Polynomial method for (distribution) testing lower bounds

Clément Canonne[*]

August 17, 2025

Let us first recap the "standard" approach to prove distribution testing lower bounds (or, maybe more generally, indistinguishability lower bounds) given i.i.d. samples: to fix notation, we are considering domain $[k] = \{1, 2, \dots, k\}$, a property $\mathcal{P}$ of distributions over $[k]$, and (possibly tolerant) testing with parameters $0 \leq \varepsilon' < \varepsilon$ (tolerant testing being the case $\varepsilon' > 0$).

1. Construct two distributions (over probability distributions) $\mathcal{Y}, \mathcal{N}$ such that

$$\Pr_{\mathbf{p}\sim\mathcal{Y}}[\,\mathbf{p} \text{ is } \varepsilon'\text{-close to } \mathcal{P}\,] = 1 - o(1), \qquad \Pr_{\mathbf{p}\sim\mathcal{N}}[\,\mathbf{p} \text{ is } \varepsilon\text{-far from } \mathcal{P}\,] = 1 - o(1)$$

   that is, $\mathcal{Y}$ is a distribution over (mostly) yes-instances, and $\mathcal{N}$ is a distribution over (mostly) no-instances.

2. Show that a randomly chosen instance $\mathbf{p}$ from $\mathcal{Y}$ is hard to distinguish from a randomly chosen instance $\mathbf{q}$ from $\mathcal{N}$, given n i.i.d. samples from $\mathbf{p}$ and $\mathbf{q}$; that is,

$$\mathrm{d}_{\mathrm{TV}}(\mathbb{E}_{\mathbf{p}\sim\mathcal{Y}}[\mathbf{p}^{\otimes n}], \mathbb{E}_{\mathbf{q}\sim\mathcal{N}}[\mathbf{q}^{\otimes n}]) = o(1)$$

   (this is sometimes referred to as *Le Cam's lemma*. Parsing the above: "the distribution of n-tuples obtained by first choosing $\mathbf{p}$ from $\mathcal{Y}$ and then drawing n i.i.d. samples from it is statistically close to the distribution of n-tuples obtained by first choosing $\mathbf{q}$ from $\mathcal{N}$ and then drawing n i.i.d. samples from it."

3. Conclude that since $\Omega(n)$ samples are necessary to distinguish $\mathcal{Y}$ from $\mathcal{N}$, this also holds for the harder problem of distinguishing "ε'-close to $\mathcal{P}$" from "ε-far from $\mathcal{P}$."

The first and second steps are of course where the main action is: coming up with (and analyzing) distributions over k-element probability distributions, and then proving that the corresponding n-sample mixtures are close in total variation (statistical) distance, can be quite daunting.

Thankfully, for *symmetric* properties $\mathcal{P}$ (also known as "label-invariant", *i.e.*, those who are closed under permutations of the domain: $\mathbf{p} \in \mathcal{P}$ implies $\mathbf{p} \circ \pi \in \mathcal{P}$ for every permutation π of $[k]$), then the second item is "morally" equivalent to having $\mathbf{p}$ and $\mathbf{q}$ matching as many *moments* as possible (exactly or approximately), that is, to having

$$\|\mathbf{p}\|_\ell^\ell \approx \|\mathbf{q}\|_\ell^\ell$$

[*] Email: clement.canonne@sydney.edu.au. Complaints or suggestions about this note are welcome.

for as many values of ℓ as possible. The reason is that these moments of the distributions capture the probability of having an ℓ-way collision, and for symmetric properties, these collision statistics are all that matters. So if the moments match up to $L-1$, then you "intuitively" would need to see at least an ℓ-way collision to have any meaningful information allowing you to distinguish between $\mathbf{p}$ and $\mathbf{q}$, and to see such an ℓ-way collision you "expect to" need

$$\Omega\left(k^{\frac{L}{L+1}}\right)$$

samples. There are many quotes and informal, wishful statements in the short discussion above, but that's the rough idea – which can be make precise.

