Topics and Techniques in Distribution Testing

Clément L. Canonne
University of Sydney
clement.canonne@sydney.edu.au

January 2, 2023

Abstract

Solutions to the exercises from Canonne (2022).

2 Testing goodness-of-fit of univariate distributions

Exercise 2.1. Prove the monotonicity of ℓ_p norms: if $1 \leq r \leq s \leq \infty$, then $\|x\|_s \leq \|x\|_r$ for every $x \in \mathbb{R}^n$.

Solution 2.1. Fix any $1 \leq r \leq s < \infty$, and any non-zero $x \in \mathbb{R}^n$ (if x is the zero vector, the inequality is trivially true). Then, since $x' := x/\|x\|_s$ has unit ℓ_s norm and $|x'_i| \leq 1$ for all i, we get

$$1 = \sum_{i=1}^n |x'_i|^s \leq \sum_{i=1}^n |x'_i|^r = \|x'\|_r^r$$

showing that $\|x\|_s^r \leq \|x\|_r^r$. Taking the r-th root on both side gives the result. Finally, the case $s = \infty$ follows from observing that $\|x\|_\infty = \max_{1\leq i\leq n} |x_i|^r \leq \sum_{i=1}^n |x_i|^r = \|x\|_r^r$.

Exercise 2.2. Prove Eq. (2.14): that is, the "unique elements" statistic Z_2 from Section 2.1.3 has expectation $\mathbb{E}_{\mathbf{p}}[Z_2] = \sum_{i\in\mathcal{X}} \mathbf{p}(i)(1-\mathbf{p}(i))^{n-1}$.

Solution 2.2. By linearity of expectation,

$$\mathbb{E}_{\mathbf{p}}[Z_2] = \frac{1}{n} \sum_{i\in\mathcal{X}} \Pr[\, N_i = 1\,]$$

where $N_i = \sum_{t=1}^n \mathbb{1}\{X_t = i\}$ follows a Binomial distribution with parameters n and $\mathbf{p}(i)$. Thus, $\Pr[\, N_i = 1\,] = n\mathbf{p}(i)(1-\mathbf{p}(i))^{n-1}$.

Exercise 2.3. Establish Claim 2.2, using (or computing) the expression for the first 4 moments of a Poisson(λ) random variable.

Solution 2.3. Let $\lambda, \mu \geq 0$, and $X \sim \text{Poisson}(\lambda)$. Then $(X-\mu)^2 = (X-\lambda)^2 + 2(\lambda-\mu)X + \mu^2 - \lambda^2$, and so

$$\begin{aligned}\mathbb{E}\left[(X-\mu)^2 - X\right] &= \text{Var}[X] - \mathbb{E}[X] + 2(\lambda-\mu)\mathbb{E}[X] + \mu^2 - \lambda^2\\ &= 2(\lambda-\mu)\lambda + \mu^2 - \lambda^2 = (\lambda-\mu)^2\end{aligned}$$

using the fact that $\text{Var}[X] = \mathbb{E}[X] = \lambda$. For the second one, we will also use the identities

$$\begin{aligned}\mathbb{E}\left[X^2\right] &= \text{Var}[X] + \mathbb{E}[X]^2 = \lambda + \lambda^2\\ \mathbb{E}\left[X^3\right] &= \lambda + 3\lambda^2 + \lambda^3\\ \mathbb{E}\left[X^4\right] &= \lambda + 7\lambda^2 + 6\lambda^3 + \lambda^4\end{aligned}$$

(which are not hard to prove by manipulating the corresponding series $\mathbb{E}[X^m] = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{m+k}}{k!}$, but are quite tedious). Then, a brute-force computation gives

$$\begin{aligned}&\mathbb{E}\left[((X-\mu)^2 - X)^2\right]\\ &\quad= \mu^4 + \mathbb{E}\left[X^4\right] - (4\mu+2)\mathbb{E}\left[X^3\right] + (6\mu^2+4\mu+1)\mathbb{E}\left[X^2\right] - (4\mu^3+2\mu^2)\mathbb{E}[X]\\ &\quad= \mu^4 + \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda - (4\mu+2)(\lambda^3+3\lambda^2+\lambda)\\ &\qquad\quad + (6\mu^2+4\mu+1)(\lambda^2+\lambda) - (4\mu^3+2\mu^2)\lambda\\ &\quad= \mu^4 + \lambda^4 + 4\lambda^3 + 2\lambda^2 - 4\mu\lambda^3 - 8\mu\lambda^2 + 6\mu^2\lambda^2 + 6\mu^2\lambda - 4\mu^3\lambda - 2\mu^2\lambda\\ &\quad= (\lambda-\mu)^4 + 2\lambda^2 + 4\lambda^3 - 8\mu\lambda^2 + 4\mu^2\lambda\\ &\quad= (\lambda-\mu)^4 + 2\lambda^2 + 4\lambda(\lambda-\mu)^2\end{aligned}$$

which is the result we wanted. There probably are more elegant ways to prove it, but this one works.

Exercise 2.4. Establish the upper bound part of Fact 2.1, by proving *via* an Hoeffding or Chernoff bound that the empirical estimator achieves the stated sample complexity. (The lower bound can be shown by considering the case $\alpha = 1/2$, but we have not seen in this chapter the information-theoretic tools to establish it: this will be in Chapter 3)

Solution 2.4. Let $X_1, \dots, X_n \sim \text{Bern}(\alpha)$ be i.i.d., and consider the empirical estimator (for α),

$$\hat{\alpha} := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

By linearity of expectation, we have $\mathbb{E}[\hat{\alpha}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \alpha$. Moreover, by a Hoeffding bound (Corollary A.4), we have, for any $\eta > 0$,

$$\Pr[|\hat{\alpha} - \alpha| > \eta] \leq 2e^{-2\eta^2 n}$$

which is at most δ for $n \geq \frac{1}{2\eta^2}\ln\frac{2}{\delta}$. Thus, having $n := \left\lceil \frac{1}{2\eta^2}\ln\frac{2}{\delta} \right\rceil = O\!\left(\frac{\log(1/\delta)}{\eta^2}\right)$ suffices.

Exercise 2.5. Establish the upper bound part of Fact 2.2, by proving *via* a Chernoff bound that appropriately thresholding the empirical estimator achieves the stated sample complexity. (For the lower bound, same remark as for Exercise 2.4.)

Solution 2.5. Let $X_1, \dots, X_n \sim \mathrm{Bern}(\alpha)$ be i.i.d., $\beta, \eta \in (0,1]$, and consider as before the empirical estimator

$$\hat{\alpha} := \frac{1}{n}\sum_{i=1}^{n} X_i$$

along with the threshold $\tau := (1+\frac{\eta}{2})\beta$. We want to argue that, for n as in the statement of the exercise:

- If $\alpha \le \beta$, then $\Pr[\hat{\alpha} \ge \tau] \le \delta$; and

- If $\alpha \ge \beta(1+\eta)$, then $\Pr[\hat{\alpha} < \tau] \le \delta$.

By a Chernoff bound (specifically, Theorem A.6, (A.7) with $\gamma := \eta/2 \in (0,1]$ and $P_H := n\beta$), in the first case we have

$$\Pr[\hat{\alpha} \ge \tau] \le e^{-\frac{n\eta^2\beta}{12}}$$

while in the second case (by (A.8), with $\gamma := \frac{\eta}{2(1+\eta)} \in (0,1]$ and $P_L := n(1+\eta)\beta$, so that $(1-\gamma)P_L = n(1+\eta/2)\beta)$ we get

$$\Pr[\hat{\alpha} < \tau] \le e^{-\frac{n\eta^2\beta}{4(1+\eta)^2}} \le e^{-\frac{n\eta^2\beta}{16}} .$$

Both are at most δ as long as $n \ge \frac{16}{\beta\eta^2}\ln\frac{1}{\delta}$. Thus, having $n := \left\lceil \frac{16}{\beta\eta^2}\ln\frac{1}{\delta} \right\rceil = O\!\left(\frac{\log(1/\delta)}{\beta\eta^2}\right)$ suffices.

Exercise 2.6. Follow the analysis of Theorem 2.1 to derive, for the bipartite collisions tester, the guarantee Eq. (2.38) from the variance bound Eq. (2.37).

Solution 2.6. We have

$$\mathrm{Var}[Z_6] \le \frac{1}{n_1 n_2}\|\mathbf{p}\|_2^2 + \frac{n_1+n_2}{n_1 n_2}(\|\mathbf{p}\|_3^3 - \|\mathbf{p}\|_2^4)\,, \tag{2.37}$$

and we want to show that, in the "far" case,

$$\Pr_{\mathbf{p}}\left[Z_6 < \frac{1+2\varepsilon^2}{k} \right] \le \frac{5k}{4\varepsilon^4 n_1 n_2} + \frac{n_1+n_2}{n_1 n_2}\left(\frac{2\sqrt{k}}{\varepsilon} + \frac{3}{\varepsilon^2} \right). \tag{2.38}$$

We will mimic the corresponding part of the proof of Theorem 2.1: Let again $\alpha^2 := k\|\mathbf{p}-\mathbf{u}_k\|_2^2 \geq 4\varepsilon^2$, so that $\mathbb{E}[Z_6] = \|\mathbf{p}\|_2^2 = \frac{1+\alpha^2}{k}$. Then

$$\begin{aligned}
\Pr\left[Z_6 < \frac{1+2\varepsilon^2}{k}\right] &= \Pr\left[Z_6 < \frac{1+2\varepsilon^2}{1+\alpha^2}\mathbb{E}[Z_6]\right] \\
&= \Pr\left[Z_6 < \left(1-\frac{\alpha^2-2\varepsilon^2}{1+\alpha^2}\right)\mathbb{E}[Z_6]\right] \\
&\leq \Pr\left[Z_6 < \left(1-\frac{\alpha^2}{2(1+\alpha^2)}\right)\mathbb{E}[Z_6]\right] && \text{(as } \alpha^2 \geq 4\varepsilon^2\text{)} \\
&\leq \frac{4(1+\alpha^2)^2}{\alpha^4}\cdot\frac{\operatorname{Var}[Z_6]}{\mathbb{E}[Z_6]^2} && \text{(Chebyshev)} \\
&\leq \frac{4(1+\alpha^2)^2}{\alpha^4 n_1 n_2\|\mathbf{p}\|_2^2} + \frac{4(1+\alpha^2)^2(n_1+n_2)}{\alpha^4 n_1 n_2}\cdot\frac{\|\mathbf{p}\|_3^3-\|\mathbf{p}\|_2^4}{\|\mathbf{p}\|_2^4}
\end{aligned}$$

the last inequality using (2.37). The first term is easily dealt with: recalling that $\|\mathbf{p}\|_2^2 = (1+\alpha^2)/k$,

$$\frac{4(1+\alpha^2)^2}{\alpha^4 n_1 n_2\|\mathbf{p}\|_2^2} = \frac{4(1+\alpha^2)k}{\alpha^4 n_1 n_2} \leq \frac{5k}{4\varepsilon^4 n_1 n_2}$$

the last inequality as in the proof of Theorem 2.1, using that $x > 0 \mapsto \frac{1+x}{x^2}$ is decreasing, $\alpha^2 \geq 4\varepsilon^2$, and $\varepsilon \leq 1$.

To handle the second, we use the inequality proven in (2.12):

$$\|\mathbf{p}\|_3^3 - \|\mathbf{p}\|_2^4 \leq \frac{\alpha^3}{k^{3/2}} + \frac{3\alpha^2}{k^2}$$

which as in (2.13) implies

$$\frac{4(1+\alpha^2)^2}{\alpha^4}\cdot\frac{\|\mathbf{p}\|_3^3-\|\mathbf{p}\|_2^4}{\|\mathbf{p}\|_2^4} \leq \frac{2\sqrt{k}}{\varepsilon} + \frac{3}{\varepsilon^2}$$

Combining the two, we get

$$\Pr_{\mathbf{p}}\left[Z_6 < \frac{1+2\varepsilon^2}{k}\right] \leq \frac{5k}{4\varepsilon^4 n_1 n_2} + \left(\frac{3\sqrt{k}}{\varepsilon} + \frac{3}{\varepsilon^2 n}\right)\frac{n_1+n_2}{n_1 n_2}$$

as we wanted.

