Sharp Constants in Uniformity Testing via the Huber Statistic

Shivam Gupta   SHIVAMGUPTA@UTEXAS.EDU
The University of Texas at Austin

Eric Price   ECPRICE@CS.UTEXAS.EDU
The University of Texas at Austin

Editors: Po-Ling Loh and Maxim Raginsky

Abstract

Uniformity testing is one of the most well-studied problems in property testing, with many known test statistics, including ones based on counting collisions, singletons, and the empirical TV distance. It is known that the optimal sample complexity to distinguish the uniform distribution on m elements from any ε-far distribution with $1 - \delta$ probability is $n = \Theta(\frac{\sqrt{m \log(1/\delta)}}{\varepsilon^2} + \frac{\log(1/\delta)}{\varepsilon^2})$, which is achieved by the empirical TV tester. Yet in simulation, these theoretical analyses are misleading: in many cases, they do not correctly rank order the performance of existing testers, even in an asymptotic regime of all parameters tending to 0 or ∞.

We explain this discrepancy by studying the *constant factors* required by the algorithms. We show that the collisions tester achieves a sharp maximal constant in the number of standard deviations of separation between uniform and non-uniform inputs. We then introduce a new tester based on the Huber loss, and show that it not only matches this separation, but also has tails corresponding to a Gaussian with this separation. This leads to a sample complexity of $(1 + o(1))\frac{\sqrt{m \log(1/\delta)}}{\varepsilon^2}$ in the regime where this term is dominant, unlike all other existing testers.

Keywords: Property testing, Sublinear algorithms

1. Introduction

Property testing of distributions is an area of study initiated in (Goldreich and Ron, 2011) and (Batu et al., 2000). The foundation of these works is a test for *uniformity*: given n samples from a distribution q on $[m]$, can we distinguish the case that q is uniform from the case that q is ε-far from uniform, with probability $1 - \delta$? The remarkable result is that this is often possible for $n \ll m$, when we cannot learn the actual distribution. Over the years, several different tests and bounds have been established for uniformity. In this paper we better understand and explain the relative performance of these testers, then introduce a new uniformity tester that outperforms all of them.

The first uniformity tester introduced was the *collisions tester* (Goldreich and Ron, 2011; Batu et al., 2000), which counts the number of collisions among the samples. It is equivalent to Pearson's χ^2 test, or any other statistic quadratic in the histogram. It succeeds with constant probability for $n = O(\sqrt{m}/\varepsilon^2)$ (Diakonikolas et al., 2019), which is optimal (Paninski, 2008).

What happens for high-probability bounds? Naive repetition gives a multiplicative $O(\log \frac{1}{\delta})$ loss, but this can be improved: Huang and Meyn (Huang and Meyn, 2013) showed that the *singletons tester* (Paninski, 2008) achieves $\sqrt{m \log \frac{1}{\delta}}/\varepsilon^2$, but only in the setting of $n = o(m)$ and $\varepsilon = \Omega(1)$. The collisions tester, however, really does involve a $\log \frac{1}{\delta}$ loss (Peebles, 2015).
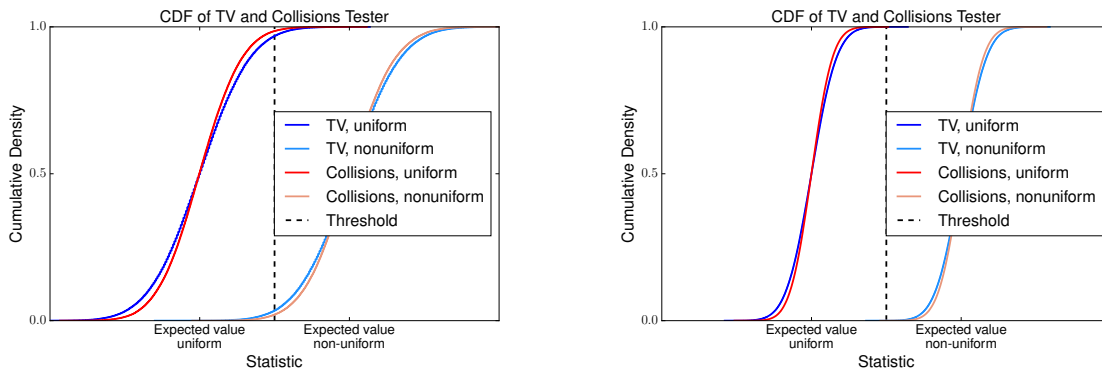
(a) Simulation of statistic for $m = n = 10000$ and $\varepsilon = 0.125$. The collisions tester has 1.7% error rate and the TV tester has 3.3%.

(b) When $m = n = 10^5$ and $\varepsilon = 0.1$, the collisions tester has 10^{-5} error rate and the TV tester has 10^{-4} error rate. So collisions can outperform TV even for tiny δ.

Figure 1: Observed performance of TV vs collisions distinguishing uniform from $\frac{1 \pm 2\varepsilon}{m}$

Achieving optimal dependence of the whole range is the empirical *TV tester* (Diakonikolas et al., 2018), which measures the TV distance between the empirical distribution and uniform. It needs

$$n = O\left(\frac{\sqrt{m \log \frac{1}{\delta}}}{\varepsilon^2} + \frac{\log \frac{1}{\delta}}{\varepsilon^2} \right) \tag{1}$$

which is optimal in all settings of parameters.

To summarize 20 years of theory, the TV tester is asymptotically optimal while the collisions tester has poor δ dependence and the singletons tester is good for $n \ll m$ but fails when $n \gg m$. This suggests that, given an actual example of a uniformity testing problem, the TV tester is as good as possible.

In Figure 4 we compare the TV tester to the collisions tester in simulation. We test the uniform distribution against the distribution that puts $\frac{1+2\varepsilon}{m}$ mass on half the bins, and $\frac{1-2\varepsilon}{m}$ mass on the remaining bins. This is the worst case ε-far distribution for both these testers (Diakonikolas et al., 2018). We find that, contrary to the theoretical prediction, the collisions tester outperforms the TV tester on the parameters we consider. In our first experiment, with $m = n = 10^4$ and $\varepsilon = 1/8$, the TV tester has twice the error rate as the collisions tester (3.3% vs 1.7%). In our second experiment, with $m = n = 10^5$ and $\varepsilon = 1/10$, the gap *widens* to a factor 10 (10^{-4} vs 10^{-5}) despite the error rate δ—the parameter the collisions tester is suboptimal in—becoming much smaller. This means that our theory is giving the *wrong advice*: a practitioner should prefer the collisions tester to the TV tester here.

To better explain this, and to develop a new tester that outperforms all existing ones, we need to start considering constant factors.

Designing a new tester. How should we design an efficient tester for uniformity? We consider "separable" testers that take as input the histogram Y_j (so Y_j is the number of samples equal to j),

Tester	Optimal variance?	Subgaussian tails?
Collisions/χ^2	**Yes** (Theorem 1)	No (Theorem 52)
TV	No (Theorem 5)	**Yes**
Huber (new)	**Yes**	**Yes** (Theorem 2)

Figure 2: Our main contributions are (1) that the collisions statistic achieves optimal variance, and (2) that the Huber statistic can get high-probability bounds matching this variance.

compute a statistic

$$S = \sum_{j=1}^{m} f(Y_j),$$

and output YES or NO based on whether S lies below some threshold τ. Existing testers are all either of this form, or use this as the main subroutine (e.g., after Poissonization or taking the median of multiple attempts). Differences lie in the choice of f. Quadratic functions $f(k) = (k - n/m)^2$ or $f(k) = \binom{k}{2}$ give the χ^2 or collisions tester (Goldreich and Ron, 2011; Batu et al., 2000), which are equivalent because $\sum_j Y_j = n$ is fixed, so the two statistics S are linearly related. The TV tester (Diakonikolas et al., 2018) uses $f(k) = |k - n/m|$, while the singletons tester (Paninski, 2008) uses $f(k) = 1_{k=1}$. But how would one design f from first principles to work well?

In this paper we introduce a natural approach to designing a test statistic with good asymptotic constants. First, we find the test statistic f that maximizes the number of standard deviations of separation between YES and NO instances; then, we modify the tails of f so that S has Gaussian tails. This approach is summarized in Figure 2.

Step 1: Optimize variance. Intuitively, S is a sum of m terms $f(Y_j)$ that are *nearly* independent, so we expect central limit-type behavior

$$S \overset{\approx}{\sim} N(\mathbb{E}[S], \text{Var}[S]).$$

That is, we expect our separable statistic to behave like a Gaussian, with expectation and variance that depend on the particular statistic. Because the hard alternative distributions q are very close to p, typically $\text{Var}_q[S] = (1 + o(1))\text{Var}_p[S]$. Then our ability to distinguish p and q depends on how this variance compares to the separation in means: we want to minimize this *normalized variance*

$$\widetilde{\text{var}}_{p,q}(S) := \frac{\text{Var}_p[S]}{(\mathbb{E}_q[S] - \mathbb{E}_p[S])^2}.$$

We can set our threshold to lie halfway between $\mathbb{E}_p[S]$ and $\mathbb{E}_q[S]$, so that, under the Gaussian approximation, the error probability will be given by

$$\delta \approx \exp\left(-\frac{((\mathbb{E}_q[S] - \mathbb{E}_p[S])/2)^2}{2\text{Var}_p[S]}\right) = \exp\left(-\frac{1}{8\,\widetilde{\text{var}}_{p,q}(S)}\right) \tag{2}$$

For any q, $\min_f \widetilde{\text{var}}_{p,q}(S_f)$ is a quadratic program in f, so we can compute the variance-minimizing f for any setting of parameters. We can also approximate it analytically in the asymptotic limit. We find that the quadratic statistics (like collisions or χ^2) are near-optimal:

Theorem 1 *Let $\varepsilon^2 \ll \frac{n}{m} \lesssim 1$ and $n, m, 1/\varepsilon \to \infty$ with $m \gtrsim 1/\varepsilon^4$. Any separable statistic S has normalized variance*

$$\widetilde{\mathrm{var}}_{p,q}(S) \geq (1 + o(1))\frac{1}{8}\frac{m}{n^2\varepsilon^4}$$

between the uniform distribution p and the balanced nonuniform distribution q with $q_k = \frac{1 \pm 2\varepsilon}{m}$.

Quadratic statistics (like collisions or χ^2) match this, getting

$$\widetilde{\mathrm{var}}_{p,q}(S) \leq (1 + o(1))\frac{1}{8}\frac{m}{n^2\varepsilon^4} \tag{3}$$

for any *ε-far distribution q.*

Theorem 1 shows that, if the Gaussian approximations hold, the collisions tester has optimal constants. Per (2), for failure probability δ we need $\widetilde{\mathrm{var}}_{p,q}(S) = \frac{1}{8\log\frac{1}{\delta}}$, or

$$n = (1 + o(1))\frac{1}{\varepsilon^2}\sqrt{m\log\frac{1}{\delta}} \tag{4}$$

samples. This matches the optimal complexity (1) in the large-m regime, but with a sharp constant of 1. (Sharp in the sense that no other separable statistic behaves better under its Gaussian approximation.) One could also trade off the false positive and false negative errors by choosing a different threshold between the means, getting

$$n = (1 + o(1))\frac{1}{\varepsilon^2}\sqrt{m} \cdot \frac{\sqrt{\log\frac{1}{\delta_+}} + \sqrt{\log\frac{1}{\delta_-}}}{2}. \tag{5}$$

for false positive/negative probabilities δ_+/δ_-.

By contrast, the TV tester has a constant factor worse normalized variance than the quadratic tester (we shall state this constant precisely later). Therefore the Gaussian approximation loses a constant factor relative to (4), and it would be very surprising if the actual statistic avoided this inefficiency.[1] So the Gaussian approximations predict the actual Figure 4 behavior.

Unfortunately, the Gaussian approximation does *not* hold in general for the collisions statistic, so it does not get (4) or (5). See Appendix E.1 for a detailed example, due to (Peebles, 2015), showing that for exponentially small δ the collisions tester does not achieve (1) for any constant much less $1 + o(1)$.

Step 2: Massage the tails. The Gaussian tail bound implying (4) and (5) is given by its moment generating function, so it would suffice to bound the MGF of S. The problem is that Y_j has roughly exponential tails for every j, so the MGF of Y_j^2 does not have a good bound. To get a good MGF for $f(Y_j)$, we need to look at an f with at most linear growth.

So this is our situation: the quadratic $f(Y_j) = Y_j^2$ has near-optimal variance but a very large MGF, while the TV statistic $f(Y_j) = |Y_j - n/m|$ has suboptimal variance but a pretty good MGF. Introducing the linear tail with $f(Y_j) = |Y_j - n/m|$ is how (Diakonikolas et al., 2018) achieved the $\sqrt{\log\frac{1}{\delta}}$ dependence, but the worse variance means it inherently performs worse than the quadratic for large δ where the Gaussian approximation holds.

---

1. Unsurprisingly, as we show in Theorem 5, the inefficiency is real.
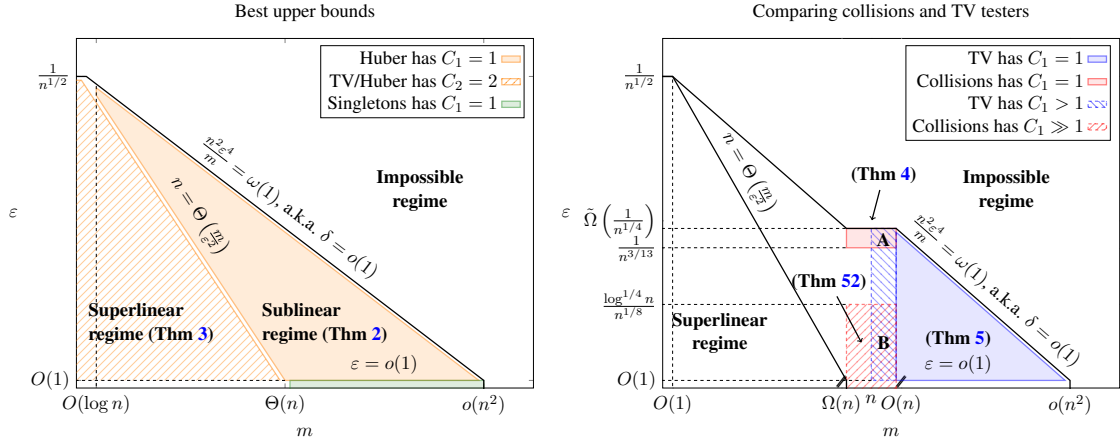
(a) The Huber statistic achieves the best constant over almost the entire region of $\varepsilon \ll 1$. The singletons statistic was previously known to achieve this for $\varepsilon = \Omega(1)$ and $m \gg n$ (Huang and Meyn, 2013).

(b) The constant C_1 in different (n, ε, δ) parameter regimes. In region **A**, the collisions tester performs better than the TV tester. In region **B**, both the collisions and TV tester have $C_1 > 1$.

Figure 3: Our results in different regimes.

To achieve both good variance and MGF, we should *start* with the good-variance quadratic statistic, then attenuate the tail behavior to get good concentration. We do this with the *Huber* loss

$$f(Y_j) = h_\beta(Y_j - n/m)$$

for

$$h_\beta(x) := \min(x^2, 2\beta|x| - \beta^2).$$

If we choose a tradeoff point $\beta \gg 1 + \sqrt{n/m}$, most Y_j will lie in the quadratic region and we still get the variance bound (3). But now the MGF is bounded. We show that, for a large range of parameters, this leads to the tester that matches a Gaussian with the optimal variance:

Theorem 2 (Huber) *The Huber statistic for appropriate β achieves (4) for $n/m \ll 1/\varepsilon^2$, $\varepsilon, \delta \ll 1$, and $m \geq C \log n$ for sufficiently large constant C. It achieves (5) under the same conditions and $\delta_-, \delta_+ \ll 1$.*

Combined with Theorem 1, Theorem 2 shows that the Huber statistic gets the optimal variance over separable statistics and matches the Gaussian concentration with this variance.

The parameter regime is illustrated in Figure 3(a). The first three asymptotic conditions for Theorem 2 delineate the boundaries of the "sublinear" regime, where testing is possible, nontrivial, and the asymptotic sample complexity (1) is dominated by the $\frac{1}{\varepsilon^2}\sqrt{m \log \frac{1}{\delta}}$ term. The last condition, that $m \geq C \log n$, is likely an artifact of our analysis but is pretty mild.

The rest of our results look at other regimes and other testers, and we summarize our results in Figure 3. If we express the sample complexity (1) as

$$n = (C_1 + o(1))\frac{\sqrt{m \log \frac{1}{\delta}}}{\varepsilon^2} + (C_2 + o(1))\frac{\log \frac{1}{\delta}}{\varepsilon^2} \tag{6}$$

5

then we can express the constants C_1, C_2 in different regimes of (n, m, ε). In the "sublinear regime" of $n/m \ll \frac{1}{\varepsilon^2}$, where we cannot reliably estimate the distribution, what matters is C_1. For $n = m$, when $\log n \ll \log \frac{1}{\delta} \ll n^{1/13}$ so that $\tilde{\Omega}(n^{-1/4}) \ll \varepsilon \ll n^{-3/13}$, that is, in regime A of small ε/large δ, collisions gets $C_1 = 1$ and TV has $C_1 > 1$; in regime B of large ε/small δ, collisions gets $C_1 \gg 1$ and TV has $1 < C_1 = O(1)$. The Huber statistic, by contrast, gets $C_1 = 1$ for almost the whole regime. In the superlinear regime, where the empirical distribution is $\varepsilon/2$-accurate, a simple union bound shows that the TV statistic (and hence Huber statistic for $\beta = 0$) gets the optimal $C_2 = 2$:

Theorem 3 (Superlinear regime) *For $n/m \gg 1/\varepsilon^2$ and $\varepsilon \ll 1$, the TV statistic achieves*

$$n = (2 + o(1)) \frac{\log \frac{1}{\delta}}{\varepsilon^2}$$

and no other tester can do better.

Analysis of collisions. While the χ^2/collisions tester does not match the Gaussian tails to achieve (4) everywhere, it still is a sum of mostly-independent variables and so looks like a Gaussian outside the extreme tails. Hence the Gaussian approximation (4) ought to hold when δ isn't too small. Indeed, we show this is true for $n = \Theta(m)$ and intermediate δ:

Theorem 4 (Collisions for large δ) *The quadratic statistic achieves* (4) *for $n/m = \Theta(1)$, $\log n \ll \log \frac{1}{\delta} \ll n^{1/13}$ and $\varepsilon \ll 1$.*

Analysis of TV. The TV tester, for $n \leq m$, is equivalent to the tester that counts empty bins. We show that this has

$$\max_{q: \|p-q\|_{TV} \geq \varepsilon} \widetilde{\mathrm{var}}_{p,q}(f) = (1 + o(1)) \frac{(e^{n/m} - 1 - n/m)}{4(n/m)^2} \frac{m}{\varepsilon^4 n^2}.$$

rather than (3). For $n \ll m$ these are equivalent, but for $n = m$ it is 44% larger, leading to about 20% more samples.

Theorem 5 (TV) *The TV statistic uses*

$$n = (1 + o(1)) \sqrt{\frac{2(e^{n/m} - 1 - n/m)}{(n/m)^2}} \frac{\sqrt{m \log \frac{1}{\delta}}}{\varepsilon^2}$$

for $n \leq m$, $n \gg 1$, and $\varepsilon, \delta \ll 1$.

Like Theorem 2, both Theorem 4 and Theorem 5 work by showing the Gaussian approximation is accurate. Thus one could also trade off false positive/negative probabilities, with a $\frac{1}{2}(\sqrt{\log \frac{1}{\delta_-}} + \sqrt{\log \frac{1}{\delta_+}})$ dependence.

Experimental performance. In Figure 4, we compare the empirical performance of the new Huber tester to the existing collisions and TV testers in a synthetic experiment. The experiment has $m = n$, $\varepsilon = .7/n^{1/8.1}$ with alternative distribution $q = \frac{1 \pm 2\varepsilon}{n}$, and varies n from 200 to 600. This is in region **B** of Figure 3(b), and as predicted we find that the Huber tester has lower failure probability than the TV or collisions testers.
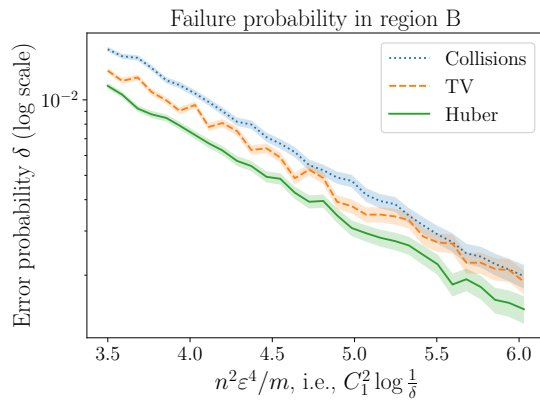
Figure 4: Empirical failure probability of different testers when $n = m = (.7/\varepsilon)^{8.1}$, which is in region **B** of Figure 3(b). The x axis is $n^2\varepsilon^4/m$, which should be linear in $\log\frac{1}{\delta}$ per (6). The shaded region shows two standard deviations of uncertainty.

1.1. Related work

The past twenty years have seen a large body of work in distribution testing; see (Goldreich, 2017; Canonne, 2020) for surveys of the area. Uniformity testing has been either the basis for, or a necessary subproblem in, many such results. Such extensions include testing identity (Batu et al., 2001; Chan et al., 2014; Goldreich, 2017; Diakonikolas and Kane, 2016; Valiant and Valiant, 2017; Diakonikolas et al., 2020), testing independence (Canonne et al., 2018), and testing uniformity over unknown domains (Batu and Canonne, 2017; Diakonikolas et al., 2017). One particularly clean relation is that you can black-box reduce testing identity to a fixed distribution p to uniformity testing with only a constant factor loss in parameters (Goldreich, 2017).

Most of the above results do not focus on the dependence on δ; exceptions include (Diakonikolas et al., 2018; Kim et al., 2020; Diakonikolas et al., 2020; Huang and Meyn, 2013) which give algorithms within constant factors of optimal for testing uniformity, identity, and independence.

Lower bounds for uniformity testing started with an $\Omega(\sqrt{m})$ bound in (Goldreich and Ron, 2011), followed by $\Omega(\sqrt{m}/\varepsilon^2)$ in (Paninski, 2008) and $\Omega(\frac{1}{\varepsilon^2}\sqrt{m\log 1/\delta} + \frac{\log 1/\delta}{\varepsilon^2})$ in (Diakonikolas et al., 2018).

When it comes to constant factors in distribution testing, the classical regime of ε, m constant and $n \to \infty$ was analyzed in (Hoeffding, 1965) and the likelihood ratio test was shown to be optimal. Alternatively, for m, δ constant and $n, 1/\varepsilon \to \infty$, Pearson's χ^2 tester—the quadratic tester—is known to be asymptotically near optimal for identity testing (see (Lehmann et al., 2005), Chapter 14).

The most closely related work to our paper is Huang and Meyn (Huang and Meyn, 2013), which (unlike the classical results) studies constant factors in a regime where all of $n, m, 1/\delta \to \infty$. They consider the singletons tester, and show that $C_1 = 1$ for constant ε and $n \ll m$. They also show that no algorithm can do better in this regime. However, for $n = \Theta(m)$ the singletons tester loses constant factors and for $n \geq O(m \log m)$ it fails with high probability.

1.2. Future work

As discussed above, uniformity testing has been the building block for many other distribution testing problems, such as identity and independence testing. Where there are direct reductions (as in testing identity to a fixed distribution (Goldreich, 2017)), these reductions lose constant factors. However, these tests still involve statistics that are the sum of mostly independent random variables. We believe that our approach to constructing a test statistic—find a statistic to optimize performance of the Gaussian approximation, then adjust it to match the Gaussian tails—could lead to higher performance testers in these problems as well.

Second, there are some settings of parameters that we have not analyzed. Most interesting would be to analyze the intermediate regime of $\frac{n}{m} = \Theta(\frac{1}{\varepsilon^2})$, where both sample complexity terms in (1) are significant.

Third, we could consider constant factors for high probability bounds in other settings. For example, it is known by the Cramér-Rao bound (Cramer, 1946) that the maximum likelihood estimator (MLE) in parametric statistics converges to a Gaussian with variance equal to the inverse of the Fisher information under a broad set of assumptions; but the tails of this estimator are less well understood, and could likely be improved by modifying the estimator to be less sensitive to outliers. Other examples lie in streaming algorithms. There has been a line of work on understanding the constants in the space complexity of cardinality estimation in streams (Flajolet et al., 2007; Ertl, 2017; Lang, 2017; Pettie and Wang, 2021), but these have focused on the constant δ regime. We believe our techniques could lead to optimal high probability bounds on the space complexity for this problem. Alternatively, for problems like heavy hitters (Charikar et al., 2002; Minton and Price, 2014; Braverman et al., 2016), the analysis has focused on the high δ regime and ignored constant factors; but the underlying algorithms involve sums of random variables that ought to converge to Gaussians.

2. Proof Overview

2.1. Variance Optimality

To show Theorem 1, we write the optimization problem

$$m^2 \, \widetilde{\mathrm{var}}_{p,q}(S) = \min_{} \mathrm{Var}_p[S_f]$$
$$\text{s.t.} \, \mathbb{E}_q[S_f] - \mathbb{E}_p[S_f] = m$$

as a quadratic program in the vector $f = (f_0, \ldots, f_n)$. For $\overline{p}_k = \mathbb{P}_p[Y_1 = k]$ and $\overline{q}_k = \mathbb{E}_{i \in [m]} \mathbb{P}_q[Y_i = k]$, the constraint is that $(\overline{q} - \overline{p}) \cdot f = 1$, and the objective is $f^T Q f$ for some matrix Q. The KKT condition (Karush, 1939; Kuhn and Tucker, 1951; Boyd and Vandenberghe, 2004) shows that the optimum is achieved when $Qf = a(\overline{q} - \overline{p})$ for some scalar a.

Solving this exactly requires the pseudoinverse Q^+, which would be tricky. Instead, we show that the quadratic statistic $f_k = k^2$ satisfies a slightly different condition

$$Qf = a(q' - \overline{p}),$$

for a different distribution $q' \in \mathbb{R}^{n+1}$ we can write explicitly. Therefore the quadratic statistic minimizes the variance subject to an expectation gap in q' relative to $\overline{p}$. Moreover, this q' turns out

to be precisely the Taylor approximation in ε to $\bar{q}$, with order ε^4 error. All that remains is to show that this $O(\varepsilon^4)$ distinction between $\bar{q}$ and q' gives $1 + o(1)$ loss in the program. That is,

$$|\underset{q'}{\mathbb{E}} f_k - \underset{\bar{p}}{\mathbb{E}} f_k| = (1 + o(1))|\underset{\bar{q}}{\mathbb{E}} f_k - \underset{\bar{p}}{\mathbb{E}} f_k|,$$

or equivalently

$$|\underset{q'}{\mathbb{E}} f_k - \underset{\bar{q}}{\mathbb{E}} f_k| \ll \frac{1}{m}|\underset{q}{\mathbb{E}}[S_f] - \underset{p}{\mathbb{E}}[S_f]|, \tag{7}$$

for any statistic that we care about. We can bound this LHS in terms of the variance of f:

$$|\underset{q'}{\mathbb{E}} f_k - \underset{\bar{q}}{\mathbb{E}} f_k| \lesssim \varepsilon^4 \sqrt{\underset{\bar{p}}{\mathbb{E}}[f_k^2]}.$$

Then we can relate the variance of f to the variance of S:

$$\underset{\bar{p}}{\mathbb{E}}[f_k^2] \lesssim \frac{1}{m}\mathrm{Var}_p[S_f]$$

using the fact that our statistic is indifferent to constant and linear terms, so we can assume WLOG $\mathbb{E}[f_k] = \mathbb{E}[kf_k] = 0$.

Combining these results, we get that (7) holds whenever

$$\widetilde{\mathrm{var}}_{p,q}(S) = \frac{\mathrm{Var}_p[S_f]}{(\mathbb{E}_q[S_f] - \mathbb{E}_p[S_f])^2} \ll \frac{1}{\varepsilon^8 m}.$$

Since the quadratic has $\widetilde{\mathrm{var}}_{p,q}(S) = \Theta(\frac{m}{\varepsilon^4 n^2})$, this holds for both the quadratic and the statistic of maximal separation $\widetilde{\mathrm{var}}_{p,q}(S)$. Therefore this maximum is within $1 + o(1)$ of the quadratic.

2.2. Concentration of Tails

Setting. In this proof overview we will focus on the Huber statistic in the regime where $1 \lesssim \frac{n}{m} \ll \frac{1}{\varepsilon^2}$, as well as $\varepsilon, \delta \ll 1$ (so $\frac{n^2}{m}\varepsilon^4 \gg 1$).

Let $X_1, \ldots, X_n$ be the n samples drawn from distribution ν supported on $[m]$, and let $Y_j^n = \sum_{i=1}^n \mathbb{1}_{\{X_i = j\}}$ be the number of balls that end up in bin j.

The Huber statistic. We consider the Huber statistic

$$S = \sum_{j=1}^m h_\beta\left(Y_j^n - \frac{n}{m}\right) \tag{8}$$

where

$$h_\beta(x) := \begin{cases} x^2 & \text{for } |x| < \beta \\ 2\beta|x| - \beta^2 & \text{otherwise} \end{cases} \tag{9}$$

is the Huber loss function, which continuously interpolates between a quadratic center and linear tails. Note that this is twice the standard definition, but the statistic's performance is invariant under affine transformations.

We will set β large enough that most bins usually lie in the quadratic regime in the uniform case. If we were to set $\beta = \infty$ (so S is an affine transformation of the collisions statistic), we would have $\mathbb{E}_p[S] = n - n/m \approx n$ for the uniform distribution p and $\mathbb{E}_q[S] \geq n - n/m + 4n(n-1)\varepsilon^2/m \approx n + 4n^2\varepsilon^2/m$ for any ε-far distribution q. This motivates us to consider the rescaled statistic:

$$\tilde{S} = \frac{m}{n^2\varepsilon^2}[S - n] \tag{10}$$

which (for $\beta = \infty$) has $\mathbb{E}[\tilde{S}]$ being $o(1)$ or $\geq 4 - o(1)$ in the uniform and far-from-uniform cases, respectively.