For instance, if your yes- and no-instances have matching first (trivial, probabilities sum to one) and second (a bit less trivial) moments, then you need to see at least one 3-way collision among your n samples, and that gives an $n=\Omega(k^{2/3})$ sample complexity lower bound.

Here comes a second idea: instead of coming up for $\mathcal{Y}$ and $\mathcal{N}$ with a way to randomly generate full-fledged *probability distributions* (which are k-length non-negative vectors summing to one) with nice properties, since we are looking at a symmetric property $\mathcal{P}$ anyway, it's enough to look at a way to randomly generate a *single* probability by defining a random variable U and V, and then draw k i.i.d. copies of each:

$$\mathbf{p} := (U^{(1)}, \dots, U^{(k)}), \qquad \mathbf{q} := (V^{(1)}, \dots, V^{(k)})$$

These may not exactly sum to one, but if $U, V \geq 0$ with

$$\mathbb{E}[U] = \mathbb{E}[V] = \frac{1}{k} \tag{1}$$

then $\mathbb{E}[\|\mathbf{p}\|_1] = \sum_{i=1}^{k} \mathbb{E}\left[U^{(i)}\right] = \mathbb{E}[\|\mathbf{q}\|_1] = 1$, and $\|\mathbf{p}\|_1 = \|\mathbf{q}\|_1 \approx 1$ with high probability. (Whether that's sufficient is not entirely obvious; but one can show, maybe with a little extra work or a small variation of (1) on a case-by-case basis, that it's indeed enough.)[1]

The above requirement already requires the first moment: for our indistinguishability, we will need more, as many as possible. The reason is provided is the next two lemma, similar in spirit, but incomparable in practice: the first requires to *exactly* match as many moments of U, V as possible, while the second allows for some *approximate* matching of moments.

Lemma 1 ([WY20, Theorem 3.3.4]). Let U, V be two random variables taking value in $[0, M]$, where $M \geq 0$. If

$$\mathbb{E}\left[U^{\ell}\right] = \mathbb{E}\left[V^{\ell}\right]$$

[1] For instance, one could argue that (by concentration of independent bounded random variables) $\mathbb{E}\left[\|\mathbf{p}\|_1\right] \in [1-O(\varepsilon), 1+O(\varepsilon)]$ with high probability, and that renormalization to 1 does not affect the rest of the argument. Or, alternatively, require instead that $\mathbb{E}[U] = \mathbb{E}[V] = \frac{1}{20k}$, so that, by Markov's inequality, $\|\mathbf{p}\|_1, \|\mathbf{q}\|_1 \leq 1$ simultaneously with probability $9/10$, and assign the remaining probability mass to a "dummy" domain element under both $\mathbf{p}$ and $\mathbf{q}$. The sky's the oyster, the world's your limit.

for all $0 \leq \ell \leq L$, then, for every $n \geq 1$

$$\mathrm{d}_{\mathrm{TV}}(\mathbb{E}_U[\mathrm{Poisson}(nU)], \mathbb{E}_V[\mathrm{Poisson}(nV)]) \leq \frac{3(nM)^{L+1}}{(L+1)!}$$

(One can relax the RHS to $\left(\frac{enM}{2L}\right)^L$, which is simpler and "typically enough" for large L).

Now, that may seem a bit odd, but to parse this: using a standard technique called *Poissonization* (taking a random number, $\mathrm{Poisson}(n)$, of independent samples from a probability distribution $\mathbf{p}$, instead of deterministically n), the number of times N_i an algorithm sees element i is a $\mathrm{Poisson}(n\mathbf{p}_i)$ random variable (instead of $\mathrm{Bin}(n, \mathbf{p}_i)$), and, more importantly, the random variables $N_1, \dots, N_k$ are now *independent*. (Also, this Poissonization is for nearly all matter and purposes without loss of generality.)