Exercise 2.7. Show that, in contrast to what we did in the empirical-distance tester case (Section 2.1.5), one cannot invoke stochastic dominance in the analysis of the bipartite collision tester to obtain the wishful variance bound Eq. (2.39) instead of Eq. (2.41). Specifically, show that it fails even for $k = 2$: if $M \sim \operatorname{Bin}(n_1, p)$, $N \sim \operatorname{Bin}(n_2, p)$ and

$M' \sim \mathrm{Bin}(n_1, q)$, $N' \sim \mathrm{Bin}(n_2, q)$ (all independent) with $1/2 \leq q < p \leq 1$, it is *not* always true that

$$MN + (n_1 - M)(n_2 - N) \succeq M'N' + (n_1 - M')(n_2 - N')$$

Hint: consider the case $n_1 = 1$, *and* $\Pr[MN + (n_1 - M)(n_2 - N) \geq 1]$ *as a function of* p.

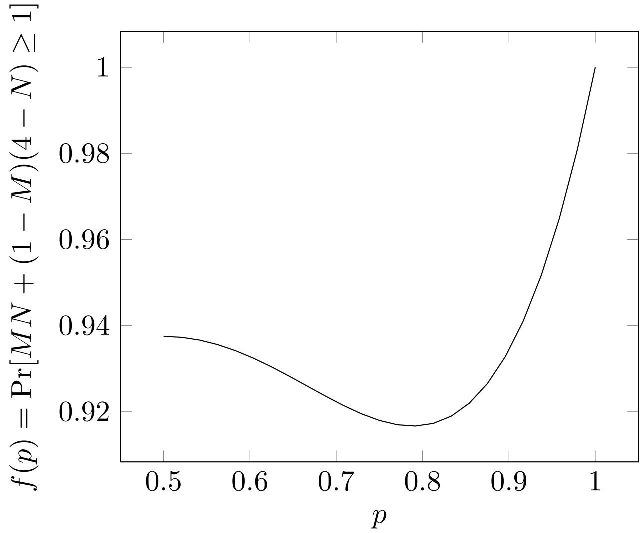
Solution 2.7. If the stochastic dominance relation held, it would imply that $f(p) := \Pr[MN + (n_1 - M)(n_2 - N) \geq 1]$ is a non-decreasing function of $p \geq 1/2$. Now, to simplify the search for a counterexample, consider the case $n_1 = 1$ (so $M \sim \mathrm{Bern}(p)$): then

$$\begin{aligned} f(p) &= \Pr[MN + (1 - M)(n_2 - N) \geq 1] \\ &= \Pr[M = 1, N > 0 \text{ or } M = 0, N < n_2] \\ &= p(1 - (1 - p)^{n_2}) + (1 - p)(1 - p^{n_2}) \end{aligned}$$

using independence of M, N. One can try to plot the corresponding function of p for various choices of n_2, or differentiate to check if it is non-decreasing on $[1/2, 1]$. Long story short: it will be non-decreasing for $n_2 \in \{1, 2, 3\}$, but for $n_2 = 4$, we get

$$f(p) = p(1 - (1 - p)^4) + (1 - p)(1 - p^4)$$

which has a local minimum at $p^* = \frac{3+\sqrt{3}}{6}$, is decreasing on $[1/2, p^*]$ and increasing on $[p^*, 1]$.



This gives a counterexample to the statement for, *e.g.*, $p = 1/2, q = \frac{3+\sqrt{3}}{6}$, $n_1 = 1$ and $n_2 = 4$.

Exercise 2.8. It is known that $x \preceq y$ if, and only if, $x = Ay$ for some doubly stochastic matrix A (Arnold, 1987, Theorem 2.1). Check that the averaging from Lemma 2.7 indeed corresponds to multiplying the pmf $\mathbf{p}$ (seen as a vector) by such a matrix.

Solution 2.8. Recall that a doubly stochastic matrix is a square matrix with non-negative entries, where each row and each column sums to one. In our case, assume for simplicity that the probability distribution $\mathbf{p}$ is non-decreasing, *i.e.*, that $\mathbf{p}(1) \geq \cdots \geq (k)$. This is without loss of generality, since one can permute the domain for this to hold, and doubly stochastic matrices are invariant to such permutations: if σ is a permutation of $[k] = \{1, 2, \dots, k\}$ and A is doubly stochastic, then for every $j \in [k]$

$$\sum_{i=1}^{k} A_{\sigma(i),\sigma(j)} = \sum_{i=1}^{k} A_{i,\sigma(j)} = 1$$

and similarly for the columns sums: for every $i \in [k]$, $\sum_{j=1}^{k} A_{\sigma(i),\sigma(j)} = \sum_{j=1}^{k} A_{\sigma(i),j} = 1$. Now, with this assumption, then the transformation to obtain $\bar{\mathbf{p}}$ is to average the probability of the first $K = \lceil k/2 \rceil$ elements, and leave the remaining $k - K$ probabilities unchanged. This is achieved by the matrix $A \in \mathbb{R}^{k \times k}$ consisting of a square $K \times K$ block with all entries equal to $1/K$ in the top left, and the rest being only $k - K$ diagonal entries equal to 1:

$$A = \begin{pmatrix} \frac{1}{K} & \cdots & \frac{1}{K} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & 0 & \cdots & 0 \\ \frac{1}{K} & \cdots & \frac{1}{K} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \cdots & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{K}\mathbf{1}_{K \times K} & \mathbf{0}_{K \times (k-K)} \\ \mathbf{0}_{(k-K) \times K} & I_{(k-K) \times (k-K)} \end{pmatrix}$$

It is now easy to check that the matrix A defined above is indeed doubly stochastic, and further that

$$A\mathbf{p} = \begin{pmatrix} \frac{1}{K} & \cdots & \frac{1}{K} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & 0 & \cdots & 0 \\ \frac{1}{K} & \cdots & \frac{1}{K} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & \cdots & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \mathbf{p}(1) \\ \vdots \\ \mathbf{p}(K) \\ \mathbf{p}(K+1) \\ \vdots \\ \mathbf{p}(k) \end{pmatrix} = \begin{pmatrix} \frac{1}{K}(\mathbf{p}(1) + \dots \mathbf{p}(K)) \\ \vdots \\ \frac{1}{K}(\mathbf{p}(1) + \dots \mathbf{p}(K)) \\ \mathbf{p}(K+1) \\ \vdots \\ \mathbf{p}(k) \end{pmatrix} = \bar{\mathbf{p}}$$

as we wanted.

Exercise 2.9 ($\star$). Generalize Lemma 2.13 to relax the condition $n_3 \leq k^{2/3}$ to $n_3 \leq k^{(s-1)/s}$, for any fixed (constant) integer $s \geq 3$, by considering s-collisions instead of 3-collisions in Algorithm 8. How does the ℓ_∞ guarantee bound in (ii) change with s?

Solution 2.9. Fix any $s \geq 4$ (Lemma 2.13 already handled $s = 3$). Recalling (2.42), we get that under the uniform distribution $\mathbf{u}_k$ the probability to observe an s-way collision among

n samples is at most

$$p(n,k,s) \leq \frac{1}{k^{s-1}}\binom{n}{s} \leq \frac{1}{k^{s-1}}\left(\frac{n}{s}\right)^s$$

which is at most $1/6$ for $n \leq \frac{k^{1-1/s}}{6^{1/s}s}$. Now, let $C = C(s)$ be a value to be determined in the course of the analysis, and assume that $\|\mathbf{p}\|_\infty > \frac{C}{n}$, so that as in the proof of Lemma 2.13, fixing an arbitrary i such that $\mathbf{p}(i) = \|\mathbf{p}\|_\infty$, the number of times i appears among the n samples, N_i, is Binomially distributed with parameters n and $\|\mathbf{p}\|_\infty$, and thus mean $n\|\mathbf{p}\|_\infty > C$. Again by a Chernoff bound (specifically, (A.8)), we have that the probability *not* to observe an s-way collision on element i is at most, choosing γ such that $(1-\gamma)C = s$ and recalling that $s \geq 4$,

$$\Pr[\, N_i < s\,] = \Pr[\, N_i < (1-\gamma)C\,] \leq e^{-\gamma^2 C/2} = e^{-\frac{\gamma^2}{2(1-\gamma)}s} \leq e^{-\frac{2\gamma^2}{1-\gamma}}$$

which is less than $1/6$ for (solving numerically), *e.g.*, $\gamma \geq 0.82$. From the above, this means that we can take any $C \geq \frac{s}{1-\gamma}$ (so for instance $C = 6s$ suffices).

Putting the two conditions together, what we get is an algorithm which, for any fixed $s \geq 4$, distinguishes with probability at least $5/6$ between (i) $\mathbf{p} = \mathbf{u}_k$ and (ii) $\|\mathbf{p}\|_\infty \geq \frac{6s}{n}$, provided that $n \leq \frac{k^{1-1/s}}{6^{1/s}s}$. The algorithm does so by checking if any s-way collision happens among the n samples, and declaring (i) if, and only if, no such collision is observed.

As an example, for $s \asymp \ln k$, we get the following:

Corollary 2.0.1. There exists an algorithm which, given n i.i.d. samples from some unknown $\mathbf{p} \in \Delta_k$, distinguishes with probability at least $5/6$ between (i) $\mathbf{p} = \mathbf{u}_k$ and (ii) $\|\mathbf{p}\|_\infty \geq \frac{\ln k}{n}$, provided that $n \lesssim \frac{k}{\ln k}$.

Exercise 2.10 ($\star$). Recall that our χ^2-based statistic (Eq. (2.19)) was analyzed under the Poissonized sampling model, which led us to define it with a $-N_i$ term in the numerator. We will show that this term is necessary: that is, under the Poissonization assumption, consider the “simpler” statistic

$$Z_3' := \sum_{i=1}^{k} \frac{(N_i - n/k)^2}{n/k}\,.$$

Show that its expectation is $nk\|\mathbf{p} - \mathbf{u}_k\|_2^2 + k$ (so the expectation *gap* remains the same), but that the variance now contains an extra term $\frac{k^2}{n}$. What sample complexity does this yield?

Solution 2.10. By linearity of expectation, we can just use the same analysis (and the expression established in Exercise 2.3), "adding back" the $\mathbb{E}[N_i] = n\mathbf{p}(i)$ terms, to get

$$\begin{aligned}
\mathbb{E}[Z_3'] &= \sum_{i=1}^{k} \frac{\mathbb{E}[(N_i - n/k)^2 - N_i] + \mathbb{E}[N_i]}{n/k} \\
&= \frac{k}{n}\sum_{i=1}^{k}\left(\left(n\mathbf{p}(i) - \frac{n}{k}\right)^2 + n\mathbf{p}(i)\right) \\
&= nk\sum_{i=1}^{k}\left(\mathbf{p}(i) - \frac{1}{k}\right)^2 + k\sum_{i=1}^{k}\mathbf{p}(i) \\
&= nk\|\mathbf{p} - \mathbf{u}_k\|_2^2 + k\,,
\end{aligned}$$

so, indeed, the expectation gap $\mathbb{E}_{\mathbf{p}}[Z_3'] - \mathbb{E}_{\mathbf{u}_k}[Z_3'] = nk\|\mathbf{p} - \mathbf{u}_k\|_2^2$ remains the same. However, the variance is now (using independence of the summands)

$$\begin{aligned}
\operatorname{Var}[Z_3'] &= \sum_{i=1}^{k} \frac{\operatorname{Var}[(N_i - n/k)^2]}{n^2/k^2} \\
&= \frac{k^2}{n^2}\sum_{i=1}^{k}\left(\mathbb{E}\left[(N_i - n/k)^4\right] - \mathbb{E}\left[(N_i - n/k)^2\right]^2\right)
\end{aligned}$$

At this point, we have to compute this horrible-looking expression. This is quite painful, but as in Exercise 2.3 (or using something like Mathematica if one would rather not go through that ordeal again) one can check that, for $X \sim \text{Poisson}(\lambda)$ and any μ,

$$\mathbb{E}\left[(X-\mu)^4\right] - \mathbb{E}\left[(X-\mu)^2\right]^2 = 2\lambda^2 + 4\lambda(\lambda-\mu)^2 + 4\lambda(\lambda-\mu) + \lambda\,. \tag{2.1}$$