Because $Y_j^n \sim B(n, 1/m)$ in the uniform case, Bernstein's inequality shows that setting

$$\beta = \omega\left(\log\left(\frac{1}{\Delta}\right) + \sqrt{\frac{n}{m}\log\left(\frac{1}{\Delta}\right)}\right) \tag{11}$$

gives that each bin lies in the quadratic regime with probability $1 - \Delta^2$, for a parameter $\Delta \ll 1$ that we will constrain later. Choosing this β leads to smaller $\mathbb{E}[S]$ than $\beta = \infty$, but the difference is only about $\beta^2\Delta^2 m$ because each of the m bins has a Δ^2 chance of lying in the linear region, and most of the differences happen at the boundary where the Huber statistic is $\Theta(\beta^2)$. This error is $O(n\Delta^2\log^2\frac{1}{\Delta}) < O(n\Delta^{1.5})$, so

$$\mathbb{E}_p[\tilde{S}] = o(1) + O(\frac{m}{n^2\varepsilon^2}\Delta^{1.5}n) = o(1)$$

as long as we have

$$\Delta = O\left(\frac{n\varepsilon^2}{m}\right) \tag{12}$$

which is $o(1)$. Similarly, this implies

$$\mathbb{E}_q[\tilde{S}] \geq 4 - o(1)$$

for any ε-far distribution q.

Finally, we will need some constraint that β is not too large/Δ too small. A third moment condition suffices, as we shall see in a few pages:

$$(\beta^2\varepsilon^2)^3 = o\left(\Delta^2\right) \tag{13}$$

One can check that β and Δ can be chosen such that the constraints (11), (12), and (13) hold in the regime we consider here.

Analyzing the Huber statistic. Our tester will pick a threshold τ, and "accept" the distribution as uniform if $\tilde{S} \leq \tau$. We therefore need to understand the false negative probability

$$\delta_- := \mathbb{P}_p[\tilde{S} \geq \tau]$$

and similarly, for any ε-far distribution q, we need to bound the false positive probability

$$\delta_+ := \mathbb{P}_q[\tilde{S} \leq \tau].$$

To bound the maximum error $\delta = \max(\delta_-, \delta_+)$, it suffices to pick $\tau = 2$, halfway between the expectation bounds in the uniform and ε-far cases.

Completeness. We start by describing how to analyze δ_-. The bulk of our analysis here is devoted to analyzing the moment generating function $M_{\tilde{S},\nu}(t) := \mathbb{E}_\nu[\exp(t\tilde{S})]$.

A careful analysis (see, e.g., (Diakonikolas et al., 2019) Lemma 3) shows that when ν is the uniform distribution, the number of collisions has variance $(1+o(1))\frac{n^2}{2m}$. For large enough β per (11), this implies

$$\mathrm{Var}[\tilde{S}] = (1+o(1))\frac{2m}{n^2\varepsilon^4}.$$

Therefore we hope that $\tilde{S}$ has MGF close to a Gaussian with this variance. In Lemma 34 we show that this is in fact the case: for the uniform distribution p,

$$M_{\tilde{S},p}\left(\frac{n^2\varepsilon^4}{m}\theta\right) = (1 + O(1/n))\exp\left\{\frac{n^2}{m}\varepsilon^4\left(\theta^2 + o\left(1\right)\right)\right\} \tag{14}$$

Here, we pulled out $\frac{n^2\varepsilon^4}{m}$ from the MGF parameter, so that we will set θ to be constant at the end. Once we have this, then standard Chernoff-type arguments imply

$$\delta_- < \inf_{\theta \geq 0} \frac{M_{\tilde{S},p}\left(\frac{n^2\varepsilon^4}{m}\theta\right)}{e^{\frac{n^2\varepsilon^4}{m}\theta(E_p[\tilde{S}]+\tau)}} < \inf_{\theta \geq 0}(1+O(1/n))\exp\left\{\frac{n^2}{m}\varepsilon^4\left[\theta^2 - \tau\theta + o\left(1\right)\right]\right\}$$

and hence

$$\delta_- \leq (1+O(1/n))\exp(-J_-(1+o(1))\frac{n^2\varepsilon^4}{m})$$

for "error exponent"

$$J_- := \sup_{\theta \geq 0}\left\{-\frac{m}{n^2\varepsilon^4}\log M_{\tilde{S},p}\left(\frac{n^2\varepsilon^4}{m}\theta\right) + \theta\tau\right\} \geq \sup_{\theta \geq 0}\{\theta\tau - \theta^2\} = \frac{\tau^2}{4} \tag{15}$$

The above is an upper bound on δ_-, but we can also get a lower bound. Because the MGF bound (14) is tightly that of a Gaussian, with both upper and lower bounds, we can apply the Gärtner-Ellis theorem (see Appendix B.1) to show that the tail bound is tightly that of a Gaussian as well: $\delta_- \gtrsim \exp(-(1+o(1))Jn^2\varepsilon^4/m)$.

Soundness. Because the Huber statistic S is convex, we can apply existing tools from (Diakonikolas et al., 2018) to analyze the statistic for uniformity testing. In particular, it is sufficient to consider alternate distributions of the form q such that

$$q_j = \begin{cases} 1/m + \frac{\varepsilon}{l}, & j \leq l \\ 1/m - \frac{\varepsilon}{m-l}, & j > l \end{cases} \tag{16}$$

for some $l \in [m]$. Our discussion of this appears in Appendix B.2. For simplicity of this exposition, suppose m is even and $l = m/2$. Using a similar procedure as in the case of the uniform distribution, we show in Lemma 34 that for this alternate distribution q,

$$M_{\tilde{S},q}\left(\frac{n^2\varepsilon^4}{m}\theta\right) = (1+O(1/n))\exp\left\{\frac{n^2\varepsilon^4}{m}\left[\theta^2 + 4\theta + o(1)\right]\right\}. \tag{17}$$

That is to say, except for a mean shift of $4 + o(1)$, $\tilde{S}$ under q concentrates as a Gaussian with the same variance as it did under p. This gives us that

$$\delta_+ \leq (1 + O(1/n)) \exp\left(-J_+(1 + o(1))\frac{n^2\varepsilon^4}{m}\right)$$

for "error exponent"

$$J_+ := \sup_{\theta \geq 0}\left\{-\frac{m}{n^2\varepsilon^4}\log M_{\tilde{S},q}\left(-\frac{n^2\varepsilon^4}{m}\theta\right) - \theta\tau\right\} \geq \sup_{\theta \geq 0}\{-\tau\theta - \theta^2 + 4\theta\} = \frac{(\tau - 4)^2}{4}$$

Setting $\tau = 2$ so that $J_- = J_+ = 1$ gives us that the error exponent achieved by the Huber tester is 1 for the uniformity testing problem in this regime.

Alternatively, we could pick a different $\tau \in (0, 4)$ to trade off δ_- and δ_+, always getting within $(1 + o(1))$ of the tradeoff given by the Gaussian approximation to S.

Analyzing the MGF.   The key question, therefore, is how to analyze the MGF. For this, we follow the structure of Huang and Meyn (2013), though with different approximations because of our different regime.

We would like to analyze the MGF M_{S_n} of our test statistic

$$S_n = h_\beta(|Y_j^n - n/m|).$$

If the Y_j^n were independent over j, this would be easy: we would simply bound the MGF of each individual term, and take the product. For the same reason, it is easy to bound the MGF $A_\lambda(\theta)$ of the *poissonized* test statistic $S_{\mathrm{Poi}(\lambda)}$, where $\mathrm{Poi}(\lambda)$ balls are drawn rather than n. We can get a Taylor approximation to A_λ that is quite accurate in our regime.

Unfortunately, we cannot just use the Poissonized MGF A_n in place of the true MGF M_{S_n}. The problem is that Poissonization inherently increases the variance: the variance of the collisions statistic is $(1 + o(1))\frac{n^2}{2m}$ before Poissonization but $(1 + o(1))(\frac{n^2}{2m} + \frac{n^3}{m^2})$ after Poissonization. For $n = \Theta(m)$ this is a constant factor we cannot afford to lose, and for $n \gg m$ it's even worse. So we need to "depoissonize" A_λ into M_{S_n}.

To depoissonize, we observe that the Poissonized MGF A_λ is a mixture of the non-Poissonized MGFs M_{S_k} for $k \geq 0$, and in fact M_{S_n} is just (up to scaling) the λ^n coefficient in the Taylor expansion of A_λ. We then use Cauchy's theorem to evaluate this coefficient.

Comparison to Huang-Meyn   Our proof structure is similar to (Huang and Meyn, 2013). Differences arise from two causes: first, (Huang and Meyn, 2013) consider the simpler singletons tester $f(k) = 1_{k=1}$, so the MGF of $f(Y_i)$ can be written in closed form. For the Huber statistic, we need to bound the terms corresponding to the higher moments of the statistic, which is done in Lemma 23. Second, they use the asymptotic regime $n/m \ll 1$ rather than $\varepsilon \ll 1$ for their Taylor series expansions to drop $o(1)$ terms, leading to a number of differences.

Finally, our proof for the alternate distributions is much simpler than the proof in (Huang and Meyn, 2013) since we make use of results from (Diakonikolas et al., 2018).

2.3. Organization of the Appendix

Appendix A shows Theorem 1, that quadratic statistics have asymptotically optimal variance. The next sections show Theorem 2, that the Huber statistic combines this variance with good concentration, in the main new regime of $1 \lesssim n/m \ll 1/\varepsilon^2$: some background is given in Appendix B, the main argument in Appendix C, and some technical computations are deferred to Appendix D.

The rest of the appendix includes our analyses of other testers and other regimes. Proof of the asymptotically poor performance of the collisions and singletons testers in some regimes is in Appendix E. The "superlinear" regime of $n/m \gg 1/\varepsilon^2$ is covered in Appendix F. Analysis of the collisions/quadratic statistic is in Appendix G, while the TV/empty bins statistic for $n < m$ is in Appendix H.

Acknowledgments

References

R.B. Ash. *Information Theory*. Dover books on advanced mathematics. Dover Publications, 1990. ISBN 9780486665214. URL https://books.google.com/books?id=yZlJZA6Wo6YC.

Tugkan Batu and Clément L Canonne. Generalized uniformity testing. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 880–889. IEEE, 2017.

Tugkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D Smith, and Patrick White. Testing that distributions are close. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 259–269. IEEE, 2000.

Tugkan Batu, Eldar Fischer, Lance Fortnow, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. Testing random variables for independence and identity. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 442–451. IEEE, 2001.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, March 2004. ISBN 0521833787. URL http://www.amazon.com/exec/obidos/redirect?tag=citeulike-20&path=ASIN/0521833787.

Vladimir Braverman, Stephen R Chestnut, Nikita Ivkin, Jelani Nelson, Zhengyu Wang, and David P Woodruff. Bptree: an ℓ_2 heavy hitters algorithm using constant memory. In *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS)*. https://arxiv.org/pdf/1603.00759, 2016.

Clément L Canonne. A survey on distribution testing: Your data is big. but is it blue? *Theory of Computing*, pages 1–100, 2020.

Clément L Canonne, Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Testing conditional independence of discrete distributions. In *2018 Information Theory and Applications Workshop (ITA)*, pages 1–57. IEEE, 2018.

Clément Canonne. A short note on poisson tail bounds. 2017. URL http://www.cs.columbia.edu/~ccanonne/files/misc/2017-poissonconcentration.pdf.

Siu-On Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 1193–1203. SIAM, 2014.

Moses Charikar, Kevin Chen, and Martin Farach-Colton. Finding frequent items in data streams. In *Automata, Languages and Programming*, pages 693–703. Springer, 2002.

Harald Cramer. *Mathematical methods of statistics*. Princeton University Press Princeton, 1946.

A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Applications of mathematics. Springer, 1998. ISBN 9780387984063. URL https://books.google.com/books?id=WmjDlhhDokIC.

Ilias Diakonikolas and Daniel M Kane. A new approach for testing properties of discrete distributions. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 685–694. IEEE, 2016.

Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Sharp bounds for generalized uniformity testing. *arXiv preprint arXiv:1709.02087*, 2017.

Ilias Diakonikolas, Themis Gouleakis, John Peebles, and Eric Price. Sample-optimal identity testing with high probability. In *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.

Ilias Diakonikolas, Themis Gouleakis, John Peebles, and Eric Price. Collision-based testers are optimal for uniformity and closeness. *Chicago Journal of Theoretical Computer Science*, 2019: 1, 2019.

Ilias Diakonikolas, Themis Gouleakis, Daniel M Kane, John Peebles, and Eric Price. Optimal testing of discrete distributions with high probability. *arXiv preprint arXiv:2009.06540*, 2020.

Otmar Ertl. New cardinality estimation methods for hyperloglog sketches. *ArXiv*, abs/1706.07290, 2017.

Philippe Flajolet, Eric Fusy, Olivier Gandouet, and Frederic Meunier. Hyperloglog: The analysis of a near-optimal cardinality estimation algorithm. In *AOFA '07: Proceedings of the 2007 International Conference on Analysis of Algorithms*, 2007.

Oded Goldreich. *Introduction to property testing*. Cambridge University Press, 2017.

Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, pages 68–75. Springer, 2011.

Wassily Hoeffding. Asymptotically Optimal Tests for Multinomial Distributions. *The Annals of Mathematical Statistics*, 36(2):369 – 401, 1965. doi: 10.1214/aoms/1177700150. URL https://doi.org/10.1214/aoms/1177700150.

Dayu Huang and Sean Meyn. Generalized error exponents for small sample universal hypothesis testing. *IEEE transactions on information theory*, 59(12):8157–8181, 2013.

William Karush. Minima of functions of several variables with inequalities as side conditions. Master's thesis, Department of Mathematics, University of Chicago, Illinois, USA, 1939.

Ilmun Kim, Sivaraman Balakrishnan, and Larry Wasserman. Minimax optimality of permutation tests. *arXiv preprint arXiv:2003.13208*, 2020.

H. W. Kuhn and A. W. Tucker. Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 481–492, Berkeley and Los Angeles, 1951. University of California Press.

Kevin J. Lang. Back to the future: an even more nearly optimal cardinality estimation algorithm. *CoRR*, abs/1708.06839, 2017. URL http://arxiv.org/abs/1708.06839.

Erich Leo Lehmann, Joseph P Romano, and George Casella. *Testing statistical hypotheses*, volume 3. Springer, 2005.

Gregory T Minton and Eric Price. Improved concentration bounds for count-sketch. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 669–686. Society for Industrial and Applied Mathematics, 2014.

Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008.

John Peebles. Personal communication, 2015.

Seth Pettie and Dingyu Wang. Information theoretic limits of cardinality estimation: Fisher meets shannon. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, page 556–569, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450380539. doi: 10.1145/3406325.3451032. URL https://doi.org/10.1145/3406325.3451032.

Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. *SIAM Journal on Computing*, 46(1):429–455, 2017.

Appendix A. Variance Optimality (Theorem 1)

Setting. Consider throwing n balls into m bins, for $\lambda = n/m = O(1)$. Suppose $m \gtrsim 1/\varepsilon^4$ (as is needed for constant success probability when $n \lesssim m$). Let k, k' be the number of balls landing in bins 1 and 2, respectively. For any f, let $\sigma^2 = \mathbb{E}_k[f_k^2]$.

A.1. Optimality under a different distribution q'

We define

$$\overline{p}_k := Bin(n, 1/m, k)$$

to be the probability any given bin has k balls in it under the uniform distribution.

Lemma 6 *For any alternative distribution $\overline{q}$, any statistic f minimizing the normalized variance*

$$\frac{Var_p[S_F]}{(\mathbb{E}_{k \sim \overline{q}}[f_k] - \mathbb{E}_{k \sim \overline{p}}[f_k])^2}$$

satisfies

$$(Qf)_k = \alpha(\overline{p}_k - \overline{q}_k)$$

for some α and all k.

Proof This is the KKT condition for minimizing the quadratic $Var_p[S_f] = \frac{1}{m} f^T Q f$ subject to $\sum_k (\overline{p}_k - \overline{q}_k) f_k = 1$. ∎

Let $\overline{q}_k := \frac{1}{2} Bin(n, (1 + 2\varepsilon)/m, k) + \frac{1}{2} Bin(n, (1 - 2\varepsilon)/m, k)$. We would like to show that a quadratic is $1 - o(1)$-close to maximizing the normalized separation between $\overline{p}$ and $\overline{q}$.

We have that

$$\overline{p}_k = \binom{n}{k} \frac{1}{m^k} (1 - 1/m)^{n-k}$$

$$\overline{q}_k = \binom{n}{k} \frac{1}{m^k} \frac{1}{2} ((1 + 2\varepsilon)^k (1 - (1 + 2\varepsilon)/m)^{n-k} + (1 - 2\varepsilon)^k (1 - (1 - 2\varepsilon)/m)^{n-k})$$

$$= \overline{p}_k \frac{(1 + 2\varepsilon)^k (1 - (1 + 2\varepsilon)/m)^{n-k} + (1 - 2\varepsilon)^k (1 - (1 - 2\varepsilon)/m)^{n-k}}{2(1 - 1/m)^{n-k}}$$

$$= \overline{p}_k \frac{1}{2} ((1 + 2\varepsilon)^k (1 - \frac{2\varepsilon}{m-1})^{n-k} + (1 - 2\varepsilon)^k (1 + \frac{2\varepsilon}{m-1})^{n-k})$$

Now, for $|a| \leq 2\varepsilon$,

$$(1 + a) = e^{a - \frac{1}{2}a^2 + \frac{1}{3}a^3 + O(\varepsilon^4)},$$

and

$$(1 + \frac{a}{m-1})^{n-k} = e^{a\lambda + O(|k - \lambda|a/m + a^2\lambda/m)} = e^{a\lambda + O(\varepsilon^4)}$$

for $k \lesssim 1/\varepsilon$ and our setting of $\lambda = O(1), m \gtrsim 1/\varepsilon^4$. Thus

$$
\begin{aligned}
\bar{q}_k &= \bar{p}_k \frac{1}{2}(e^{2\varepsilon k - 2\varepsilon^2 k + \frac{8}{3}\varepsilon^3 k + O(\varepsilon^4 k)} e^{-2\varepsilon\lambda + O(\varepsilon^4)} + e^{-2\varepsilon k - 2\varepsilon^2 k - \frac{8}{3}\varepsilon^3 k + O(\varepsilon^4 k)} e^{2\varepsilon\lambda + O(\varepsilon^4)}) \\
&= \bar{p}_k e^{-2\varepsilon^2 k + O(\varepsilon^4(k+1))} \cosh(2\varepsilon(k-\lambda) + \frac{8}{3}\varepsilon^3 k) \\
&= \bar{p}_k (1 + 2\varepsilon^2((k-\lambda)^2 - k) + O(\varepsilon^4(k^2 + 1 + (k-\lambda)^4)))
\end{aligned}
$$

as long as the final error term is $o(1)$ We now define

$$
\alpha_k := (k-\lambda)^2 - k + \lambda/m \tag{18}
$$

so that

$$
\bar{q}_k = \bar{p}_k (1 + 2\varepsilon^2 \alpha_k + O(\varepsilon^4(k^4 + 1))) \tag{19}
$$

under $k \lesssim \frac{1}{\varepsilon}$ and our assumptions. We make the following simple observations:

Lemma 7

$$
\mathbb{E}_{k \sim \bar{p}}[k] = \lambda \tag{20}
$$

$$
\mathbb{E}_{k \sim \bar{p}}[(k-\lambda)^2] = \lambda(1 - 1/m) \tag{21}
$$

$$
\mathbb{E}_{k \sim \bar{p}}[\alpha_k] = 0. \tag{22}
$$

Proof The first two equations are just the mean and variance of a binomial random variable, and the third follows trivially. ∎

Define

$$
q'_k := \bar{p}_k(1 + 2\varepsilon^2 \alpha_k)
$$

which is also a probability distribution, since $\mathbb{E}_{\bar{p}}[\alpha_k] = 0$ and $\alpha_k \geq -O(\lambda)$ so it is positive. For q', the quadratic statistics are *exactly* optimal:

Lemma 8 *Quadratic statistics $f_k = ak^2 + bk + c$ minimize*

$$
\frac{Var_p[S]}{(\mathbb{E}_{k \sim q'}[mf_k] - \mathbb{E}_{k \sim \bar{p}}[mf_k])^2}
$$

over all f, attaining value

$$
(1 + o(1))\frac{1}{8\varepsilon^4\lambda^2 m}. \tag{23}
$$

Proof Value. We first measure the value obtained by the quadratic statistic. The quadratic statistic $f(k) = (k-\lambda)^2$ has four times the variance of the collisions statistic $\binom{k}{2}$, so Lemma 3 of (Diakonikolas et al., 2019) shows that

$$
\text{Var}_p[S] = 4\binom{n}{2}(\frac{1}{m} - \frac{1}{m^2}) = (1 + o(1))2\lambda^2 m.
$$

We also have, using the moments of a binomial, that

$$
\begin{aligned}
\mathop{\mathbb{E}}_{k \sim q'}[f_k] - \mathop{\mathbb{E}}_{k \sim \bar{p}}[f_k] &= \sum_k 2\varepsilon^2 \alpha_k \bar{p}_k f_k \\
&= 2\varepsilon^2 \mathop{\mathbb{E}}_{\bar{p}}[(k - \lambda)^4 - (k - \lambda)^3 - (1 - 1/m)\lambda(k - \lambda)^2] \\
&= 2\varepsilon^2(1 - 1/m)(\lambda(1 + \frac{3n-6}{m}(1 - 1/m)) - \lambda(1 - 2/m) - \lambda^2(1 - 1/m)) \\
&= 2\varepsilon^2(1 + o(1))(\lambda + 3\lambda^2 - \lambda - \lambda^2) \\
&= 4\varepsilon^2\lambda^2(1 + o(1))
\end{aligned}
$$

Hence

$$
\frac{\mathrm{Var}_p[S]}{(\mathbb{E}_{k \sim q'}[f_k] - \mathbb{E}_{k \sim \bar{p}}[f_k])^2} = (1 + o(1))\frac{m}{8\varepsilon^4\lambda^2}.
$$

Scaling by m^2 gives the result.

Optimality. We now show that it is optimal. For any statistic f, we have that

$$
\mathrm{Var}[S_f] = mf^T QF
$$

for a matrix Q defined by

$$
Q_{k,k} = \bar{p}_k + (m - 1)\bar{p}_k\bar{p}_{k'|k} - m\bar{p}_k^2
$$

and

$$
Q_{k,k'} = (m - 1)\bar{p}_k\bar{p}_{k'|k} - m\bar{p}_k\bar{p}_{k'},
$$

where $\bar{p}_{k'|k} = \mathbb{P}[Y_2 = k' \mid Y_1 = k]$.

For any statistic $S = \sum f(Y_i)$, we have that

$$
(Qf)_k = \bar{p}_k(\mathbb{E}[S \mid Y_1 = k] - \mathbb{E}[S]).
$$

We also have that

$$
\mathop{\mathbb{E}}_{k \sim q'}[f_k] - \mathop{\mathbb{E}}_{k \sim \bar{p}}[f_k] = \sum_k 2\varepsilon^2 \alpha_k f_k.
$$

Therefore we can express the optimization as

$$
\begin{aligned}
\min_f \quad & f^T Q f \\
\text{s.t.} \quad & 2\varepsilon^2 \sum_k \alpha_k f_k = 1/m.
\end{aligned}
\tag{24}
$$

The KKT condition for optimality is then that $Qf = a\alpha$ for some constant a.

Now, the quadratic function $f_k = k^2$ satisfies

$$
\mathbb{E}[S] = \frac{n^2}{m} + n(1 - \frac{1}{m}).
$$

Therefore

$$
\mathbb{E}[S \mid Y_1 = k] = k^2 + \frac{(n - k)^2}{m - 1} + (n - k)(1 - \frac{1}{m - 1}).
$$

19

so

$$\mathbb{E}[S \mid Y_1 = k] - \mathbb{E}[S] = \frac{m}{m-1}k^2 - k(\frac{2n}{m-1} + (1 - \frac{1}{m-1})) + (h_1(n,m))$$
$$= \frac{m}{m-1}\alpha_k + h_2(n,m)$$

for some functions h_1, h_2 of n and m but not k. But since the LHS is zero in expectation over $k \sim \overline{p}$, and so is α_k by Lemma 7, we have $h_2 = 0$. Thus:

$$Qf = \frac{m}{m-1}\alpha.$$

Hence the quadratic satisfies the KKT condition, so it optimizes (24) when scaled appropriately. ■

We also note that the error in approximating $\overline{q}$ by q' has low moments:

Lemma 9 *In our setting,*
$$\mathbb{E}_{\overline{p}}[(\frac{q'_k - \overline{q}_k}{\overline{p}_k})^2] \lesssim \varepsilon^8.$$

Proof

For $k \leq 1/\varepsilon$, we have by (19) that

$$|\frac{q'_k - \overline{q}_k}{\overline{p}_k}| = |\alpha_k - \frac{\overline{q}_k}{\overline{p}_k}| \lesssim \varepsilon^4(k^4 + 1)$$

such that

$$\mathbb{E}_{\overline{p}}[(\frac{q'_k - \overline{q}_k}{\overline{p}_k})^2 1_{k \leq 1/\varepsilon}] \lesssim \mathbb{E}_{\overline{p}}[\varepsilon^8(k^8 + 1)] \lesssim \varepsilon^8.$$

On the other hand, for $k > 1/\varepsilon$,

$$\mathbb{E}_{\overline{p}}[(\frac{q'_k - \overline{q}_k}{\overline{p}_k})^2 1_{k>1/\varepsilon}] \lesssim \mathbb{E}_{\overline{p}}[(\frac{\overline{q}_k}{\overline{p}_k})^2 1_{k>1/\varepsilon}].$$

Now, for $k > 1/\varepsilon$, the $\lambda(1 + \varepsilon)$ part of $\overline{q}$ is more likely than the $\lambda(1 - \varepsilon)$ part. Thus

$$\frac{\overline{q}_k}{\overline{p}_k} \leq \frac{\binom{n}{k}((1+\varepsilon)/m)^k(1-(1+\varepsilon)/m)^{n-k}}{\binom{n}{k}(1/m)^k(1-1/m)^{n-k}} \leq (1+\varepsilon)^k$$

while

$$\overline{p}_k \leq (\frac{e\lambda}{k})^k = e^{O(k) - k \log k}$$

so

$$\mathbb{E}_{\overline{p}}[(\frac{q'_k - \overline{q}_k}{\overline{p}_k})^2 1_{k>1/\varepsilon}] \leq \sum_{k \geq 1/\varepsilon} e^{2\varepsilon k} e^{O(k) - k \log k} \lesssim \varepsilon^{\Omega(1/\varepsilon)} < \varepsilon^8$$

giving the result. ■

A.2. Relating the covariance of one bin to the whole

Recall that $\sigma^2 = \mathbb{E}[f_k^2]$, for $k \sim \bar{p}$.

Lemma 10 *For any $B > 2\lambda$, we have*

$$\mathbb{E}[f_{k'} f_k 1_{k'>B}] \lesssim \sigma^2 \sqrt{\mathbb{P}[k' > B]}.$$

Proof For any $t > B$, we have

$$\mathbb{P}[k \mid k' = t] \lesssim \mathbb{P}[k]$$

for all k. This is trivially true for small $k \leq O(1)$ because $\mathbb{P}[k] = \Omega(1)$, and for large k—since t is above average—$\mathbb{P}[k \mid k' = t] < \mathbb{P}[k]$.

This implies

$$\mathbb{E}[|f_k| \mid k' = t] \lesssim \mathbb{E}_k[|f_k|] \leq \sigma.$$

So

$$\mathbb{E}[f_{k'} f_k 1_{k'>B}] = \sum_{k'>B} p_{k'} f_{k'} \, \mathbb{E}[f_k \mid k'] \lesssim \sigma \sum_{k'>B} p_{k'} |f_{k'}|$$

Of course, by Cauchy-Schwarz,

$$\sum_{k'>B} p_{k'} |f_{k'}| \leq \sqrt{\left(\sum_{k'>B} p_{k'}\right)\left(\sum_{k'>B} p_{k'} f_{k'}^2\right)} \leq \sigma \sqrt{\mathbb{P}[k' > B]}$$

and hence

$$\mathbb{E}[f_{k'} f_k 1_{k'>B}] \lesssim \sigma^2 \sqrt{\mathbb{P}[k' > B]}.$$

∎

Lemma 11 *Let f_k satisfy $\mathbb{E}_p[f_k] = \mathbb{E}_p[k f_k] = 0$. For sufficiently large n, m we have*

$$\sigma^2 \lesssim \frac{1}{m} Var[S_f].$$

Proof We can expand

$$Var[S_f] = m\sigma^2 + m(m-1) \mathbb{E}_{k,k'}[f_k f_{k'}].$$

The lemma statement would be implied by

$$|\mathbb{E}_{k,k'}[f_k f_{k'}]| \leq \frac{1}{2(m-1)} \sigma^2, \tag{25}$$

where k is the number of balls in bin 1 and k' is the number in bin 2. The probability that $k > B$ is at most

$$2\binom{n}{B} \frac{1}{m^B} \leq 2\left(\frac{e\lambda}{B}\right)^B < \frac{1}{n^2 m^4}$$

for $B = O(\log m)$. By Lemma 10,

$$|\mathbb{E}_{k,k'}[f_k f_{k'} 1_{k>B \cup k'>B}]| \leq 2|\mathbb{E}_{k,k'}[f_k f_{k'} 1_{k>B}]| \lesssim \sqrt{\frac{1}{m^4}} \sigma^2.$$

Therefore it would suffice to show

$$|\underset{k,k'}{\mathbb{E}}[f_k f_{k'} 1_{k<B\cap k'<B}]| \ll \frac{\sigma^2}{m}. \tag{26}$$

Let $\mathcal{B}$ be the event that $k < B \cap k' < B$.