So for our lower bounds, we can assume the algorithms we are trying to defeat use Poissonization, and, combined with our construction of $\mathbf{p}, \mathbf{q}$ based on k i.i.d. copies of U, V, we get that the number of times element i of the domain is observed, for a fixed realization of U, V, is a $\mathrm{Poisson}(nU)$ under $\mathbf{p}$, and $\mathrm{Poisson}(nV)$ under $\mathbf{q}$. *That's for a fixed realization of U, V*, which represent the probabilities of our given element i of the domain. Since these U, V are themselves random variables, when we take into account their randomness, we get that

The number of times N_i any given element of the domain, say i, is observed under $\mathbf{p}$ (resp. M_i under $\mathbf{q}$) is a mixture of Poisson random variables, distributed as

$$N_i \sim \mathbb{E}_U[\mathrm{Poisson}(nU)]$$

under $\mathbf{p}$, and $\mathbb{E}_V[\mathrm{Poisson}(nV)]$ under $\mathbf{q}$. Moreover, the $N_1, \dots, N_k$ are mutually independent (and same for $M_1, \dots, M_k$).

But going back to the very beginning, to show indistinguishability given n samples, we want to show that

$$\mathrm{d}_{\mathrm{TV}}(\mathbb{E}_{\mathbf{p}\sim\mathcal{Y}}[\mathbf{p}^{\otimes n}], \mathbb{E}_{\mathbf{q}\sim\mathcal{N}}[\mathbf{q}^{\otimes n}]) = o(1)$$

i.e., that the distributions of the tuples $(N_1, \dots, N_k)$ and $(M_1, \dots, M_k)$ (over the choices of $\mathbf{p}, \mathbf{q}$ and the drawing of the samples) are very close in total variation distance. Well, by the above, this is now equivalent to showing that

$$\mathrm{d}_{\mathrm{TV}}\left(\mathbb{E}_U[\mathrm{Poisson}(nU)]^{\otimes k}, \mathbb{E}_V[\mathrm{Poisson}(nV)]^{\otimes k}\right) = o(1)$$

and, by subbaditivity of total variation distance for independent random variables, it is *enough* to show that

$$\mathrm{d}_{\mathrm{TV}}(\mathbb{E}_U[\mathrm{Poisson}(nU)], \mathbb{E}_V[\mathrm{Poisson}(nV)]) = o\left(\frac{1}{k}\right) \tag{2}$$

which is great, since that's *exactly* what Lemma 1 gives us:

Match the first L moments of $U, V \in [0, M]$ so that

$$\frac{(nM)^{L+1}}{(L+1)!} \ll \frac{1}{k},$$

conclude that $\Omega(n)$ samples are necessary to distinguish $\mathcal{Y}$ from $\mathcal{N}$.

The recipe above is very successful: in some cases, one may want to use a variant of that moment-indistinguishability lemma mentioned earlier, which provides slightly different guarantees:

Lemma 2 ([Han19, Theorem 4])**.** Let U', V' be two random variables taking value in $[-\nu, \infty)$, where $\nu \geq 0$. Then

$$\mathrm{d_{TV}}(\mathbb{E}_{U'}[\mathrm{Poisson}(n(\nu+U'))], \mathbb{E}_{V'}[\mathrm{Poisson}(n(\nu+V'))]) \leq \frac{1}{2}\left(\sum_{\ell=0}^{\infty} \frac{n^\ell \left(\mathbb{E}[U'^\ell] - \mathbb{E}[V'^\ell]\right)^2}{\nu^\ell \ell!}\right)^{1/2}.$$

Moreover, if $\mathbb{E}[U'] = 0$ and $|U'| \leq M$, then

$$\chi^2(\mathbb{E}_{U'}[\mathrm{Poisson}(n(\nu+U'))] \mid\mid \mathbb{E}_{V'}[\mathrm{Poisson}(n(\nu+V'))]) \leq e^{nM} \sum_{\ell=0}^{\infty} \frac{n^\ell \left(\mathbb{E}[U'^\ell] - \mathbb{E}[V'^\ell]\right)^2}{\nu^\ell \ell!}.$$