This leads to

$$\begin{aligned}
&\operatorname{Var}[Z_3'] \\
&\quad= \frac{k^2}{n^2}\sum_{i=1}^{k}\left(2n^2\mathbf{p}(i)^2 + 4n^3\mathbf{p}(i)\left(\mathbf{p}(i) - \frac{1}{k}\right)^2 + 4n^2\mathbf{p}(i)\left(\mathbf{p}(i) - \frac{1}{k}\right) + n\mathbf{p}(i)\right) \\
&\quad= \operatorname{Var}[Z_3] + \frac{k^2}{n^2}\sum_{i=1}^{k}\left(4n^2\mathbf{p}(i)\left(\mathbf{p}(i) - \frac{1}{k}\right) + n\mathbf{p}(i)\right) \\
&\quad= \operatorname{Var}[Z_3] + 4k^2\sum_{i=1}^{k}\mathbf{p}(i)\left(\mathbf{p}(i) - \frac{1}{k}\right) + \frac{k^2}{n} \\
&\quad= \operatorname{Var}[Z_3] + 4k^2\left(\|\mathbf{p}\|_2^2 - \frac{1}{k}\right) + \frac{k^2}{n} \\
&\quad= \operatorname{Var}[Z_3] + 4k^2\|\mathbf{p} - \mathbf{u}_k\|_2^2 + \frac{k^2}{n}
\end{aligned}$$

where we recognized (and took out) the expression of $\mathrm{Var}[Z_3]$ from Section 2.1.4. To see the sample complexity this would lead to, note that to get "variance $\ll$ (expectation gap)^2" we will need, considering the three terms of $\mathrm{Var}[Z_3']$ above,

$$\begin{aligned}\mathrm{Var}[Z_3] &\ll (nk\|\mathbf{p}-\mathbf{u}_k\|_2^2)^2,\\ k^2\|\mathbf{p}-\mathbf{u}_k\|_2^2 &\ll (nk\|\mathbf{p}-\mathbf{u}_k\|_2^2)^2,\\ \frac{k^2}{n} &\ll (nk\|\mathbf{p}-\mathbf{u}_k\|_2^2)^2\end{aligned}$$

The first one is exactly what we had in Section 2.1.4, and leads to the condition $n \gg \sqrt{k}/\varepsilon^2$. The second is not too problematic: after simplication, and recalling that in the "far" case we have $\|\mathbf{p}-\mathbf{u}_k\|_2 \geq 2\varepsilon/\sqrt{k}$, this results in the (weaker) condition $n \gg \sqrt{k}/\varepsilon$. The third one, however, is the bottleneck: again using the bound on $\|\mathbf{p}-\mathbf{u}_k\|_2$ in the "far" case (the only handle we have on this quantity), it results in the condition

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

which means, reorganizing, that n must satisfy $n \gg \frac{k^{2/3}}{\varepsilon^{4/3}}$. Altogether, and once made formal via, *e.g.*, Chebyshev's inequality as usual, these 3 conditions will yield the sample complexity

$$n = O\left(\max\left(\frac{k^{2/3}}{\varepsilon^{4/3}}, \frac{\sqrt{k}}{\varepsilon^2}\right)\right)$$

which is suboptimal (and, as a side note, is the expression for the sample complexity of *closeness* testing, the harder testing problem where both $\mathbf{p}$ and $\mathbf{q}$ are unknown).

Exercise 2.11 ($\star$). Combine the doubling search technique discussed in Section 1.1 with the sample complexity of uniformity testing given in Eq. (2.50) to prove the following. There is an adaptive uniformity testing algorithm which, on input k and $\varepsilon \in (0,1]$, and access to samples from an unknown distribution $\mathbf{p} \in \Delta_k$:

- correctly distinguishes between (1) $\mathbf{p} = \mathbf{u}_k$ and (2) $\varepsilon(\mathbf{p}) := \mathrm{d_{TV}}(\mathbf{p}, \mathbf{u}_k) > \varepsilon$, with probability at least 2/3;
- always takes at most

  $$O\left(\frac{1}{\varepsilon^2}\left(\sqrt{k\log\log\frac{1}{\varepsilon}} + \log\log\frac{1}{\varepsilon}\right)\right)$$

  samples; but also
- if $\varepsilon(\mathbf{p}) > \varepsilon$, takes at most

  $$O\left(\frac{1}{\varepsilon(\mathbf{p})^2}\left(\sqrt{k\log\log\frac{1}{\varepsilon(\mathbf{p})}} + \log\log\frac{1}{\varepsilon(\mathbf{p})}\right)\right)$$

  samples, with probability at least 2/3; and, finally,

- show that this constant-probability bound on the number of samples also holds *in expectation.*

That is, in the "far" case this algorithm never does much worse (up to a $\log\log$ factor) than an ideal algorithm provided with the exact value $\varepsilon(\mathbf{p})$ and asked to distinguish between $\mathbf{p} = \mathbf{u}_k$ and $\mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{u}_k) = \varepsilon(\mathbf{p})$.

Solution 2.11. As discussed in Section 1.1, the overall idea is to run a uniformity tester sequentially, with varying parameters (the j-th instance, for $0 \leq j \leq L$, being run with parameters $k, \varepsilon_j, \delta_j$, for decreasing values of ε_j), and to stop and return 0 if any of the tester's invocations returns 0. If all of the $L + 1$ invocations returns 1, then we return 1.

Specifically, we set $L := \lceil \log(1/\varepsilon) \rceil$, and for $0 \leq j \leq L$ choose

$$\varepsilon_j := 2^{-j}, \qquad \delta_j := \frac{2}{\pi^2 (j+1)^2}$$

If the unknown distribution $\mathbf{p}$ *is* uniform, then by a union bound all $L + 1$ invocations of the uniformity tester will return 1 with overall probability at least

$$1 - \sum_{j=0}^{L} \delta_j \geq 1 - \sum_{j=0}^{\infty} \delta_j = 1 - \frac{2}{\pi^2} \sum_{j=0}^{\infty} \frac{1}{(j+1)^2} = \frac{2}{3}\,.$$

However, if $\varepsilon(\mathbf{p}) := \mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{u}_k) > \varepsilon$, then there exists $1 \leq j(\mathbf{p}) \leq L$ such that

$$\frac{1}{2^{j(\mathbf{p})}} < \varepsilon(\mathbf{p}) \leq \frac{1}{2^{j(\mathbf{p})-1}}$$

and so either the algorithm rejects before reaching invocation $j(\mathbf{p})$ (which is alright) or reaches invocation $j(\mathbf{p})$, when it then rejects with probability at least

$$1 - \delta_{j(\mathbf{p})} = 1 - \frac{2}{\pi^2 (j(\mathbf{p}) + 1)^2} \geq 1 - \frac{1}{2\pi^2} \geq \frac{2}{3}.$$

This deals with the correctness; we still need to establish the 3 components of the sample complexity (worst-case as a function of ε, with high probability as a function of $\varepsilon(\mathbf{p})$ in the non-uniform case, and on expectation as a function of $\varepsilon(\mathbf{p})$ in the non-uniform case). Let $n(k, \varepsilon, \delta)$ denote the optimal sample complexity from Eq. (2.50).

- Since we are running at most $L + 1$ invocations of the uniformity tester, the sample complexity is at most the sum of these $L + 1$ sample complexities, and so is bounded by

$$\sum_{j=0}^{L} n(k,\varepsilon_j,\delta_j) \asymp \sum_{j=0}^{L} \frac{\sqrt{k\log(1/\delta_j)}+\log(1/\delta_j)}{\varepsilon_j^2}$$
$$\asymp \sum_{j=0}^{L} 2^{2j}\left(\sqrt{k\log(j+1)}+\log(j+1)\right)$$
$$\asymp 2^{2L}\left(\sqrt{k\log(L+1)}+\log(L+1)\right)$$

which, recalling our choice of $L = \lceil \log(1/\varepsilon) \rceil$, is

$$O\left(\frac{\sqrt{k\log\log(1/\varepsilon)}+\log\log(1/\varepsilon)}{\varepsilon^2}\right)$$

as claimed.

- In the non-uniform case, where $\varepsilon(\mathbf{p}) > \varepsilon$, it suffices to note that with probability at least $1-\delta_{j(\mathbf{p})}$ the algorithm with stop at the $j(\mathbf{p})$ invocation (where $j(\mathbf{p})$ is as defined above, within a factor two of $\varepsilon(\mathbf{p})$). We thus can reuse the above analysis of the sample complexity, but stopping at $j(\mathbf{p})$, to obtain that with probability at least $1-\delta_{j(\mathbf{p})} \geq 2/3$ the number of samples taken will be

$$O\left(2^{2j(\mathbf{p})}\left(\sqrt{k\log(j(\mathbf{p})+1)}+\log(j(\mathbf{p})+1)\right)\right),$$

which is $O\left(\frac{\sqrt{k\log\log(1/\varepsilon(\mathbf{p}))}+\log\log(1/\varepsilon(\mathbf{p}))}{\varepsilon(\mathbf{p})^2}\right)$.

- Finally, to get the same guarantee *on expectation* in the non-uniform case, let T (a random variable) denote the index of the last invocation of the tester, so that $0 \leq T \leq L$. We have seen that $T = j(\mathbf{p})$ with probability at least $1-\delta_{j(\mathbf{p})}$; but we have much stronger guarantees! Namely, for any $j > j(\mathbf{p})$,

$$\Pr[\,T \geq j\,] \leq \prod_{i=j(\mathbf{p})}^{j-1} \delta_i \leq \delta_{j(\mathbf{p})}^{j-j(\mathbf{p})}$$

since this means that all invocations from $j(\mathbf{p})$ onwards must have failed (*i.e.*, did not reject even though they should have). The expected sample complexity is then at most

$$\sum_{j=0}^{L} \Pr[\,T \geq j\,]\cdot n(k,\varepsilon_j,\delta_j) \leq \sum_{j=0}^{j(\mathbf{p})} n(k,\varepsilon_j,\delta_j) + \sum_{j=j(\mathbf{p})+1}^{L} \delta_{j(\mathbf{p})}^{j-j(\mathbf{p})} n(k,\varepsilon_j,\delta_j)\,.$$

We have already analyzed the first term of the RHS, showing it was $O\left(\frac{\sqrt{k\log\log(1/\varepsilon(\mathbf{p}))}+\log\log(1/\varepsilon(\mathbf{p}))}{\varepsilon(\mathbf{p})^2}\right)$. The second term is at most (ignoring constants and recalling the setting of δ_j)

$$\begin{aligned}
\sum_{j=j(\mathbf{p})+1}^{L} \frac{1}{(2\pi^2)^{j-j(\mathbf{p})}} \cdot 2^{2j}\left(\sqrt{k\log(j+1)}+\log(j+1)\right) \\
&\leq 2^{2j(\mathbf{p})}\sum_{j=1}^{\infty}\left(\frac{2}{\pi^2}\right)^{j}\left(\sqrt{k\log(j+j(\mathbf{p})+1)}+\log(j+j(\mathbf{p})+1)\right) \\
&\lesssim 2^{2j(\mathbf{p})}\left(\sqrt{k\log(j(\mathbf{p})+1)}+\log(j(\mathbf{p})+1)\right)
\end{aligned}$$

which is $O\left(\frac{\sqrt{k\log\log(1/\varepsilon(\mathbf{p}))}+\log\log(1/\varepsilon(\mathbf{p}))}{\varepsilon(\mathbf{p})^2}\right)$ as well. Here, we relied on the fact that $\frac{2}{\pi^2}<1$ to be able to bound the converging series by its first term: at its core, this is possible because the probability $\Pr[\,T\geq j\,]$ to keep running the tests after reaching $j(\mathbf{p})$ decreases exponentially quickly with j.

Exercise 2.12. Given two probability distributions $\mathbf{p},\mathbf{q}$, an integer $n\geq 1$, and a parameter $\alpha\in[0,1]$, consider the following two sampling processes:

- Sample $N\sim\operatorname{Poisson}(n)$, and draw N i.i.d. samples from the mixture $(1-\alpha)\mathbf{p}+\alpha\mathbf{q}$.
- Sample $N\sim\operatorname{Poisson}(n)$, and draw N i.i.d. samples from $\mathbf{p}$. Then, for each $1\leq i\leq N$, independently sample $B_i\sim\operatorname{Bern}(\alpha)$: if $B_i=1$, replace the i-th sample by a new (and independent from everything else) sample drawn from $\mathbf{q}$.

Show that these two processes result in the same distribution.