Let $\lambda' := (n - k')/(m - 1) = \lambda(1 + \varepsilon')$ for $\varepsilon' = \frac{1}{m-1}\lambda(\lambda - k')$, which under $\mathcal{B}$ satisfies $|\varepsilon'| \lesssim m^{-2/3}$. Then $(k \mid k')$ is $b(n - k', 1/(m - 1))$, which is well approximated by $Poi(\lambda')$. This Poisson approximation gives

$$
\begin{aligned}
p_k' &= \frac{(\lambda')^k e^{-\lambda'}}{k!} = p_k (1 + \varepsilon')^k e^{-\lambda\varepsilon'} \\
&= p_k e^{(k-\lambda)\varepsilon' + O((\varepsilon')^2 k)} \\
&= p_k (1 + (k - \lambda)\varepsilon' + O((k + (k - \lambda)^2)(\varepsilon')^2))
\end{aligned}
$$

for $k \lesssim 1/|\varepsilon'|$, which holds given $\mathcal{B}$. Since $\mathbb{E}_p[f_k] = \mathbb{E}_p[k f_k] = 0$, we have that

$$
\begin{aligned}
|\sum_k p_k (1 + (k - \lambda)\varepsilon')f_k 1_{k\leq B}| &= |\sum_k p_k (1 + (k - \lambda)\varepsilon')f_k 1_{k>B}| \\
&\leq \sqrt{(\sum_{k>B} p_k)\underset{k}{\mathbb{E}}(1 + (k - \lambda)\varepsilon')^2 f_k^2} \\
&\leq \sqrt{\mathbb{P}[k > B]}\sigma n\varepsilon' \leq \frac{1}{m^2}\sigma.
\end{aligned}
$$

Therefore, for any $k' \leq B$,

$$
\begin{aligned}
|\underset{k|k'}{\mathbb{E}}[f_k 1_{k\leq B}]| &\leq \frac{\sigma}{m^2} + |\underset{k}{\mathbb{E}}[O(k + (k - \lambda)^2)(\varepsilon')^2 f_k 1_{k\leq B}]| \\
&\lesssim \frac{\sigma}{m^2} + (\varepsilon')^2 \sqrt{\underset{k}{\mathbb{E}}[(k + (k - \lambda)^2)^2]\underset{k}{\mathbb{E}}[f_k^2]} \\
&\approx (\frac{1}{m} + (\varepsilon')^2)\sigma \\
&\lesssim \sigma/m^{4/3}.
\end{aligned}
$$

Therefore

$$|\underset{k',k}{\mathbb{E}}[f_{k'} f_k 1_{k,k'\leq B}]| = \underset{k'}{\mathbb{E}}[|f_{k'} 1_{k'\leq B}\underset{k|k'}{\mathbb{E}}[f_k 1_{k\leq B}]|] \lesssim \underset{k'}{\mathbb{E}}[|f_{k'}|1_{k'\leq B}\frac{\sigma}{m^{4/3}}] \lesssim \frac{\sigma^2}{m^{4/3}}$$

which gives (26) as needed. ∎

A.3. Putting it together

Theorem 1 *Let $\varepsilon^2 \ll \frac{n}{m} \lesssim 1$ and $n, m, 1/\varepsilon \to \infty$ with $m \gtrsim 1/\varepsilon^4$. Any separable statistic S has normalized variance*

$$\widetilde{\mathrm{var}}_{p,q}(S) \geq (1 + o(1))\frac{1}{8}\frac{m}{n^2\varepsilon^4}$$

between the uniform distribution p and the balanced nonuniform distribution q with $q_k = \frac{1 \pm 2\varepsilon}{m}$.

Quadratic statistics (like collisions or χ^2) match this, getting

$$\widetilde{\mathrm{var}}_{p,q}(S) \leq (1 + o(1)) \frac{1}{8} \frac{m}{n^2 \varepsilon^4} \tag{3}$$

for any ε-far distribution q.

Proof Because the normalized separation is invariant to adding any linear function $ak + b$ to f_k, we can add use this degree of freedom to WLOG satisfy any two linear constraints. We require that

$$\mathop{\mathbb{E}}_{\overline{p}}[f_k] = 0$$

and

$$\mathop{\mathbb{E}}_{\overline{p}}[k f_k] = 0.$$

Let $\widehat{f}_k = k^2 + ak + b$ be the quadratic test statistic with a and b set to satisfy these two constraints.. By Lemma 8, $\widehat{f}$ is optimal under q', so we have that

$$OPT := (1 + o(1)) \frac{1}{8\varepsilon^4 \lambda^2 m} = \frac{\mathrm{Var}[S_{\widehat{f}}]}{\mathbb{E}_{q'}[m\widehat{f}_k]^2} \geq \frac{\mathrm{Var}[S_f]}{\mathbb{E}_{q'}[mf_k]^2}. \tag{27}$$

We have that

$$\begin{aligned}
(\mathop{\mathbb{E}}_{q'}[f_k] - \mathop{\mathbb{E}}_{\overline{q}}[f_k])^2 &= \left(\sum_k (q'_k - \overline{q}_k) f_k\right)^2 \\
&= (\mathop{\mathbb{E}}_{\overline{p}}[\frac{q'_k - \overline{q}_k}{\overline{p}_k} f_k])^2 \\
&\leq \mathop{\mathbb{E}}_{\overline{p}}[(\frac{q'_k - \overline{q}_k}{\overline{p}_k})^2] \mathop{\mathbb{E}}_{\overline{p}}[f_k^2] \\
&\lesssim \varepsilon^8 \sigma_f^2 \qquad\qquad\qquad\text{(Lemma 9)} \\
&\lesssim \varepsilon^8 \frac{1}{m} \mathrm{Var}[S_f]. \qquad\quad\text{(Lemma 10)}
\end{aligned}$$

The same holds for $\widehat{f}$, where we also have by (27) that

$$\mathop{\mathbb{E}}_{q'}[\widehat{f}_k]^2 = (1 + o(1)) \frac{1}{m} \mathrm{Var}[S_{\widehat{f}}] \cdot 8\varepsilon^4 \lambda^2$$

so

$$(\mathop{\mathbb{E}}_{q'}[\widehat{f}_k] - \mathop{\mathbb{E}}_{\overline{q}}[\widehat{f}_k])^2 \lesssim \varepsilon^8 \frac{1}{m} \mathrm{Var}[S_{\widehat{f}}] = \varepsilon^4 \cdot (1 + o(1)) \frac{1}{8\lambda^2} \mathop{\mathbb{E}}_{q'}[\widehat{f}_k]^2 \ll \mathop{\mathbb{E}}_{q'}[\widehat{f}_k]^2$$

and hence

$$\mathop{\mathbb{E}}_{\overline{q}}[\widehat{f}_k]^2 = (1 + o(1)) \mathop{\mathbb{E}}_{q'}[\widehat{f}_k]^2$$

so

$$\widetilde{\mathrm{var}}_{p,q}(S_{\widehat{f}}) = \frac{\mathrm{Var}_p[S_{\widehat{f}}]}{\mathbb{E}_{\overline{q}}[S_{\widehat{f}}]^2} = (1 + o(1))OPT.$$

For any alternative f, we split into two cases:

Reasonably good f. When

$$\frac{\text{Var}[S_f]}{\mathbb{E}_{q'}[mf_k]^2} \leq 100OPT,$$

we again have

$$\varepsilon^8 \frac{1}{m} \text{Var}[S_f] \ll \mathbb{E}_{q'}[\widehat{f}_k]^2$$

so

$$\mathbb{E}_{\overline{q}}[f_k]^2 = (1 + o(1)) \mathbb{E}_{q'}[\widehat{f}_k]^2$$

and

$$\widetilde{\text{var}}_{p,q}(S_f) = (1 + o(1)) \frac{\text{Var}[S_f]}{\mathbb{E}_{q'}[mf_k]^2} \geq (1 + o(1))OPT.$$

Bad f. When

$$\frac{\text{Var}[S_f]}{\mathbb{E}_{q'}[mf_k]^2} \geq 100OPT,$$

we use $(a + b)^2 \leq 2a^2 + 2b^2$ to observe that

$$\begin{aligned}
\widetilde{\text{var}}_{p,q}(S_f) = \frac{\text{Var}[S_f]}{\mathbb{E}_q[mf_k]^2} &\geq \frac{1}{2} \frac{\text{Var}[S_f]}{\mathbb{E}_{q'}[mf_k]^2 + m^2(\mathbb{E}_{q'}[f_k] - \mathbb{E}_{\overline{q}}[f_k])^2} \\
&= \frac{1}{2} \frac{\text{Var}[S_f]}{\mathbb{E}_{q'}[mf_k]^2 + O(\varepsilon^8 m \text{Var}[S_f])} \\
&\geq \frac{1}{2} \frac{1}{\frac{1}{100OPT} + \varepsilon^8 m} \\
&> OPT
\end{aligned}$$

Thus, the quadratic tester achieves near-optimal separation for this q.

Finally, for arbitrary distributions q ε-far from p in TV, we note that the collisions tester satisfies

$$\mathbb{E}_q[S] = \binom{n}{2} \|q\|_2^2.$$

By convexity the ε-far q minimizing this has its values above and below $1/m$ all equal; if there are k values above $1/m$ this gives

$$\begin{aligned}
\frac{1}{\binom{n}{2}} \mathbb{E}_q[S] &= k(\frac{1}{m} + \frac{\varepsilon}{k})^2 + (m - k)(\frac{1}{m} - \frac{\varepsilon}{m-k})^2 \\
&= \frac{1}{m} + \varepsilon^2 \left(\frac{1}{k} + \frac{1}{m-k} \right)
\end{aligned}$$

which is minimized at $k = m - k = m/2$, precisely the q considered above. ∎

Appendix B. Background for Tester Analysis

B.1. Gärtner-Ellis Theorem

The statements in this section are taken from Dembo and Zeitouni (1998).

Consider a sequence of random variables $Z_n \sim p_n$ and let the logarithmic moment generating function of Z_n be

$$\Lambda_n(\theta) := \log \mathbb{E}[e^{\theta Z_n}]$$

Assumption 1 *Suppose that for each $\theta \in \mathbb{R}$, the logarithmic moment generating function, defined as the limit*

$$\Lambda(\theta) = \lim_{n \to \infty} \frac{1}{n} \Lambda_n(n\theta)$$

exists as an extended real number, and that the origin lies in the interior of the set $\mathcal{D}_\Lambda := \{\theta \in \mathbb{R} : \Lambda(\theta) < \infty\}$.

Let

$$\Lambda^*(\tau) = \sup_{\theta \geq 0} \{\theta\tau - \Lambda(\theta)\} \tag{28}$$

be the Fenchel-Legendre transform of Λ.

Definition 12 $\tau \in \mathbb{R}$ *is an exposed point of Λ^* if for some $\lambda \in \mathbb{R}$, for every $x \neq y$,*

$$\lambda\tau - \Lambda^*(\tau) > \lambda x - \Lambda^*(x)$$

Then, λ is called an exposing hyperplane.

Theorem 13 (Gärtner–Ellis) *Let Assumption 1 hold.*

(a) For any closed set F,

$$\limsup_{n \to \infty} \frac{1}{n} \log p_n(F) \leq - \inf_{x \in F} \Lambda^*(x)$$

(b) For any open set G,

$$\liminf_{n \to \infty} \frac{1}{n} \log p_n(G) \geq - \inf_{x \in G \cap F} \Lambda^*(x)$$

where F is the set of exposed points of Λ^* whose exposing hyperplane belongs to D_Λ^o, where D_Λ^o is the interior of D_Λ.

B.2. Worst Case Distributions for Uniformity Testing

In this section, we study the worst-case ε-far distributions for test statistics that are convex symmetric functions of the histogram (i.e., the number of times each domain element is sampled) of an arbitrary random variable Y. This is an extension of the results in (Diakonikolas et al., 2018), which we recap below.

Prior work. We start with the following definition:

Definition 14 *Let $p = (p_1, \ldots, p_n), q = (q_1, \ldots, q_n)$ be probability distributions and $p^{\downarrow}, q^{\downarrow}$ denote the vectors with the same values as p and q respectively, but sorted in non-increasing order. We say that p majorizes q (denoted by $p \succ q$) if*

$$\forall k : \sum_{i=1}^{k} p_i^{\downarrow} \geq \sum_{i=1}^{k} q_i^{\downarrow} . \tag{29}$$

A proof of the following simple fact can be found in (Diakonikolas et al., 2018):

Fact 15 *Let p be a probability distribution over $[n]$ and $S \subseteq [n]$. Let q be the distribution which is identical to p on $[n] \setminus S$, and for every $i \in S$ we have $q_i = \frac{p(S)}{|S|}$, where $|S|$ denotes the cardinality of S. Then, we have that $p \succ q$.*

We also use the following standard terminology: we say that a real random variable A *stochastically dominates* a real random variable B if for all $x \in \mathbb{R}$ it holds $\mathbb{P}[A > x] \geq \mathbb{P}[B > x]$.

We say that a test statistic S is "convex symmetric" if it is a convex function of the histogram $(Y_1, \ldots, Y_m)$ and invariant under permutation of the Y_i. A "test" is given by a test statistic S and threshold τ, and outputs "uniform" if $S \leq \tau$ and "non-uniform" otherwise.

It was shown in (Diakonikolas et al., 2018) that if p majorizes q, then a convex symmetric test statistic of p stochastically dominates one from q:

Lemma 16 (Lemma 19 of (Diakonikolas et al., 2018)) *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a symmetric convex function and p be a distribution over $[n]$. Suppose that we draw m samples from p, and let X_i denote the number of times we sample element i. Let $g(p)$ be the random variable $f(X_1, X_2, \ldots, X_n)$. Then, for any distribution q over $[n]$ such that $p \succ q$, we have that $g(p)$ stochastically dominates $g(q)$.*

New claims. We will show that it suffices to consider distributions that are "flat", meaning that p_i takes only two values:

Definition 17 *We say a probability distribution p over $[n]$ is an γ-skewed flat distribution if it takes the form:*

$$p_i = \begin{cases} a & i \in T \\ b & i \notin T \end{cases}$$

for some reals a, b and set $T \subseteq [n]$ with $|T| \in [\gamma n, (1 - \gamma)n]$.

We make the following generalization of Lemma 21 in (Diakonikolas et al., 2018) (which is the $\gamma = 1/2$ case):

Lemma 18 *Let p be a probability distribution. For any $0 < \gamma < 1/2$, there exists an γ-skewed flat distribution p' such that $p \succ p'$ and*

$$(1 - \gamma) \cdot \|p - U_n\|_{TV} \leq \|p' - U_n\|_{TV} \leq \|p - U_n\|_{TV}.$$

Proof Let $T = \{i : p_i > 1/n\}$, so

$$\|p - U_n\|_{TV} = \sum_{i \in T}(p_i - 1/n) = \sum_{i \in [n] \setminus T}(1/n - p_i). \tag{30}$$

If $|T| \in [\gamma n, (1 - \gamma)n]$, we can simply choose p' to be p averaged over T, and p averaged over $[n] \setminus T$—this is γ-skewed and flat, has $p \succ p'$ by Lemma 15, and has $\|p' - U_n\|_{TV} = \|p - U_n\|_{TV}$. The only remaining cases have $|T| \notin [\gamma n, (1 - \gamma)n]$, since this approach would not be γ-skewed.

Let $T' \subset [n]$ contain the largest either γn or $(1 - \gamma)n$ coordinates of p, depending on whether $|T| < \gamma n$ or $|T| > (1 - \gamma)n$, and let p' average p over T' and over $[n] \setminus T$. This is γ-skewed and flat, and has $p \succ p'$ by Lemma 15, so the only question is the TV bound.

We have that

$$\|p' - U_n\|_{TV} = \sum_{i \in T'}(p'_i - 1/n) = \sum_{i \in T'}(p_i - 1/n) = \|p - U_n\|_{TV} - \sum_{i \in T \setminus T'}(p_i - 1/n) - \sum_{i \in T' \setminus T}(1/n - p_i).$$

Every term in the right two sums is nonnegative, so $\|p' - U_n\|_{TV} \leq \|p - U_n\|_{TV}$.

Now, if $|T| < \gamma n$, then $T \setminus T'$ is empty and, since T' takes the largest coordinates in p, p_i is larger on average for $i \in T' \setminus T$ than for $i \in [n] \setminus T$:

$$\sum_{i \in T' \setminus T}(1/n - p_i) \leq \frac{|T' \setminus T|}{|[n] \setminus T|}\sum_{i \in [n] \setminus T}(1/n - p_i) = \frac{\gamma n - |T|}{n - |T|}\|p - U_n\|_{TV} \leq \gamma\|p - U_n\|_{TV}$$

so

$$\|p' - U_n\|_{TV} \geq (1 - \gamma)\|p - U_n\|_{TV}.$$

Similarly, if $|T| > (1 - \gamma)n$, then $T' \setminus T$ is empty and

$$\sum_{i \in T \setminus T'}(p_i - 1/n) \leq \frac{|T \setminus T'|}{|T|}\sum_{i \in T}(p_i - 1/n) = \frac{|T| - (1 - \gamma)n}{|T|}\|p - U_n\|_{TV} \leq \gamma\|p - U_n\|_{TV},$$

again giving

$$\|p' - U_n\|_{TV} \geq (1 - \gamma)\|p - U_n\|_{TV}$$

as desired. $\blacksquare$

The above results mean that it suffices to prove that our algorithm can distinguish the uniform distribution from γ-skewed flat distributions. The inefficiency from not considering extremely skewed distributions is only $1 + O(\gamma)$:

Lemma 19 *Suppose a convex symmetric test statistic S and threshold has the property that, when applied to any ε-far γ-skewed flat distribution p, the false negative rate is at most δ. Then the same statistic and threshold, when applied to any $\frac{1}{1-\gamma}\varepsilon$-far distribution p', also has false negative rate at most δ.*

Proof For any such p', Lemma 18 states that there exists a p that is γ-skewed, ε-far from U_n in TV, and with $p' \succ p$. Lemma 16 then states that S on p' stochastically dominates S on p, so the chance of falling below the threshold is smaller for p' than for p—and the latter is δ by assumption. $\blacksquare$

Implication for Error Exponents. Let $\varepsilon = \varepsilon(n)$, $m = m(n)$, and $\tau = \tau(n)$ be functions of n. Let p be uniform on $[m]$. The false positive error exponent $c_+ = c_+(\varepsilon, m)$ of a test (S, τ) is

$$c_+ = \lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \mathbb{P}_p[S > \tau].$$

For a *particular* family of distributions q, the false negative error exponent $c_-^{(q)} = c_-^{(q)}(\varepsilon, m)$ is

$$c_-^{(q)} = \lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \mathbb{P}_q[S \le \tau].$$

The false negative error exponent c_- is the worst such exponent over all ε-far distributions q:

$$c_- = \inf_{q : \|p - q\|_{TV} \ge \varepsilon} c_-^{(q)}.$$

Varying τ allows for a tradeoff between false negatives and false positives. Balancing the two gives us the *error exponent* $c = c(\varepsilon, m)$ for a test statistic S:

$$c = \sup_\tau \min(c_+, c_-).$$

If a test statistic has error exponent c, it can distinguish the uniform distribution from any non-uniform distribution with probability $1 - \exp(-(1 + o(1))c\varepsilon^4 n^2 / m)$. Equivalently, it gets error probability δ where

$$n = \frac{1 + o(1)}{\sqrt{c}} \cdot \frac{\sqrt{m \log \frac{1}{\delta}}}{\varepsilon^2}.$$

We define $\bar{c} = \bar{c}(\varepsilon, m, \gamma)$ to denote an alternative to c where we only consider ε-far distributions q that are γ-skewed and flat.

Lemma 20 *For any functions ε, m, γ,*

$$(1 - 4\gamma) \cdot \bar{c}((1 - \gamma)\varepsilon, m, \gamma) \le c(\varepsilon, m) \le \bar{c}(\varepsilon, m, \gamma).$$

Proof The upper bound on c is trivial: as an infimum over a larger set of q, $c_- \le \bar{c}_-$, so $c \le \bar{c}$.

For the lower bound on c, we note by Lemma 19 that for any q with $\|p - q\|_{TV} \ge \varepsilon$ that there exists a γ-skewed flat distribution q' with $\|p - q'\|_{TV} \ge (1 - \gamma)\varepsilon$ such that

$$\mathbb{P}_q[S \le \tau] \le \mathbb{P}_{q'}[S \le \tau]. \tag{31}$$

This implies that

$$\frac{m}{n^2 \varepsilon^4} \log \mathbb{P}_q[S \le \tau] \le (1 - \gamma)^4 \frac{m}{n^2 (1 - \gamma)^4 \varepsilon^4} \log \mathbb{P}_{q'}[S \le \tau].$$

so

$$c_-^{(q)}(\varepsilon, m) \ge (1 - \gamma)^4 c_-^{(q')}((1 - \gamma)\varepsilon, m),$$

and hence $c \ge (1 - 4\gamma)\bar{c}((1 - \gamma)\varepsilon, m)$. ∎

Lemma 21 *Let S by a convex symmetric test statistic. Consider any family of parameters (n, ε, m). Suppose that there exists a constant γ' such that, for any $\gamma = \Omega(1) > \gamma'$ and ε' that uniformly satisfies $(1 - \gamma')\varepsilon(n) \leq \varepsilon'(n) \leq \varepsilon(n)$,*

$$\overline{c}(\varepsilon', m, \gamma) = c^*$$

for a fixed value c^* [that depends on the family (n, ε, m) but not on the value of n or γ, γ'].
    Then

$$c(\varepsilon, m) = c^*.$$

Proof By Lemma 20,

$$c(\varepsilon, m) \leq \overline{c}(\varepsilon, m, \gamma) = c^*.$$

Moreover, for any C we have

$$c(\varepsilon, m) \geq (1 - 4\gamma)c^*.$$

where c is a limit as $n \to \infty$ independent of C. But this means that $c(\varepsilon, m) = c^*$, because it is larger than any fixed value less than c^*. ∎

Appendix C. Huber Statistic in Sublinear Regime

C.1. Regime

The Huber statistic is given by

$$S = \sum_{j=1}^{m} h_\beta \left(Y_j^n - \frac{n}{m} \right) \tag{8}$$

where

$$h_\beta(x) := \begin{cases} x^2 & \text{for } |x| < \beta \\ 2\beta|x| - \beta^2 & \text{otherwise} \end{cases} \tag{9}$$

is the Huber loss function. Here $Y_j^n = \sum_{i=1}^{n} \mathbb{1}_{\{X_i = j\}}$ and $X_1, \ldots, X_n$ are the n samples drawn from distribution ν supported on $[m]$.

Assumption 2 $n = \Omega(m), n/m \ll \frac{1}{\varepsilon^2}, \varepsilon \ll 1, \frac{n^2}{m}\varepsilon^4 \gg 1$, and $m \geq C \log n$ for sufficiently large constant C. In addition, we have the following constraints on β, the Huber parameter, and Δ.

$$\beta = \omega \left(\log \left(\frac{1}{\Delta} \right) + \sqrt{\frac{n}{m} \log \left(\frac{1}{\Delta} \right)} \right) \tag{11}$$

$$\Delta = O \left(\frac{n\varepsilon^2}{m} \right) \tag{12}$$

$$(\beta^2 \varepsilon^2)^3 = o \left(\Delta^2 \right) \tag{13}$$

We will assume that Assumption 2 holds throughout this section.

Note that since $\Delta = o(1)$, (13) implies that

$$\beta^2 \varepsilon^2 = o(1) \tag{32}$$

Our goal is to compute an upper bound on the asymptotic expansion of the cumulant generating function (also called the logarithmic moment generating function) of this statistic.

For ease, instead of analyzing S directly, we will analyze the statistic

$$\tilde{S} = \frac{m}{n^2 \varepsilon^2} [S - n] \tag{33}$$

Note that this has the same error probability as S since it simply applies a translation and scaling to S.

Consider the moment generating function (MGF) of $\tilde{S}$ with respect to distribution ν, given by

$$M_{\tilde{S},\nu}(\theta) = \mathbb{E}_\nu \left[\exp(\theta \tilde{S}) \right]$$

The logarithmic moment generating function of $\tilde{S}$ with respect to distribution ν is given by

$$\Lambda_{n,\nu}(\theta) := \log \left(M_{\tilde{S},\nu}(\theta) \right) \tag{34}$$

We will compute an asymptotic expansion of the limiting logarithmic moment generating function of $\tilde{S}$, given by

$$\Lambda_\nu(\theta) = \lim_{n \to \infty} \frac{m}{n^2 \varepsilon^4} \Lambda_{n,\nu} \left(\frac{n^2 \varepsilon^4}{m} \theta \right)$$

For ease of exposition, we define a centering function

$$\phi(k) := \left| k - \frac{n}{m} \right| \tag{35}$$

C.2. Poissonization

Define $\tilde{S}_{Poi(\lambda)}$ to be the Poissonized statistic, that is the statistic $\tilde{S}$ when the number of balls is chosen according to the Poisson distribution with mean λ.

We begin by computing the MGF of $\tilde{S}_{Poi(\lambda)}$ with MGF parameter $\frac{n^2 \varepsilon^4}{m} \theta$. That is, let

$$A_\lambda(\theta) := \mathbb{E}\left[\exp\left(\frac{n^2 \varepsilon^4}{m} \theta \tilde{S}_{Poi(\lambda)} \right) \right] = \exp(-\varepsilon^2 \theta n) \, \mathbb{E}\left[\exp\left(\varepsilon^2 \theta \sum_{j=1}^{m} h_\beta \left(Z_j - \frac{n}{m} \right) \right) \right] \tag{36}$$

where $Z_j \sim Poi(\lambda \nu_j)$ and are independent. Due to this independence,

$$A_\lambda(\theta) = \exp(-\varepsilon^2 \theta n) \prod_{j=1}^{m} \mathbb{E}\left[\exp\left(\varepsilon^2 \theta h_\beta \left(Z_j - \frac{n}{m} \right) \right) \right]$$

Define

$$f(k) := 1 + \varepsilon^2 \theta \phi(k)^2 + \frac{\varepsilon^4 \theta^2}{2} \phi(k)^4 \tag{37}$$

We will first show the following

Lemma 22

$$\varepsilon^2 \theta \, \mathbb{E}\left[h_\beta \left(Z_j - \frac{n}{m} \right) \right] = \varepsilon^2 \theta \, \mathbb{E}[\phi(Z_j)^2] + o(\Delta^2)$$

$$\frac{\varepsilon^4 \theta^2}{2} \mathbb{E}\left[h_\beta \left(Z_j - \frac{n}{m} \right)^2 \right] = \frac{\varepsilon^4 \theta^2}{2} \mathbb{E}[\phi(Z_j)^4] + o(\Delta^2)$$

$$\sum_{l=3}^{\infty} \frac{(\varepsilon^2 \theta)^l}{l!} \mathbb{E}\left[h_\beta \left(Z_j - \frac{n}{m} \right)^l \right] = o(\Delta^2)$$

where h_β is defined in (9), and $Z_j \sim Poi(\lambda \nu_j)$, for $\lambda = n(1 + O(\varepsilon^2))$ and $\nu_j = 1/m + O(\varepsilon/m)$ for all j.

Proof

$$\varepsilon^2 \theta \, \mathbb{E}\left[h_\beta \left(Z_j - \frac{n}{m} \right) \right] = \varepsilon^2 \theta \left\{ \mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j) \leq \beta\}} \phi(Z_j)^2 \right] + \mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j) > \beta\}} \beta(2\phi(Z_j) - \beta) \right] \right\}$$

$$= \varepsilon^2 \theta \, E[\phi(Z_j)^2] - \varepsilon^2 \theta \, \mathbb{E}[\mathbb{1}_{\{\phi(Z_j) > \beta\}} \phi(Z_j)^2] + \varepsilon^2 \theta \, \mathbb{E}[\mathbb{1}_{\{\phi(Z_j) > \beta\}} \beta(2\phi(Z_j) - \beta)]$$

31

By Lemma 45, the second term is $o(\Delta^2)$. For the third term,

$$\varepsilon^2 \theta \, \mathbb{E}[\mathbb{1}_{\{\phi(Z_j) > \beta\}} \beta(2\phi(Z_j) - \beta)] \leq \mathbb{E}[\mathbb{1}_{\{\phi(Z_j) > \beta\}} \exp(\varepsilon^2 |\theta| \beta(2\phi(Z_j) - \beta))]$$

By Lemma 47, this is $o(\Delta)^2$. So, we have the first claim. The second claim can be proved in a similar way. For the third claim,

$$\sum_{l=3}^{\infty} \frac{(\varepsilon^2 \theta)^l}{l!} \, \mathbb{E}\left[h_\beta \left(Z_j - \frac{n}{m} \right)^l \right] = \sum_{l=3}^{\infty} \frac{(\varepsilon^2 \theta)^l}{l!} \, \mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j) \leq \beta\}} \phi(Z_j)^{2l} \right]$$

$$+ \sum_{l=3}^{\infty} \frac{(\varepsilon^2 \theta)^l}{l!} \, \mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j) > \beta\}} (\beta(2\phi(Z_j) - \beta))^l \right]$$

By Lemma 42, the first term is $o(\Delta^2)$. For the second term, in a similar fashion as before,

$$\sum_{l=3}^{\infty} \frac{(\varepsilon^2 \theta)^l}{l!} \, \mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j) > \beta\}} (\beta(2\phi(Z_j) - \beta))^l \right] \leq \mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j) > \beta\}} \exp(\varepsilon^2 |\theta| \beta(2\phi(Z_j) - \beta)) \right]$$

By Lemma 47, this is $o(\Delta^2)$. ∎

Lemma 23 *We have*

$$\mathbb{E}\left[\exp\left(\varepsilon^2 \theta h_\beta \left(Z_j - \frac{n}{m} \right) \right) \right] = \mathbb{E}\left[f(Z_j) \right] + o\left(\Delta^2 \right) \tag{38}$$

where h_β is defined in (9), and $Z_j \sim Poi(\lambda \nu_j)$, for $\lambda = n(1 + O(\varepsilon^2))$ and $\nu_j = 1/m + O(\varepsilon/m)$ for all j.

Proof Follows from Lemma 22. ∎

C.3. Depoissonization

First, we will show that $A_\lambda(\theta)$ is analytic in λ.