In this case, note that if the first L moments match, then the first L terms of the series are zero. (Again, this is somewhat more robust, as one does not require them to *exactly* match to obtain a meaningful bound.) The key difference, however, is that this allows to use ν as an "offset" when U, V are both centered around a common value: we will see an example shortly. Another advantages is that the part after the "Moreover" allows us to bound the χ^2 divergence instead of total variation distance, which can be better than relying on the subadditivity of total variation: for any two product distributions $P^{\otimes k}, Q^{\otimes k}$,

$$\mathrm{d_{TV}}\left(P^{\otimes k}, Q^{\otimes k}\right) \leq \sqrt{\chi^2(P^{\otimes k} \mid\mid Q^{\otimes k})} \lesssim \sqrt{k} \cdot \sqrt{\chi^2(P \mid\mid Q)} \tag{3}$$

so to conclude the TV distance is $o(1)$ is suffices to show that $\chi^2(P \mid\mid Q) = o(1/k)$, instead of showing as before that $\mathrm{d_{TV}}(P, Q) = o(1/k)$. This can be better by up to a $\sqrt{k}$ factor, as "morally" $\chi^2(P \mid\mid Q) \simeq \mathrm{d_{TV}}(P, Q)^2$. As said before, *we'll see an example soon.*

But before this: we have almost all the ingredients, except for one crucial one:

How do we design these random variables U, V with many matching moments?

The answer (surprise!) is *polynomials.* Specifically, polynomials with simple roots, leveraging the following "standard" fact.[2]

[2] I don't have a good reference for this lemma, except "homework." If anybody has one, that'd be welcome.

Lemma 3. Let P be a degree-d univariate polynomial with *distinct* roots $\lambda_1, \dots, \lambda_d$. Then, for every $0 \leq \ell \leq d-2$,

$$\sum_{i=1}^{d} \frac{\lambda_i^\ell}{P'(\lambda_i)} = 0\,.$$

Also, one can check that the signs of $P'(\lambda_1), \dots, P'(\lambda_d)$ alternate. One easy consequence is that

$$\sum_{\substack{1\leq i\leq d\\ P'(\lambda_i)>0}} \frac{\lambda_i^\ell}{P'(\lambda_i)} = \sum_{\substack{1\leq i\leq d\\ P'(\lambda_i)<0}} \frac{\lambda_i^\ell}{|P'(\lambda_i)|}$$

which, if one squints for long enough, starts looking like matching the first d moments of some strange random variables: say, U, V with take value λ_i with probability proportional to $\mathbb{1}_{\{P'(\lambda_i)>0\}}/P'(\lambda_i)$ and $\mathbb{1}_{\{P'(\lambda_i)<0\}}/|P'(\lambda_i)|$, respectively!

Recipe. Given a degree-d univariate polynomial P with distinct roots $\lambda_1, \dots, \lambda_d \geq 0$, let Λ_+, Λ_- be defined as

$$\Lambda_+ := \{\, 1 \leq i \leq d :\; P'(\lambda_i) > 0 \,\}, \qquad \Lambda_- := \{\, 1 \leq i \leq d :\; P'(\lambda_i) < 0 \,\}$$

and let U, V be the random variables defined on Λ_+ and Λ_-, respectively, and given by

$$\begin{aligned} \Pr[\, U = \lambda_i \,] &= \frac{C_P}{|P'(\lambda_i)|}, \qquad i \in \Lambda_+ \\ \Pr[\, V = \lambda_i \,] &= \frac{C_P}{|P'(\lambda_i)|}, \qquad i \in \Lambda_-\,, \end{aligned}$$

where $C_P := \sum_{i\in\Lambda_+} \frac{1}{|P'(\lambda_i)|} = \sum_{i\in\Lambda_-} \frac{1}{|P'(\lambda_i)|}$ (by Lemma 3). Then $0 \leq U, V \leq M := \max_{1\leq i\leq d} \lambda_i$, and

$$\mathbb{E}\Big[U^\ell\Big] = \mathbb{E}\Big[V^\ell\Big], \qquad 0 \leq \ell \leq L := d-2\,.$$