Solution 2.12 (Sketch). Condition on a given value of N. In both cases, the N samples are mutually independent, so it is enough to show that the marginal distribution of a single sample is the same in both cases. This part is then quite straightforward: suppose X is a sample from the mixture $(1-\alpha)\mathbf{p}+\alpha\mathbf{q}$, and Y obtained by the second process (draw independently $Y'\sim\mathbf{p}$, $Y''\sim\mathbf{q}$, and $B\sim\sim\operatorname{Bern}(\alpha)$, and set $Y=(1-B)\cdot Y'+B\cdot Y''$). Then, for all x,

$$\Pr[X=x]=(1-\alpha)\mathbf{p}(x)+\alpha\mathbf{q}(x)$$

while

$$\begin{aligned}
\Pr[Y=x] &= \Pr[\,Y=x\mid B=0\,]\cdot\Pr[B=0]+\Pr[\,Y=x\mid B=1\,]\cdot\Pr[B=1] \\
&= \Pr[\,Y'=x\mid B=0\,]\cdot(1-\alpha)+\Pr[\,Y''=x\mid B=1\,]\cdot\alpha \\
&= \mathbf{p}(x)(1-\alpha)+\mathbf{q}(x)\alpha\,.
\end{aligned}$$

Exercise 2.13 ($\star$)**.** Establish the analogue of Theorem 2.22 for the *two-distribution* case (when both $\mathbf{p}, \mathbf{q}$ are unknown, and you are given n i.i.d. samples from each). Specifically, consider the statistic $Z' = \sum_{i=1}^{k}((X_i - Y_i)^2 - X_i - Y_i)$ for which you will have to establish the following counterpart of Claim 2.2:

Claim 2.0.1. If $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$ are independent, then $\mathbb{E}\left[(X-Y)^2 - X - Y\right] = (\lambda-\mu)^2$ and $\mathbb{E}\left[((X-Y)^2 - X - Y)^2\right] = (\lambda-\mu)^4 + 2(\lambda+\mu)^2 + 4(\lambda+\mu)(\lambda-\mu)^2$.

Show that the sample complexity is $O(\max(\|\mathbf{p}\|_2, \|\mathbf{q}\|_2)/\varepsilon^2)$. Try to establish the (incomparable) bound $O(\min(\|\mathbf{p}\|_2, \|\mathbf{q}\|_2)/\varepsilon^2 + 1/\varepsilon)$.

Solution 2.13. We will not here establish Claim 2.1, which can be checked by tedious computations *(if you have an elegant proof of this claim, let me know!)*. Consider the following algorithm, the analogue of Algorithm 11 for two unknown distributions: The proof will be

```
Require: Multisets of $n$ samples each, $x_1, \dots, x_n \in \mathcal{X}$ and $y_1, \dots, y_n \in \mathcal{X}$, parameters
    $\varepsilon \in (0, 1]$.                                      ▷ Assumes Poissonization
Set $\tau \leftarrow 3n^2\varepsilon^2$
Compute
    $$Z = \sum_{j \in \mathcal{X}} \left( (N_j - N'_j)^2 - N_j - N'_j \right)$$
    where $N_j \leftarrow \sum_{t=1}^{n} \mathbb{1}\{x_t = j\}$, $N'_j \leftarrow \sum_{t=1}^{n} \mathbb{1}\{y_t = j\}$.
if $Z \geq \tau$ then return 0                                     ▷ $\mathbf{p}, \mathbf{q}$ far (in $\ell_2$)
else return 1                                                      ▷ $\mathbf{p}, \mathbf{q}$ close (in $\ell_2$)
```

Algorithm 1: Robust ℓ_2 Tester (for Closeness)

very similar to that of Theorem 2.22. Assuming Poissonization, if $x_1, \dots, x_n$ are independent samples from $\mathbf{p}$ and $y_1, \dots, y_n$ from $\mathbf{q}$, we get that, for all $j \in \mathcal{X}$, $N_j \sim \text{Poisson}(n\mathbf{p}(j))$ and $N'_j \sim \text{Poisson}(n\mathbf{q}(j))$ and so, by Claim 2.0.1 above,

$$\mathbb{E}_{\mathbf{p}}[Z] = \sum_{j=1}^{k} (n\mathbf{p}(j) - n\mathbf{q}(j))^2 = n^2 \|\mathbf{p} - \mathbf{q}\|_2^2 \tag{2.2}$$

and

$$\begin{aligned}
\operatorname{Var}_{\mathbf{p}}[Z] &= \sum_{j=1}^{k} \operatorname{Var}\Big[\Big((N_j - N'_j)^2 - N_j - N'_j\Big)\Big] \\
&= \sum_{j=1}^{k} \Big(\mathbb{E}\Big[((N_j - N'_j)^2 - N_j - N'_j)^2\Big] - n^4(\mathbf{p}(j) - \mathbf{q}(j))^4\Big) \\
&= \sum_{j=1}^{k} \Big(2n^2(\mathbf{p}(j) + \mathbf{q}(j))^2 + 4n^3(\mathbf{p}(j) + \mathbf{q}(j))(\mathbf{p}(j) - \mathbf{q}(j))^2\Big) \\
&\leq 4n^2\Big(\|\mathbf{p}\|_2^2 + \|\mathbf{q}\|_2^2\Big) + 4n^3(\|\mathbf{p}\|_2 + \|\mathbf{q}\|_2)\|\mathbf{p} - \mathbf{q}\|_2^2 \qquad (2.3)
\end{aligned}$$

the last step again using $\mathbf{p}(j) + \mathbf{q}(j) \leq \|\mathbf{p}\|_\infty + \|\mathbf{q}\|_\infty \leq \|\mathbf{p}\|_2 + \|\mathbf{q}\|_2$.

- If $\|\mathbf{p} - \mathbf{q}\|_2 \leq \varepsilon$, then $\mathbb{E}_{\mathbf{p}}[Z] \leq n^2\varepsilon^2$ and by Markov's inequality

$$\Pr\Big[Z \geq 3n^2\varepsilon^2\Big] \leq \frac{\mathbb{E}_{\mathbf{p}}[Z]}{3n^2\varepsilon^2} \leq \frac{1}{3}$$

- If $\|\mathbf{p} - \mathbf{q}\|_2 \geq 2\varepsilon$, then $\mathbb{E}_{\mathbf{p}}[Z] \geq 4n^2\varepsilon^2$ and by Chebyshev's

$$\begin{aligned}
\Pr\Big[Z < 3n^2\varepsilon^2\Big] &\leq \frac{16 \operatorname{Var}_{\mathbf{p}}[Z]}{\mathbb{E}_{\mathbf{p}}[Z]^2} \leq \frac{64(\|\mathbf{p}\|_2^2 + \|\mathbf{q}\|_2^2)}{n^2\|\mathbf{p} - \mathbf{q}\|_2^4} + \frac{64(\|\mathbf{p}\|_2 + \|\mathbf{q}\|_2)}{n\|\mathbf{p} - \mathbf{q}\|_2^2} \\
&\leq \frac{8\max(\|\mathbf{p}\|_2, \|\mathbf{q}\|_2)^2}{n^2\varepsilon^4} + \frac{32\max(\|\mathbf{p}\|_2, \|\mathbf{q}\|_2)}{n\varepsilon^2}
\end{aligned}$$

  which is less than $1/3$ for $n \geq 100\max(\|\mathbf{p}\|_2, \|\mathbf{q}\|_2)/\varepsilon^2$.

This establishes the first bound from the exercise.

The second is a little trickier, and we here only provide an outline.[1] It will relies on the following lemma (which will not be proven here, but can be shown by, *e.g.*, adapting Batu and Canonne (2017, Lemma 3.1)):

Lemma 2.0.2 (Constant-factor estimate of the ℓ_2 norm)**.** There exists an algorithm which, given n i.i.d. samples from an unknown distribution $\mathbf{p} \in \Delta_{\mathbb{N}}$ and a parameter $\tau \in (0, 1]$, outputs a value $\hat{\rho} \in (\tau, 1] \cup \{\perp\}$ and has the following guarantees:

- if $\|\mathbf{p}\|_2 \geq \tau$, then $\hat{\rho} \neq \perp$ with probability at least $9/10$;
- if $\hat{\rho} \neq \perp$, then $\|\mathbf{p}\|_2/2 \leq \hat{\rho} \leq 2\|\mathbf{p}\|_2$ with probability at least $9/10$;

[1] If you have a fully written solution you'd like to include, or if you think something isn't quite right, do contact me!

as long as $n = \Theta(1/\tau)$.

Essentially, the above state that it is possible to use $O(1/\tau)$ samples to obtain (with high constant probability) a constant-factor estimate of the ℓ_2 norm of a distribution if we are promised that this norm is at least τ (and, if that norm is much less than τ, then we will detect it).

With this at our disposal, we can proceed as follows:

1. Set $\tau \asymp \varepsilon$, and use the above algorithm with $O(1/\varepsilon)$ samples from both $\mathbf{p}$ and $\mathbf{q}$ to try and get estimates of their ℓ_2 norms.

2. If we get $\perp$ for both, then we know that $\|\mathbf{p} - \mathbf{q}\|_2 \leq \|\mathbf{p}\|_2 + \|\mathbf{q}\|_2 \leq 2\tau \leq \varepsilon$ and we are done;

3. If we get estimates $\hat{\rho}_{\mathbf{p}}$, $\hat{\rho}_{\mathbf{q}} \geq \tau$ for both, then

    (a) we check that those two values are within a (sufficiently) large constant factor of each other: if not, we know that $\|\mathbf{p} - \mathbf{q}\|_2 \geq 2\varepsilon$ and we are done;

    (b) if they are within constant factors, we get that $\|\mathbf{p}\|_2 \asymp \|\mathbf{q}\|_2$, and so $\min(\|\mathbf{p}\|_2, \|\mathbf{q}\|_2) \asymp \max(\|\mathbf{p}\|_2, \|\mathbf{q}\|_2)$ and we can use the algorithm from the first part, with sample complexity $O(\max(\|\mathbf{p}\|_2, \|\mathbf{q}\|_2)/\varepsilon^2) = O(\min(\|\mathbf{p}\|_2, \|\mathbf{q}\|_2)/\varepsilon^2)$, and we are done;

4. If we get an estimate for one (say $\mathbf{p}$ but $\perp$ for the other (say $\mathbf{q}$), then we know that $\|\mathbf{q}\|_2 \leq \tau$ and have a constant-factor estimate $\hat{\rho}_{\mathbf{p}}$ for $\|\mathbf{p}\|_2$.

    (a) If $\hat{\rho}_{\mathbf{p}} \gg \varepsilon$, then we know that $\|\mathbf{p} - \mathbf{q}\|_2 \geq \|\mathbf{p}\|_2 - \tau \gg 2\varepsilon$ and we are done.

    (b) If $\hat{\rho}_{\mathbf{p}} \asymp \varepsilon$, then we know that $\max(\|\mathbf{p}\|_2, \|\mathbf{q}\|_2 \asymp \varepsilon$, so we can use the algorithm from the first part, with sample complexity $O(\max(\|\mathbf{p}\|_2, \|\mathbf{q}\|_2)/\varepsilon^2) = O(1/\varepsilon)$, and we are done.

Overall, the sample complexity is $O\big(\max\big(\min(\|\mathbf{p}\|_2, \|\mathbf{q}\|_2)/\varepsilon^2\big), 1/\varepsilon\big)$, establishing the second bound from the exercise.

Exercise 2.14. Show that the transformation Φ from Section 2.2.2 (Eq. (2.59)) "maps χ^2 divergence to ℓ_2 distance" in the following, approximate way: for any $\mathbf{p}, \mathbf{q} \in \Delta_k$,

$$\|\Phi_{\mathbf{q}}(\mathbf{p}) - \Phi_{\mathbf{q}}(\mathbf{q})\|_2^2 = \sum_{i \in \mathcal{X}} \frac{(\mathbf{p}(i) - \mathbf{q}(i))^2}{1 + \lfloor k\mathbf{q}(i) \rfloor}.$$

Conclude that, assuming $\min_i \mathbf{q}(i) \geq 1/(2k)$ (as we could in Section 2.2.1 after using the "mixture trick" of Eq. (2.52)),

$$\frac{1}{2}\chi^2(\mathbf{p} \mid\mid \mathbf{q}) \leq k\|\Phi_{\mathbf{q}}(\mathbf{p}) - \Phi_{\mathbf{q}}(\mathbf{q})\|_2^2 \leq \chi^2(\mathbf{p} \mid\mid \mathbf{q})$$

for every $\mathbf{p} \in \Delta_k$.