Lemma 24 $A_\lambda(\theta)$ *is analytic in λ.*

Proof We will show that $\mathbb{E}[\exp\left(\varepsilon^2 \theta h_\beta \left(Z_j - \frac{n}{m} \right) \right)]$ can be written as a finite sum of analytic functions in λ. Since the sum and product of analytic functions also analytic, this will show that $A_\lambda(\theta)$ is analytic. Let

$$A := \sum_{k:\phi(k) \leq \beta} \left[\frac{(\lambda \nu_j)^k}{k!} e^{-\lambda \nu_j} \exp\left(\varepsilon^2 \theta \phi(k)^2 \right) \right]$$

$$B := \mathbb{E}\left[\exp\left\{ \varepsilon^2 \theta \beta \left(2 \left(Z_j - \frac{n}{m} \right) - \beta \right) \right\} \right]$$

$$C := \sum_{k:k < \frac{n}{m} - \beta} \left[\frac{(\lambda \nu_j)^k}{k!} e^{-\lambda \nu_j} \left[\exp\left\{ \varepsilon^2 \theta \beta \left(2 \left(\frac{n}{m} - k \right) - \beta \right) \right\} - \exp\left\{ \varepsilon^2 \theta \beta \left(2 \left(k - \frac{n}{m} \right) - \beta \right) \right\} \right] \right]$$

Then, the expectations in B and C can be expressed in terms of the moments and MGF of the Poisson distribution, and so, are analytic. Each of A, B, C is a finite sum of analytic functions, and so, is analytic. It is easy to verify that

$$\mathbb{E}\left[\exp\left(\varepsilon^2\theta h_\beta\left(Z_j - \frac{n}{m}\right)\right)\right] = A + B + C$$

is thus analytic.

∎

Now, we want depoissonize $A_\lambda(\theta)$. For ease of exposition, we will prove a more general result. First we assume the following.

Assumption 3 *Suppose ξ is a function such that*

$$A_\lambda(\theta) = \exp(-\varepsilon^2\theta n)\prod_{j=1}^m \mathbb{E}[\exp(\varepsilon^2\theta\xi(Z_j))]$$

where $A_\lambda(\theta)$ is analytic in λ, and $Z_j \sim Poi(\lambda\nu_j)$.
We assume that, for $\lambda = n(1 + O(\varepsilon^2))$ and $\nu_j = 1/m + O(\varepsilon/m)$ for all j, we have

$$\varepsilon^2\theta\,\mathbb{E}[\xi(Z_j)] = \varepsilon^2\theta\,\mathbb{E}[\phi(Z_j)^2] + o(\Delta^2) \tag{39}$$

$$\frac{\varepsilon^4\theta^2}{2}\,\mathbb{E}[\xi(Z_j)^2] = \frac{\varepsilon^4\theta^2}{2}\,\mathbb{E}[\phi(Z_j)^4] + o(\Delta^2) \tag{40}$$

$$\sum_{l=3}^\infty \frac{(\varepsilon^2\theta)^l}{l!}\,\mathbb{E}[\xi(Z_j)^l] = o(\Delta^2) \tag{41}$$

so that

$$\mathbb{E}[\exp(\varepsilon^2\theta\xi(Z_j))] = \mathbb{E}[f(Z_j)] + o\left(\Delta^2\right)$$

where f is defined in (37).
Let $Y_j^n = \sum_{i=1}^n \mathbb{1}_{\{X_i=j\}}$ and $X_1,\ldots,X_n$ be n samples drawn from distribution ν supported on $[m]$.

We will show the following:

Lemma 25 *Suppose Assumption 3 holds. Then, if ν is the uniform distribution such that $\nu_j = 1/m$ for all j, we have*

$$\exp(-\varepsilon^2\theta n)\,\mathbb{E}\left[\exp\left(\varepsilon^2\theta\sum_{j=1}^m \xi(Y_j^n)\right)\right] = (1 + O(1/n))\exp\left\{\frac{n^2\varepsilon^4}{m}(\theta^2 + o(1))\right\}$$

If ν is an alternate distribution such that $\nu_j = \frac{1}{m} + \frac{\varepsilon}{\gamma m}$ for $j \leq \gamma m$, and $\nu_j = \frac{1}{m} - \frac{\varepsilon}{(1-\gamma)m}$ for $j > \gamma m$, for $\gamma = \Theta(1), 1 - \gamma = \Theta(1)$, we have

$$\exp(-\varepsilon^2\theta n)\,\mathbb{E}\left[\exp\left(\varepsilon^2\theta\sum_{j=1}^m \xi(Y_j^n)\right)\right] = (1+O(1/n))\exp\left\{\frac{n^2\varepsilon^4}{m}\left(\theta^2 + \theta\frac{1}{\gamma(1-\gamma)} + o(1)\right)\right\}$$

First, we have that our expression stated can be written as an integral.

Lemma 26 *Consider any function $f : \mathbb{R} \to \mathbb{R}$. If we draw $Z_j \sim Poi(\lambda \nu_j)$ for $j \in [m]$, and*

$$\prod_{j=1}^{m} \mathbb{E}[f(Z_j)]$$

is analytic in λ, we have

$$\mathbb{E}\left[\prod_{j=1}^{m} f(Y_j^n)\right] = \frac{n!}{2\pi i} \oint \frac{e^\lambda}{\lambda^{n+1}} \prod_{j=1}^{m} \mathbb{E}[f(Z_j)] d\lambda$$

where $Y_1^n, \ldots, Y_m^n$ are n samples drawn according to ν.

Proof Conditioning on $\sum_{j=1}^{m} Z_j = k$, we have

$$\prod_{j=1}^{m} \mathbb{E}\left[f(Z_j)\right] = \mathbb{E}\left[\prod_{j=1}^{m} f(Z_j)\right] = \sum_{k=0}^{\infty} \mathbb{P}\left[\sum_{j=1}^{m} Z_j = k\right] \mathbb{E}\left[\prod_{j=1}^{m} f(Z_j) \middle| \sum_{j=1}^{m} Z_j = k\right]$$

$$= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \mathbb{E}\left[\prod_{j=1}^{m} f(Y_j^n)\right]$$

Now, for any analytic function $\phi(\lambda)$ with power series expansion given by

$$\phi(\lambda) = \sum_{k=0}^{\infty} a_k \lambda^k$$

we have by Cauchy's theorem that

$$a_n = \frac{1}{2\pi i} \oint \phi(\lambda) \frac{1}{\lambda^{n+1}} d\lambda.$$

By assumption, $\prod_{j=1}^{m} \mathbb{E}[f(Z_j)]$ is analytic in λ. Therefore,

$$\mathbb{E}\left[\prod_{j=1}^{m} f(Y_j^n)\right] = \frac{n!}{2\pi i} \oint \frac{e^\lambda}{\lambda^{n+1}} \prod_{j=1}^{m} \mathbb{E}[f(Z_j)] d\lambda$$

which is the desired bound.

■

Corollary 27 *Under Assumption 3,*

$$\exp(-\varepsilon^2 \theta n) \mathbb{E}\left[\exp\left(\varepsilon^2 \theta \sum_{j=1}^{m} \xi(Y_j^n)\right)\right] = \frac{n!}{2\pi i} \oint e^\lambda A_\lambda(\theta) \frac{d\lambda}{\lambda^{n+1}} \tag{42}$$

We will choose a contour passing through a particular λ_0, and this will make it easy to evaluate the integral. We carry out the integration along the contour given by $\lambda = \lambda_0 e^{i\psi}$, where

$$\lambda_0 = n(1 - \varepsilon^2\theta)$$

We substitute $\lambda_0 e^{i\psi}$ into (42) to get that

$$\exp(-\varepsilon^2\theta n)\,\mathbb{E}\left[\exp\left(\varepsilon^2\theta\sum_{j=1}^{m}\xi(Y_j^n)\right)\right] = e^{-\varepsilon^2\theta n}\frac{n!}{2\pi}\lambda_0^{-n}Re\left[\int_{-\pi}^{\pi}g(\psi)d\psi\right] \qquad (43)$$

with

$$g(\psi) := e^{-in\psi}\prod_{j=1}^{m}\left\{\sum_{k=0}^{\infty}\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}\exp(\varepsilon^2\theta\xi(k))\right\} \qquad (44)$$

We will split this integral into 3 parts. Let

$$I_1 = Re\left[\int_{-\pi/3}^{\pi/3}g(\psi)d\psi\right]$$

$$I_2 = Re\left[\int_{-\pi}^{-\pi/3}g(\psi)d\psi\right] \qquad (45)$$

$$I_3 = Re\left[\int_{\pi/3}^{\pi}g(\psi)d\psi\right]$$

We will show that I_1 dominates. We show this by bounding $g(\psi)$ in the region $\psi \in [-\pi, -\pi/3]\cup[\pi/3, \pi]$ as follows.

Lemma 28 *Under Assumption 3, and $m \geq C\log n$ for sufficiently large constant C, for $\psi \in [-\pi, -\pi/3] \cup [\pi/3, \pi]$,*

$$|g(\psi)|\leq O\left(\frac{e^n}{n}\right)$$

Proof By definition of g from (44), and using the assumption on ξ from Assumption 3, we have that,

$$|g(\psi)| = \left|e^{-in\psi}\prod_{j=1}^{m}\left\{\sum_{k=0}^{\infty}\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}\exp(\varepsilon^2\theta\xi(k))\right\}\right|$$

$$\leq \left|\prod_{j=1}^{m}\left\{\sum_{k=0}^{\infty}\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}\right\}\right| + \left|\prod_{j=1}^{m}\left\{\sum_{k=0}^{\infty}\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}\left(\sum_{l=1}^{\infty}\frac{(\varepsilon^2\theta)^l}{l!}\xi(k)^l\right)\right\}\right|$$

Now, for choice of $\lambda_0 = n(1 - \varepsilon^2\theta)$, and $\psi \in [-\pi, -\pi/3] \cup [\pi/3, \pi]$,

$$\left|\prod_{j=1}^{m}\left\{\sum_{k=0}^{\infty}\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}\right\}\right| = \left|\prod_{j=1}^{m}e^{\lambda_0\nu_j e^{i\psi}}\right| = \left|e^{\lambda_0 e^{i\psi}}\right| = |e^{n(1-\varepsilon^2\theta)e^{i\psi}}|\leq O(e^{0.5n})$$

35

For the second term, by Assumption 3, for our $\lambda_0 = n(1 - \varepsilon^2\theta)$, and $\nu_j = 1/m + O(\varepsilon/m)$ for all m, this is

$$e^{\lambda_0} \prod_{j=1}^{m} \left\{ \sum_{l=1}^{\infty} \frac{(\varepsilon^2\theta)^l}{l!} \mathbb{E}[\xi(Z_j)^l] \right\} = e^{n(1-\varepsilon^2\theta)} \prod_{j=1}^{m} \left\{ \varepsilon^2\theta \, \mathbb{E}[\phi(Z_j)^2] + \frac{\varepsilon^4}{2}\theta^2 \, \mathbb{E}[\phi(Z_j)^4] + o(\Delta^2) \right\}$$

By Lemma 38, and since $n = o(m/\varepsilon^2)$, this is

$$e^{n(1-\varepsilon^2\theta)} \prod_{j=1}^{m} O\left(\frac{n\varepsilon^2}{m}\right) \leq e^{n(1-\varepsilon^2\theta) - \Omega(m)}$$

Since $m \geq C \log n$ for sufficiently large constant C, the claim follows. ■

Note that this implies that for the integrals defined in (45) that

$$I_2 + I_3 = O\left(\frac{e^n}{n}\right) \tag{46}$$

Now, we will compute I_1. Define $G(\psi) := \log(g(\psi))$. Then, by definition of g in (44),

$$G(\psi) = -in\psi + \sum_{j=1}^{m} \log\left\{ \sum_{k=0}^{\infty} \frac{(\lambda_0\nu_j e^{i\psi})^k}{k!} \exp(\varepsilon^2\theta\xi(k)) \right\} \tag{47}$$

Note that

$$\text{Im}(G(0)) = 0 \tag{48}$$

Then, applying Lemma 51,

$$\text{Re}(G'(0)) = 0 \tag{49}$$

Computing the asymptotic expansion of $G''(\psi)$ by Lemma 41, we have

$$G''(\psi) = -ne^{i\psi} + O\left(\frac{n^2\varepsilon^2}{m}\right) + o(1) \tag{50}$$

Now, by Taylor's theorem, for any $\psi \in [-\pi/3, \pi/3]$ there exists $\tilde{\psi} \in (0, \psi)$ such that

$$G(\psi) = G(0) + G'(0)\psi + \frac{G''(\tilde{\psi})}{2}\psi^2 \tag{51}$$

But, by (50), $Re[G''(\psi)] \leq -0.4n$ for any $\psi \in [-\pi/3, \pi/3]$. So, for $\psi \in [-\pi/3, \pi/3]$,

$$Re(G(\psi)) \leq G(0) - 0.2n\psi^2 \tag{52}$$

Now, we have the following upper bound on I_1.

Lemma 29

$$I_1 \leq e^{G(0)} \frac{\sqrt{\pi}}{\sqrt{0.2n}}$$

Proof

$$I_1 = Re\left[\int_{-\pi/3}^{\pi/3} e^{G(\psi)} d\psi\right] \leq \int_{-\pi/3}^{\pi/3} e^{Re(G(\psi))} d\psi \leq e^{G(0)} \int_{-\pi/3}^{\pi/3} e^{-0.2n\psi^2} d\psi \qquad (53)$$

∎

$$\leq e^{G(0)} \int_{-\infty}^{\infty} e^{-0.2n\psi^2} d\psi = e^{G(0)} \frac{\sqrt{\pi}}{\sqrt{0.2n}}$$

The next lemma shows that I_1 is also lower bounded by the above quantity (up to constants).

Lemma 30

$$I_1 \geq e^{G(0)} \frac{0.5\sqrt{\pi}}{\sqrt{1.1n}} (1 + o(1))$$

where I_1 is defined in (45)

Proof By (50), $\text{Im}(G''(\psi)) = -n\sin(\psi) + O\left(\frac{n^2 \varepsilon^2}{m}\right)$. So, for large enough n, since $|\sin(\psi)| \leq |\psi|$, for any $\psi \in [-\pi/3, \pi/3]$, $|\text{Im}(G''(\psi))| \leq 1.1n|\psi|$. So, by (51), (48) and (49), we have that for constant $c > 0$ and $\psi \in [-\pi/3, \pi/3]$,

$$|\text{Im}(G(\psi))| \leq 1.1n|\psi|^3 + cn\varepsilon^2\psi^2$$

Also, $\text{Re}(G''(\psi)) \geq -1.1n$ by a similar argument. So, by (51) and (49), for $\psi \in [-\pi/3, \pi/3]$,

$$\text{Re}(G(\psi)) \geq G(0) - 1.1n\psi^2$$

Now, for $t_n = 0.1 \min\{n^{-1/3}, \frac{1}{\varepsilon\sqrt{cn}}\}$, we have that for $\psi \in [-t_n, t_n]$, $\cos(\text{Im}(G(\psi))) \geq 0.5$ so that $\text{Re}(e^{G(\psi)}) \geq 0.5 e^{\text{Re}(G(\psi))}$. We can split I_1 further into 3 parts:

$$I_1 = \text{Re}\left[\int_{-\pi/3}^{-t_n} e^{G(\psi)} d\psi\right] + \text{Re}\left[\int_{t_n}^{\pi/3} e^{G(\psi)} d\psi\right] + \text{Re}\left[\int_{-t_n}^{t_n} e^{G(\psi)} d\psi\right]$$

Now, by (52),

$$\left|\int_{-\pi/3}^{-t_n} e^{G(\psi)} d\psi\right| \leq e^{G(0)} \int_{-\infty}^{-t_n} e^{-0.2n\psi^2} d\psi = t_n e^{G(0)} \int_{-\infty}^{-1} e^{-0.2nt_n^2\bar{\psi}^2} d\bar{\psi}$$

$$\leq t_n e^{G(0)} \int_{-\infty}^{-1} e^{-0.2nt_n^2|\bar{\psi}|} d\bar{\psi} = e^{G(0)} O\left(\frac{1}{nt_n}\right) = e^{G(0)} o\left(\frac{1}{\sqrt{n}}\right)$$

In a similar way, we can bound the second term. For the third term, we have

$$\text{Re}\left[\int_{-t_n}^{t_n} e^{G(\psi)} d\psi\right] \geq \int_{-t_n}^{t_n} 0.5 e^{\text{Re}(G(\psi))} d\psi \geq 0.5 e^{G(0)} \int_{-t_n}^{t_n} e^{-1.1n\psi^2} d\psi$$

$$\geq 0.5 e^{G(0)} \left[\int_{-\infty}^{\infty} e^{-1.1n\psi^2} d\psi - 2 \int_{-\infty}^{-t_n} e^{-1.1n\psi^2} d\psi\right]$$

$$\geq 0.5 e^{G(0)} \left(\frac{\sqrt{\pi}}{\sqrt{1.1n}} + O\left(\frac{1}{nt_n}\right) \right) = 0.5 e^{G(0)} \frac{\sqrt{\pi}}{\sqrt{1.1n}} (1 + o(1))$$

Combining the bounds, we get that

$$I_1 \geq e^{G(0)} \frac{0.5\sqrt{\pi}}{\sqrt{1.1n}} (1 + o(1))$$

∎

Combining the upper bound on I_1 from Lemma 29 and the lower bound from Lemma 30, we have

$$I_1 = e^{G(0)} \frac{1}{\sqrt{n}} e^{O(1)}$$

So, by (43) and (46),

$$\exp(-\varepsilon^2 \theta n) \, \mathbb{E}_\nu \left[\xi(Y_j^n) \right] = e^{-\varepsilon^2 \theta n} \frac{n!}{2\pi} \lambda_0^{-n} e^{G(0)} \frac{\sqrt{\pi}}{\sqrt{0.2n}} (1 + o(1)) \tag{54}$$

So, it remains to compute $G(0)$.

Lemma 31 *Under Assumption 3,*

$$
\begin{aligned}
G(0) = \lambda_0 + \sum_{j=1}^{m} \Bigg\{ & \varepsilon^2 \theta \left[(\lambda_0 \nu_j)^2 + \lambda_0 \nu_j - 2\lambda_0 \nu_j \frac{n}{m} + \frac{n^2}{m^2} \right] \\
& + \frac{\varepsilon^4 \theta^2}{2} \left[4(\lambda_0 \nu_j)^3 + 6(\lambda_0 \nu_j)^2 + (\lambda_0 \nu_j) - 8(\lambda_0 \nu_j)^2 \frac{n}{m} - 4\frac{n}{m}(\lambda_0 \nu_j) + 4\frac{n^2}{m^2}(\lambda_0 \nu_j) \right] \\
& \hspace{8cm} + o\left(\Delta^2\right) \Bigg\}
\end{aligned}
\tag{55}
$$

Proof Using equation (47), we have

$$G(0) = \lambda_0 + \sum_{j=1}^{m} \log \left\{ \mathbb{E} \left[\exp(\varepsilon^2 \theta \xi(Z_j)) \right] \right\}$$

where $Z_j \sim Poi(\lambda_0 \nu_j)$. By Assumption 3, we have that this is

$$\lambda_0 + \sum_{j=1}^{m} \log \left\{ \mathbb{E}\left[f(Z_j) \right] + o\left(\Delta^2\right) \right\}$$

Using the definition of f from (37), this is

$$\lambda_0 + \sum_{j=1}^{m} \log \left\{ 1 + \varepsilon^2 \theta \, \mathbb{E}[\phi(Z_j)^2] + \frac{\varepsilon^4 \theta^2}{2} \mathbb{E}[\phi(Z_j)^4] + o\left(\Delta^2\right) \right\}$$

Since by Lemma 38, $\varepsilon^2\theta\,\mathbb{E}[\phi(Z_j)^2]$ and $\varepsilon^4\frac{\theta^2}{2}\,\mathbb{E}[\phi(Z_j)^4]$ are $o(1)$, using the fact that $\log(1+x) = x - x^2/2 + O(x^3)$ for $x \to 0$, we have

$$G(0) = \lambda_0 + \sum_{j=1}^{m}\left\{\varepsilon^2\theta\,\mathbb{E}[\phi(Z_j)^2] + \frac{\varepsilon^4\theta^2}{2}\left(\mathbb{E}[\phi(Z_j)^4] - \mathbb{E}[\phi(Z_j)^2]^2\right) + o\left(\Delta^2\right)\right\}$$

Now, using Lemma 48, we have that

$$G(0) = \lambda_0 + \sum_{j=1}^{m}\left\{\varepsilon^2\theta\left[(\lambda_0\nu_j)^2 + \lambda_0\nu_j - 2\lambda_0\nu_j\frac{n}{m} + \frac{n^2}{m^2}\right]\right.$$
$$+ \frac{\varepsilon^4\theta^2}{2}\left[4(\lambda_0\nu_j)^3 + 6(\lambda_0\nu_j)^2 + (\lambda_0\nu_j) - 8(\lambda_0\nu_j)^2\frac{n}{m} - 4\frac{n}{m}(\lambda_0\nu_j) + 4\frac{n^2}{m^2}(\lambda_0\nu_j)\right]$$
$$\left. + o\left(\Delta^2\right)\right\}$$

∎

Lemma 32 *For the uniform distribution so that $\nu_j = 1/m$ for all j, and $\lambda_0 = n(1 - \varepsilon^2\theta)$,*

$$G(0) = \lambda_0 + \varepsilon^2\theta n + \varepsilon^4\theta^2\left(-\frac{n}{2} + \frac{n^2}{m}\right) + o\left(\frac{n^2\varepsilon^4}{m}\right)$$

Proof Substituting $\nu_j = 1/m$ for all j, and $\lambda_0 = n(1 - \varepsilon^2\theta)$, we have

$$\sum_{j=1}^{m}\left[(\lambda_0\nu_j)^2 + \lambda_0\nu_j - 2\lambda_0\nu_j\frac{n}{m} + \frac{n^2}{m^2}\right] = n - n\varepsilon^2\theta + \frac{n^2\varepsilon^4}{m}\theta^2$$

$$\sum_{j=1}^{m}\left[4(\lambda_0\nu_j)^3 + 6(\lambda_0\nu_j)^2 + (\lambda_0\nu_j) - 8(\lambda_0\nu_j)^2\frac{n}{m} - 4\frac{n}{m}(\lambda_0\nu_j) + 4\frac{n^2}{m^2}(\lambda_0\nu_j)\right]$$
$$= n + 2\frac{n^2}{m} + O\left(\frac{n^2\varepsilon^2}{m}\right)$$

So, by Lemma 31, we have

$$G(0) = \lambda_0 + \varepsilon^2\theta n + \varepsilon^4\theta^2\left(\frac{-n}{2} + \frac{n^2}{m}\right) + o\left(\frac{n^2\varepsilon^4}{m}\right)$$

∎

Lemma 33 *For alternate distributions such that $\nu_j = \frac{1}{m} + \frac{\varepsilon}{\gamma m}$ for $j \le \gamma m$, and $\nu_j = \frac{1}{m} - \frac{\varepsilon}{(1-\gamma)m}$ for $j > \gamma$, for $\gamma = \Theta(1), 1 - \gamma = \Theta(1)$, and $\lambda_0 = n(1 - \varepsilon^2\theta)$, we have*

$$G(0) = \lambda_0 + \varepsilon^2\theta n + \varepsilon^4\theta^2\left(\frac{n}{2} + \frac{n^2}{m}\right) + \varepsilon^4\theta\frac{n(\gamma m^2\theta - \gamma^2 m^2\theta - mn)}{\gamma(\gamma - 1)m^2} + o\left(\frac{n^2\varepsilon^4}{m}\right)$$

Proof We have

$$\sum_{j=1}^{m}\left[(\lambda_0\nu_j)^2 + \lambda_0\nu_j - 2\lambda_0\nu_j\frac{n}{m} + \frac{n^2}{m^2}\right] = n - \frac{n(mn + \gamma^2 m^2\theta - \gamma m^2\theta)\varepsilon^2}{\gamma(\gamma-1)m^2} + o\left(\frac{n^2\varepsilon^2}{m}\right)$$

$$\sum_{j=1}^{m}\left[4(\lambda_0\nu_j)^3 + 6(\lambda_0\nu_j)^2 + (\lambda_0\nu_j) - 8(\lambda_0\nu_j)^2\frac{n}{m} - 4\frac{n}{m}(\lambda_0\nu_j) + 4\frac{n^2}{m^2}(\lambda_0\nu_j)\right]$$
$$= n + 2\frac{n^2}{m} + O\left(\frac{n^2\varepsilon^2}{m}\right)$$

Thus, for this ν, from Lemma 31,

$$G(0) = \lambda_0 + \varepsilon^2\theta n + \varepsilon^4\theta^2\left(\frac{n}{2} + \frac{n^2}{m}\right) + \varepsilon^4\theta\frac{n(\gamma m^2\theta - \gamma^2 m^2\theta - mn)}{\gamma(\gamma-1)m^2} + o\left(\frac{n^2\varepsilon^4}{m}\right)$$

$\blacksquare$

Finally, we prove the main lemma.

Proof [Proof of Lemma 25] By (54), substituting $\lambda_0 = n(1 - \varepsilon^2\theta)$ and $G(0)$ from Lemma 32, we have for uniform ν,

$$\exp(-\varepsilon^2\theta n)\mathop{\mathbb{E}}_{\nu}\left[\xi(Y_j^n)\right] = e^{-\varepsilon^2\theta n}\frac{e^n n!}{\sqrt{2\pi n}}(n(1 - \varepsilon^2\theta))^{-n}\exp\left\{-n\varepsilon^2\theta + n\varepsilon^2\theta\right.$$
$$\left. + \varepsilon^4\theta^2\left(-\frac{n}{2} + \frac{n^2}{m}\right) + o\left(\frac{n^2\varepsilon^4}{m}\right)\right\}$$

$$= e^{-\varepsilon^2\theta n}\frac{e^n n!}{n^n\sqrt{2\pi n}}\exp\left\{-n\left(-\varepsilon^2\theta - \frac{\varepsilon^4\theta^2}{2}\right)\right\}\exp\left\{\varepsilon^4\theta^2\left(-\frac{n}{2} + \frac{n^2}{m}\right) + o\left(\frac{n^2\varepsilon^4}{m}\right)\right\}$$

$$= \frac{e^n n!}{n^n\sqrt{2\pi n}}\exp\left\{\varepsilon^4\theta^2\left(\frac{n^2}{m}\right) + o\left(\frac{n^2\varepsilon^4}{m}\right)\right\}$$

$$= (1 + O(1/n))\exp\left\{\frac{n^2\varepsilon^4}{m}\left(\theta^2 + o(1)\right)\right\}$$

By (54), substituting $\lambda_0 = n(1 - \varepsilon^2\theta)$ and $G(0)$ from Lemma 33, for ν such that $\nu_j = \frac{1}{m} + \frac{\varepsilon}{\gamma m}$ for $j \leq \gamma m$ and $\nu_j = \frac{1}{m} - \frac{\varepsilon}{(1-\gamma)m}$ for $j > (1-\gamma)m$, with $\gamma = \Theta(1), (1-\gamma) = \Theta(1)$, we have

$$\exp(-\varepsilon^2\theta n)\mathop{\mathbb{E}}_{\nu}[\xi(Y_j^n)] = (1 + O(1/n))\exp\left\{\frac{n^2\varepsilon^4}{m}\left(\theta^2 + \theta\frac{1}{\gamma(1-\gamma)} + o(1)\right)\right\}$$

$\blacksquare$

Finally, this gives us the MGF of the Huber statistic.

Lemma 34 *We have that for uniform ν such that $\nu_j = 1/m$ for all j,*

$$\mathbb{E}\left[\exp\left(\frac{n^2\varepsilon^4}{m}\theta\tilde{S}\right)\right] = \exp(-\varepsilon^2\theta n)\,\mathbb{E}\left[\exp\left(\varepsilon^2\theta\sum_{j=1}^{m}h_\beta\left(Y_j^n - \frac{n}{m}\right)\right)\right]$$

$$= (1 + O(1/n))\exp\left\{\frac{n^2\varepsilon^4}{m}(\theta^2 + o(1))\right\}$$

and for alternate ν such that $\nu_j = \frac{1}{m} + \frac{\varepsilon}{\gamma m}$ for $j \le \gamma m$ and $\nu_j = \frac{1}{m} - \frac{\varepsilon}{(1-\gamma)m}$ for $j > (1-\gamma)m$, for $\gamma = \Theta(1), 1 - \gamma = \Theta(1)$,

$$\mathbb{E}\left[\exp\left(\frac{n^2\varepsilon^4}{m}\theta\tilde{S}\right)\right] = \exp(-\varepsilon^2\theta n)\,\mathbb{E}\left[\exp\left(\varepsilon^2\theta\sum_{j=1}^{m}h_\beta\left(Y_j^n - \frac{n}{m}\right)\right)\right]$$

$$= (1 + O(1/n))\exp\left\{\frac{n^2\varepsilon^4}{m}\left(\theta^2 + \theta\frac{1}{\gamma(1-\gamma)} + o(1)\right)\right\}$$

Proof Note that Assumption 3 holds for $A_\lambda(\theta)$ as defined in (36) due to Lemma 22. So, by Lemma 25, the claim holds. ∎

C.4. Application of the Gärtner-Ellis Theorem

In this section, we apply the Gärtner-Ellis Theorem to obtain the probability that our statistic crosses a threshold, under the uniform distribution, and under one of the worst-case ε-far distributions.