(In case $\alpha := \mathbb{E}[U] = \mathbb{E}[V] \neq \frac{1}{k}$, normalize by considering instead $\tilde{U} = \frac{U}{\alpha k}$, $\tilde{V} = \frac{V}{\alpha k}$, *i.e.*, replacing $P(X)$ by $\tilde{P}(X) = P(X/(\alpha k))$. The moments still match, but M scales by $1/(\alpha k)$.)

For this recipe to work well, what do we need from our polynomial P?

1. To have simple roots, and a "high" degree d (this will give us L)
2. To have "small" roots, so that M is small (after renormalization/scaling of $\mathbb{E}[U], \mathbb{E}[V]$)
3. Its positive-derivative roots (for U) to be consistent with a yes-instance, and the negative-derivative roots (for V) to be consistent with a no-instance (what "consistent" means depending on which task you are trying to prove a lower bound against)

This is very abstract, so let us illustrate this with two examples.

Example 1: The Usual Suspect (Uniformity Testing) We want to test whether an arbitrary distribution over $[k] = \{1, 2, \dots, k\}$ is *the* uniform distribution $\mathbf{u}_k$ over the domain, is is ε-far from it in total variation distance. In this case, the property to be tested is $\mathcal{P} := \{\mathbf{u}_k\}$ (this is clearly symmetric!), we have $\varepsilon' = 0$, and the (tight) sample complexity lower bound we want to establish is

$$\Omega(\sqrt{k}/\varepsilon^2)\,.$$

This lower bound has been shown already in several ways, starting with [**Paninski08**] for $\varepsilon \geq 1/ab^{1/4}$ (see also [Can22, Chapter 3]). Which does not mean we cannot prove it differently! Now, Paninski's $\texttt{no}$-instances were the distributions with half of the domain elements with probability $\frac{1+2\varepsilon}{k}$ and the other half with probability $\frac{1-2\varepsilon}{k}$ (and, obviously, the only $\texttt{yes}$-instance is the uniform distribution itself): to emulate this, we should have U taking value $1/k$ with probability one (this will give us the uniform distribution), and V taking values $\frac{1\pm2\varepsilon}{k}$ with equal probability $1/2$.

Since we want a polynomial which encodes this, let's take

$$P(X) = -\left(X - \frac{1-2\varepsilon}{k}\right)\left(X - \frac{1}{k}\right)\left(X - \frac{1+2\varepsilon}{k}\right)$$

This has degree $d = 3$, three distinct roots $\lambda_1 = \frac{1-2\varepsilon}{k}, \lambda_2 = \frac{1}{k}, \lambda_3 = \frac{1+2\varepsilon}{k}$, and one can check that[3]

$$\Lambda_- = \{1, 3\}, \qquad \Lambda_+ = \{2\}$$

Moreover, $L = d - 2 = 1$, $M = \frac{1+2\varepsilon}{k} \leq \frac{2}{k}$, and (always a good exercise to check)

$$C_P = \sum_{i \in \Lambda_+} \frac{1}{P'(\lambda_i)} = \frac{1}{P'(1/k)} = \frac{k^2}{4\varepsilon^2}$$

so that we get

$$\Pr[\,U = 1/k\,] = 1$$
$$\Pr[\,V = (1 \pm 2\varepsilon)/k\,] = \frac{1}{2},$$

which is what we wanted. It's easy to see that setting

$$\mathbf{p} := (U^{(1)}, \dots, U^{(k)}), \qquad \mathbf{q} := (V^{(1)}, \dots, V^{(k)})$$