Solution 2.14. From the expression of Φ and with the notation of Section 2.2.2, we have

$$\begin{aligned}
\|\Phi_{\mathbf{q}}(\mathbf{p}) - \Phi_{\mathbf{q}}(\mathbf{q})\|_2^2 &= \sum_{j=1}^{k'} (\Phi_{\mathbf{q}}(\mathbf{p}) - \Phi_{\mathbf{q}}(\mathbf{q}))^2 \\
&= \sum_{j=1}^{k'} \left(\sum_{i=1}^{k} \left(\frac{\mathbf{p}(i)}{k_i} - \frac{\mathbf{q}(i)}{k_i} \right) \mathbb{1}\{j \in S_i\} \right)^2 \\
&= \sum_{j=1}^{k'} \sum_{i=1}^{k} \mathbb{1}\{j \in S_i\} \left(\frac{\mathbf{p}(i)}{k_i} - \frac{\mathbf{q}(i)}{k_i} \right)^2 && (*) \\
&= \sum_{i=1}^{k} \frac{1}{k_i^2} (\mathbf{p}(i) - \mathbf{q}(i))^2 \sum_{j=1}^{k'} \mathbb{1}\{j \in S_i\} \\
&= \sum_{i=1}^{k} \frac{(\mathbf{p}(i) - \mathbf{q}(i))^2}{k_i}
\end{aligned}$$

where for $(*)$ we used that exactly one term of the inner sum in non-zero (as each $j \in [k]$ belongs to exactly one set S_i), allowing us to bring the sum and indicator outside the square; and later that $\sum_{j=1}^{k'} \mathbb{1}\{j \in S_i\} = |S_i| = k_i$. The result then follows from the definition of $k_i = 1 + \lfloor k\mathbf{q}(i) \rfloor$ (for $i \in [k]$).
Finally, if $\min_i \mathbf{q}(i) \geq 1/(2k)$ we get that

$$k\mathbf{q}(i) \leq 1 + \lfloor k\mathbf{q}(i) \rfloor \leq 2k\mathbf{q}(i), \qquad i \in [k]$$

where the RHS is obtained by observing that $\max_{x \geq 1/2} \frac{1+\lfloor x \rfloor}{x} = 1/2$. This directly implies

$$\underbrace{\frac{1}{2} \sum_{i=1}^{k} \frac{(\mathbf{p}(i) - \mathbf{q}(i))^2}{\mathbf{q}(i)}}_{\frac{1}{2}\chi^2(\mathbf{p}||\mathbf{q})} \leq \underbrace{k \sum_{i=1}^{k} \frac{(\mathbf{p}(i) - \mathbf{q}(i))^2}{1 + \lfloor k\mathbf{q}(i) \rfloor}}_{k\|\Phi_{\mathbf{q}}(\mathbf{p}) - \Phi_{\mathbf{q}}(\mathbf{q})\|_2^2} \leq \underbrace{\sum_{i=1}^{k} \frac{(\mathbf{p}(i) - \mathbf{q}(i))^2}{\mathbf{q}(i)}}_{\chi^2(\mathbf{p}||\mathbf{q})}$$

as wanted.

Exercise 2.15 $(\star\star)$**.** Generalize the transformation Φ from Section 2.2.3 in two ways: first, by replacing the mixture $\Phi_{\mathbf{q}}^{(3)}(\mathbf{p}) = \frac{1}{2}\mathbf{p} + \frac{1}{2}\mathbf{u}_k$ by $\alpha\mathbf{p} + (1-\alpha)\mathbf{u}_k$, where $\alpha \in (0,1)$. Second, by replacing the choice $k' = 4k$ in $\Phi^{(2)}(\mathbf{p})$ by $k' = \beta k$, for some integer β such that $\beta(1-\alpha) \geq 1$.

1. By tracking down the various restrictions on α, β and their use across $\Phi^{(1)}$, $\Phi^{(2)}$, and $\Phi^{(3)}$, show that doing so now maps identity testing with parameters (k, ε) to uniformity testing with parameters

$$\left(\beta k, \alpha \left(1 - \frac{1}{\beta(1-\alpha)} \right) \varepsilon \right)$$

2. Check that setting $(\alpha, \beta) = (1/2, 4)$ as in Section 2.2.3 recovers Theorem 2.28, and the blowup factor of 32 discussed at the end of the section.

3. Recalling that the sample complexity scales as $\sqrt{k}/\varepsilon^2$, optimize over (α, β) to find the optimal choice of parameters, and prove that the resulting blowup is ≈ 12.2.

4. What would be the optimal choice of (α, β), and the corresponding blowup, in a setting where the sample complexity of uniformity testing scales as k/ε^2 instead of $\sqrt{k}/\varepsilon^2$? (This is not that far-fetched: we will see in Section 4.3 an example of such a setting.)

Solution 2.15. The solution below follows the aalysis of Section 2.2.3.

1. By setting things as suggested, we will be able to assume in our analysis of $\Phi^{(2)}$ that $k'\mathbf{q}(i) \geq k'(1-\alpha)/k = \beta(1-\alpha)$ for all $i \in [k]$, and then

$$\min_i \frac{\lfloor k'\mathbf{q}(i) \rfloor}{k'\mathbf{q}(i)} \geq \frac{\beta(1-\alpha)-1}{\beta(1-\alpha)} \tag{2.4}$$

(the $\lfloor k'\mathbf{q}(i) \rfloor$ in the numerator is the reason we need to enforce $\beta(1-\alpha) \geq 1$). We then get the analogue of Eq. (2.65): for every $\mathbf{p}_1, \mathbf{p}_2 \in \Delta_k$,

$$\mathrm{d_{TV}}\Big(\Phi^{(2)}(\mathbf{p}_1), \Phi^{(2)}(\mathbf{p}_2)\Big) \geq \frac{\beta(1-\alpha)-1}{\beta(1-\alpha)} \mathrm{d_{TV}}(\mathbf{p}_1, \mathbf{p}_2)\,, \tag{2.5}$$

We then only have to take into account the last piece, *i.e.*, that $\Phi^{(3)}$ does allow us to assume that Eq. (2.4) holds, but doing so comes at the cost of shrinking the total variation distance by a factor α as well; so that, when combining $\Phi^{(1)}$, $\Phi^{(2)}$, and $\Phi^{(1)}$ all together, we get a mapping Φ such that, for every $\mathbf{p}_1, \mathbf{p}_2 \in \Delta_k$,

$$\mathrm{d_{TV}}(\Phi(\mathbf{p}_1), \Phi(\mathbf{p}_2)) \geq \alpha\left(1 - \frac{1}{\beta(1-\alpha)}\right) \mathrm{d_{TV}}(\mathbf{p}_1, \mathbf{p}_2)\,. \tag{2.6}$$

Overall, for any choice of $\alpha \in [0, 1]$ and $\beta \geq 1$ such that $\beta(1-\alpha) \geq 1$, we can convert an identity testing instance with parameters (k, ε) to a uniformity testing instance with parameters $\left(k' = \beta k, \varepsilon' = \alpha\left(1 - \frac{1}{\beta(1-\alpha)}\right)\right)$.

2. For $\alpha = 1/2$ and $\beta = 4$, we get $\alpha\left(1 - \frac{1}{\beta(1-\alpha)}\right) = 1/4$, and so

$$\frac{\sqrt{k'}}{\varepsilon'^2} = \frac{\sqrt{\beta}}{\left(\alpha\left(1 - \frac{1}{\beta(1-\alpha)}\right)\right)^2} \cdot \frac{\sqrt{k}}{\varepsilon^2} = \frac{\sqrt{4}}{(1/4)^2} \cdot \frac{\sqrt{k}}{\varepsilon^2} = \boxed{32} \cdot \frac{\sqrt{k}}{\varepsilon^2}$$

3. From the above, we want to choose α, β to minimize the factor

$$\frac{\sqrt{\beta}}{\left(\alpha\left(1 - \frac{1}{\beta(1-\alpha)}\right)\right)^2} \tag{2.7}$$

subject to $\beta(1-\alpha) \geq 1$, $\alpha \in [0,1]$, $\beta \geq 1$. For the sake of the minimization, it is easier to minimize the square of this quantity, and to set $\gamma := \beta(1-\alpha)$, so that we seek to find the minimizer of

$$\frac{\beta}{\alpha^4\left(1-\frac{1}{\beta(1-\alpha)}\right)^4} = \frac{1}{\alpha^4(1-\alpha)} \cdot \frac{\gamma^5}{(\gamma-1)^4}$$

subject to $\alpha \in [0,1]$ and $\gamma \geq 1$. The variables are now separated, and we can minimize separately in α and in γ. This leads to minimizers $\alpha^* = 4/5$ and $\gamma^* = 5$. Getting back to our original quantity, we therefore get that it is minimized for $(\alpha, \beta) = (4/5, 25)$, for which the blowup factor in Eq. (2.7) is approximately $\boxed{12.21}$.

4. If the sample complexity were to scale as k/ε^2 instead of $\frac{\sqrt{k}}{\varepsilon^2}$, then the blowup factor would become

$$\frac{k'/\varepsilon'^2}{k/\varepsilon^2} = \frac{\beta}{\left(\alpha\left(1-\frac{1}{\beta(1-\alpha)}\right)\right)^2} \tag{2.8}$$

and minimizing this, as before subject to $\beta(1-\alpha) \geq 1$, $\alpha \in [0,1]$, $\beta \geq 1$, leads in the same way to minimizers $(\alpha, \beta) = (2/3, 9)$, for which the blowup factor in Eq. (2.8) is approximately $\boxed{45.56}$.

3 Information-theoretic lower bounds

Exercise 3.1. Combine (the second part of) Lemma B.4 with (the first part of) Lemma B.5 to obtain Eq. (3.16) from Eq. (3.7). Use it to derive Theorem 3.3.

Solution 3.1. Invoking Lemma B.4 (specifically, Eq. (B.9)) and Lemma B.5, we get

$$\begin{aligned}
1 - 2\delta &\leq \mathrm{d_{TV}}\left(\mathbb{E}_\theta\left[\mathbf{p}_\theta^{\otimes n}\right], \mathbf{p}_1^{\otimes n}\right) && \text{(Eq. (3.7))}\\
&\leq 1 - \frac{1}{2}e^{-\mathrm{D}\left(\mathbb{E}_\theta\left[\mathbf{p}_\theta^{\otimes n}\right] \| \mathbf{p}_1^{\otimes n}\right)} && \text{(Eq. (B.9))}\\
&\leq 1 - \frac{1}{2(1+\chi^2\left(\mathbb{E}_\theta\left[\mathbf{p}_\theta^{\otimes n}\right] \mid\mid \mathbf{p}_1^{\otimes n}\right)} && \text{(Lemma B.5)}
\end{aligned}$$

which, reorganizing the inequality, yields Eq. (3.16):

$$\frac{1}{4\delta} \leq 1 + \chi^2\left(\mathbb{E}_\theta\left[\mathbf{p}_\theta^{\otimes n}\right] \mid\mid \mathbf{p}_1^{\otimes n}\right).$$

Now, since in Eq. (3.14) we had derived

$$\chi^2\left(\mathbb{E}_\theta\left[\mathbf{p}_\theta^{\otimes n}\right] \mid\mid \mathbf{p}_1^{\otimes n}\right) = \mathbb{E}_{\theta,\theta'}\left[(1+H(\theta,\theta'))^n\right] - 1 \leq e^{\frac{81\varepsilon^4 n^2}{k}} - 1$$

we can plug this in Eq. (3.16) to obtain the necessary inequality

$$\frac{1}{4\delta} \le e^{\frac{81\varepsilon^4 n^2}{k}}, \tag{3.1}$$

that is, $n \ge \frac{\sqrt{k \log(1/(4\delta))}}{9\varepsilon^2}$; proving Theorem 3.3. *(As a side note: one can similarly prove a lower bound of $n = \Omega(\log(1/\delta)/\varepsilon^2)$ using Lemma B.4 instead of Pinsker's inequality in Eq. (3.4); and combining the two leads to the following bound for uniformity testing,* **E:** Try it!

$$n = \Omega\left(\frac{\sqrt{k \log(1/(\delta))} + \log(1/\delta)}{\varepsilon^2}\right), \tag{3.2}$$

which is optimal.

Erratum: The first part of the next exercise (about the distance to $\mathcal{P}_k$) was incorrectly stated in the published version.