Lemma 35 *Under the uniform distribution p, we have that for $\tau > 0$,*

$$\lim_{n\to\infty} -\frac{m}{n^2\varepsilon^4}\log\left(\mathbb{P}_p\left[\tilde{S} \ge \tau\right]\right) = \frac{\tau^2}{4}$$

Under an ε-far distribution q of the form $q_j = \frac{1}{m} + \frac{\varepsilon}{\gamma m}$ for $j \le l$ and $q_j = \frac{1}{m} - \frac{\varepsilon}{(1-\gamma)m}$ for $j > l$, and $\gamma = \Theta(1), 1 - \gamma = \Theta(1)$, for $\tau < \frac{1}{\gamma(1-\gamma)}$,

$$\lim_{n\to\infty} -\frac{m}{n^2\varepsilon^4}\log\left(\mathbb{P}_q\left[\tilde{S} \le \tau\right]\right) = \frac{(\tau\gamma(\gamma-1)+1)^2}{4\gamma^2(\gamma-1)^2}$$

Proof Note that by Lemma 34, the limiting logarithmic moment generating function with respect to the uniform distribution p is given by

$$\Lambda_p(\theta) = \lim_{n\to\infty} \frac{m}{n^2\varepsilon^4}\log\left(\mathbb{E}_p\left[\exp\left(\frac{n^2\varepsilon^4}{m}\theta\tilde{S}\right)\right]\right) = \theta^2$$

Thus, Assumption 1 holds for $\mathcal{D}_{\Lambda_p} = \mathbb{R}$. Furthermore, the Fenchel-Legendre Transform (defined in equation 28) of Λ_p is given by

$$\Lambda_p^*(\tau) = \sup_\theta\{\theta\tau - \theta^2\} = \frac{\tau^2}{4}$$

41

This is a strongly convex function of τ, so the set of exposed points of Λ_p^* whose exposing hyperplane belongs to $\mathcal{D}_{\Lambda_p}^o$ is all of $\mathbb{R}$. Thus, by the Theorem 13 (Gärtner-Ellis), for $\tau > 0$,

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_p \left[\tilde{S} \geq \tau \right] \right) = \inf_{x \geq \tau} \Lambda_p^*(x) = \frac{\tau^2}{4}$$

Similarly, the limiting logarithmic moment generating function with respect to an alternate distribution q is given by

$$\Lambda_q(\theta) = \lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{E}_q \left[\exp \left(\frac{n^2 \varepsilon^4}{m} \theta \tilde{S} \right) \right] \right) = \theta^2 + \frac{1}{\gamma(1 - \gamma)} \theta$$

The Fenchel-Legendre transform is given by

$$\Lambda_q^*(\tau) = \sup_\theta \left\{ \theta \tau - \theta^2 - \frac{1}{\gamma(1 - \gamma)} \theta \right\} = \frac{(\tau \gamma (\gamma - 1) + 1)^2}{4 \gamma^2 (\gamma - 1)^2}$$

Again, applying the Gärtner-Ellis Theorem gives, for $\tau < \frac{1}{\gamma(1-\gamma)}$,

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_q \left[\tilde{S}_n^* \leq \tau \right] \right) = \inf_{x \leq \tau} \Lambda_q^*(x) = \frac{(\tau \gamma (\gamma - 1) + 1)^2}{4 \gamma^2 (\gamma - 1)^2}$$

$\blacksquare$

C.5. Setting the threshold

We need to set our threshold τ so that the minimum of the error probability under the uniform distribution p, and any ε-far distribution q is maximized. Note that by Lemma 35, it is sufficient to consider a threshold τ such that $0 < \tau < \frac{1}{\gamma(1-\gamma)}$, since otherwise, the error probability in one of the two cases is at least constant. To set our threshold, we will first observe that for any τ in this range, the "error exponent" under ε-far distributions is minimized for a particular ε-far distribution. Then, we will set the threshold to maximize the minimum of the error exponent under the uniform distribution, and under this ε-far distribution.

Lemma 36 *Setting the threshold $\tau = 2$, we have for the uniform distribution p,*

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_p \left[\tilde{S} \geq \tau \right] \right) = 1$$

and for any ε-far distribution q such that $q_j = \frac{1}{m} + \frac{\varepsilon}{\gamma m}$ for $j \leq \gamma m$ and $q_j = \frac{1}{m} - \frac{\varepsilon}{(1-\gamma)m}$ for $j > \gamma m$ and $\gamma = \Theta(1), 1 - \gamma = \Theta(1)$,

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_q \left[\tilde{S} \leq \tau \right] \right) \geq 1$$

with equality for q such that $q_j = \frac{1}{m} + \frac{2\varepsilon}{m}$ for $j \leq m/2$ and $q_j = \frac{1}{m} - \frac{2\varepsilon}{m}$ for $j > m/2$.

Proof By Lemma 35, for $0 < \tau < \frac{1}{\gamma(1-\gamma)}$,

$$\lim_{n \to \infty} \frac{m}{n^2 \varepsilon^4} \log \left(\underset{q}{\mathbb{P}} \left[\tilde{S} \leq \tau \right] \right) = \frac{(\tau \gamma (\gamma - 1) + 1)^2}{4 \gamma^2 (\gamma - 1)^2}$$

Now, the numerator of the right hand side is minimized when $\gamma = 1/2$, and the denominator is maximized when $\gamma = 1/2$. Thus, the right hand side is minimized when $\gamma = 1/2$. So, we have that,

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\underset{q}{\mathbb{P}} \left[\tilde{S} \leq \tau \right] \right) \geq \frac{1}{4} (\tau - 4)^2$$

with equality for distribution q such that $q_j = 1/m + 2\varepsilon/m$ for $j \leq m/2$ and $q_j = 1/m - 2\varepsilon/m$ for $j > m/2$.

Then, the claim follows by substituting in $\tau = 2$ in the expression for the uniform distribution in Lemma 35 and in the above expression. ∎

Recall that our target sample complexity is

$$n = (1 + o(1)) \sqrt{m \log \frac{1}{\delta}} / \varepsilon^2 \tag{4}$$

We have our result for the Huber tester.

Theorem 37 (Huber with $n/m \gtrsim 1$) *The Huber statistic for appropriate β achieves (4) for $1 \lesssim n/m \ll 1/\varepsilon^2$, $\varepsilon, \delta \ll 1$ and $m \geq C \log n$ for sufficiently large constant C.*

Proof First we need to show that for every (n, m, ε) that satisfy our conditions, there is a β, Δ that satisfies (12), (11) and (13). We will set

$$\Delta = \frac{n \varepsilon^2}{m}$$

so that (12) is satisfied. Then, observe that (11) and (13) can be satisfied as long as

$$\log \left(\frac{1}{\Delta} \right) + \sqrt{\frac{n}{m} \log \left(\frac{1}{\Delta} \right)} = o \left(\frac{\Delta^{1/3}}{\varepsilon} \right)$$

Now, since $n = \Omega(m)$,

$$\log \left(\frac{1}{\Delta} \right) = \log \left(\frac{m}{n \varepsilon^2} \right) = o \left(\log \left(\frac{1}{\varepsilon} \right) \right) = o \left(\frac{1}{\varepsilon^{1/3}} \right) = o \left(\frac{n^{1/3}}{m^{1/3} \varepsilon^{1/3}} \right) = o \left(\frac{\Delta^{1/3}}{\varepsilon} \right)$$

and since $n = o \left(\frac{m}{\varepsilon^2 \log^3 \left(\frac{m}{n \varepsilon^2} \right)} \right)$,

$$\sqrt{\frac{n}{m} \log \left(\frac{1}{\Delta} \right)} = \sqrt{\frac{n}{m} \log \left(\frac{m}{n \varepsilon^2} \right)} = o \left(\frac{n^{1/3}}{m^{1/3} \varepsilon^{1/3}} \right) = o \left(\frac{\Delta^{1/3}}{\varepsilon} \right)$$

43

By Lemma 36, we have that $\bar{c}(\varepsilon, m, C) = 1$ for every ε that satisfies our assumptions, and every $C > 2$. In particular, any ε' such that $\left(1 - \frac{1}{C'}\right) \varepsilon(n) \le \varepsilon'(n) \le \varepsilon(n)$ has $\bar{c}(\varepsilon', m, C) = 1$ for every $C > 2$. Thus, by Lemma 21, we have that $c(\varepsilon, m) = 1$ for every ε that satisfies our assumptions. The claim follows. ∎

Since the Huber statistic for $\beta = 0$ is equivalent to the TV statistic, Theorem 37 and Theorem 5 together give us the main result.

Theorem 2 (Huber) *The Huber statistic for appropriate β achieves* (4) *for $n/m \ll 1/\varepsilon^2$, $\varepsilon, \delta \ll 1$, and $m \ge C \log n$ for sufficiently large constant C. It achieves* (5) *under the same conditions and $\delta_-, \delta_+ \ll 1$.*

Appendix D. MGF computation Lemmas

D.1. Huber Lemmas

Lemma 38 *Suppose Assumption 2 holds. For $Z_j \sim Poi(\lambda_0 \nu_j)$ and $\lambda_0 = n(1 - \varepsilon^2\theta)$, $\nu_j = \frac{1}{m} + O\left(\frac{\varepsilon}{m}\right)$, and for $n = \Omega(m)$*

$$\varepsilon^2\theta \, \mathbb{E}[\phi(Z_j + c)^2] = O\left(\frac{n\varepsilon^2}{m}\right)$$

$$\frac{\varepsilon^4\theta^2}{2} \, \mathbb{E}[\phi(Z_j + c)^4] = O\left(\left(\frac{n\varepsilon^2}{m}\right)^2\right)$$

for any $0 \le c \le 4$

Proof

Note that $\lambda_0 \nu_j = \frac{n}{m} + O(\frac{n\varepsilon}{m})$. Then, using Lemma 48,

$$\varepsilon^2\theta \, \mathbb{E}[\phi(Z_j)^2] = \varepsilon^2\theta \left(\frac{n}{m} + O\left(\frac{n\varepsilon}{m}\right) + O\left(\frac{n^2\varepsilon^2}{m^2}\right)\right) = O\left(\frac{n\varepsilon^2}{m}\right)$$

since $n = o(m/\varepsilon^2)$.

Similarly,

$$\varepsilon^4 \frac{\theta^2}{2} \, \mathbb{E}[\phi(Z_j)^4] = \varepsilon^4 \frac{\theta^2}{2} \left[\frac{n}{m} + O\left(\frac{n^2}{m^2}\right) + O\left(\frac{n^3\varepsilon^2}{m^3}\right) + O\left(\frac{n^4\varepsilon^4}{m^4}\right)\right] = O\left(\left(\frac{n\varepsilon^2}{m}\right)^2\right)$$

Now,

$$\varepsilon^2\theta \, \mathbb{E}[\phi(Z_j + c)^2] = \varepsilon^2\theta \, \mathbb{E}\left[\left(Z_j + c - \frac{n}{m}\right)^2\right] = \varepsilon^2\theta \, \mathbb{E}\left[\phi(Z_j)^2 + 2c\left(Z_j - \frac{n}{m}\right) + c^2\right]$$

$$= O\left(\frac{n\varepsilon^2}{m}\right) + O(\varepsilon^2) = O\left(\frac{n\varepsilon^2}{m}\right)$$

since $\varepsilon^2 = O\left(\frac{n\varepsilon^2}{m}\right)$.

Similarly,

$$\frac{\varepsilon^4\theta^2}{2} \, \mathbb{E}[\phi(Z_j + c)^4] = \frac{\varepsilon^4\theta^2}{2} \, \mathbb{E}\left[\phi(Z_j)^4 + 4c\left(Z_j - \frac{n}{m}\right)^3 + 6c^2\phi(Z_j)^2 + 4c^3\left(Z_j - \frac{n}{m}\right) + c^4\right]$$

Now,

$$\mathbb{E}\left[4c\left(Z_j - \frac{n}{m}\right)^3\right] = 4\,\mathbb{E}\left[Z_j^3 - 3Z_j^2\frac{n}{m} + 3Z_j\frac{n^2}{m^2} - \frac{n^3}{m^3}\right]$$

By Lemma 50, this is

$$4\left[(\lambda_0\nu_j)^3 + 3(\lambda_0\nu_j)^2 + \lambda_0\nu_j - 3((\lambda_0\nu_j)^2 + \lambda_0\nu_j)\frac{n}{m} + 3\lambda_0\nu_j\frac{n^2}{m^2} - \frac{n^3}{m^3}\right]$$

45

Now, since $\lambda_0 \nu_j = \frac{n}{m} + O\left(\frac{n\varepsilon}{m}\right)$, this is

$$\frac{n}{m} + O\left(\frac{n^2\varepsilon}{m^2}\right) + O\left(\frac{n^3\varepsilon^3}{m^3}\right)$$

so that

$$\frac{\varepsilon^4\theta^2}{2} \mathbb{E}\left[4c\left(Z_j - \frac{n}{m}\right)^3\right] = O\left(\left(\frac{n\varepsilon^2}{m}\right)^2\right) = O\left(\left(\frac{n\varepsilon^2}{m}\right)^2\right)$$

So, finally

$$\frac{\varepsilon^4\theta^2}{2} \mathbb{E}[\phi(Z_j + c)^4] = O\left(\left(\frac{n\varepsilon^2}{m}\right)^2\right)$$

as required. ∎

Lemma 39 *Suppose Assumption 2 holds. For $Z_j \sim Poi(\lambda_0\nu_j)$, and $\lambda_0 = n(1 - \varepsilon^2\theta)$, $\nu_j = \frac{1}{m} + O(\frac{\varepsilon}{m})$ for all j,*

$$\prod_{j=1}^{m} \left\{\mathbb{E}\left[f(Z_j)\right] + o\left(\Delta^2\right)\right\} = \exp\{O(n\varepsilon^2)\}$$

where f is defined in (37)

Proof

$$\prod_{j=1}^{m} \left\{\mathbb{E}\left[f(Z_j)\right] + o\left(\Delta^2\right)\right\} = \exp\left\{\sum_{j=1}^{m} \log\left[\mathbb{E}\left[f(Z_j)\right] + o\left(\Delta^2\right)\right]\right\}$$

Note that due to Lemma 38, $\Delta = O\left(\frac{n\varepsilon^2}{m}\right)$, and the fact that $n = o(m/\varepsilon^2)$, the above is

$$\exp\left\{\sum_{j=1}^{m} \log\left[1 + O\left(\frac{n\varepsilon^2}{m}\right)\right]\right\}$$

We can Taylor expand the log since it is of form $\log(1 + o(1))$. The above is then

$$\exp\left\{\sum_{j=1}^{m} \left[O\left(\frac{n\varepsilon^2}{m}\right)\right]\right\} = \exp\{O(n\varepsilon^2)\}$$

Setting $\eta \geq 2$, this is

$$\exp(O(n\varepsilon^2))$$

∎

46

Lemma 40 *Suppose Assumption 2 holds. For $\lambda_0 = n(1 - \varepsilon^2\theta)$ and $\nu_j = \frac{1}{m} + O\left(\frac{\varepsilon}{m}\right)$ for all j, and for ξ that satisfies Assumption 3, where $Z_j \sim Poi(\lambda_0\nu_j)$, we have*

$$
e^{-2\lambda_0\nu_j} \left| \left(\sum_{k=0}^{\infty} \frac{(\lambda_0\nu_j e^{i\psi})^k}{k!} \exp(\varepsilon^2\theta\xi(k+1)) \right)^2 \right.
$$
$$
\left. - \left(\sum_{k=0}^{\infty} \frac{(\lambda_0\nu_j e^{i\psi})^k}{k!} \exp(\varepsilon^2\theta\xi(k)) \right) \left(\sum_{k=0}^{\infty} \frac{(\lambda_0\nu_j e^{i\psi})^k}{k!} \exp(\varepsilon^2\theta\xi(k+2)) \right) \right| = O(\varepsilon^2) + O\left(\left(\frac{n\varepsilon^2}{m} \right)^2 \right)
$$

and

$$
e^{-2\lambda_0\nu_j} \left| \left(\sum_{k=0}^{\infty} \frac{(\lambda_0\nu_j e^{i\psi})^k}{k!} \exp(\varepsilon^2\theta\xi(k)) \right) \left(\sum_{k=0}^{\infty} \frac{(\lambda_0\nu_j e^{i\psi})^k}{k!} \exp(\varepsilon^2\theta\xi(k+1)) \right) \right| = 1 + O\left(\frac{n\varepsilon^2}{m} \right)
$$

Proof Notation: For simplicity, let $\tilde{\mathbb{E}}[f(W_j)] = \sum_{k=0}^{\infty} \frac{(\lambda_0\nu_j e^{i\psi})^k}{k!} f(k)$ Note that

$$
\tilde{\mathbb{E}}[f(W_j) + g(W_j)] = \tilde{\mathbb{E}}[f(W_j)] + \tilde{\mathbb{E}}[g(W_j)]
$$

Also, note that $\left| e^{-\lambda_0\nu_j} \tilde{\mathbb{E}}[f(W_j)] \right| = \mathbb{E}[f(Z_j)]$ where $Z_j \sim Poi(\lambda_0\nu_j)$. We have that,

$$
\tilde{\mathbb{E}}[\exp(\varepsilon^2\theta\xi(W_j + 1))]^2 = \tilde{\mathbb{E}} \left[1 + \varepsilon^2\theta\xi(W_j + 1) + \sum_{l=2}^{\infty} \frac{(\varepsilon^2\theta\xi(W_j + 1))^l}{l!} \right]^2
$$
$$
= \tilde{\mathbb{E}} \left[1 + \varepsilon^2\theta\xi(W_j + 1) \right]^2 + 2\tilde{\mathbb{E}} \left[1 + \varepsilon^2\theta\xi(W_j + 1) \right] \tilde{\mathbb{E}} \left[\sum_{l=2}^{\infty} \frac{(\varepsilon^2\theta\xi(W_j + 1))^l}{l!} \right] + \tilde{\mathbb{E}} \left[\sum_{l=2}^{\infty} \frac{(\varepsilon^2\theta\xi(W_j + 1))^l}{l!} \right]^2
$$

Similarly,

$$
\tilde{\mathbb{E}}[\exp(\varepsilon^2\theta\xi(W_j))] \tilde{\mathbb{E}}[\exp(\varepsilon^2\theta\xi(W_j + 2))] = \tilde{\mathbb{E}}[1 + \varepsilon^2\theta\xi(W_j)] \tilde{\mathbb{E}}[1 + \varepsilon^2\theta\xi(W_j + 2)]
$$
$$
+ \tilde{\mathbb{E}}[1 + \varepsilon^2\theta\xi(W_j)] \tilde{\mathbb{E}} \left[\sum_{l=2}^{\infty} \frac{(\varepsilon^2\theta\xi(W_j + 2))^l}{l!} \right]
$$
$$
+ \tilde{\mathbb{E}} \left[\sum_{l=2}^{\infty} \frac{(\varepsilon^2\theta\xi(W_j))^l}{l!} \right] \tilde{\mathbb{E}}[1 + \varepsilon^2\theta\xi(W_j + 1)] + \tilde{\mathbb{E}} \left[\sum_{l=2}^{\infty} \frac{(\varepsilon^2\theta\xi(W_j))^l}{l!} \right] \tilde{\mathbb{E}} \left[\sum_{l=2}^{\infty} \frac{(\varepsilon^2\theta\xi(W_j + 2))^l}{l!} \right]
$$

So, using the properties of ξ from Assumption 3, this first expression is

$$
e^{-2\lambda_0\nu_j} \left| \tilde{\mathbb{E}}[\exp(\varepsilon^2\theta\xi(W_j + 1))]^2 - \tilde{\mathbb{E}}[\exp(\varepsilon^2\theta\xi(W_j))] \tilde{\mathbb{E}}[\exp(\varepsilon^2\theta\xi(W_j + 2))] \right|
$$
$$
\leq e^{-2\lambda_0\nu_j} \left| \tilde{\mathbb{E}}[1 + \varepsilon^2\theta\xi(W_j + 1)]^2 - \tilde{\mathbb{E}}[1 + \varepsilon^2\theta\xi(W_j)] \tilde{\mathbb{E}}[1 + \varepsilon^2\theta\xi(W_j + 2)] \right| + o(\Delta^2)
$$
$$
= e^{-2\lambda_0\nu_j} \left| \varepsilon^2\theta\tilde{\mathbb{E}}[1] \left\{ 2\tilde{\mathbb{E}}[\xi(W_j + 1)] - \tilde{\mathbb{E}}[\xi(W_j)] - \tilde{\mathbb{E}}[\xi(W_j + 2)] \right\} \right.
$$
$$
\left. + \varepsilon^4\theta^2 \left\{ \tilde{\mathbb{E}}[\xi(W_j + 1)]^2 - \tilde{\mathbb{E}}[\xi(W_j)] \tilde{\mathbb{E}}[\xi(W_j + 2)] \right\} \right| + o(\Delta^2)
$$
$$
= \varepsilon^2\theta \left\{ 2\,\mathbb{E}[\xi(Z_j + 1)] - \mathbb{E}[\xi(Z_j)] - \mathbb{E}[\xi(Z_j + 2)] \right\} + \varepsilon^4\theta^2 \left\{ \mathbb{E}[\xi(Z_j + 1)]^2 - \mathbb{E}[\xi(Z_j)]\,\mathbb{E}[\xi(Z_j + 2)] \right\} + o(\Delta^2)
$$

Using properties of ξ, this is

$$\varepsilon^2\theta\left\{2\,\mathbb{E}[\phi(Z_j+1)^2]-\mathbb{E}[\phi(Z_j)^2]-\mathbb{E}[\phi(Z_j+2)^2]\right\}$$
$$+\varepsilon^4\theta^2\left\{\mathbb{E}[\phi(Z_j+1)^2]^2-\mathbb{E}[\phi(Z_j)^2]\,\mathbb{E}[\phi(Z_j+2)^2]\right\}+o(\Delta^2)$$

Simplifying using the definition of ϕ, and applying Lemma 38, this is

$$O(\varepsilon^2)+O\left(\left(\frac{n\varepsilon^2}{m}\right)^2\right)$$

This gives us the first claim. For the second claim, we have the expression

$$e^{-2\lambda_0\nu_j}|\tilde{\mathbb{E}}[\exp(\varepsilon^2\theta\xi(W_j))]\tilde{\mathbb{E}}[\exp(\varepsilon^2\theta\xi(W_j+1))]|=\mathbb{E}[\exp(\varepsilon^2\theta\xi(Z_j))]\,\mathbb{E}[\exp(\varepsilon^2\theta\xi(Z_j+1))]$$

By properties of ξ from Assumption 3, and using Lemma 38, this is

$$1+O\left(\frac{n\varepsilon^2}{m}\right)$$

<div style="text-align:right">∎</div>

Lemma 41 *Suppose Assumption 2 holds. For $\lambda_0=n(1-\varepsilon^2\theta)$, $\nu_j=\frac{1}{m}+O\left(\frac{\varepsilon}{m}\right)$, and ξ that satisfies Assumption 3,*
and

$$G(\psi)=-in\psi+\sum_{j=1}^m\log\left\{\sum_{k=0}^\infty\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}\xi(k)\right\}$$

we have

$$G''(\psi)=-ne^{i\psi}+O\left(\frac{n^2\varepsilon^2}{m}\right)$$

Proof By Lemmas 38, 40 and 51 we have

$$G''(\psi)=\sum_{j=1}^m\frac{(\lambda_0\nu_j e^{i\psi})^2\left\{O(\varepsilon^2)+O\left(\left(\frac{n\varepsilon^2}{m}\right)^2\right)\right\}-(\lambda_0\nu_j e^{i\psi})(1+O(\frac{n\varepsilon^2}{m}))}{(1+O(\frac{n\varepsilon^2}{m}))}$$

By Lemma 38, this is,

$$\sum_{j=1}^m\left\{O\left(\frac{n^2\varepsilon^2}{m^2}\right)-(\lambda_0\nu_j e^{i\psi})\left(1+O\left(\frac{n\varepsilon^2}{m}\right)\right)\right\}\left(1+O\left(\frac{n\varepsilon^2}{m}\right)\right)$$

$$=-(\lambda_0 e^{i\psi})\left(1+O\left(\frac{n\varepsilon^2}{m}\right)\right)+O\left(\frac{n^2\varepsilon^2}{m}\right)=-ne^{i\psi}+O\left(\frac{n^2\varepsilon^2}{m}\right)$$

<div style="text-align:right">∎</div>

Lemma 42

Suppose Assumption 2 holds. For $\lambda_0 = n(1 + O(\varepsilon^2))$ and ν such that $\nu_j = 1/m + O(\varepsilon/m)$ for all j, and $\theta = O(1)$, we have

$$\left| \mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j+c)\leq\beta\}} \left\{ \sum_{l=3}^{\infty} \frac{(\varepsilon^2\theta\phi(Z_j+c)^2)^l}{l!} \right\} \right] \right| = o\left(\Delta^2\right)$$

where $Z_j \sim Poi(\lambda_0\nu_j)$

Proof We have that

$$\left| \mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j+c)\leq\beta\}} \left\{ \sum_{l=3}^{\infty} \frac{(\varepsilon^2\theta\phi(Z_j+c)^2)^l}{l!} \right\} \right] \right| \leq \left| \mathbb{E}\left[\sum_{l=3}^{\infty} \frac{(\varepsilon^2\theta\beta^2)^l}{l!} \right] \right|$$

$$\leq \sum_{l=3}^{\infty} \frac{(\varepsilon^2\theta\beta^2)^l}{l!}$$

$$= O\left((\varepsilon^2\theta\beta^2)^3 \sum_{l=0}^{\infty} \frac{(\varepsilon^2\theta\beta^2)^l}{l!} \right)$$

Also,

$$\sum_{l=0}^{\infty} \frac{(\varepsilon^2\theta\beta^2)^l}{l!} = \exp(\varepsilon^2\theta\beta^2)$$

But, by (32),

$$\varepsilon^2\theta\beta^2 = o(1)$$

Thus, we have that

$$\exp(\varepsilon^2\theta\beta^2) = e^{o(1)} = O(1)$$

Putting the above together gives us the claim. ∎

Lemma 43 *Suppose Assumption 2 holds. For $\lambda = n(1 + O(\varepsilon^2) + O(\frac{n}{m}\varepsilon^2))$ and ν such that $\nu_j = 1/m + O(\varepsilon/m)$ for all j, for any β,*

$$\mathbb{P}[\phi(Z_j+c) > \beta] \leq 2\exp\left\{ -\frac{\Omega(\beta^2)}{O(\frac{n}{m})} \right\} + 2\exp\left\{-\Omega(\beta)\right\}$$

for integer $0 \leq c \leq 3$ and any constant $\eta > 0$.

Proof Note that for the conditions given,

$$\lambda\nu_j = \frac{n}{m} + O(\frac{n\varepsilon}{m})$$

Now, since $Z_j \sim Poi(\lambda\nu_j)$ and $\lambda = n(1 + O(\varepsilon^2) + O(\frac{n}{m}\varepsilon^2))$,

$$\mathbb{E}[\mathbb{1}_{\phi(Z_j+c)>\beta}] = \mathbb{P}[\phi(Z_j+c) > \beta] = \mathbb{P}\left[\left| Z_j + c - \frac{n}{m} \right| > \beta \right]$$

$$\leq \mathbb{P}\left[|Z_j - \lambda\nu_j| > \beta + \left|\lambda\nu_j + c - \frac{n}{m}\right|\right] \leq \mathbb{P}\left[|Z_j - \lambda\nu_j| > \beta + O(\frac{n\varepsilon}{m})\right]$$

Using Poisson concentration bounds (Canonne, 2017), this is at most

$$2\exp\left\{-\frac{\left(\beta + O(\frac{n\varepsilon}{m})\right)^2}{\lambda\nu_j + \left(\beta + O(\frac{n\varepsilon}{m})\right)}\right\} = 2\exp\left\{-\frac{\left(\beta + O(\frac{n\varepsilon}{m})\right)^2}{\frac{n}{m} + \beta + O(\frac{n\varepsilon}{m})}\right\}$$

Now, if $\beta = O(\frac{n}{m})$, this is

$$2\exp\left\{-\frac{(\beta + O(\frac{n\varepsilon}{m}))^2}{O(\frac{n}{m}) + O(\frac{n\varepsilon}{m})}\right\} = 2\exp\left\{-\frac{\Omega(\beta^2)}{O(\frac{n}{m})}\right\}$$

Similarly, if instead $\frac{n}{m} = O(\beta)$, the bound is

$$2\exp\left\{-\frac{(\beta + O(\frac{n\varepsilon}{m}))^2}{O(\beta) + O(\frac{n\varepsilon}{m})}\right\} = 2\exp\left\{-\Omega(\beta)\right\}$$

The claim follows.

∎

Lemma 44 *Suppose Assumption 2 holds. For $\lambda = n(1 + O(\varepsilon^2) + O(\frac{n}{m}\varepsilon^2))$ and ν such that $\nu_j = 1/m + O(\varepsilon/m)$ for all j, for β such that (11) is satisfied,*

$$\mathbb{E}[\mathbb{1}_{\phi(Z_j+c)>\beta}] = O\left(\Delta^{2\eta}\right)$$

for integer $0 \leq c \leq 3$ and any constant $\eta > 0$.

Proof

By Lemma 43,

$$\mathbb{E}[\mathbb{1}_{\{\phi(Z_j+c)>\beta\}}] \leq 2\exp\left\{-\frac{\Omega(\beta^2)}{O(\frac{n}{m})}\right\} + 2\exp\left\{-\Omega(\beta)\right\}$$

But by (11), this is

$$4\exp\left\{-\omega\left(\log\left(\frac{1}{\Delta}\right)\right)\right\} = O\left(\Delta^{2\eta}\right)$$

∎

Lemma 45 *Suppose Assumption 2 holds. For $\lambda = n(1 + O(\varepsilon^2))$, $\nu_j = 1/m + O(\varepsilon/m)$ for all j, we have*

$$\left|\mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j+c)>\beta\}}\varepsilon^2\theta\phi(Z_j + c)^2\right]\right| = o\left(\Delta^2\right)$$

$$\left|\mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j+c)>\beta\}}\frac{\varepsilon^4\theta^2}{2}\phi(Z_j + c)^4\right]\right| = o\left(\Delta^2\right)$$

for integer $0 < c \leq 3$ and $Z_j \sim Poi(\lambda\nu_j)$

Proof

We have

$$\left| \mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j+c)>\beta\}} \varepsilon^2\theta\phi(Z_j+c)^2 \right]\right| \leq \sum_{l=0}^{\infty} \left| \mathbb{P}\left[\phi(Z_j+c) > 2^l\beta \right] \varepsilon^2\theta(2^{l+1}\beta)^2 \right|$$

By Lemma 43, this is

$$\sum_{l=0}^{\infty} \left| \left(2\exp\left\{ -\frac{\Omega(2^{2l}\beta^2)}{O(\frac{n}{m})} \right\} + 2\exp\left\{ -\Omega(2^l\beta) \right\} \right) \varepsilon^2\theta(2^{l+1}\beta)^2 \right|$$

By (11), this is

$$\sum_{l=0}^{\infty} \left| O\left(\Delta^{2^l\eta} \right) \varepsilon^2\theta(2^{l+1}\beta)^2 \right|$$

Note that by (13), $\varepsilon^2\beta^2 = o(1)$. Thus, this is

$$\sum_{l=0}^{\infty} \left| O\left(\Delta^{2^l\eta} \right) o(1)2^{2(l+1)} \right|$$

Setting $\eta \geq 3$, since $2^{2(l+1)} \leq O(2^{2^l})$ we have that this is this is

$$O\left(\Delta^3 \right) \sum_{l=0}^{\infty} \left| o(1)^{2^l} \right| = O\left(\Delta^3 \right)$$

The first claim follows. The second claim can be proved in a similar way.