will result in (1) $\mathbf{p} = \mathbf{u}_k$, and (2) $\|\mathbf{q} - \mathbf{u}_k\|_1 = 2\varepsilon$ (as each coordinate contributes $2\varepsilon/k$ to the ℓ_1 distance). Since we also have

$$\mathbb{E}[U] = \mathbb{E}[V] = \frac{1}{k}$$

and (*e.g.*, by Hoeffding) $\|\mathbf{q}\|_1 \in [1 - O(\varepsilon), 1 + O(\varepsilon)]$ with probability $e^{-\Omega(k)}$, by renormalizing we'll have a valid probability distribution, $\Theta(\varepsilon)$-far from uniform, with overwhelming probability.

For the lower bound itself, if we appply Lemma 1 with $L = 1$, $M \leq\leq \frac{2}{k}$, we get a lower bound for any n such that

$$\left(\frac{n}{k}\right)^2 \asymp \frac{(nM)^{L+1}}{(L+1)!} \ll \frac{1}{k}\,,$$

[3] The only reason I chose to put a minus sign in the definition of $P(X)$ was to be consistent with my choice of Λ_+ for U. Besides that, it's inconsequential, and one could have picked $P(X) = \left(X - \frac{1-2\varepsilon}{k}\right)\left(X - \frac{1}{k}\right)\left(X - \frac{1+2\varepsilon}{k}\right)$ instead.

that is, $n \ll \sqrt{k}$. A lower bound of $\Omega(\sqrt{k})$, which is correct, but *seriously underwhelming*: no dependence on ε?!

The reason is that here, the bound on M loses the ε: or, put differently, the "baseline $1/k$" is hiding the "perturbation $\pm\varepsilon/k$" around it. But that's OK: we have our second option, Lemma 2, which can handle specifically this case!

Set $U' := U - 1/k$, $V' := V - 1/k$, $\nu := 1/k$, $M := 0$: since $\mathbb{E}[U'] = 0$ (trivially), we can use the second conclusion to get[4], using $\mathbb{E}[|V'|^{\ell}] = (2\varepsilon/k)^{\ell}$, that

$$\sum_{\ell=0}^{\infty} \frac{n^{\ell}\left(\mathbb{E}[U'^{\ell}] - \mathbb{E}[V'^{\ell}]\right)^2}{\nu^{\ell}\ell!} = \sum_{\ell=L+1}^{\infty} \frac{n^{\ell}\left(0 - \mathbb{E}[V'^{\ell}]\right)^2}{(1/k)^{\ell}\ell!} \leq \sum_{\ell=2}^{\infty} \frac{4^{\ell}n^{\ell}\varepsilon^{2\ell}}{k^{\ell}\ell!} \lesssim \frac{n^2\varepsilon^4}{k^2}$$

(the last inequality as long as $n\varepsilon^2/k \ll 1$, recalling that $\sum_{\ell=2}^{\infty} \frac{x^{\ell}}{\ell!} = e^x - x - 1 = O(x^2)$ for $x \ll 1$). So by the discussion after Lemma 2 get our indistinguishability lower bound as long as the RHS is $o(1/k)$, that is, whenever

$$\frac{n^2\varepsilon^4}{k^2} \ll \frac{1}{k}$$

which, reorganizing, gives us our lower bound of $n = \Omega(\sqrt{k}/\varepsilon^2)$ for distinguishability.

Example 2: The (Un)usual Suspect (Tolerant Uniformity Testing) The above example was meant to show that the polynomial method described in this note can be used for "standard" testing ($\varepsilon' = 0$): but we have other methods for these. For *tolerant* testing ($\varepsilon' > 0$), however, we have much fewer arrows in our toolbox (hammers in our quiver): but this method is one.