Exercise 3.2. Fix a property $\mathcal{P}_k \subseteq \Delta_k$ of distributions, and denote by $\tilde{\mathcal{P}}_k$ its "extension to probability measures" (not just probability distributions) defined as follows:

$$\tilde{\mathcal{P}}_k := \{\, \alpha\mathbf{q} : \ \mathbf{q} \in \mathcal{P}_k, \alpha \ge 0 \,\} \tag{3.3}$$

(for instance, for uniformity, $\mathcal{P}_k = \{\mathbf{u}_k\}$ and $\tilde{\mathcal{P}}_k = \{\alpha\mathbf{1}_k\}_{\alpha\ge 0}$.) Let $\mathbf{p}$ be a measure (not necessarily a probability measure) such that the ℓ_1 distance between $\mathbf{p}$ and $\tilde{\mathcal{P}}_k$ satisfies $\ell_1(\mathbf{p}, \tilde{\mathcal{P}}_k) > 2\varepsilon$, and $1/2 \le \|\mathbf{p}\|_1 \le 3/2$. Defining $\mathbf{p}' := \mathbf{p}/\|\mathbf{p}\|_1$ (an actual probability distribution), provide a lower bound on $\mathrm{d}_{\mathrm{TV}}(\mathbf{p}', \mathcal{P}_k)$. Moreover, show that obtaining n "samples" from the Poisson process with measure $\mathbf{p}$ is equivalent to getting $\mathrm{Poisson}(n\|\mathbf{p}\|_1)$ samples from the distribution $\mathbf{p}'$.

Conclude with how one could use a testing algorithm $\mathcal{A}$ for property $\mathcal{P}_k$ given $\mathrm{Poisson}(n)$ samples (*i.e.*, in the Poissonized sampling model) to distinguish between two families of measures (yes- and no-instances) far in ℓ_1 distance, thus justifying the relaxed assumption from Section 3.2.

Solution 3.2. Let $\mathbf{p}$, $\mathcal{P}_k$, $\tilde{\mathcal{P}}_k$, and $\mathbf{p}'$ as above. Fix any $\mathbf{q} \in \mathcal{P}_k$; we have

$$\begin{aligned} \|\mathbf{p}' - \mathbf{q}\|_1 &= \left\| \frac{\mathbf{p}}{\|\mathbf{p}\|_1} - \mathbf{q} \right\|_1 = \frac{1}{\|\mathbf{p}\|_1} \|\mathbf{p} - \|\mathbf{p}\|_1 \mathbf{q}\|_1 \\ &\ge \frac{2}{3} \|\mathbf{p} - \|\mathbf{p}\|_1 \mathbf{q}\|_1 > \frac{2}{3} \cdot 2\varepsilon \end{aligned}$$

using for the last inequality that $\|\mathbf{p}\|_1 \mathbf{q} \in \tilde{\mathcal{P}}_k$, as a (positive) rescaling of $\mathbf{q} \in \mathcal{P}_k$. As $\mathbf{q}$ was arbitrary, this implies $\mathrm{d}_{\mathrm{TV}}(\mathbf{p}', \mathcal{P}_k) > \frac{2}{3}\varepsilon$.

Then, the output of a Poisson process with parameter $n\mathbf{p}$ (*i.e.*, parameter n and underlying measure $\mathbf{p}$) is by definition a set of mutually independent values $N_1, \dots, N_k$, where N_i is distributed as $\operatorname{Poisson}(n\mathbf{p}(i)) = \operatorname{Poisson}(n\|\mathbf{p}\|_1 \mathbf{p}'(i))$. Which is exactly what one get by drawing $\operatorname{Poisson}(n')$ i.i.d. samples from $\mathbf{p}'$ for $n' := n\|\mathbf{p}\|_1$ (not necessarily an integer).

Finally, suppose we have a testing algorithm $\mathcal{A}$ for property $\mathcal{P}_k$, which succeeds when given $\operatorname{Poisson}(n)$ samples for $n = n(k, \varepsilon, \delta)$. Then we can use it to distinguish whether the unknown measure $\mathbf{p}$ is a yes- ($\mathbf{p}/\|\mathbf{p}\|_1 \in \mathcal{P}_k$) or a no- instance ($\mathbf{p}/\|\mathbf{p}\|_1$ 2ε-far in ℓ_1 distance, or equivalently ε-far in TV distance) in our "relaxed," "Poisson process" setting by feeding the output of our Poisson process with parameter $n'\mathbf{p}$ for $n' := 2n(k, \frac{2}{3}\varepsilon, \delta)$. By the above, this corresponds to $\operatorname{Poisson}(n'\|\mathbf{p}\|_1)$ samples from $\mathbf{p}'$, which (i) in the yes case is in $\mathcal{P}_k$, and (ii) in no case will be $\frac{2}{3}\varepsilon$-far from $\mathcal{P}_k$, and

$$n'\|\mathbf{p}\|_1 = 2\|\mathbf{p}\|_1 \cdot n(k, \frac{2}{3}\varepsilon, \delta) \geq n(k, \frac{2}{3}\varepsilon, \delta)$$

since $\|\mathbf{p}\|_1 \geq 1/2$; so $\mathcal{A}$ will be correct with probability at least $1 - \delta$. This implies that any lower bound in this "relaxed" setting carries over, with only constant-factor losses in the parameters, to the Poissonized setting.

Exercise 3.3 ($\star$)**.** Recall that we defined the no-instances in Section 3.2 by Eq. (3.17) (measures, instead of *bona fide* probability measures) in order to guarantee mutual independence of $N_1, \dots, N_k$ (conditioned on $\mathbf{b}$. Check the argument to see which part of the argument would fail if we had used Eq. (3.11) instead. Then, modify the argument to fix this, and obtain the same sample complexity lower bound. *(Hint: we still have mutual independence of the $k/2$ random variables $(N_1, N_2), \dots, (N_{k-1}, N_k)$ conditioned on $\mathbf{b}$. Establish the analogue of Eq. (3.20) with $N_1 = j, N_2 = \ell$ instead of $N_1 = j$, and proceed from there.*

Solution 3.3. We can proceed as in Section 3.2 to bound $I(\mathbf{b} \wedge X)$, but now keeping in mind (as per the hint) that only the pairs (N_{2i-1}, N_{2i}) are mutually independent conditioned on $\mathbf{b}$, not the N_i themselves. We can adapt the argument, and write

$$\begin{aligned}
I(\mathbf{b} \wedge X) &= H(N_1, \dots, N_k) - H((N_1, \dots, N_k) \mid \mathbf{b}) \\
&= H(N_1, \dots, N_k) - \sum_{i=1}^{k/2} H((N_{2i-1}, N_{2i}) \mid \mathbf{b}) && \text{(conditional independence)} \\
&\leq \sum_{i=1}^{k/2} H((N_{2i-1}, N_{2i})) - \sum_{i=1}^{k/2} H((N_{2i-1}, N_{2i}) \mid \mathbf{b}) && \text{(subadditivity)} \\
&= \sum_{i=1}^{k/2} I(\mathbf{b} \wedge (N_{2i-1}, N_{2i}))\,,
\end{aligned}$$

leading to an analogue of Eq. (3.19):

$$I(\mathbf{b} \wedge X) \leq \frac{k}{2} I(\mathbf{b} \wedge (N_1, N_2))\,. \tag{3.4}$$

Building towards the counterpart of Eq. (3.20), we then have

$$
\begin{aligned}
&I(\mathbf{b} \wedge (N_1, N_2)) \\
&= \mathbb{E}_{\mathbf{b}}\Big[\mathrm{D}\Big(P_{(N_1,N_2)|\mathbf{b}} \| P_{(N_1,N_2)}\Big)\Big] \\
&\leq \mathbb{E}_{\mathbf{b}}\Big[\chi^2\Big(P_{(N_1,N_2)|\mathbf{b}} \mid\mid P_{(N_1,N_2)}\Big)\Big] \\
&= \frac{1}{2}\Big(\chi^2\Big(P_{(N_1,N_2)|\mathbf{b}=1} \mid\mid P_{(N_1,N_2)}\Big) + \chi^2\Big(P_{(N_1,N_2)|\mathbf{b}=0} \mid\mid P_{(N_1,N_2)}\Big)\Big) \\
&= \frac{1}{2}\Big(\sum_{j=0}^{\infty}\sum_{\ell=0}^{\infty} \frac{(\Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 1] - \Pr[N_1 = j, N_2 = \ell])^2}{\Pr[N_1 = j, N_2 = \ell]} \\
&\qquad + \sum_{j=0}^{\infty}\sum_{\ell=0}^{\infty} \frac{(\Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 0] - \Pr[N_1 = j, N_2 = \ell])^2}{\Pr[N_1 = j, N_2 = \ell]}\Big) \\
&= \frac{1}{2}\sum_{j=0}^{\infty}\sum_{\ell=0}^{\infty} \frac{(\Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 0] - \Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 1])^2}{\Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 0] + \Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 1]} \\
&\leq \frac{1}{2}\sum_{j=0}^{\infty}\sum_{\ell=0}^{\infty} \Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 1]\Big(1 - \frac{\Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 0]}{\Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 1]}\Big)^2
\end{aligned}
$$

using as before for the second-to-last equality that, $\mathbf{b}$ being a uniform bit, $\Pr[N_1 = j, N_2 = \ell] = \frac{1}{2}(\Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 1] + \Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 0])$ for all $j, \ell \geq 0$.

This gives us Eq. (3.5), our "new Eq. (3.20):"

$$
I(\mathbf{b} \wedge X) \leq \frac{k}{4}\sum_{j=0}^{\infty}\sum_{\ell=0}^{\infty} \Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 1]\Big(1 - \tfrac{\Pr[N_1=j,N_2=\ell|\mathbf{b}=0]}{\Pr[N_1=j,N_2=\ell|\mathbf{b}=1]}\Big)^2 . \tag{3.5}
$$

To bound it, we need to compute $\Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 1]$ and $\Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 0]$ for arbitrary integers $j, \ell \geq 0$. Thankfully, this is not too difficult (recall that we still work in the Poissonized sampling model):

$$
\begin{aligned}
\Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 1] &= e^{-\frac{n}{k}}\frac{(n/k)^j}{j!} \cdot e^{-\frac{n}{k}}\frac{(n/k)^\ell}{\ell!} = e^{-\frac{2n}{k}}\frac{(n/k)^{j+\ell}}{j!\ell!} \\
\Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 0]& \\
&= \frac{1}{2}e^{-\frac{n(1+3\varepsilon)}{k}}\frac{(n(1+3\varepsilon)/k)^j}{j!}e^{-\frac{n(1-3\varepsilon)}{k}}\frac{(n(1-3\varepsilon)/k)^\ell}{\ell!} \\
&\quad + \frac{1}{2}e^{-\frac{n(1-3\varepsilon)}{k}}\frac{(n(1-3\varepsilon)/k)^j}{j!}e^{-\frac{n(1+3\varepsilon)}{k}}\frac{(n(1+3\varepsilon)/k)^\ell}{\ell!} \\
&= e^{-\frac{2n}{k}}\frac{(n/k)^{j+\ell}}{j!\ell!} \cdot \frac{(1+3\varepsilon)^j(1-3\varepsilon)^\ell + (1-3\varepsilon)^j(1+3\varepsilon)^\ell}{2} ,
\end{aligned}
$$

and so, for all $j, \ell \geq 0$,

$$
\frac{\Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 0]}{\Pr[N_1 = j, N_2 = \ell \mid \mathbf{b} = 1]} = \frac{1}{2}\Big((1+3\varepsilon)^j(1-3\varepsilon)^\ell + (1-3\varepsilon)^j(1+3\varepsilon)^\ell\Big) \tag{3.6}
$$

(Compare this to Eq. (3.22)!). Plugging this into Eq. (3.5) leads to

$$\begin{aligned} I(\mathbf{b} \wedge X) &\leq \frac{k}{4} \sum_{j=0}^{\infty} \sum_{\ell=0}^{\infty} e^{-\frac{2n}{k}} \frac{(n/k)^{j+\ell}}{j!\ell!} \left(1 - \tfrac{(1+3\varepsilon)^j(1-3\varepsilon)^\ell+(1-3\varepsilon)^j(1+3\varepsilon)^\ell}{2}\right)^2 \\ &= \frac{k}{2} \sinh^2 \frac{9n\varepsilon^2}{k} \\ &\leq \frac{81n^2\varepsilon^4}{2k} \cdot e^{\frac{3n\varepsilon^2}{k}} \end{aligned} \tag{3.7}$$

the last equality again following either by somewhat tedious series computations, or a symbolic computation system such as Julia, Mathematica, or Maple; and the inequality being $\sinh u \leq u e^{u^2/6}$ (for $u \in \mathbb{R}$.)[1] We can then conclude as in Section 3.2: since we must have $I(\mathbf{b} \wedge X) \gtrsim 1$ by Fact 3.1, we must have $\frac{n^2\varepsilon^4}{k} \gtrsim 1$, showing the $n = \Omega\left(\sqrt{k}/\varepsilon^2\right)$ lower bound.