■

Lemma 46 *Suppose Assumption 2 holds. For $\lambda = n(1 + O(\varepsilon^2) + O(\frac{n}{m}\varepsilon^2))$ and ν such that $\nu_j = 1/m + O(\varepsilon/m)$ for all j,*

$$E\left[\exp(\varepsilon^2\theta\beta(2\phi(Z_j+c) - \beta)) \right] = O(1)$$

for integer $0 \leq c \leq 3$

Proof

$$\exp(\varepsilon^2\theta\beta(2\phi(k) - \beta)) = \exp\left\{ \varepsilon^2\theta\beta\left(2\left| k - \frac{n}{m} \right| - \beta \right) \right\}$$

$$= \exp\left\{ \varepsilon^2\theta\beta\left(\mathbb{1}_{\{k\leq\frac{n}{m}\}}2\left(\frac{n}{m} - k \right) + \mathbb{1}_{\{k>\frac{n}{m}\}}2\left(k - \frac{n}{m} \right) - \beta \right) \right\}$$

$$= \mathbb{1}_{\{k\leq\frac{n}{m}\}}\exp\left\{ \varepsilon^2\theta\beta\left(2\left(\frac{n}{m} - k \right) - \beta \right) \right\} + \mathbb{1}_{\{k>\frac{n}{m}\}}\exp\left\{ \left(2\theta\beta\left(2\left(k - \frac{n}{m} \right) - \beta \right) \right\}$$

Since $\exp(x) \geq 0$ for all x, this is at most

$$\exp\left\{ \varepsilon^2\theta\beta\left(2\left(\frac{n}{m} - k \right) - \beta \right) \right\} + \exp\left\{ \varepsilon^2\theta\beta\left(2\left(k - \frac{n}{m} \right) - \beta \right) \right\}$$

51

Thus,

$$\mathbb{E}\left[\exp(\varepsilon^2\theta\beta(2\phi(Z_j+c)-\beta))\right] \le \mathbb{E}\left[\exp\left\{\varepsilon^2\theta\beta\left(2\left(\frac{n}{m}-Z_j-c\right)-\beta\right)\right\}\right]$$
$$+ \mathbb{E}\left[\exp\left\{\varepsilon^2\theta\beta\left(2\left(Z_j+c-\frac{n}{m}\right)-\beta\right)\right\}\right]$$

Now,

$$E\left[\exp\left\{\varepsilon^2\theta\beta\left(2\left(\frac{n}{m}-Z_j-c\right)-\beta\right)\right\}\right] = \exp\left\{\varepsilon^2\theta\beta\left(\frac{2n}{m}-2c-\beta\right)\right\}\mathbb{E}\left[\exp\left\{-2\varepsilon^2\theta\beta Z_j\right\}\right]$$
(56)

But, since $Z_j \sim Poi(\lambda\nu_j)$, we have

$$\mathbb{E}[\exp\{-2\varepsilon^2\theta\beta Z_j\}] = \exp\left\{\lambda\nu_j\left(e^{-2\varepsilon^2\theta\beta}-1\right)\right\}$$

Note that

$$\lambda\nu_j = \frac{n}{m} + O(\frac{n\varepsilon}{m})$$

So,

$$\mathbb{E}[\exp\{-2\varepsilon^2\theta\beta Z_j\}] = \exp\left\{\frac{n}{m}(1+O(\varepsilon))\left(e^{-2\varepsilon^2\theta\beta}-1\right)\right\}$$

Using the fact that $e^x \le 1+x$ for all x, the above is at most

$$\exp\left\{\frac{n}{m}(1+O(\varepsilon))(-2\varepsilon^2\theta\beta)\right\} = \exp\left\{-2\varepsilon^2\theta\beta\frac{n}{m}+O(\frac{n}{m}\varepsilon\theta\beta)\right\}$$

Since $n = o(m/\varepsilon^2)$, and $\theta = \Theta(\varepsilon^2)$,

$$O\left(\frac{n}{m}\varepsilon\theta\beta\right) = o\left(\beta\varepsilon\right)$$

Using (32), this is $o(1)$.
Thus,

$$\mathbb{E}[\exp\{-2\varepsilon^2\theta\beta Z_j\}] = \exp\left\{-2\varepsilon^2\theta\beta\frac{n}{m}+o(1)\right\}$$

Thus, using this in (56), we have that

$$\mathbb{E}\left[\exp\left\{\varepsilon^2\theta\beta\left(2\left(\frac{n}{m}-Z_j-c\right)-\beta\right)\right\}\right] = \exp\left\{-\varepsilon^2\theta\beta^2-2c\varepsilon^2\theta\beta+o(1)\right\}$$

Equation (32) tells us that

$$\varepsilon^2\theta\beta^2 = o(1)$$

Similarly,

$$\varepsilon^2\theta\beta = o(\varepsilon)$$

So,

$$\mathbb{E}\left[\exp\left\{\varepsilon^2\theta\beta\left(2\left(\frac{n}{m}-Z_j-c\right)-\beta\right)\right\}\right] = \exp\left\{o(1)\right\} = O(1)$$

Using a very similar argument, we can show that

$$\mathbb{E}\left[\exp\left\{\varepsilon^2\theta\beta\left(2\left(Z_j+c-\frac{n}{m}\right)-\beta\right)\right\}\right] = O(1)$$

.

So, by the above and (D.1),

$$\mathbb{E}\left[\exp(\varepsilon^2\theta\beta(2\phi(Z_j+c)-\beta))\right] = O(1)$$

∎

Lemma 47 *Suppose Assumption 2 holds. For $\lambda = n(1 + O(\varepsilon^2) + O(\frac{n}{m}\varepsilon^2))$ and ν such that $\nu_j = 1/m + O(\varepsilon/m)$ for all j,*

$$\mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j+c)>\beta\}}\exp\left\{\varepsilon^2\theta\beta(2\phi(Z_j+c)-\beta)\right\}\right] = O\left(\Delta^\eta\right)$$

with integer $0 \le c \le 3$ and any constant $\eta > 0$.

Proof By the Cauchy-Schwarz inequality,

$$\mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j+c)>\beta\}}\exp\left\{\varepsilon^2\theta\beta(2\phi(Z_j+c)-\beta)\right\}\right]$$
$$\le \sqrt{\mathbb{E}[\mathbb{1}_{\{\phi(Z_j+c)>\beta\}}]\,\mathbb{E}\left[\exp\left\{\varepsilon^2\theta\beta(2\phi(Z_j+c)-\beta)\right\}\right]}$$

By Lemmas 44 and 46, this is $O\left(\Delta^\eta\right)$. ∎

D.2. General Lemmas

Lemma 48 *For $Z_j \sim Poi(\lambda\nu_j)$, and ϕ defined in (35)*

$$\mathbb{E}[\phi(Z_j)^2] = (\lambda\nu_j)^2 + \lambda\nu_j - 2\lambda\nu_j\frac{n}{m} + \frac{n^2}{m^2}$$
$$\mathbb{E}[\phi(Z_j)^4] = (\lambda\nu_j)^4 + 6(\lambda\nu_j)^3 + 7(\lambda\nu_j)^2 + \lambda\nu_j - 4\frac{n}{m}\left[(\lambda\nu_j)^3 + 3(\lambda\nu_j)^2 + \lambda\nu_j\right]$$
$$+ 6\left(\frac{n}{m}\right)^2\left[(\lambda\nu_j)^2 + \lambda\nu_j\right] - 4\left(\frac{n}{m}\right)^3\lambda\nu_j + \left(\frac{n}{m}\right)^4$$

Proof

$$\mathbb{E}[\phi(Z_j)^2] = \mathbb{E}\left[\left(Z_j - \frac{n}{m}\right)^2\right] = \mathbb{E}\left[Z_j^2 - 2Z_j\frac{n}{m} + \frac{n^2}{m^2}\right]$$
$$= (\lambda\nu_j)^2 + \lambda\nu_j - 2\lambda\nu_j\frac{n}{m} + \frac{n^2}{m^2}$$

$$\mathbb{E}[\phi(Z_j)^4] = \mathbb{E}\left[\left(Z_j - \frac{n}{m}\right)^4\right] = \mathbb{E}\left[Z_j^4 - 4\frac{n}{m}Z_j^3 + 6\left(\frac{n}{m}\right)^2 Z_j^2 - 4\left(\frac{n}{m}\right)^3 Z_j + \left(\frac{n}{m}\right)^4\right]$$
$$= (\lambda\nu_j)^4 + 6(\lambda\nu_j)^3 + 7(\lambda\nu_j)^2 + \lambda\nu_j - 4\frac{n}{m}\left[(\lambda\nu_j)^3 + 3(\lambda\nu_j)^2 + \lambda\nu_j\right]$$
$$+ 6\left(\frac{n}{m}\right)^2\left[(\lambda\nu_j)^2 + \lambda\nu_j\right] - 4\left(\frac{n}{m}\right)^3\lambda\nu_j + \left(\frac{n}{m}\right)^4$$

∎

Lemma 49 *For $Z_j \sim Poi(\lambda\nu_j)$, and ϕ defined in* (35)

$$\mathbb{E}\left[\phi(Z_j + 1)^2 - \phi(Z_j)^2\right] = 2\lambda\nu_j + 1 - \frac{2n}{m}$$

$$\mathbb{E}\left[\phi(Z_j + 1)^4 - \phi(Z_j)^4\right] = 4(\lambda\nu_j)^3 + 18(\lambda\nu_j)^2 + 14(\lambda\nu_j) + 1$$
$$- \frac{4n}{m}(3(\lambda\nu_j)^2 + 6\lambda\nu_j + 1) + \frac{6n^2}{m^2}(2\lambda\nu_j + 1) - \frac{4n^3}{m^3}$$

Proof

$$\mathbb{E}[\phi(Z_j + 1)^2 - \phi(Z_j)^2] = \mathbb{E}\left[2Z_j + 1 - \frac{2n}{m}\right] = 2\lambda\nu_j + 1 - \frac{2n}{m}$$

$$\mathbb{E}[\phi(Z_j + 1)^4 - \phi(Z_j)^4]$$
$$= \mathbb{E}\left[(4Z_j^3 + 6Z_j^2 + 4Z_j + 1) - \frac{4n}{m}(3Z_j^2 + 3Z_j + 1) + 6\frac{n^2}{m^2}(2Z_j + 1) - 4\frac{n^3}{m^3}\right]$$
$$= 4(\lambda\nu_j)^3 + 18(\lambda\nu_j)^2 + 14(\lambda\nu_j) + 1 - \frac{4n}{m}(3(\lambda\nu_j)^2 + 6\lambda\nu_j + 1) + \frac{6n^2}{m^2}(2\lambda\nu_j + 1) - \frac{4n^3}{m^3}$$

∎

Lemma 50 *The first four moments of the Poisson distribution are given by*

$$\mathbb{E}[X] = \lambda$$

$$\mathbb{E}[X^2] = \lambda^2 + \lambda$$

$$\mathbb{E}[X^3] = \lambda^3 + 3\lambda^2 + \lambda$$

$$\mathbb{E}[X^4] = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda$$

for $X \sim Poi(\lambda)$

Proof Computation of moments. ∎

Lemma 51 *For any function f and*

$$G(\psi) = -in\psi + \sum_{j=1}^{m} \log\left\{\sum_{k=0}^{\infty}\left[\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}f(k)\right]\right\}$$

we have

$$G'(\psi) = -in + i\sum_{j=1}^{m}(\lambda_0\nu_j e^{i\psi})\frac{\sum_{k=0}^{\infty}\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}f(k+1)}{\sum_{k=0}^{\infty}\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}f(k)}$$

so that $Re(G'(0)) = 0$, and

$$G''(\psi) = \sum_{j=1}^{m} \left\{ \frac{1}{\left(\sum_{k=0}^{\infty} \frac{(\lambda_0\nu_j e^{i\psi})^k}{k!} f(k)\right)^2} \left[(\lambda_0\nu_j e^{i\psi})^2 \left\{ \left(\frac{\sum_{k=0}^{\infty}(\lambda_0\nu_j e^{i\psi})^k}{k!} f(k+1)\right)^2 \right. \right. \right.$$

$$\left. - \left(\sum_{k=0}^{\infty} \frac{(\lambda_0\nu_j e^{i\psi})^k}{k!} f(k)\right) \left(\sum_{k=0}^{\infty} \frac{(\lambda_0\nu_j e^{i\psi})^k}{k!} f(k+2)\right) \right\}$$

$$\left. \left. - (\lambda_0\nu_j e^{i\psi}) \left(\sum_{k=0}^{\infty} \frac{(\lambda_0\nu_j e^{i\psi})^k}{k!} f(k)\right) \left(\sum_{k=0}^{\infty} \frac{(\lambda_0\nu_j e^{i\psi})^k}{k!} f(k+1)\right) \right] \right\}$$

Proof Follows from taking derivatives. ∎

Appendix E. Lower bounds

E.1. The collisions tester is asymptotically bad when $n = \Theta(m)$ and $\varepsilon = \omega\left(\frac{\log^{1/4} n}{n^{1/8}}\right)$

The following lower bound showing the collisions tester is asymptotically suboptimal is based on a note of Peebles (Peebles, 2015).

Theorem 52 *When $n = \Theta(m)$ and $\varepsilon = \omega\left(\frac{\log^{1/4} n}{n^{1/8}}\right)$, the collisions tester has error probability* $\exp\left(-o\left(\frac{n^2\varepsilon^4}{m}\right)\right)$*, and so, takes $\omega(\sqrt{m\log(1/\delta)}/\varepsilon^2)$ samples to distinguish between the uniform distribution and an ε-far distribution with error probability δ.*

[Due to Peebles.] When $n = \Theta(m)$ and $\varepsilon = \omega\left(\frac{\log^{1/4} n}{n^{1/8}}\right)$, the collisions tester has error probability $\exp\left(-o\left(\frac{n^2\varepsilon^4}{m}\right)\right)$, and so, takes $\omega(\sqrt{m\log(1/\delta)}/\varepsilon^2)$ samples to distinguish between the uniform distribution and an ε-far distribution with error probability δ.

Proof First, let $X_1, \ldots, X_n$ be the elements sampled from our distribution. Let $E_{i,j}$ be the event that $X_i = X_j$. Under the uniform distribution p, we have that the probability that $X_i = X_j$ is

$$\mathbb{P}_p[E_{i,j}] = \sum_{j=1}^m p_j^2 = \frac{1}{m}$$

Thus, the expected number of collisions under the uniform distribution is

$$\mathbb{E}_p\left[\sum_{i<j} E_{i,j}\right] = \sum_{i<j} \mathbb{E}_p[E_{i,j}] = \binom{n}{2}/m$$

Now, consider the ε-far distribution q such that $q_j = \frac{1}{m} + \frac{2\varepsilon}{m}$ for $j \le m/2$ and $q_j = \frac{1}{m} - \frac{2\varepsilon}{m}$ for $j > m/2$. We have that the probability that $X_i = X_j$ under q is

$$\mathbb{P}_q[E_{i,j}] = \sum_{j=1}^m q_j^2 = \frac{1}{m} + \sum_{j=1}^m \left(q_j - \frac{1}{m}\right)^2 = \frac{1+4\varepsilon^2}{m}$$

Thus, the expected number of collisions under q is

$$\mathbb{E}_q\left[\sum_{i<j} E_{i,j}\right] = \sum_{i<j} \mathbb{E}[E_{i,j}] = \binom{n}{2}\frac{1+4\varepsilon^2}{m}$$

Now, under p, if we sample the first element at least $4n/\sqrt{m}$ times, then we will have at least $\binom{4n/\sqrt{m}}{2} > \binom{n}{2}\frac{1+4\varepsilon^2}{m}$ collisions for large enough n, m. In this case, since the number of collisions under p is more than the expected number of collisions under q, and our threshold will be less than the expected number of collisions under q, we will output q, and make a mistake. This happens with probability at least

$$\frac{1}{n^{4n/\sqrt{m}}} = \exp\left(-\frac{4n}{\sqrt{m}}\log n\right)$$

This is bigger than $\exp(-\Omega(\frac{n^2\varepsilon^4}{m}))$, the error probability of the TV tester, as long as

$$\frac{4n}{\sqrt{m}}\log n = o\left(\frac{n^2\varepsilon^4}{m}\right)$$

Since $n = \Theta(m)$, this happens as long as

$$\varepsilon = \omega\left(\frac{\log^{1/4}n}{n^{1/8}}\right)$$

Thus, for error probability δ, we require $\omega(\sqrt{m\log(1/\delta)}/\varepsilon^2)$ samples in the regime stated. ■

E.2. Paninski tester is asymptotically bad when $n \geq \Theta(m\log m)$

When $n \geq 48m\log m$, the Paninski tester fails to distinguish between the uniform distribution on m and an ε-far distribution with error probability $\Omega(1)$ for $\varepsilon < 1/3$.

Proof Recall that the Paninski tester counts the number of bins that see exactly one sample. Let $E_j = \mathbb{1}_{\{Y_j=1\}}$ be the event that the bin j has exactly 1 sample. Now, for p the uniform distribution on $[m]$, the expected number of samples that land in the jth bin is

$$\mathbb{E}_p[Y_j] = \frac{n}{m} = 48\log m$$

Thus, by the Bernstein's inequality,

$$\mathbb{P}_p[Y_j \leq \log m] = \mathbb{P}_p\left[Y_j \leq \left(1 - \frac{47}{48}\right)\mathbb{E}_p[Y_j]\right] \leq e^{-\left(\frac{47}{48}\right)^2 \times 12\log m} = e^{-3\log m} \leq \frac{1}{m^3}$$

So, by union bound,

$$\mathbb{P}_p[\exists j, Y_j \leq \log m] \leq \sum_{j=1}^{m}\mathbb{P}_p[Y_j \leq \log m] \leq \frac{1}{m^2}$$

So, with probability $1 - 1/m^2$, every bin has at least $\log m$ balls, which means that the Paninski statistic is 0 with probability $1 - 1/m^2$.

Now, under ε-far distribution q such that $q_j = \frac{1}{m} + \frac{2\varepsilon}{m}$ for $j \leq m/2$ and $q_j = \frac{1}{m} - \frac{2\varepsilon}{m}$ for $j > m/2$, we have, for $j \leq m/2$

$$\mathbb{E}_q[Y_j] = \frac{n}{m} = (1 + 2\varepsilon)48\log m$$

and for $j > m/2$,

$$\mathbb{E}_q[Y_j] = \frac{n}{m} = (1 - 2\varepsilon)48\log m$$

Then, for $j \leq m/2$, since $\varepsilon > 0$,

$$\mathbb{P}_q[Y_j \leq \log m] = \mathbb{P}_q\left[Y_j \leq \left(1 - \frac{47 + 96\varepsilon}{48 + 96\varepsilon}\right)\mathbb{E}_q[Y_j]\right] \leq \mathbb{P}_q\left[Y_j \leq \left(1 - \frac{47}{48}\right)\mathbb{E}_q[Y_j]\right] \leq \frac{1}{m^3}$$

On the other hand, for $j > m/2$

$$\mathbb{P}_q[Y_j \leq \log m] = \mathbb{P}_q\left[Y_j \leq \left(1 - \frac{47 - 96\varepsilon}{48(1 - 2\varepsilon)} \mathbb{E}_q[Y_j]\right)\right] \leq e^{-\left(\frac{47 - 96\varepsilon}{48(1 - 2\varepsilon)}\right)^2 \times (1 - 2\varepsilon) 12 \log m}$$

Since $\varepsilon < 1/3$, this is at most

$$e^{-\frac{15^2 \times 3}{48^2} \times 12 \log m} \leq \frac{1}{m^3}$$

Thus, we have that

$$\mathbb{P}_q[\exists j, Y_j \leq \log m] \leq \sum_{j=1}^{m} \mathbb{P}_q[Y_j \leq \log m] \leq \frac{1}{m^2}$$

So again, with probability $1 - 1/m^2$, every bin has at least $\log m$ balls under q, so that the Paninski statistic is 0 with $1 - \frac{1}{m^2}$. Putting the above together gives us that we will fail with probability $\Omega(1)$.

∎

Appendix F. TV Tester in the Superlinear Regime

Lemma 53 *The TV tester can distinguish between the uniform distribution on $[m]$ and an ε-far distribution with failure probability $e^{-\frac{1}{2}\varepsilon^2 n(1+o(1))}$ when $\varepsilon = o(1)$ and $n = \omega(m/\varepsilon^2)$.*

Proof Let Y be the empirical distribution when n samples are drawn from the distribution. The TV tester compares the empirical distribution Y to the uniform distribution and outputs uniform if and only if

$$\|Y - p\|_{TV} < \frac{\varepsilon}{2}$$

where p is the uniform distribution on $[m]$.

Since by the definition of TV distance,

$$\|Y - p\|_{TV} = \max_{S \subseteq [m]} |Y_S - p_S|$$

we have that under the uniform distribution p, the probability of failure is

$$\mathbb{P}_p[\|Y - p\|_{TV} \geq \tau^*] = \mathbb{P}_p\left[\max_{S \subseteq [m]} |Y_S - p_S| \geq \frac{\varepsilon}{2}\right]$$

$$\leq \sum_{S \subseteq [m]} \mathbb{P}_p[|Y_S - p_S| \geq \frac{\varepsilon}{2}]$$

Note that the summand is the probability that the empirical mean of n samples from a coin with probability of heads $\frac{|S|}{m}$ deviates from its expectation by at least $\frac{\varepsilon}{2}$. By the Chernoff bound, this is at most

$$e^{-D\left(\frac{|S|}{m} + \frac{\varepsilon}{2} \| \frac{|S|}{m}\right)n} \leq e^{-\frac{1}{2}\varepsilon^2 n}$$

since for any r, τ, $D(r + \tau \| r) \geq 2\tau^2$.

Thus,

$$\mathbb{P}_p\left[\|Y - p\|_{TV} \geq \frac{\varepsilon}{2}\right] \leq 2^m e^{-\frac{1}{2}\varepsilon^2 n} = e^{-\frac{1}{2}\varepsilon^2 n(1+o(1))}$$

since $n = \omega(m/\varepsilon^2)$.

Now, under ε-far distribution q, since $\|p - q\|_{TV} \geq \varepsilon$, the set $S = \{j | q_j > p_j\}$ has $|q_S - p_S| \geq \varepsilon$ so that $\sum_{j \in S} q_j \geq \frac{|S|}{m} + \varepsilon$. Now, since $\|Y - p\|_{TV} \geq |Y_S - p_S|$, we have

$$\mathbb{P}_q\left[\|Y - p\|_{TV} \leq \frac{\varepsilon}{2}\right] \leq \mathbb{P}_q\left[|Y_S - p_S| \leq \frac{\varepsilon}{2}\right]$$

Now, the RHS of the above is at most the probability that the empirical mean of n samples from a coin with probability of heads at least $\frac{1}{2} + \varepsilon$ is less than $\frac{1}{2} + \frac{\varepsilon}{2}$. By the Chernoff bound, this is at most

$$e^{-D(\frac{1}{2} + \frac{\varepsilon}{2} \| \frac{1}{2} + \varepsilon)n} \leq e^{-\frac{1}{2}n\varepsilon^2}$$

Thus, the TV tester fails with probability at most $e^{-\frac{1}{2}n\varepsilon^2(1+o(1))}$ as required. ∎

Lemma 54 *When $n = \omega\left(\frac{m}{\varepsilon^2}\right)$ and $\varepsilon = o(1)$, any tester that distinguishes between the uniform distribution on $[m]$ and an ε-far distribution fails with probability at least $e^{-\frac{1}{2}n\varepsilon^2(1+o(1))}$.*

Proof Let p be the uniform distribution on $[m]$, and let q be the ε-far distribution such that $q_j = \frac{1}{m} + \frac{2\varepsilon}{m}$ for $j \leq m/2$ and $q_j = \frac{1}{m} - \frac{2\varepsilon}{m}$ for $j > m/2$. Let Y be the empirical distribution from the samples drawn, and let $B = \{\sum_{j=1}^{m/2} Y_j \geq \frac{1}{2} + \frac{\varepsilon}{2}\}$. Under the uniform distribution, by Lemma 56,

$$\mathbb{P}_p[B] \geq e^{-\frac{1}{2}\varepsilon^2 n(1+o(1))} \tag{57}$$

Now, the likelihood ratio of $Y \in B$ is given by

$$\frac{q}{p}(Y) \geq (1+2\varepsilon)^{(\frac{1}{2}+\frac{\varepsilon}{2})n}(1-2\varepsilon)^{(\frac{1}{2}-\frac{\varepsilon}{2})n}$$

$$= e^{(\frac{1}{2}+\frac{\varepsilon}{2})\log(1+2\varepsilon)n + (\frac{1}{2}-\frac{\varepsilon}{2})\log(1-2\varepsilon)n}$$

Now,

$$\left(\frac{1}{2}+\frac{\varepsilon}{2}\right)\log(1+2\varepsilon) + \left(\frac{1}{2}-\frac{\varepsilon}{2}\right)\log(1-2\varepsilon) = \left(\frac{1}{2}+\frac{\varepsilon}{2}\right)\left\{\log\left(\frac{\frac{1}{2}+\frac{\varepsilon}{2}}{\frac{1}{2}}\right) - \log\left(\frac{\frac{1}{2}+\frac{\varepsilon}{2}}{\frac{1}{2}+\varepsilon}\right)\right\}$$

$$+ \left(\frac{1}{2}-\frac{\varepsilon}{2}\right)\left\{\log\left(\frac{\frac{1}{2}-\frac{\varepsilon}{2}}{\frac{1}{2}}\right) - \log\left(\frac{\frac{1}{2}-\frac{\varepsilon}{2}}{\frac{1}{2}-\varepsilon}\right)\right\}$$

$$= D\left(\frac{1}{2}+\frac{\varepsilon}{2}\|\frac{1}{2}\right) - D\left(\frac{1}{2}+\frac{\varepsilon}{2}\|\frac{1}{2}+\varepsilon\right)$$

$$= o(\varepsilon^2)$$

since $D(\frac{1}{2}+\frac{\varepsilon}{2}\|\frac{1}{2}) = \frac{\varepsilon^2}{2}(1+o(1))$ and $D(\frac{1}{2}+\frac{\varepsilon}{2}\|\frac{1}{2}+\varepsilon) = \frac{\varepsilon^2}{2}(1+o(1))$. Thus,

$$\frac{q}{p}(X) \geq e^{o(\varepsilon^2)} \tag{58}$$

Now, suppose there is a test ϕ that uses n samples and distinguishes between the two cases with failure probability $e^{-(\frac{1}{2}+\xi)n\varepsilon^2}$ for $\xi > 0$, so that $\phi = 0$ denotes that the test outputs uniform, and $\phi = 1$ denotes that the test outputs far from uniform.

By assumption,

$$\mathbb{P}_p[\phi = 1] \leq e^{-(\frac{1}{2}+\xi)n\varepsilon^2}$$

Now,

$$\mathbb{P}_q[\phi = 0] \geq \mathbb{P}_q[\{\phi = 0\} \cap B] \geq \frac{q}{p}(\{\phi = 0\} \cap B)\,\mathbb{P}_p[\{\phi = 0\} \cap B]$$

But by (57) and our assumption,

$$\mathbb{P}_p[\{\phi = 0\} \cap B] \geq \mathbb{P}_p[B] - \mathbb{P}_p[\phi = 1] \geq e^{-\frac{1}{2}\varepsilon^2 n(1+o(1))}$$

so that, using (58)

$$\mathbb{P}_q[\phi = 0] \geq e^{-\frac{1}{2}\varepsilon^2 n(1+o(1))}$$

Thus, ϕ has error probability at least $e^{-\frac{1}{2}\varepsilon^2 n(1+o(1))}$, and we have a contradiction, so that any test fails with probability at least $e^{-\frac{1}{2}n\varepsilon^2(1+o(1))}$. $\blacksquare$

60

Theorem 3 (Superlinear regime) *For $n/m \gg 1/\varepsilon^2$ and $\varepsilon \ll 1$, the TV statistic achieves*

$$n = (2 + o(1)) \frac{\log \frac{1}{\delta}}{\varepsilon^2}$$

and no other tester can do better.

Proof By Lemmas 53 and 54, the claim follows. ∎

F.1. Anticoncentration of a Binomial Variable

In this section we show that the Chernoff bound has sharp constants for binomial random variables $B(n, \frac{1}{2})$.

We use the following lower bound on binomial coefficients, found in Lemma 4.7.1 of (Ash, 1990):

Lemma 55 *If n and np are integers, then*

$$\binom{n}{p} \geq \frac{1}{\sqrt{8np(1-p)}} 2^{nh(p)}$$

where $h(p) := -p \log_2 p - (1-p) \log_2(1-p)$ is the binary entropy function.

Lemma 56 *Let $X \sim B(n, 1/2)$, and $\frac{1}{\sqrt{n}} \ll \varepsilon \ll 1$. Then*

$$\mathbb{P}[X > \frac{n}{2} + \varepsilon n] = \exp(-2\varepsilon^2 n (1 + o(1))).$$

Proof The upper bound is the standard Chernoff bound, so we focus on the lower bound. For any $\varepsilon' \in [\varepsilon, \varepsilon + 3/\sqrt{n}]$ with $\frac{n}{2} + \varepsilon' n$ integral we have by Lemma 55 that

$$\mathbb{P}[X = \frac{n}{2} + \varepsilon' n] = 2^{-n} \binom{n}{(1/2 + \varepsilon')n}$$
$$\geq \frac{1}{\sqrt{2n(1 + o(1))}} 2^{-n(1 - h(\frac{1}{2} + \varepsilon'))}.$$

Now, the binary entropy function has Taylor series

$$h(\frac{1}{2} + \varepsilon') = 1 - \frac{2}{\ln 2}(\varepsilon')^2 + O((\varepsilon')^4).$$

Since $\varepsilon' = \varepsilon(1 + o(1))$ by construction, this means

$$\mathbb{P}[X = \frac{n}{2} + \varepsilon' n] \geq \frac{1}{\sqrt{2n(1 + o(1))}} e^{-2n\varepsilon^2(1 + o(1))}.$$

Summing over the $3\sqrt{n}$ such ε', we have

$$\mathbb{P}[X \geq \frac{n}{2} + \varepsilon n] \geq e^{-2n\varepsilon^2(1 + o(1))}$$

as desired. ∎

Appendix G. Squared Statistic in Sublinear Regime

Define a centering function

$$\phi(k) := \left| k - \frac{n}{m} \right| \tag{35}$$

We will analyze the squared statistic S given by

$$S = \sum_{j=1}^{m} \phi(Y_j^n)^2$$

where $Y_j^n = \sum_{i=1}^{n} 1_{X_i=j}$ and $X_1, \ldots, X_n$ are the n samples drawn from distribution ν supported on $[m]$. Note that this is equivalent to the collisions statistic, since it simply applies a translation and scaling. We will set

$$\beta := \kappa \frac{n^2 \varepsilon^4}{m} \tag{59}$$

for constant $\kappa > 0$ to be set later. We also define a parameter Δ.