Consider for simplicity the setting where $\varepsilon', \varepsilon = \Theta(1)$, and we want to distinguish, say, distributions 0.1-*close* to $\mathbf{u}_k$ from distributions 0.7-*far* from $\mathbf{u}_k$. This is known to require

$$n = \Omega\left(\frac{k}{\log k}\right)$$

samples (and this is tight) [VV11]. To prove such a lower bound, we we try the same recipe: a "natural" idea is to start with the same type of instances, uniform or small perturbations around uniform, as in the previous example of uniformity testing: so, starting with a polynomial like

$$-\left(X - \frac{1}{2k}\right)\left(X - \frac{1}{k}\right)\left(X - \frac{3}{2k}\right)$$

But of course, to get the right lower bound, we need to match a lot more moments than this, so we need a degree d much larger than a measly 3. So we should look at a polynomial of the form

$$P(X) = -\left(X - \frac{1}{2k}\right)\left(X - \frac{1}{k}\right)\left(X - \frac{3}{2k}\right)\tilde{P}(X)$$

where (1) $\tilde{P}(X)$ has simple roots, and very large degree (d goes up!), but also (2) small enough roots (M stays small!), and (3) the derivative of $\tilde{P}'$ at these "extra" roots is large (not putting a lot of probability mass for U, V on these!)

[4] I'll let the interested reader check why the first of the two conclusions of Lemma 2 does not suffice, and why using the χ^2 bound is necessary here.

To choose the best polynomial $\tilde{P}(X)$, one option is to invoke the theory of best polynomial approximation or phrase this as an optimization problem and argue about the existence of a solution. This is a valid, clean way to do this: see for instance [WY19].

The *other*, maybe more instructive way, is to come up with an explicit construction, recalling that to have a polynomial satisfying extremal properties such as (1), (2), and (3) above, *a good guess is always to start with the Chebyshev polynomials.*[5]

Letting T_d be the Chebyshev polynomial (of the first kind) of degree d, given a parameter $M > 2/k$ of our choosing we can set

$$P(X) = -\left(X - \frac{1}{2k}\right)\left(X - \frac{1}{k}\right)\left(X - \frac{3}{2k}\right)T_d\left(1 - \frac{X}{M}\right)$$

which is a simple polynomial of degree $d+3$, with roots in $[0, M]$.
With this, one can (and should) verify that[6]

- the "intended" roots of P are

  $$\lambda_1 = \frac{1}{2k},\ \lambda_2 = \frac{1}{k},\ \lambda_3 = \frac{3}{2k}$$

  while the "extra" roots satisfy

  $$\lambda_{3+r} = M\left(1 - \cos\left(\frac{2r-1}{2d}\pi\right)\right) = \Theta\left(\frac{Mr^2}{d^2}\right), \qquad 1 \leq r \leq d$$

- Turning to the derivative P', for the "intended" roots,

  $$|P'(\lambda_1)|, |P'(\lambda_2)|, |P'(\lambda_3)| = \Theta\left(\frac{1}{k^2}\right)$$

  while the "extra" roots have

  $$|P'(\lambda_{3+r})| = \Theta\left(\frac{M^2 r^5}{d^4}\right)$$

  using that $|T_d{}'(1 - M\lambda_{3+r})| = \frac{d}{|\sin\frac{2r-1}{2d}|} = \Theta(d^2/r)$.

Going through the calculations, this means that

$$2C_P = \underbrace{\frac{1}{|P'(\lambda_1)|} + \frac{1}{|P'(\lambda_2)|} + \frac{1}{|P'(\lambda_3)|}}_{\Theta(k^2)} + \sum_{r=1}^{d} \frac{1}{|P'(\lambda_{r+3})|} \asymp k^2 + \frac{d^4}{M^2}\sum_{r=1}^{d}\frac{1}{r^5} \asymp k^2 + \frac{d^4}{M^2}$$

which is $\Theta(k^2)$ as long as

$$M \gg d^2/k$$

in which case the total probability mass but by U, V on the "extra" roots will be negligible in front of the probability of the "intended" roots. *Why is that important?* This tells us that the yes-instances

[5] In particular, (3) corresponds to saturating the Markov Brothers' inequality, tight for… the Chebyshev polynomials.