Exercise 3.4. Verify that applying Theorem 3.9 to (i) the uniform distribution $\mathbf{u}_k$ and (ii) the "Zipf" distribution $\mathbf{q} \in \Delta_k$ such that $\mathbf{q}(i) \propto 1/\sqrt{i}$ leads, in both cases, to an $\Omega(\sqrt{k}/\varepsilon^2)$ sample complexity lower bound for identity testing.

Solution 3.4. (i) When applying Theorem 3.9 to the uniform distribution (*i.e.*, $\mathbf{q} = \mathbf{u}_k$), for any given $\varepsilon \in (0, 1/4)$ the vector $\tilde{\mathbf{q}}_{-4\varepsilon}^{-\max}$ is the m-dimensional vector with all coordinates equal to $1/k$, where $m := k - (1 + \lfloor 4\varepsilon k \rfloor) \geq (1 - 4\varepsilon - 1/k)k$ (since we removed the first coordinate, as well as the last $\lfloor 4\varepsilon k \rfloor$). Then,

$$\left\|\tilde{\mathbf{q}}_{-4\varepsilon}^{-\max}\right\|_{2/3} = \left(m \cdot \frac{1}{k^{2/3}}\right)^{3/2} = \frac{m^{3/2}}{k} \geq (1/2 - 4\varepsilon)^{3/2}\sqrt{k} \geq \frac{\sqrt{k}}{4^{3/2}}$$

the first inequality as $k \geq 2$, and the second for $\varepsilon \leq 1/16$. This gives the lower bound $\Omega\left(\left\|\tilde{\mathbf{q}}_{-4\varepsilon}^{-\max}\right\|_{2/3}/\varepsilon^2\right) = \Omega\left(\sqrt{k}/\varepsilon^2\right)$ we wanted.

(ii) When applying the theorem with $\mathbf{q}$ set to be the Zipf distribution, that is, $\mathbf{q}(i) = \frac{1}{H_{k,1/2}\sqrt{i}}$ for every $i \in [k]$ with $H_{k,1/2} = \sum_{i=1}^{k} \frac{1}{\sqrt{i}} = \Theta\left(\sqrt{k}\right)$, for any given $\varepsilon \in (0, 1/4)$ we get

$$\left\|\tilde{\mathbf{q}}_{-4\varepsilon}^{-\max}\right\|_{2/3} = H_{k,1/2}^{-1}\left(\sum_{i=2}^{\ell} \frac{1}{i^{1/3}}\right)^{3/2} \asymp \frac{\left(\ell^{2/3}\right)^{3/2}}{H_{k,1/2}} \asymp \frac{\ell}{\sqrt{k}}$$

where ℓ is the smallest integer such that $H_{k,1/2}^{-1} \sum_{i=\ell}^{k} 1/\sqrt{i} \leq 4\varepsilon$. Computing the asymptotics of the sum shows that ℓ satisfies $\sqrt{k} - \sqrt{\ell} \sim 4\varepsilon\sqrt{k}$, leading to $\ell \sim (1 - 4\varepsilon)^2 k$. For $\varepsilon \leq 1/5$ (for instance), this again gives $\left\|\tilde{\mathbf{q}}_{-4\varepsilon}^{-\max}\right\|_{2/3} = \Theta\left(\sqrt{k}\right)$, and Theorem 3.9 then yields the same (tight, in view of the identity testing upper bound) sample complexity lower bound $\Omega\left(\sqrt{k}/\varepsilon^2\right)$.

[1] See, *e.g.*, https://math.stackexchange.com/a/1759409/75808.

Exercise 3.5. Check that you can express several of the algorithms in Section 2.1 as a function of F only (as defined in Section 3.3). Specifically, verify this for Algorithms 1, 2 and 4. Verify this also for Algorithm 3, keeping in mind that this algorithm was stated and analyzed in the Poissonized setting: what does it change?

Solution 3.5. We can rewrite Z_1, from Algorithm 1, as

$$\begin{aligned} Z_1 &= \frac{1}{\binom{n}{2}} \sum_{i\in\mathcal{X}} \binom{N_i}{2} = \frac{1}{\binom{n}{2}} \sum_{i\in\mathcal{X}} \sum_{j=0}^{n} \binom{j}{2} \mathbb{1}\{N_i = j\} = \frac{1}{\binom{n}{2}} \sum_{j=0}^{n} \binom{j}{2} \sum_{i\in\mathcal{X}} \mathbb{1}\{N_i = j\} \\ &= \frac{1}{\binom{n}{2}} \sum_{j=0}^{n} \binom{j}{2} F_j \,. \end{aligned}$$

Similarly, Z_2, from Algorithm 2, is given by

$$Z_2 = \frac{1}{n} F_1$$

while Z_4, from Algorithm 4, can be written as

$$Z_4 = \frac{1}{2} \sum_{j=0}^{n} \left| \frac{j}{n} - \frac{1}{k} \right| F_j \,.$$

We can also rewrite Z_3 as a function of the fingerprint F; however, in the Poissonized setting, $F \in \mathbb{N}^{\mathbb{N}}$ (it is no longer a finite-dimensional vector, but a sequence) and does not necessarily sum to n anymore. With this in mind, we get

$$\begin{aligned} Z_3 &= \sum_{i\in\mathcal{X}} \frac{(N_i - n/k)^2 - N_i}{n/k} \\ &= \sum_{i\in\mathcal{X}} \sum_{j=0}^{\infty} \frac{(j - n/k)^2 - j}{n/k} \mathbb{1}\{N_i = j\} \\ &= \sum_{j=0}^{\infty} \frac{(j - n/k)^2 - j}{n/k} F_j \,. \end{aligned}$$

Note that we now have to sum over all integers, not just up to n.

Exercise 3.6 ($\star$). Prove that the mapping Φ defined in Eq. (3.40) does satisfy the requirements of a reduction, for $k' = 2k$ and $\varepsilon' = \varepsilon/2$. That is, if $\mathbf{p} \in \Delta_k$ is ε-far from $\mathbf{u}_k$, then $\Phi(\mathbf{p}) \in \Delta_{2k}$ is ε'-far from every distribution $\mathbf{q} \in \mathcal{P}^{\searrow}_{2k}$. *(Hint: for any given monotone* $\mathbf{q}$*, analyse the distance* $\mathrm{d}_{\mathrm{TV}}(\Phi(\mathbf{p}), \mathbf{q})$ *according to whether* $\mathbf{q}(k) > 1/(2k)$ *or not, relating this to the set* $S \subseteq [k]$ *on which* $\mathbf{p}$ *is greater than* $\mathbf{u}_k$*.)* Moreover, show that this loss by a factor 1/2 in the distance is necessary.

Solution 3.6. Fix any $\mathbf{p} \in \Delta_k$ such that $\mathrm{d_{TV}}(\mathbf{p}, \mathbf{u}_k) > \varepsilon$, and let $S \subseteq [k]$ the set which witnesses it: $S := \{\ 1 \leq i \leq k\ :\ \mathbf{p}(i) > 1/k\ \}$ which satisfies

$$\mathbf{p}(S) - \mathbf{u}_k(S) = \mathbf{u}_k([k] \setminus S) - \mathbf{p}([k] \setminus S) > \varepsilon$$

Further define $T := [k] \setminus S \subseteq \{1, 2, \dots, k\}$, and $S + k = \{\ i + k\ :\ i \in S\ \} \subseteq \{k+1, \dots, 2k\}$. By definition of $\Phi(\mathbf{p})$, we then have

$$\mathbf{u}_{2k}(T) - \Phi(\mathbf{p})(T) = \Phi(\mathbf{p})(S+k) - \mathbf{u}_{2k}(S+k) = \frac{1}{2}(\mathbf{p}(S) - \mathbf{u}_k(S) > \frac{\varepsilon}{2}\,.$$

Now, fix any monotone distribution $\mathbf{q} \in \mathcal{P}^{\searrow}_{2k}$. We have two cases:

- If $\mathbf{q}(k) > 1/(2k)$, then, since $\mathbf{q}$ is monotone, $\mathbf{q}(i) > 1/(2k)$ for every $i \leq k$. This implies
$$\mathbf{q}(T) - \Phi(\mathbf{p})(T) \geq \mathbf{u}_{2k}(T) - \Phi(\mathbf{p})(T) > \frac{\varepsilon}{2}$$
- If $\mathbf{q}(k) \leq 1/(2k)$, then $\mathbf{q}(i) \leq 1/(2k)$ for every $i \geq k$. This implies
$$\Phi(\mathbf{p})(S+k) - \mathbf{q}(S+k) \geq \Phi(\mathbf{p})(S+k) - \mathbf{u}_{2k}(S+k) > \frac{\varepsilon}{2}$$

This shows that $\mathrm{d_{TV}}(\Phi(\mathbf{p}), \mathbf{q}) > \varepsilon/2$. For the "necessary" part, consider $\mathbf{p}$ such that $\mathbf{p}(1) = 1$.

Exercise 3.7. A *Poisson Binomial Distribution* (PBD) with parameters k and $\vec{p} = (p_1, \dots, p_k)$ is the distribution of the sum of k independent Bernoulli random variables $X_1, \dots, X_k$, where $X_i \sim \mathrm{Bern}(p_i)$. (This is a generalization of Binomial distributions, which correspond to $p_1 = \dots = p_k$.) Let $\mathcal{P}^{\fish}_k$ denote the class of all PBDs with parameter k. Using the facts that (1) $\mathcal{P}^{\fish}_k$ can be agnostically learned with $O(\log^2(1/\varepsilon)/\varepsilon^2)$ samples (independent of k) (Daskalakis *et al.*, 2015), and (2) the "standard" Binomial distribution $\mathrm{Bin}(k, 1/2)$ is a PBD, show that testing $\mathcal{P}^{\fish}_k$ has sample complexity $\Omega(k^{1/4}/\varepsilon^2)$ (as long as $\varepsilon \geq 1/2^{O(k^{1/8})}$). *(Hint: combine the results of Sections 3.4 and 3.5.)*

Solution 3.7. For convenience, denote by $\mathbf{q}$ the Binomial $\mathrm{Bin}(k, 1/2)$, and fix $\delta = 1/3$. We will apply Theorem 3.11 with $\mathcal{P}' := \{\mathbf{q}\}$ and $\mathcal{P} := \mathcal{P}^{\fish}_k$, so that $\mathcal{P}' \subseteq \mathcal{P}$.

From the statement of the exercise, from (Daskalakis *et al.*, 2015)[2] we know that the sample complexity of agnostically learning $\mathcal{P}^{\fish}_k$ is

$$n_{\mathcal{L}}(k, \varepsilon, 1/3) = O\left(\frac{\log^2(1/\varepsilon)}{\varepsilon^2}\right)$$

[2] This is not explicit in the paper, but their cover-based approach implies agnostic learning, for a (large, but constant) $C \geq 1$.

so to apply the theorem it remains to give a lower bound on $n_{\mathcal{T}}(k, \varepsilon, 1/3)$ (sample complexity of testing $\mathcal{P}'$) and comparing the two. But testing our $\mathcal{P}'$ is by definition testing identity to $\mathbf{q}$, for which we can apply Theorem 3.9: for all $\varepsilon \in (0, 1/4)$,

$$n_{\mathcal{T}}(k, \varepsilon, 1/3) = \Omega\left(\left\|\tilde{\mathbf{q}}_{-4\varepsilon}^{-\max}\right\|_{2/3} / \varepsilon^2\right),$$

where $\tilde{\mathbf{q}}_{-4\varepsilon}^{-\max}$ is as defined in the statement of Theorem 3.9. In our case, to get a hold of $\left\|\tilde{\mathbf{q}}_{-4\varepsilon}^{-\max}\right\|_{2/3}$, we will make three observations. *(The argument below is heuristic and informal, but can be made rigorous).* First, the maximum probability value of $\mathbf{q}$ is $\frac{1}{2^n}\binom{k}{k/2} = \Theta\left(\frac{1}{\sqrt{k}}\right)$, so removing it will be negligible. Second, to remove the 4ε probability mass from the tails of $\mathbf{q}$, we can use Hoeffding's inequality (Corollary A.4) as heuristic, since even though it provides an upper bound instead of a lower bound, Hoeffding's inequality is essentially tight for the standard Binomial. Working it out, this then tells us that

$$\Pr\left[\left|X - \frac{k}{2}\right| \geq m\right] \approx 4\varepsilon$$

for $m \asymp \sqrt{k \log \frac{1}{\varepsilon}}$, which means that we only keep $2m = \Theta\left(\sqrt{k \log \frac{1}{\varepsilon}}\right)$ coordinates in our vector $\tilde{\mathbf{q}}_{-4\varepsilon}^{-\max}$ (since $\mathbf{q}$ is symmetric around its expectation, those are the $2m$ middle elements of the domain, centered at $k/2$). That is quite difficult to bound, though, so since we are only aiming to *lower bound* $\left\|\tilde{\mathbf{q}}_{-4\varepsilon}^{-\max}\right\|_{2/3}$ we can remove more elements, and only keep the middle $2m'$ for say $m' := \sqrt{k} \leq m$.