Assumption 4 $1 \lesssim n/m \ll \frac{1}{\varepsilon^2}$, $\varepsilon \ll \frac{1}{n^{3/13}}$, $\frac{n^2}{m}\varepsilon^4 \gg \log m$ and $m \geq C \log n$ for sufficiently large constant C. In addition, we have the following constraints on β and Δ.

$$\beta = \omega \left(\log \left(\frac{1}{\Delta} \right) + \sqrt{\frac{n}{m} \log \left(\frac{1}{\Delta} \right)} \right) \tag{11}$$

$$\Delta = O \left(\frac{n\varepsilon^2}{m} \right) \tag{12}$$

$$(\beta^2 \varepsilon^2)^3 = o \left(\Delta^2 \right) \tag{13}$$

We assume that Assumption 4 holds throughout this section. Observe that Assumption 4 implies Assumption 2. Note that since $n = \Theta(m)$ and $\frac{n^2 \varepsilon^4}{m} = \omega(\log m)$, we have that

$$\varepsilon = \omega \left(\frac{\log m}{n^{1/4}} \right) \tag{60}$$

For ease of exposition, we will analyze

$$\tilde{S} = \frac{m}{n^2 \varepsilon^2}[S - n] \tag{61}$$

We will study the log MGF of this statistic conditioned on all $\phi(Y_j^n)$'s at most β, given by

$$\Lambda_{n,\nu} := \log \left(\mathop{\mathbb{E}}_{\nu} \left[\exp(\theta \tilde{S}) | \forall j, \phi(Y_j^n) \leq \beta \right] \right) \tag{62}$$

We will compute an asymptotic expansion of the limiting log MGF conditioned on all $\phi(Y_j^n)$ at most β given by

$$\Lambda_\nu(\theta) := \lim_{n \to \infty} \frac{m}{n^2 \varepsilon^4} \Lambda_{n,\nu} \left(\frac{n^2 \varepsilon^4}{m} \theta \right) \tag{63}$$

G.1. Poissonization

Let $\tilde{S}_{Poi(\lambda)}$ be the Poissonized statistic. We compute the conditional MGF of $\tilde{S}_{Poi(\lambda)}$ with MGF parameter $\frac{n^2\varepsilon^4}{m}\theta$ conditioned on it being the case that $\phi(Z_j) \leq \beta$ for all j. That is,

$$A_\lambda(\theta) := \exp(-\varepsilon^2\theta n)\, \mathbb{E}\left[\mathbb{1}_{\{\forall j, \phi(Z_j)\leq\beta\}} \exp\left(\varepsilon^2\theta \sum_{j=1}^m \phi(Z_j)^2\right)\right] \tag{64}$$

where $Z_j \sim Poi(\lambda\nu_j)$ and are independent. Due to this independence,

$$A_\lambda(\theta) = \exp(-\varepsilon^2\theta n) \prod_{j=1}^m \mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j)\leq\beta\}} \exp\left(\varepsilon^2\theta\phi(Z_j)^2\right)\right]$$

Lemma 57 *Suppose Assumption 4 holds. We have,*

$$\varepsilon^2\theta\, \mathbb{E}[\mathbb{1}_{\{\phi(Z_j)\leq\beta\}}\phi(Z_j)^2] = \varepsilon^2\theta\, \mathbb{E}[\phi(Z_j)^2] + o(\Delta^2)$$

$$\frac{\varepsilon^4\theta^2}{2}\, \mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j)\leq\beta\}}\phi(Z_j)^4\right] = \frac{\varepsilon^4\theta^2}{2}\, \mathbb{E}[\phi(Z_j)^4] + o(\Delta^2)$$

$$\sum_{l=3}^\infty \frac{(\varepsilon^2\theta)^l}{l!}\, \mathbb{E}\left[\mathbb{1}_{\{\phi(Z_j)\leq\beta\}}\phi(Z_j)^{2l}\right] = o(\Delta^2)$$

when $Z_j \sim Poi(\lambda\nu_j)$ for $\lambda = n(1 + O(\varepsilon^2))$ and $\nu_j = 1/m + O(\varepsilon/m)$ for all j.

Proof For the first claim,

$$\varepsilon^2\theta\, \mathbb{E}[\mathbb{1}_{\{\phi(Z_j)\leq\beta\}}\phi(Z_j)^2] = \varepsilon^2\theta\, \mathbb{E}[\phi(Z_j)^2] - \varepsilon^2\theta\, \mathbb{E}[\mathbb{1}_{\{\phi(Z_j)>\beta\}}\phi(Z_j)^2]$$

By Lemma 45, the second term is $o(\Delta^2)$. Thus, we have the first claim. The second claim can be proved in a similar way. The third claim follows by Lemma 42. ∎

G.2. Depoissonization

As before, we will first show that $A_\lambda(\theta)$ is analytic in λ.

Lemma 58 $A_\lambda(\theta)$ *is analytic in λ.*

Proof We will show that $\mathbb{E}[\mathbb{1}_{\{\phi(Z_j)\leq\beta\}}\exp(\varepsilon^2\theta\phi(Z_j)^2)]$ is analytic. We have that it can be written as

$$\sum_{k:\phi(k)\leq\beta}\left[\frac{(\lambda\nu_j)^k}{k!}e^{-\lambda\nu_j}\exp(\varepsilon^2\theta\phi(k)^2)\right]$$

which is a finite sum of analytic functions, and is thus analytic. The claim follows. ∎

Note that for $A_\lambda(\theta)$ defined in (64), by Lemmas 57 and 58, Assumption 3 holds for

$$\xi(k) = \mathbb{1}_{\{\phi(k)\leq\beta\}}\exp(\varepsilon^2\theta\phi(k)^2)$$

Then, we have the following

Lemma 59 *We have that for uniform distribution p such that $p_j = 1/m$ for all j,*

$$\underset{p}{\mathbb{E}} \left[\exp\left(\frac{n^2 \varepsilon^4}{m} \theta \tilde{S} \right) | \forall j, \phi(Y_j^n) \leq \beta \right] = \exp(-\varepsilon^2 \theta n) \underset{p}{\mathbb{E}} \left[\exp\left(\varepsilon^2 \theta \sum_{j=1}^{m} \mathbb{1}_{\{\phi(Y_j^n) \leq \beta\}} \phi(Y_j^n)^2 \right) \right]$$

$$= (1 + O(1/n)) \exp\left\{ \frac{n^2 \varepsilon^4}{m} (\theta^2 + o(1)) \right\}$$

and for alternate distributions q such that $q_j = \frac{1}{m} + \frac{\varepsilon}{\gamma m}$ for $j \leq l$ and $\nu_j = \frac{1}{m} - \frac{\varepsilon}{(1-\gamma)m}$ for $j > l$,

$$\underset{q}{\mathbb{E}} \left[\exp\left(\frac{n^2 \varepsilon^4}{m} \theta \tilde{S} \right) | \forall j, \phi(Y_j^n) \leq \beta \right] = \exp(-\varepsilon^2 \theta n) \underset{q}{\mathbb{E}} \left[\exp\left(\varepsilon^2 \theta \sum_{j=1}^{m} \mathbb{1}_{\{\phi(Y_j^n) \leq \beta\}} \phi(Y_j^n)^2 \right) \right]$$

$$= (1 + O(1/n)) \exp\left\{ \frac{n^2 \varepsilon^4}{m} \left[\theta^2 + \theta \frac{1}{\gamma(1-\gamma)} + o(1) \right] \right\}$$

Proof By Lemma 25, the claim holds. ∎

G.3. Application of the Gärtner-Ellis Theorem

In this section, we apply the Gärtner-Ellis Theorem to obtain the probability that our statistic crosses a threshold, under the uniform distribution, and under one of the worst-case ε-far distributions.

Lemma 60 *Under the uniform distribution p, we have that for $\tau > 0$,*

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log\left(\underset{p}{\mathbb{P}} \left[\tilde{S} \geq \tau | \forall j, \phi(Y_j^n) \leq \beta \right] \right) = \frac{\tau^2}{4}$$

Under an ε-far distribution q of the form $q_j = \frac{1}{m} + \frac{\varepsilon}{\gamma m}$ for $j \leq l$ and $q_j = \frac{1}{m} - \frac{\varepsilon}{(1-\gamma)m}$ for $j > l$, and $\gamma = \Theta(1), 1 - \gamma = \Theta(1)$, for $\tau < \frac{1}{\gamma(1-\gamma)}$,

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log\left(\underset{q}{\mathbb{P}} \left[\tilde{S} \leq \tau | \forall j, \phi(Y_j^n) \leq \beta \right] \right) = \frac{(\tau\gamma(\gamma - 1) + 1)^2}{4\gamma^2(\gamma - 1)^2}$$

Proof Note that by Lemma 59, the limiting logarithmic moment generating function with respect to the uniform distribution p is given by

$$\Lambda_p(\theta) = \lim_{n \to \infty} \frac{m}{n^2 \varepsilon^4} \log\left(\underset{p}{\mathbb{E}} \left[\exp\left(\frac{n^2 \varepsilon^4}{m} \theta \tilde{S} \right) | \forall j, \phi(Y_j^n) \leq \beta \right] \right) = \theta^2$$

Thus, Assumption 1 holds for $\mathcal{D}_{\Lambda_p} = \mathbb{R}$. Furthermore, the Fenchel-Legendre Transform (defined in equation 28) of Λ_p is given by

$$\Lambda_p^*(\tau) = \sup_{\theta}\{\theta\tau - \theta^2\} = \frac{\tau^2}{4}$$

64

This is a strongly convex function of τ, so the set of exposed points of Λ_p^* whose exposing hyperplane belongs to $\mathcal{D}_{\Lambda_p}^o$ is all of $\mathbb{R}$. Thus, by the Theorem 13 (Gärtner-Ellis), for $\tau > 0$,

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_p \left[\tilde{S} \geq \tau | \forall j, \phi(Y_j^n) \leq \beta \right] \right) = \inf_{x \geq \tau} \Lambda_p^*(x) = \frac{\tau^2}{4}$$

Similarly, the limiting logarithmic moment generating function with respect to an alternate distribution q is given by

$$\Lambda_q(\theta) = \lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{E}_q \left[\exp \left(\frac{n^2 \varepsilon^4}{m} \theta \tilde{S} \right) | \forall j, \phi(Y_j^n) \leq \beta \right] \right) = \theta^2 + \frac{1}{\gamma(1-\gamma)} \theta$$

The Fenchel-Legendre transform is given by

$$\Lambda_q^*(\tau) = \sup_\theta \left\{ \theta \tau - \theta^2 - \frac{1}{\gamma(1-\gamma)} \theta \right\} = \frac{(\tau \gamma(\gamma - 1) + 1)^2}{4\gamma^2(\gamma - 1)^2}$$

Again, applying the Gärtner-Ellis Theorem gives, for $\tau < \frac{1}{\gamma(1-\gamma)}$,

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_q \left[\tilde{S}_n^* \leq \tau | \forall j, \phi(Y_j^n) \leq \beta \right] \right) = \inf_{x \leq \tau} \Lambda_q^*(x) = \frac{(\tau \gamma(\gamma - 1) + 1)^2}{4\gamma^2(\gamma - 1)^2}$$

$\blacksquare$

G.4. Removing the conditioning

So far, we have analyzed the probability that $\tilde{S}$ crosses a threshold τ under the uniform distribution, and under a family of ε-far distributions conditioned on the event that $\phi(Y_j^n) \leq \beta$ for all j. We will now remove this conditioning.

Lemma 61 *Under the uniform distribution p, we have that for $\tau > 0$,*

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_p \left[\tilde{S} \geq \tau \right] \right) = \frac{\tau^2}{4}$$

Under an ε-far distribution q for the form $q_j = \frac{1}{m} + \frac{\varepsilon}{\gamma m}$ for $j \leq l$ and $q_j = \frac{1}{m} - \frac{\varepsilon}{(1-\gamma)m}$ for $j > l$, and $\gamma = \Theta(1)$, $1 - \gamma = \Theta(1)$, for $\tau < \frac{1}{\gamma(1-\gamma)}$,

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_q \left[\tilde{S} \leq \tau \right] \right) = \frac{(\tau \gamma(\gamma - 1) + 1)^2}{4\gamma^2(\gamma - 1)^2}$$

Proof For the uniform distribution we will upper and lower bound $\mathbb{P}_p[\tilde{S} \geq \tau]$ by

$$e^{\frac{n^2 \varepsilon^4}{m} \frac{\tau^2}{4}(1 + o(1))}$$

The claim for the uniform distribution will follow.

Under any distribution ν,

$$\mathbb{P}_{\nu}\left[\tilde{S} \geq \tau\right] = \mathbb{P}_{\nu}\left[\tilde{S} \geq \tau | \forall j, \phi(Y_j^n) \leq \beta\right] \mathbb{P}_{\nu}[\forall j, \phi(Y_j^n) \leq \beta]$$

$$+ \mathbb{P}_{\nu}\left[\tilde{S} \geq \tau | \exists j, \phi(Y_j^n) > \beta\right] \mathbb{P}_{\nu}[\exists j, \phi(Y_j^n) > \beta]$$

Now, by Bernstein's inequality, since $n\varepsilon^4 = \omega(1)$ by (60) and $\beta = \kappa\frac{n^2\varepsilon^4}{m}$, we have, for the uniform distribution p, for every j,

$$\mathbb{P}_{p}[\phi(Y_j^n) > \beta] \leq 2e^{-\frac{\kappa}{4}\frac{n^2\varepsilon^4}{m}}$$

Thus, by union bound,

$$\mathbb{P}_{p}\left[\exists j, \phi(Y_j^n) > \beta\right] \leq \sum_{j=1}^{m} \mathbb{P}_{p}[\phi(Y_j^n) > \beta] \leq 2me^{-\frac{\kappa n^2\varepsilon^4}{4m}} = 2e^{-\frac{\kappa n^2\varepsilon^4}{4m}+\log m} = 2e^{-\frac{\kappa n^2\varepsilon^4}{4m}(1+o(1))}$$

since $\frac{n^2\varepsilon^4}{m} = \omega(\log m)$. For $\mathbb{P}_p[\forall j, \phi(Y_j^n) \leq \beta]$ and $\mathbb{P}_p[\tilde{S} \geq \tau | \exists j, \phi(Y_j^n) > \beta]$, we apply the trivial upper bound of 1. So, for $\kappa \geq 2\tau^2$, by Lemma 60,

$$\mathbb{P}_{p}[\tilde{S} \geq \tau] \leq e^{-\frac{\tau^2}{4}\frac{n^2\varepsilon^4}{m}} + 2e^{-\frac{\tau^2}{2}\frac{n^2\varepsilon^4}{m}(1+o(1))} = e^{-\frac{\tau^2}{4}\frac{n^2\varepsilon^4}{m}}(1+o(1))$$

For the lower bound, note that

$$\mathbb{P}_{p}[\forall j, \phi(Y_j^n) \leq \beta] \geq 1 - \sum_{j} \mathbb{P}_{p}[\phi(Y_j^n) > \beta] \geq 1 - 2me^{-\frac{\kappa}{4}\frac{n^2\varepsilon^4}{m}} = 1 - 2e^{-\frac{\kappa}{4}\frac{n^2\varepsilon^4}{m}(1+o(1))}$$

For the second term, i.e., $\mathbb{P}_p\left[\tilde{S} \geq \tau | \exists j, \phi(Y_j^n) > \beta\right] \mathbb{P}_p[\exists j, \phi(Y_j^n) > \beta]$, we apply the trivial lower bound of 0. Again, setting $\kappa \geq 2\tau^2$, we get

$$\mathbb{P}_{p}\left[\tilde{S} \geq \tau\right] \geq e^{-\frac{n^2\varepsilon^4}{m}\frac{\tau^2}{4}(1+o(1))}$$

So, we have the claim for the uniform distribution. We will upper and lower bound the probability of crossing τ for ε-far distribution q in the same way. For ease, let

$$c_\tau := \frac{(\tau\gamma(\gamma-1)+1)^2}{4\gamma^2(\gamma-1)^2}$$

By Bernstein's inequality, we have that since $\gamma = \Theta(1)$ and $1 - \gamma = \Theta(1)$ so that $q_j = 1/m + O(\varepsilon/m)$ for every j, and since $n\varepsilon^4 = \omega(1)$ by (60) and $\beta = \kappa\frac{n^2\varepsilon^4}{m}$, for every j, under ε-far distribution q,

$$\mathbb{P}_{q}[\phi(Y_j^n) > \beta] \leq 2e^{-\frac{\kappa}{4}\frac{n^2\varepsilon^4}{m}(1+o(1))}$$

Then, setting $\kappa \geq 8c_\tau$, and repeating the same argument as in the uniform case, we obtain the required upper and lower bounds on $\mathbb{P}_q[\tilde{S} \leq \tau]$. The claim follows. ∎

G.5. Setting the threshold

We need to set our threshold τ so that the minimum of the error probability under the uniform distribution p, and any ε-far distribution q is maximized. Note that, it is sufficient to consider a threshold τ such that $0 < \tau < \frac{1}{\gamma(1-\gamma)}$, since otherwise, the error probability in one of the two cases is at least constant. To set our threshold, we will first observe that for any τ in this range, the "error exponent" under ε-far distributions is minimized for a particular ε-far distribution. Then, we will set the threshold to maximize the minimum of the error exponent under the uniform distribution, and under this ε-far distribution.

Lemma 62 *Setting the threshold $\tau = 2$, we have for the uniform distribution p,*

$$\lim_{n\to\infty} -\frac{m}{n^2\varepsilon^4} \log\left(\mathbb{P}_p\left[\tilde{S} \geq \tau\right]\right) = 1$$

and for any ε-far distribution q such that $q_j = \frac{1}{m} + \frac{\varepsilon}{\gamma m}$ for $j \leq l$ and $q_j = \frac{1}{m} - \frac{\varepsilon}{(1-\gamma)m}$ for $j > l$ and $\gamma = \Theta(1), 1 - \gamma = \Theta(1)$,

$$\lim_{n\to\infty} -\frac{m}{n^2\varepsilon^4} \log\left(\mathbb{P}_q\left[\tilde{S} \leq \tau\right]\right) \geq 1$$

with equality for q such that $q_j = 1/m + 2\varepsilon/m$ for $j \leq m/2$ and $q_j = 1/m - 2\varepsilon/m$ for $j > m/2$.

Proof By Lemma 61, for $0 < \tau < \frac{1}{\gamma(1-\gamma)}$,

$$\lim_{n\to\infty} \frac{m}{n^2\varepsilon^4} \log\left(\mathbb{P}_q\left[\tilde{S} \leq \tau\right]\right) = \frac{(\tau\gamma(\gamma-1)+1)^2}{4\gamma^2(\gamma-1)^2}$$

Now, the numerator of the right hand side is minimized when $\gamma = 1/2$, and the denominator is maximized when $\gamma = 1/2$. Thus, the right hand side is minimized when $\gamma = 1/2$. So, we have that,

$$\lim_{n\to\infty} -\frac{m}{n^2\varepsilon^4} \log\left(\mathbb{P}_q\left[\tilde{S} \leq \tau\right]\right) \geq \frac{1}{4}(\tau-4)^2$$

with equality for distribution q such that $q_j = 1/m + 2\varepsilon/m$ for $j \leq m/2$ and $q_j = 1/m - 2\varepsilon/m$ for $j > m/2$.

Then, the claim follows by substituting in $\tau = 2$ in the expression for the uniform distribution in Lemma 35 and in the above expression. ∎

Recall that our target sample complexity is

$$n = (1 + o(1))\sqrt{m \log \frac{1}{\delta}}/\varepsilon^2 \tag{4}$$

We have our result for the squared/collisions tester.

Theorem 4 (Collisions for large δ) *The quadratic statistic achieves* (4) *for $n/m = \Theta(1), \log n \ll \log \frac{1}{\delta} \ll n^{1/13}$ and $\varepsilon \ll 1$.*

Proof We need to show that we can satisfy (12), (11) and (13) for the β that we chose in (59), and some Δ. Set

$$\Delta = \frac{n\varepsilon^2}{m} = \Theta(\varepsilon^2)$$

so that (12) is satisfied. Now,

$$\beta = \kappa \frac{n^2 \varepsilon^4}{m} = \omega(\log m) = \omega \left(\log \left(\frac{1}{\Delta} \right) \right)$$

and since $n = \Theta(m)$,

$$\beta = \kappa \frac{n^2 \varepsilon^4}{m} = \Theta \left(\sqrt{\frac{n}{m}} \frac{n^2 \varepsilon^4}{m} \right) = \omega \left(\sqrt{\frac{n}{m}} \log m \right) = \omega \left(\sqrt{\frac{n}{m}} \log \left(\frac{1}{\Delta} \right) \right)$$

Thus, (11) is satisfied. For (13), note that

$$\left(\beta^2 \varepsilon^2 \right)^3 = \Theta \left(\left(\frac{n^4 \varepsilon^{10}}{m^2} \right)^3 \right) = \Theta \left(\frac{n^{12} \varepsilon^{30}}{m^6} \right) = \Theta \left(n^6 \varepsilon^{30} \right) = o(\varepsilon^4) = o(\Delta^2)$$

since $\varepsilon = o \left(\frac{1}{n^{3/13}} \right)$. So, the conditions are satisfied.

By Lemma 62, we have that $\bar{c}(\varepsilon, m, C) = 1$ for every ε that satisfies our assumptions, and every $C > 2$. In particular, any ε' such that $\left(1 - \frac{1}{C'} \right) \varepsilon(n) \leq \varepsilon'(n) \leq \varepsilon(n)$ has $\bar{c}(\varepsilon', m, C) = 1$ for every $C > 2$. Thus, by Lemma 21, we have that $c(\varepsilon, m) = 1$ for every ε that satisfies our assumptions. The claim follows. ∎

Appendix H. Empty bins statistic in Sublinear Regime

Assumption 5 $n/m \lesssim 1$, $\varepsilon, \delta \ll 1$, $n \gg 1$.

We will assume that Assumption 5 holds throughout this section. Let our statistic be

$$S := \sum_{j=1}^{m} \mathbb{1}_{\{Y_j^n = 0\}}$$

where $Y_j^n = \sum_{i=1}^{n} \mathbb{1}_{\{X_i = j\}}$, and $X_1, \ldots, X_n$ are the n samples drawn from the distribution ν supported on $[m]$. In other words, it is the number of empty bins. Note that this has the same error probability as the total variation statistic when $n \leq m$. For ease of exposition, we will analyze the statistic

$$\tilde{S} = \frac{m}{n^2 \varepsilon^2}[S - m e^{-n/m}] \tag{65}$$

Note that this has the same error probability as S. Consider the moment generating function (MGF) of $\tilde{S}$ with respect to distribution ν, given by

$$M_{\tilde{S}, \nu}(\theta) = \mathbb{E}_{\nu}\left[\exp(\theta \tilde{S})\right]$$

Note that the logarithmic moment generating function of this statistic is given by

$$\Lambda_{n,\nu}(\theta) := \log(\mathbb{E}_{\nu}[\exp(\theta \tilde{S})]) \tag{66}$$

We wish to study the error exponent (Huang and Meyn, 2013) of this statistic with respect to normalization $\frac{n^2 \varepsilon^4}{m}$. To do this, we will compute the asymptotic expansion of the limiting logarithmic moment generating function of the statistic, given by

$$\Lambda_{\nu}(\theta) := \lim_{n \to \infty} \frac{m}{n^2 \varepsilon^4} \Lambda_{n,\nu}\left(\frac{n^2 \varepsilon^4}{m} \theta\right) \tag{67}$$

H.1. Poissonization

Define $\tilde{S}_{Poi(\lambda)}$ to be the Poissonized statistic, that is the statistic $\tilde{S}$ when the number of samples is chosen according to the Poisson distribution with mean λ. We begin by computing the MGF of $\tilde{S}_{Poi(\lambda)}$ with MGF parameter $\frac{n^2 \varepsilon^4}{m} \theta$. Let this be

$$A_\lambda(\theta) := \mathbb{E}\left[\exp\left(\frac{n^2 \varepsilon^4}{m} \theta \tilde{S}_{Poi(\lambda)}\right)\right] = \exp\left(-m e^{-n/m} \varepsilon^2 \theta\right) \mathbb{E}\left[\exp\left(\varepsilon^2 \theta \sum_{j=1}^{m} \mathbb{1}_{\{Z_j = 0\}}\right)\right]$$

where $Z_j \sim Poi(\lambda \nu_j)$ are independent. Due to this independence,

$$A_\lambda(\theta) = \exp\left(-m e^{-n/m} \varepsilon^2 \theta\right) \prod_{j=1}^{m} \mathbb{E}\left[\exp\left(\varepsilon^2 \theta \mathbb{1}_{\{Z_j = 0\}}\right)\right] \tag{68}$$

Now, we have the following.

Lemma 63

$$\mathbb{E}\left[\exp\left(\varepsilon^2\theta\mathbb{1}_{\{Z_j=0\}}\right)\right] = e^{-\lambda\nu_j}e^{\varepsilon^2\theta} - e^{-\lambda\nu_j} + 1$$

when $Z_j \sim Poi(\lambda\nu_j)$.

Proof

$$\mathbb{E}\left[\exp\left(\varepsilon^2\theta\mathbb{1}_{\{Z_j=0\}}\right)\right] = \exp(\varepsilon^2\theta)\,\mathbb{P}[Z_j=0] + \sum_{k=1}^{\infty}\mathbb{P}[Z_j=k]$$

$$= \exp(\varepsilon^2\theta)\,\mathbb{P}[Z_j=0] + 1 - \mathbb{P}[Z_j=0] = e^{-\lambda\nu_j}e^{\varepsilon^2\theta} - e^{-\lambda\nu_j} + 1$$

$\blacksquare$

H.2. Depoissonization

First, we will show that $A_\lambda(\theta)$ is analytic in λ.

Lemma 64 $A_\lambda(\theta)$ *is analytic in λ.*

Proof By Lemma 63, $\mathbb{E}[\exp(\varepsilon^2\theta\mathbb{1}_{\{Z_j=0\}})]$ is analytic λ since it is the finite sum of analytic functions. Then, $A_\lambda(\theta)$ is the finite product of analytic functions, and is hence analytic. The claim follows. $\blacksquare$

Lemma 65

$$M_{\tilde{S},\nu}\left(\frac{n^2\varepsilon^4}{m}\theta\right) = \frac{n!}{2\pi i}\oint\frac{e^\lambda}{\lambda^{n+1}}A_\lambda(\theta)d\lambda \tag{69}$$

Proof Follows from Lemma 26 since $A_\lambda(\theta)$ is analytic in λ. $\blacksquare$

We will choose a contour passing through a particular λ_0, and this will make it easy to evaluate the integral. We carry out the integration along the contour given by $\lambda = \lambda_0 e^{i\psi}$, where

$$\lambda_0 = n(1 + e^{-n/m}\varepsilon^2\theta) \tag{70}$$

We substitute $\lambda = \lambda_0 e^{i\psi}$ and $A_\lambda(\theta)$ from (68) into (69) to get that

$$M_{\tilde{S},\nu}\left(\frac{n^2\varepsilon^4}{m}\theta\right) = \exp\left(-me^{-n/m}\varepsilon^2\theta\right)\frac{n!}{2\pi}\lambda_0^{-n}Re\left[\int_{-\pi}^{\pi}g(\psi)d\psi\right] \tag{71}$$

with

$$g(\psi) := e^{-in\psi}e^{\lambda_0 e^{i\psi}}\prod_{j=1}^{m}\left\{e^{-\lambda_0\nu_j e^{i\psi}}e^{\varepsilon^2\theta} - e^{-\lambda_0\nu_j e^{i\psi}} + 1\right\} \tag{72}$$

We will split this integral into 3 parts. Let

$$
\begin{aligned}
I_1 &= Re\left[\int_{-\pi/3}^{\pi/3} g(\psi)d\psi\right] \\
I_2 &= Re\left[\int_{-\pi}^{-\pi/3} g(\psi)d\psi\right] \\
I_3 &= Re\left[\int_{\pi/3}^{\pi} g(\psi)d\psi\right]
\end{aligned}
\tag{73}
$$

We will show that I_1 dominates. We show this by bounding $g(\psi)$ in the region $\psi \in [-\pi, -\pi/3] \cup [\pi/3, \pi]$ as follows.