[6] See, *e.g.*, [Can+22, Appendix B] for an example.

defined by U will indeed be close to $\mathbf{u}_k$ (as most of the probability of U is on $1/k$), while the no-instances defined by V will be close to the "Paninski-type" instances (as most of the probability of V is on $(1 \pm (1/2))/k$): these "Paninski-type" are far from $\mathbf{u}_k$ – and so will be our no-instances themselves.

Similarly, one can also check that $\mathbb{E}[U], \mathbb{E}[V] = \frac{1+o(1)}{k}$, which tells us that we can renormalize our yes- and no-instances to get *bona fide* probability distributions. We're almost there!

How to pick d, M in our construction, subject to the condition $M \gg d^2/k$?

Recall that, in view of Lemma 1, we want to minimize the quantity

$$\left(\frac{enM}{2L}\right)^{L}$$

specifically to make it $o(1/k)$. (We are in the regime where $L = (d+3) - 2 = d+1$ should be large, so we should be alright using the simpler expression at the end of the lemma.) Since our condition above requires $M \gg L^2/k$, we may as well set

$$\frac{M}{L^2} = \frac{C}{k}$$

for a sufficiently large constant $C > 0$, in which case we get

$$\left(\frac{enM}{2L}\right)^{L} = \left(\frac{CenL}{2k}\right)^{L} \overset{\text{(want)}}{\ll} \frac{1}{k}\,. \tag{4}$$

Equivalently, we want L such that $\frac{nLk^{1/L}}{k}$ is minimized (to get the inequality for the largest n possible). One can check that $L = \Theta(\log k)$ does the job, for which we then get (4) as long as $n \ll \frac{k}{\log k}$. This gives us the $\Omega\left(\frac{k}{\log k}\right)$ lower bound.

Concluding note. This short exposition was meant to provide a clear and helpful introduction to this technique of using polynomials to obtain indistinguishability lower bounds, and as a result sample complexity lower bounds for testing symmetric properties of dicrete distributions. For more, in more depth, and more insights, the reader is encouraged to consult the monograph [WY20] by Yihong Wu and Pengkun Yang.

References

[Can+22] Clément L. Canonne et al. "Nearly-Tight Bounds for Testing Histogram Distributions". In: *NeurIPS*. Also available at arXiv:2207.06596. 2022.

[Can22] Clément L. Canonne. "Topics and Techniques in Distribution Testing: A Biased but Representative Sample". In: *Found. Trends Commun. Inf. Theory* 19.6 (2022), pp. 1032–1198.

[Han19] Yanjun Han. *Lecture 7: Mixture vs. Mixture and Moment Matching*. 2019. URL: https://web.archive.org/web/20250324045136/https://theinformaticists.wordpress.com/2019/08/28/lecture-7-mixture-vs-mixture-and-moment-matching/.

[VV11] Gregory Valiant and Paul Valiant. “Estimating the unseen: an n/log(n)-sample estimator for entropy and support size, shown optimal via new CLTs”. In: *STOC*. ACM, 2011, pp. 685–694.

[WY19] Yihong Wu and Pengkun Yang. “Chebyshev polynomials, moment matching, and optimal estimation of the unseen”. In: *Ann. Statist.* 47.2 (2019), pp. 857–883. ISSN: 0090-5364. DOI: 10.1214/17-AOS1665. URL: https://doi.org/10.1214/17-AOS1665.

[WY20] Yihong Wu and Pengkun Yang. “Polynomial Methods in Statistical Inference: Theory and Practice”. In: *Found. Trends Commun. Inf. Theory* 17.4 (2020). Available at https://arxiv.org/abs/2104.07317, pp. 402–586.