{1}{\alpha} \leq \underbrace{2^{2^{2^{\cdot^{\cdot^{\cdot}}}}}}_{O(\delta^{-5})}.$$

This was recently improved by Fox to the substantially better, but still awful, bound of

$$\frac{1}{\alpha} \leq \underbrace{2^{2^{2^{\cdot^{\cdot^{\cdot}}}}}}_{O(\log(1/\delta))}.$$

One might believe that despite these enormous bounds, the correct dependence of k, and thus α, on δ should be polynomial, as it was in the case of ordinary polling. However, this turns

[4]Strictly speaking they're weakly dependent, but this problem can easily be fixed by slightly modifying the algorithm.

out to be not the case; Alon and Shapira proved that there exist graphs that are δ-far from being triangle-free, but testing this requires a number of queries that is super-polynomial in k. This is closely related to a lower bound of Alon's for the triangle removal lemma, which shows that one must take

$$\frac{1}{\alpha} \geq \left(\frac{1}{\delta}\right)^{\Omega(\log(1/\delta))},$$

a function that grows faster than any polynomial in $1/\delta$. Nevertheless, note that our best upper and lower bounds for $1/\alpha$ are miles apart, between barely super-polynomial and tower-type; it is a major open problem to improve either of them.

Despite the fact that even testing for triangle-freeness is so complicated (and requires a number of queries that's super-polynomial in $1/\delta$), many other properties are actually much better-behaved. Here are, with few details, some results on these topics.

- Some properties, like connectivity, can be tested with *zero* queries. How? Note that any graph on n vertices can be made connected by adding at most $n-1$ edges, namely by connecting some chosen vertex to all others. Thus, if $\delta > 0$ is fixed and $n \geq 1/\delta$, then *every* graph on n vertices will be δ-close to connected. Thus, the promise that our graph is either connected or δ-far from connected becomes equivalent to the promise that our graph is connected, so our testing algorithm can simply always be correct without making any queries.

- In their original 1996 paper, Goldreich, Goldwasser, and Ron, proved that a large family of properties can be tested in polynomially many queries. This family includes the properties of being k-colorable, for any k. This is somewhat surprising, because determining k-colorability for any $k \geq 3$ is an NP-complete problem, whereas determining if a graph contains triangles can be done in polynomial time. This demonstrates that property testing (and in particular, the approximate nature of δ-closeness) is a very different problem from the exact decision problems in P and NP.

- One might imagine that since testing for the triangle C_3 is so hard, testing for the longer cycle C_4 is even harder. However, this is not the case: the property of C_4-freeness can be tested in very few queries, namely $k = O(\delta^{-1}\log(1/\varepsilon))$. This follows from (a special case of) the Kővári–Sós–Turán theorem, which says that a C_4-free graph on n vertices can have only $O(n^{3/2})$ edges. Observe that by forgetting the graph structure and simply polling edges, we can approximate the number of edges in G. If this number is less than $\delta\binom{n}{2}$, then our graph is δ-close to the empty graph, and in particular is δ-close to being C_4-free. If not, then as long as n is sufficiently large, $\delta\binom{n}{2} \gg n^{3/2}$, so our graph must be δ-far from C_4-free.

  In fact, this exact same algorithm works to test H-freeness for any bipartite H, using only $O(\delta^{-1}\log(1/\varepsilon))$ queries, since the Kővári–Sós–Turán theorem implies that any H-free graph must have a sub-quadratic number of edges. However, for testing H-freeness for non-bipartite H, the situation is essentially the same as for triangles: we know that it can be done using a number of queries that doesn't depend on n, but our understanding of the δ-dependence is extremely bad.

- A remarkable theorem of Alon, Fischer, Newman, and Shapira (2006) characterizes exactly the set of graph properties that can be tested with k independent of n, i.e. for which graph properties the situation is akin to what we saw for simple polling. The precise statement is somewhat technical, but the gist is that Szemerédi's regularity lemma is crucial: a graph property is testable independently of n if and only if it can be described in the language of regularity.

- Even among the properties that can be tested independently of n, we saw that the dependence on δ can be rather complicated. It's sometimes very simple (e.g. zero, or linear, or polynomial), and sometimes it's super-polynomial but we don't know much else. A recent result of Gishboliner and Shapira (2018) shows that this bizarreness is unavoidable: they prove that for any super-polynomial function $f : (0, 1) \to \mathbb{N}$, there is a property P_f that is testable with k queries, where k is independent of n but $k = f(\Theta(\delta)) \log(1/\varepsilon)$. In other words, the δ dependence can be arbitrarily complicated.

3 Coda: the PCP theorem

I'd like to end on one final polling-like result, which is one of the most surprising theorems I know. Let's think of a proof as a sequence of logical formulas, starting from some axioms and ending in some desired theorem. Each line in the proof is supposed to follow from some previous lines, together with a specified set of deduction rules (e.g. modus ponens, or the law of the excluded middle). Suppose we are given a proof of length n, and wish to verify that it's correct, but we're too lazy to check all n deductions; instead, we pick k random lines and check that their deductions are valid.

Of course, if we do this, we will not, in general, find the bug in a faulty proof; there might be only one wrong deduction, and if $k \ll n$, we won't find it. However, the PCP[5] theorem of Arora, Lund, Motwani, Sudan, and Szegedy (1998) says that every proof can be converted into a new proof for which this lazy verification algorithm *does* work, for a *constant* k. More precisely, there is a universal constant k and a polynomial-time algorithm that takes in a (putative) proof of length n and outputs a new proof of the same result of length $n^{O(1)}$, such that

- If the original proof was correct, then so is the new one (and in particular, the lazy verifier will accept it as correct).

- If the original proof was incorrect, then a constant fraction of the deductions in the new proof will be wrong. Thus, with probability at least 99%, among k randomly chosen lines, at least one will be wrong, so the lazy verifier will reject the proof.

[5]Short for *probabilistically checkable proofs.*