Finally, since $\mathbf{q}$ is unimodal, among those $2\sqrt{k}$ elements the minimum probability is $\mathbf{q}(k/2 + \sqrt{k})$, which by Stirling can be bounded as

$$\mathbf{q}(k/2 + m') = \frac{1}{2^k}\binom{k}{\frac{k}{2} + \sqrt{k}} \asymp \frac{1}{\sqrt{k}}.$$

Thus, we can bound

$$\begin{aligned}\left\|\tilde{\mathbf{q}}_{-4\varepsilon}^{-\max}\right\|_{2/3} &\geq \left(2m' \cdot \mathbf{q}(k/2 + m')^{2/3}\right)^{3/2} \\ &\asymp k^{3/4} \cdot \mathbf{q}(k/2 + \sqrt{k}) \\ &\asymp k^{1/4}\end{aligned}$$

and Theorem 3.9 then tells us that $n_{\mathcal{T}}(k, \varepsilon, 1/3) = \Omega\left(k^{1/4}/\varepsilon^2\right)$. *(Intuitively: the standard Binomial distribution is "roughly flat" within a few standard deviations of its expectation, which means it is sort-of-uniform on a domain of size $\Theta(\sqrt{k})$. So we get "the uniformity testing lower bound" but on a domain of size $\Theta(\sqrt{k})$, not k: hence the $k^{1/4}$.*

To conclude and apply Theorem 3.11, we just need to check the third item, and see if there is a range of parameters for which

$$n_{\mathcal{L}}(k, C\varepsilon, 1/3) \leq \frac{1}{2} n_{\mathcal{T}}(k, 3C\varepsilon, 1/3)$$

(where again, $C \geq 1$ is a constant, implicit in (Daskalakis *et al.*, 2015)). This boils down to checking when

$$\frac{\log^2(1/\varepsilon)}{\varepsilon^2} \ll \frac{k^{1/4}}{\varepsilon^2}$$

which is true for $\varepsilon \geq 1/2^{O(k^{1/8})}$. Theorem 3.11 then yields the $\Omega(k^{1/4}/\varepsilon^2)$ lower bound for testing $\mathcal{P}_{k}^{\text{🐟}}$ in that parameter regime.

4 Testing with Constrained Measurements

Exercise 4.1. Verify that the error amplification technique discussed in Lemma 1.1 still goes through in the communication-constrained distributed setting.

Solution 4.1 (Sketch)**.** Regardless of the specific distributed setting (private-coin, public-coin, or sequentially interactive), one can still run $m = O(\log(1/\delta))$ independent instances of the protocol on m disjoint sets of users, and use the same argument as in Lemma 1.1.

Exercise 4.2. Verify that the reduction from identity to uniformity testing discussed in Section 2.2.3 still goes through in the communication-constrained distributed setting, both in the private- and public-coin settings. Do the users need to know the reference distribution $\mathbf{q}$?

Solution 4.2. As described in Acharya *et al.* (2020, Proposition A.16), the reduction does go through, as each user can "locally" and independently apply the mapping $\Psi_{\mathbf{q}}$ to their sample, to get a draw from the distribution $\Phi_{\mathbf{q}}(\mathbf{p})$. This preserves the setting of randomness (private- or public-coin), as it only require private randomness; however, as mentioned in Acharya *et al.* (2020, Remark A.17), all users *do* need to know the reference distribution $\mathbf{q}$ (in order to be able to compute the mapping $\Psi_{\mathbf{q}}$.

However, the two specific approaches seen in Section 4, for private-coin protocols (Section 4.2) and public-coin protocols (Section 4.3) can be used to perform identity testing *even if the users do not know the reference* $\mathbf{q}$. Namely:

- In the private-coin case, distributed simulation (Theorem 4.2) can still be used to simulate $n' \asymp n2^{\ell}/k$ samples from $\mathbf{p}$ at the server. The server (which *does* know $\mathbf{q}$, even if it is the only one to do so) can then convert these n' i.i.d. samples into n' i.i.d. samples from $\Phi_{\mathbf{q}}(\mathbf{p})$, performing the reduction to uniformity testing in a centralized fashion.

- In the public-coin case, domain compression (Theorem 2.12) is used to convert the n i.i.d. samples from $\mathbf{p}$ into n i.i.d. samples from some (randomly chosen) $\mathbf{p}_\Pi$ on a domain of size $L := 2^\ell$. This still works and can be done without knowledge of $\mathbf{q}$; then, the guarantee of Theorem 2.12 ensures that $\mathrm{d}_{\mathrm{TV}}(\mathbf{p}_\Pi, \mathbf{q}_\Pi) \gtrsim \sqrt{L/k} \cdot \mathrm{d}_{\mathrm{TV}}(\mathbf{p}, \mathbf{q})$. This means that the server (which knows $\mathbf{q}$ as well as the public randomness used to choose Π, and thus can compute what the "new reference" $\mathbf{q}_\Pi$ is) can apply the reduction from identity testing (with reference $\mathbf{q}_\Pi$ over $[L]$) to uniformity testing locally, since it has all it needs for this: n i.i.d. samples from $\mathbf{p}_\Pi$, and knowledge of $\mathbf{q}_\Pi$ and $\varepsilon' \asymp \sqrt{L/k} \cdot \varepsilon$.

Exercise 4.3 ($\star$)**.** Extend the argument of Lemma 4.3 to $\ell \geq 1$, to establish the more general Theorem 4.2. *(Hint: suppose that $2^\ell - 1$ divides k, and partition the domain in $m := k/(2^\ell - 1)$ sets. Each pair of users is now "assigned" one of these sets.)*

Solution 4.3. As suggested, partition the domain into $m := \left\lceil k/2^\ell - 1 \right\rceil$ sets (*e.g.*, intervals) $S_1, \dots, S_m$, each of size $k' := 2^\ell - 1$ except at most one possibly smaller (if $2^\ell - 1$ does not divide k). As in the proof of Lemma 4.3, we first show how to generate one sample from $2m$ users.

Divide these $2m$ users into m pairs, where users of the pair $(2i-1, 2i)$ are "assigned" set S_i. The ℓ-bit message these two users send is defined as follows: either the all-zero sequence $\mathbf{0}_\ell$ if their sample did not fall in S_i, or, if it did, the exact value of the sample (they can do so, as $|S_i| \leq 2^\ell - 1$, as long as the protocol specified an encoding beforehand):

$$Y_{2i-1} = \begin{cases} \mathrm{encode}(X_{2i-1}) & \text{if } X_{2i-1} \in S_i \\ \mathbf{0}_\ell & \text{otherwise.} \end{cases}$$

$$Y_{2i} = \begin{cases} \mathrm{encode}(X_{2i}) & \text{if } X_{2i} \in S_i \\ \mathbf{0}_\ell & \text{otherwise.} \end{cases}$$

As in the single-bit ($\ell = 1$) case, the server, upon receiving these $2m$ messages, will check the following two conditions:

- there exists one, and only one, pair $(2i-1, 2i)$ of users for which the "even" user sent a non-zero value (that is, $Y_{2i} \neq \mathbf{0}_\ell$); and

- for this pair $(2i-1, 2i)$, the "odd" user sent the all-zero sequence ($Y_{2i} = \mathbf{0}_\ell$).

If those two conditions do not simultaneously hold, then the server aborts (does not output any sample, but instead the special symbol $\perp$). Otherwise, the server outputs (the decoding of) Y_{2i}, which is X_{2i}, as its sample. Then, analogously to the proof of Lemma 4.3, for any

$j \in [k]$, letting $i(j)$ be the index of the set such that $j \in S_{i(j)}$, the probability to output j is

$$\Pr[\text{output is } j] = \mathbf{p}(j) \cdot (1 - \mathbf{p}(S_{i(j)})) \prod_{\substack{1 \leq i \leq m \\ i \neq i(j)}} (1 - \mathbf{p}(S_i)) = \mathbf{p}(j) \prod_{i=1}^{m} (1 - \mathbf{p}(S_i)) \,.$$

We have $\Pr[\text{output is } j] \propto \mathbf{p}(j)$ for all $j \in [k]$, so all we need to prove, as before, is that $\Pr[\text{output is } \perp]$ is not too close to one, or, equivalently, that $\prod_{i=1}^{m}(1 - \mathbf{p}(S_i))$ is not vanishingly small. The exact same argument as the one leading to Eq. (4.4) shows that

$$\prod_{i=1}^{m}(1 - \mathbf{p}(S_i)) \geq \frac{1}{4}$$

as long as $\max_{1 \leq i \leq m} \mathbf{p}(S_i) \leq \frac{1}{2}$, which is implied by $\|\mathbf{p}\|_\infty \leq 1/2$ and thus can be ensured by using the same trick as in Lemma 4.3 (only losing, as we did there, a factor 2 in the number of users). Altogether, this establishes that we can generate (on expectation) *one* sample from $\mathbf{p}$ using the ℓ-bits messages from

$$4 \cdot 4m = 16 \left\lceil \frac{k}{2^\ell - 1} \right\rceil$$

users. By repeating this on disjoint groups of users, this yields Theorem 4.2, showing that we can generate an expected

$$n' \geq \frac{1}{16} \cdot \frac{n}{\left\lceil \frac{k}{2^\ell - 1} \right\rceil} \asymp \frac{2^\ell n}{k}$$

i.i.d. samples, using the ℓ-bits messages from n users. For a more detailed proof and discussion of Theorem 4.2, see Acharya *et al.* (2020, Theorem IV.9).

Exercise 4.4 (⋆⋆)**.** Extend the argument of Theorem 4.2 further to apply to the case where user has a communication constraint ℓ_i (heterogeneous constraints among users). Establish an analogous bound, with 2^ℓ replaced by $\frac{1}{n}\sum_{j=1}^{n} 2^{\ell_j}$. *(Hint: consider a dyadic partition of the domain* $[k]$*. It* should *work.)*

Solution 4.4. *No solution for this one (for now at least). Feel free to contact me if you've tried and are stuck!*

References

Acharya, J., C. L. Canonne, and H. Tyagi. (2020). “Inference under information constraints II: Communication constraints and shared randomness”. *IEEE Trans. Inform. Theory.* 66(12): 7856–7877. ISSN: 0018-9448. DOI: 10.1109/TIT.2020.3028439. URL: https://doi.org/10.1109/TIT.2020.3028439.

Arnold, B. C. (1987). *Majorization and the Lorenz order: a brief introduction*. Vol. 43. *Lecture Notes in Statistics*. Springer-Verlag, Berlin. vi+122. ISBN: 3-540-96592-0. DOI: 10.1007/978-1-4615-7379-1. URL: https://doi.org/10.1007/978-1-4615-7379-1.

Batu, T. and C. L. Canonne. (2017). “Generalized uniformity testing”. In: *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017*. IEEE Computer Soc., Los Alamitos, CA. 880–889.

Canonne, C. L. (2022). “Topics and Techniques in Distribution Testing: A Biased but Representative Sample”. *Foundations and Trends® in Communications and Information Theory*. 19(6): 1032–1198. ISSN: 1567-2190. DOI: 10.1561/0100000114. URL: http://dx.doi.org/10.1561/0100000114.

Daskalakis, C., I. Diakonikolas, and R. A. Servedio. (2015). “Learning Poisson Binomial Distributions”. *Algorithmica*. 72(1): 316–357.