Lemma 66 *For $\psi \in [-\pi, -\pi/3] \cup [\pi/3, \pi]$,*

$$
|g(\psi)| \le \exp\{0.5n(1 + O(\varepsilon^2)) + 0.5me^{-n/m}\varepsilon^2\theta(1 + O(\varepsilon^2))\}
$$

Proof By definition of g from (72) and Lemma 77, we have that for $Z_j \sim Poi(\lambda_0 \nu_j)$,

$$
\begin{aligned}
|g(\psi)| &= \left| e^{-in\psi} e^{\lambda_0 e^{i\psi}} \prod_{j=1}^{m} \left\{ e^{-\lambda_0 \nu_j e^{i\psi}} e^{\varepsilon^2 \theta} - e^{-\lambda_0 \nu_j e^{i\psi}} + 1 \right\} \right| \\
&= \left| e^{-in\psi} e^{\lambda_0 e^{i\psi}} \exp\left\{ me^{-\frac{n}{m} e^{i\psi}} \varepsilon^2 \theta(1 + O(\varepsilon^2)) + O(n\varepsilon^3) \right\} \right|
\end{aligned}
$$

Since $\lambda_0 = n(1 + O(\varepsilon^2))$, this is

$$
\left| e^{-in\psi} \exp\{ ne^{i\psi} + me^{-\frac{n}{m} e^{i\psi}} \varepsilon^2 \theta(1 + O(\varepsilon^2)) + O(n\varepsilon^2) \} \right|
$$

The claim follows since $|e^{-in\psi}| = 1$ and

$$
|\exp(e^{i\psi})| = |\exp(\cos(\psi) + i\sin(\psi))| = \exp(\cos(\psi)) \le \exp(0.5)
$$

for ψ in the range stated.   ∎

Note that this implies that for the integrals defined in (73) that

$$
I_2 + I_3 = O(e^{0.6n + 0.6me^{-n/m}\varepsilon^2\theta})
\tag{74}
$$

Now, we will compute I_1. Define $G(\psi) := \log(g(\psi))$. Then, by definition of g in (72),

$$
G(\psi) = -in\psi + \lambda_0 e^{i\psi} + \sum_{j=1}^{m} \log\left\{ e^{-\lambda_0 \nu_j e^{i\psi}} e^{\varepsilon^2 \theta} - e^{-\lambda_0 \nu_j e^{i\psi}} + 1 \right\}
\tag{75}
$$

Note that

$$
Im(G(0)) = 0
\tag{76}
$$

Then, applying Lemma 51,

$$
Re(G'(0)) = 0
\tag{77}
$$

71

Now, computing the asymptotic expansion of $G''(\psi)$ by Lemma 78, we have

$$G''(\psi) = -ne^{i\psi} + O\left(n\varepsilon^2\right) \tag{78}$$

Now, by Taylor's theorem, for any $\psi \in [-\pi/3, \pi/3]$ there exists $\tilde{\psi} \in (0, \psi)$ such that

$$G(\psi) = G(0) + G'(0)\psi + \frac{G''(\tilde{\psi})}{2}\psi^2 \tag{79}$$

But, by (78), $Re[G''(\psi)] \leq -0.4n$ for any $\psi \in [-\pi/3, \pi/3]$. So, for $\psi \in [-\pi/3, \pi/3]$,

$$Re(G(\psi)) \leq G(0) - 0.2n\psi^2 \tag{80}$$

Now, we have the following upper bound on I_1.

Lemma 67

$$I_1 \leq e^{G(0)}\frac{\sqrt{\pi}}{\sqrt{0.2n}}$$

Proof

$$I_1 = Re\left[\int_{-\pi/3}^{\pi/3} e^{G(\psi)}d\psi\right] \leq \int_{-\pi/3}^{\pi/3} e^{Re(G(\psi))}d\psi \leq e^{G(0)}\int_{-\pi/3}^{\pi/3} e^{-0.2n\psi^2}d\psi \tag{81}$$

∎

$$\leq e^{G(0)}\int_{-\infty}^{\infty} e^{-0.2n\psi^2}d\psi = e^{G(0)}\frac{\sqrt{\pi}}{\sqrt{0.2n}}$$

The next lemma shows that I_1 is also lower bounded by the above quantity (up to constants).

Lemma 68

$$I_1 \geq e^{G(0)}\frac{0.5\sqrt{\pi}}{\sqrt{1.1n}}(1 + o(1))$$

where I_1 is defined in (73)

Proof By (78), $Im(G''(\psi)) = -n\sin(\psi) + O\left(n\varepsilon^2\right)$. So, for large enough n, since $|\sin(\psi)| \leq |\psi|$, for any $\psi \in [-\pi/3, \pi/3]$, $|Im(G''(\psi))| \leq 1.1n|\psi|$. So, by (79), (76) and (77), we have that for constant $c > 0$ and $\psi \in [-\pi/3, \pi/3]$,

$$|Im(G(\psi))| \leq 1.1n|\psi|^3 + cn\varepsilon^2\psi^2$$

Also, $Re(G''(\psi)) \geq -1.1n$ by a similar argument. So, by (79) and (77), for $\psi \in [-\pi/3, \pi/3]$,

$$Re(G(\psi)) \geq G(0) - 1.1n\psi^2$$

Now, for $t_n = 0.1\min\{n^{-1/3}, \frac{1}{\varepsilon\sqrt{cn}}\}$, we have that for $\psi \in [-t_n, t_n]$, $\cos(Im(G(\psi))) \geq 0.5$ so that $Re(e^{G(\psi)}) \geq 0.5e^{Re(G(\psi))}$. We can split I_1 further into 3 parts:

$$I_1 = Re\left[\int_{-\pi/3}^{-t_n} e^{G(\psi)}d\psi\right] + Re\left[\int_{t_n}^{\pi/3} e^{G(\psi)}d\psi\right] + Re\left[\int_{-t_n}^{t_n} e^{G(\psi)}d\psi\right]$$

Now, by (80),

$$\left| \int_{-\pi/3}^{-t_n} e^{G(\psi)} d\psi \right| \le e^{G(0)} \int_{-\infty}^{-t_n} e^{-0.2n\psi^2} d\psi = t_n e^{G(0)} \int_{-\infty}^{-1} e^{-0.2nt_n^2 \bar{\psi}^2} d\bar{\psi}$$

$$\le t_n e^{G(0)} \int_{-\infty}^{-1} e^{-0.2nt_n^2 |\bar{\psi}|} d\bar{\psi} = e^{G(0)} O\left(\frac{1}{nt_n}\right) = e^{G(0)} o\left(\frac{1}{\sqrt{n}}\right)$$

In a similar way, we can bound the second term. For the third term, we have

$$\text{Re}\left[\int_{-t_n}^{t_n} e^{G(\psi)} d\psi \right] \ge \int_{-t_n}^{t_n} 0.5 e^{\text{Re}(G(\psi))} d\psi \ge 0.5 e^{G(0)} \int_{-t_n}^{t_n} e^{-1.1n\psi^2} d\psi$$

$$\ge 0.5 e^{G(0)} \left[\int_{-\infty}^{\infty} e^{-1.1n\psi^2} d\psi - 2 \int_{-\infty}^{-t_n} e^{-1.1n\psi^2} d\psi \right]$$

$$\ge 0.5 e^{G(0)} \left(\frac{\sqrt{\pi}}{\sqrt{1.1n}} + O\left(\frac{1}{nt_n}\right) \right) = 0.5 e^{H(0)} \frac{\sqrt{\pi}}{\sqrt{1.1n}} (1 + o(1))$$

Combining the bounds, we get that

$$I_1 \ge e^{G(0)} \frac{0.5\sqrt{\pi}}{\sqrt{1.1n}} (1 + o(1))$$

∎

Combining the upper bound on I_1 from Lemma 67 and the lower bound from Lemma 68, we have

$$I_1 = e^{G(0)} \frac{1}{\sqrt{n}} e^{O(1)}$$

So, by (71) and (74),

$$M_{\tilde{S}, \nu}\left(\frac{n^2 \varepsilon^4}{m} \theta \right) = \exp(-me^{-n/m}\varepsilon^2\theta) \frac{n!}{2\pi} \lambda_0^{-n} e^{G(0)} \frac{\sqrt{\pi}}{\sqrt{0.2n}} (1 + o(1)) \tag{82}$$

So, it remains to compute $G(0)$.

Lemma 69 *Under the uniform distribution given by $\nu_j = 1/m$ for all j, and $\lambda_0 = n(1 + e^{-n/m}\varepsilon^2\theta)$,*

$$G(0) = \lambda_0 + me^{-n/m}\varepsilon^2\theta + \varepsilon^4\theta^2[-ne^{-2n/m} + m\frac{e^{-n/m}}{2} - m\frac{e^{-2n/m}}{2}] + O(m\varepsilon^6)$$

Proof By definition of $G(\psi)$ in equation (75) and Lemma 63,

$$G(0) = \lambda_0 + \sum_{j=1}^{m} \log[1 + e^{-\lambda_0 \nu_j}(e^{\varepsilon^2\theta} - 1)]$$

73

Now, since $e^{\varepsilon^2\theta} - 1 = \varepsilon^2\theta + \frac{\varepsilon^4\theta^2}{2} + O(\varepsilon^6)$, this is

$$\lambda_0 + \sum_{j=1}^{m}[e^{-\lambda_0\nu_j}(\varepsilon^2\theta + \frac{\varepsilon^4\theta^2}{2}) - \frac{\varepsilon^4\theta^2}{2}e^{-2\lambda_0\nu_j} + O(\varepsilon^6)]$$

Substituting in λ_0 from (70), and $\nu_j = 1/m$ for all j, this is

$$= \lambda_0 + \sum_{j=1}^{m}[e^{-\frac{n}{m}}e^{-\frac{n}{m}}e^{-\frac{n}{m}\varepsilon^2\theta}(\varepsilon^2\theta + \frac{\varepsilon^4\theta^2}{2}) - \frac{\varepsilon^4\theta^2}{2}e^{-2\frac{n}{m}(1+O(\varepsilon^2))}) + O(\varepsilon^6)]$$

$$= \lambda_0 + \sum_{j=1}^{m}[e^{-\frac{n}{m}}(1 - \frac{n}{m}e^{-\frac{n}{m}\varepsilon^2\theta})(\varepsilon^2\theta + \frac{\varepsilon^4\theta^2}{2}) - \frac{\varepsilon^4\theta^2}{2}e^{-2\frac{n}{m}} + O(\varepsilon^6)]$$

$$= \lambda_0 + \sum_{j=1}^{m}[e^{-\frac{n}{m}}\varepsilon^2\theta - \frac{n}{m}e^{-2\frac{n}{m}}\varepsilon^4\theta^2 + \frac{e^{-\frac{n}{m}}}{2}\varepsilon^4\theta^2 - \frac{\varepsilon^4\theta^2}{2}e^{-2\frac{n}{m}} + O(\varepsilon^6)]$$

$$= \lambda_0 + me^{-\frac{n}{m}}\varepsilon^2\theta + \varepsilon^4\theta^2[-ne^{-2\frac{n}{m}} + m\frac{e^{-\frac{n}{m}}}{2} - m\frac{e^{-2\frac{n}{m}}}{2}] + O(m\varepsilon^6)$$

∎

Lemma 70 *Under distribution ν such that $\nu_j = \frac{1}{m} + \frac{\varepsilon}{\gamma m}$ for $j \le \gamma m$ and $\nu_j = \frac{1}{m} - \frac{\varepsilon}{(1-\gamma)m}$ for $j > l$, for $\gamma = \Theta(1)$, $(1-\gamma) = \Theta(1)$, and $\lambda_0 = n(1 + e^{-n/m}\varepsilon^2\theta)$, we have*

$$G(0) = \lambda_0 + me^{-\frac{n}{m}}\varepsilon^2\theta + \varepsilon^4\theta^2\left[-ne^{-2n/m} + m\frac{e^{-n/m}}{2} - m\frac{e^{-2n/m}}{2}\right] + \varepsilon^4\theta\frac{e^{-n/m}n^2}{2m\gamma(1-\gamma)}$$

Proof By definition of $G(\psi)$ in equation (75) and Lemma 63,

$$G(0) = \lambda_0 + \sum_{j=1}^{m}\log[1 + e^{-\lambda_0\nu_j}(e^{\varepsilon^2\theta} - 1)]$$

Now, since $e^{\varepsilon^2\theta} - 1 = \varepsilon^2\theta + \frac{\varepsilon^4\theta^2}{2} + O(\varepsilon^6)$, this is

$$\lambda_0 + \sum_{j=1}^{m}[e^{-\lambda_0\nu_j}(\varepsilon^2\theta + \frac{\varepsilon^4\theta^2}{2}) - \frac{\varepsilon^4\theta^2}{2}e^{-2\lambda_0\nu_j} + O(\varepsilon^6)]$$

Substituting in λ_0 and ν, this is

$$= \lambda_0 + me^{-\frac{n}{m}}\varepsilon^2\theta + \varepsilon^4\theta^2\left[-ne^{-2n/m} + m\frac{e^{-n/m}}{2} - m\frac{e^{-2n/m}}{2}\right] + \varepsilon^4\theta\frac{e^{-n/m}mn^2}{2m\gamma(1-\gamma)}$$

∎

Finally, we compute the MGF.

Lemma 71 *Under the uniform distribution p over $[m]$,*

$$M_{\tilde{S},p}\left(\frac{n^2\varepsilon^4}{m}\theta\right) = (1+O(1/n))\exp\left\{\frac{n^2\varepsilon^4}{m}\theta^2\left[-\frac{m}{n}\frac{e^{-\frac{2n}{m}}}{2}+\frac{m^2}{n^2}\frac{e^{-n/m}}{2}-\frac{m^2}{n^2}\frac{e^{-\frac{2n}{m}}}{2}\right]\right\}$$

and over distribution q such that $q_j = \frac{1}{m}+\frac{\varepsilon}{\gamma m}$ for $j \leq \gamma m$ and $\nu_j = \frac{1}{m}-\frac{\varepsilon}{(1-\gamma)m}$ for $j > \gamma m$, for $\gamma = \Theta(1)$ and $1-\gamma = \Theta(1)$, we have

$$M_{\tilde{S},q}\left(\frac{n^2\varepsilon^4}{m}\theta\right) = (1+O(1/n))\exp\left\{\frac{n^2\varepsilon^4}{m}\theta^2\left[-\frac{m}{n}\frac{e^{-\frac{2n}{m}}}{2}+\frac{m^2}{n^2}\frac{e^{-n/m}}{2}-\frac{m^2}{n^2}\frac{e^{-\frac{2n}{m}}}{2}\right]+\varepsilon^4\theta\frac{e^{-n/m}n^2}{2m\gamma(1-\gamma)}\right\}$$

Proof For the uniform distribution p, by (82), substituting $\lambda_0 = n(1+e^{-n/m}\varepsilon^2\theta)$ and $G(0)$ from Lemma 69, we have

$$M_{\tilde{S},p}\left(\frac{n^2\varepsilon^4}{m}\theta\right) = \exp(-me^{-n/m}\varepsilon^2\theta)\frac{e^n n!}{\sqrt{2\pi n}}(n(1+e^{-n/m}\varepsilon^2\theta))^{-n}$$

$$\exp\left(ne^{-n/m}\varepsilon^2\theta + me^{-n/m}\varepsilon^2\theta + \varepsilon^4\theta^2\left[-ne^{-2\frac{n}{m}}+m\frac{e^{-\frac{n}{m}}}{2}-m\frac{e^{-2\frac{n}{m}}}{2}\right]+O(m\varepsilon^6)\right)$$

$$= \frac{e^n n!}{\sqrt{2\pi n}}\exp(-n(\log n + \log(1+e^{-n/m}\varepsilon^2\theta)))$$

$$\exp\left(ne^{-n/m}\varepsilon^2\theta + \varepsilon^4\theta^2\left[-ne^{-2n/m}+m\frac{e^{-n/m}}{2}-m\frac{e^{-2n/m}}{2}\right]+O(m\varepsilon^6)\right)$$

$$= \frac{e^n n!}{n^n\sqrt{2\pi n}}\exp\left\{-n(e^{-n/m}\varepsilon^2\theta - \frac{e^{-2n/m}\varepsilon^4\theta^2}{2})\right\}$$

$$\exp\left\{ne^{-n/m}\varepsilon^2\theta + \varepsilon^4\theta^2\left[-ne^{-2n/m}+m\frac{e^{-n/m}}{2}-m\frac{e^{-2n/m}}{2}\right]+O(m\varepsilon^6)\right\}$$

$$= (1+O(1/n))\exp\left\{\frac{n^2\varepsilon^4}{m}\theta^2\left[-\frac{m}{n}\frac{e^{-\frac{2n}{m}}}{2}+\frac{m^2}{n^2}\frac{e^{-n/m}}{2}-\frac{m^2}{n^2}\frac{e^{-\frac{2n}{m}}}{2}\right]\right\}$$

by Stirling's approximation.

Similarly, for q as stated, we have the claim. ∎

H.3. Application of the Gärtner-Ellis Theorem

In this section, we apply the Gärtner-Ellis Theorem to obtain the probability that our statistic crosses a threshold, under the uniform distribution, and under one of the worst-case ε-far distributions.

Lemma 72 *When $n = \Theta(m)$, let $\alpha = n/m$. Then, under the uniform distribution p, we have that for $\tau > 0$,*

$$\lim_{n\to\infty} -\frac{m}{n^2\varepsilon^4}\log\left(\mathbb{P}_p\left[\tilde{S}\geq\tau\right]\right) = \frac{\tau^2\alpha^2 e^{2\alpha}}{2e^\alpha - 2 - 2\alpha}$$

Under an ε-far distribution q of the form $q_j = \frac{1}{m} + \frac{\varepsilon}{\gamma m}$ for $j \leq \gamma m$ and $q_j = \frac{1}{m} - \frac{\varepsilon}{(1-\gamma)m}$ for $j > \gamma m$, and $\gamma = \Theta(1), 1 - \gamma = \Theta(1)$, for $\tau < \frac{e^{-\alpha}}{2\gamma(1-\gamma)}$,

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_q \left[\tilde{S} \leq \tau \right] \right) = \frac{\alpha^2 (2\tau e^{\alpha} \gamma(\gamma - 1) + 1)^2}{8(-1 - \alpha + e^{\alpha})\gamma^2(\gamma - 1)^2}$$

Proof Note that by Lemma 71, the limiting logarithmic moment generating function with respect to the uniform distribution p is given by

$$\Lambda_p(\theta) = \lim_{n \to \infty} \frac{m}{n^2 \varepsilon^4} \log \left(M_{\tilde{S},p} \left(\frac{n^2 \varepsilon^4}{m} \theta \right) \right) = \theta^2 \left[-\frac{e^{-2\alpha}}{2\alpha} + \frac{e^{-\alpha}}{2\alpha^2} - \frac{e^{-2\alpha}}{2\alpha^2} \right]$$

Thus, Assumption 1 holds for $\mathcal{D}_{\Lambda_p} = \mathbb{R}$. Furthermore, the Fenchel-Legendre Transform (defined in equation 28) of Λ_p is given by

$$\Lambda_p^*(\tau) = \sup_\theta \left\{ \theta\tau - \theta^2 \left[-\frac{e^{-2\alpha}}{2\alpha} + \frac{e^{-\alpha}}{2\alpha^2} - \frac{e^{-2\alpha}}{2\alpha^2} \right] \right\} = \frac{\tau^2 \alpha^2 e^{2\alpha}}{2e^{\alpha} - 2 - 2\alpha}$$

This is a strongly convex function of τ, so the set of exposed points of Λ_p^* whose exposing hyperplane belongs to $\mathcal{D}_{\Lambda_p}^o$ is all of $\mathbb{R}$. Thus, by the Theorem 13 (Gärtner-Ellis), for $\tau > 0$,

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_p \left[\tilde{S} \geq \tau \right] \right) = \inf_{x \geq \tau} \Lambda_p^*(x) = \frac{\tau^2 \alpha^2 e^{2\alpha}}{2e^{\alpha} - 2 - 2\alpha}$$

Similarly, the limiting logarithmic moment generating function with respect to an alternate distribution q is given by

$$\Lambda_q(\theta) = \lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{E}_q \left[\exp \left(\frac{n^2 \varepsilon^4}{m} \theta \tilde{S} \right) \right] \right) = \theta^2 \left[-\frac{e^{-2\alpha}}{2\alpha} + \frac{e^{-\alpha}}{2\alpha^2} - \frac{e^{-2\alpha}}{2\alpha^2} \right] + \theta \frac{e^{-\alpha}}{2\gamma(1-\gamma)}$$

The Fenchel-Legendre transform is given by

$$\Lambda_q^*(\tau) = \sup_\theta \left\{ \theta\tau - \theta^2 \left[-\frac{e^{-2\alpha}}{2\alpha} + \frac{e^{-\alpha}}{2\alpha^2} - \frac{e^{-2\alpha}}{2\alpha^2} \right] - \theta \frac{e^{-\alpha}}{2\gamma(1-\gamma)} \right\} = \frac{\alpha^2 (2\tau e^{\alpha} \gamma(\gamma - 1) + 1)^2}{8(-1 - \alpha + e^{\alpha})\gamma^2(\gamma - 1)^2}$$

Again, applying the Gärtner-Ellis Theorem gives, for $\tau < \frac{e^{-\alpha}}{2\gamma(1-\gamma)}$,

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_q \left[\tilde{S}_n^* \leq \tau \right] \right) = \inf_{x \leq \tau} \Lambda_q^*(x) = \frac{\alpha^2 (2\tau e^{\alpha} \gamma(\gamma - 1) + 1)^2}{8(-1 - \alpha + e^{\alpha})\gamma^2(\gamma - 1)^2}$$

∎

Lemma 73 *When $n = o(m)$, under the uniform distribution p, we have for $\tau > 0$,*

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_p \left[\tilde{S} \geq \tau \right] \right) = \tau^2$$

and under an ε-far distribution q such that $q_j = \frac{1}{m} + \frac{\varepsilon}{\gamma m}$ for $j \leq \gamma m$ and $q_j = \frac{1}{m} - \frac{\varepsilon}{(1-\gamma)m}$ for $j > \gamma m$, for $\gamma = \Theta(1), 1 - \gamma = \Theta(1)$, and $\tau < \frac{1}{2\gamma(1-\gamma)}$,

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_q \left[\tilde{S} \leq \tau \right] \right) = \frac{(2\tau\gamma(\gamma - 1) + 1)^2}{4\gamma^2(\gamma - 1)^2}$$

Proof Since $n = o(m)$, we have by Lemma 71, Taylor expanding,

$$M_{\tilde{S},p}\left(\frac{n^2\varepsilon^4}{m}\right) = \exp\left\{\frac{n^2\varepsilon^4}{m}\theta^2\left[\frac{1}{4} + O(n/m)\right]\right\}$$

and similarly,

$$M_{\tilde{S},q}\left(\frac{n^2\varepsilon^4}{m}\right) = \exp\left\{\frac{n^2\varepsilon^4}{m}\theta\left[\frac{1}{4}\theta + \frac{1}{2\gamma(1-\gamma)} + O(n/m)\right]\right\}$$

Then, by a similar argument as in 72, the claim follows. ∎

H.4. Setting the threshold

We need to set our threshold τ so that the minimum of the error probability under the uniform distribution p, and any ε-far distribution q is maximized. Note that by Lemmas 72 and 73, it is sufficient to consider a threshold τ such that $0 < \tau < \frac{e^{-\alpha}}{2\gamma(1-\gamma)}$ when $n = \Theta(m)$ and $0 < \tau < \frac{1}{2\gamma(1-\gamma)}$ when $n = o(m)$, since otherwise, the error probability in one of the two cases is at least constant. To set our threshold, we will first observe that for any τ in this range, the "error exponent" under ε-far distributions is minimized for a particular ε-far distribution. Then, we will set the threshold to maximize the minimum of the error exponent under the uniform distribution, and under this ε-far distribution.

Lemma 74 *When $n = \Theta(m)$ so that $\alpha = n/m$, setting the threshold $\tau = e^{-\alpha}$, we have for the uniform distribution p,*

$$\lim_{n\to\infty} -\frac{m}{n^2\varepsilon^4}\log\left(\mathbb{P}_p\left[\tilde{S} \geq \tau\right]\right) = \frac{\alpha^2}{2(-1-\alpha+e^\alpha)}$$

and for any ε-far distribution q such that $q_j = 1/m + \frac{\varepsilon}{\gamma m}$ for $j \leq \gamma m$ and $q_j = 1/m - \frac{\varepsilon}{(1-\gamma)m}$ for $j > \gamma m$ and $\gamma = \Theta(1), 1-\gamma = \Theta(1)$,

$$\lim_{n\to\infty} -\frac{m}{n^2\varepsilon^4}\log\left(\mathbb{P}_q\left[\tilde{S} \leq \tau\right]\right) \geq \frac{\alpha^2}{2(-1-\alpha+e^\alpha)}$$

with equality for q such that $q_j = 1/m + 2\varepsilon/m$ for $j \leq m/2$ and $q_j = 1/m - 2\varepsilon/m$ for $j > m/2$.

Proof By Lemma 72, for $0 < \tau < \frac{e^{-\alpha}}{2\gamma(1-\gamma)}$,

$$\lim_{n\to\infty} \frac{m}{n^2\varepsilon^4}\log\left(\mathbb{P}_q\left[\tilde{S} \leq \tau\right]\right) = \frac{\alpha^2(2\tau e^\alpha\gamma(\gamma-1)+1)^2}{8(-1-\alpha+e^\alpha)\gamma^2(\gamma-1)^2}$$

Now, the numerator of the right hand side is minimized when $\gamma = 1/2$, and the denominator is maximized when $\gamma = 1/2$. Thus, the right hand side is minimized when $\gamma = 1/2$. So, we have that,

$$\lim_{n\to\infty} -\frac{m}{n^2\varepsilon^4}\log\left(\mathbb{P}_q\left[\tilde{S} \leq \tau\right]\right) \geq \frac{2\alpha^2(1-\frac{1}{2}\tau e^\alpha)^2}{(-1-\alpha+e^\alpha)}$$

with equality for distribution q such that $q_j = 1/m + 2\varepsilon/m$ for $j \leq m/2$ and $q_j = 1/m - 2\varepsilon/m$ for $j > m/2$.

Then, the claim follows by substituting in $\tau = e^{-\alpha}$ in the expression for the uniform distribution in Lemma 72 and in the above expression. ∎

Lemma 75 *When $n = o(m)$, setting the threshold $\tau = 1$, we have for the uniform distribution p,*

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_p \left[\tilde{S} \geq \tau \right] \right) = 1$$

and for any ε-far distribution q such that $q_j = 1/m + \frac{\varepsilon}{\gamma m}$ for $j \leq \gamma m$ and $q_j = 1/m - \frac{\varepsilon}{(1-\gamma)m}$ for $j > \gamma m$, and $\gamma = \theta(1), 1 - \gamma = \Theta(1)$,

$$\lim_{n \to \infty} -\frac{m}{n^2 \varepsilon^4} \log \left(\mathbb{P}_q \left[\tilde{S} \leq \tau \right] \right) \geq 1$$

with equality for distribution q such that $q_j = 1/m + 2\varepsilon/m$ for $j \leq m/2$ and $q_j = 1/m - 2\varepsilon/m$ for $j > m/2$.

Proof Follows from Lemma 73 and using a similar argument as in Lemma 74. ∎

Finally, we have our results.

Theorem 5 (TV) *The TV statistic uses*

$$n = (1 + o(1)) \sqrt{\frac{2(e^{n/m} - 1 - n/m)}{(n/m)^2}} \frac{\sqrt{m \log \frac{1}{\delta}}}{\varepsilon^2}$$

for $n \leq m$, $n \gg 1$, and $\varepsilon, \delta \ll 1$.

Proof Recall that our statistic is equivalent to the TV tester when $n \leq m$. When $n = \Theta(m)$, by Lemma 74, we have that $\bar{c}(\varepsilon, m, C) = 1$ for every ε that satisfies our assumptions, and every $C > 2$. In particular, any ε' such that $\left(1 - \frac{1}{C'}\right) \varepsilon(n) \leq \varepsilon'(n) \leq \varepsilon(n)$ has $\bar{c}(\varepsilon', m, C) = 1$ for every $C > 2$. Thus, by Lemma 21, we have that $c(\varepsilon, m) = 1$ for every ε that satisfies our assumptions.

So, we have shown that our tester fails with probability $e^{-\frac{n^2 \varepsilon^4}{m}(\xi + o(1))}$ for

$$\xi = \frac{\alpha^2}{2(-1 - \alpha + e^\alpha)}$$

where $\alpha = n/m$. This implies that the TV tester uses $n = \frac{\sqrt{m \log(1/\delta)}}{\varepsilon^2}(c + o(1))$ samples for failure probability δ, where

$$c = \frac{\sqrt{2(e^\alpha - 1 - \alpha)}}{\alpha}$$

as required. Now, $c > 1$ for $0 \leq \alpha \leq 1$. The claim for the $n = \Theta(m)$ case follows. By a similar argument using Lemma 75, the claim for the $n = o(m)$ case follows. ∎

H.5. Empty bins lemmas

We assume that Assumption 5 holds throughout this section.

Lemma 76 *Suppose Assumption 5 holds.*

$$\sum_{k=0}^{\infty} \left\{ \frac{(\lambda_0 \nu_j e^{i\psi})^k}{k!} e^{-\lambda_0 \nu_j e^{i\psi}} \exp(\varepsilon^2 \theta \mathbb{1}_{\{k=0\}}) \right\} = 1 + O(\varepsilon^2)$$

for $\lambda_0 = n(1 + O(\varepsilon))$ and $\nu_j = 1/m + O(\varepsilon/m)$ for all j,

Proof By Lemma 63,

$$\sum_{k=0}^{\infty} \left\{ \frac{(\lambda_0 \nu_j e^{i\psi})^k}{k!} e^{-\lambda_0 \nu_j e^{i\psi}} \exp(\varepsilon^2 \theta \mathbb{1}_{\{k=0\}}) \right\} = e^{-\lambda_0 \nu_j e^{i\psi}} e^{\varepsilon^2 \theta} - e^{-\lambda_0 \nu_j e^{i\psi}} + 1$$

$$= e^{-\lambda_0 \nu_j e^{i\psi}} (\varepsilon^2 \theta + O(\varepsilon^4)) + 1$$

Substituting λ_0 and ν_j, this is

$$1 + e^{-\frac{n}{m} e^{i\psi}(1+O(\varepsilon))} (\varepsilon^2 \theta + O(\varepsilon^4))$$

$$= 1 + O(\varepsilon^2)$$

∎

Lemma 77 *Suppose Assumption 5 holds. For $\lambda_0 = n(1 + e^{-n/m}\varepsilon^2\theta)$, $\nu_j = \frac{1}{m} + O\left(\frac{\varepsilon}{m}\right)$ for all j,*

$$\prod_{j=1}^{m} \left\{ e^{-\lambda_0 \nu_j e^{i\psi}} e^{\varepsilon^2 \theta} - e^{-\lambda_0 \nu_j e^{i\psi}} + 1 \right\} = \exp \left\{ m e^{-\frac{n}{m} e^{i\psi}} \varepsilon^2 \theta (1 + O(\varepsilon^2)) + O(n\varepsilon^3) \right\}$$

Proof By Lemma 63,

$$\prod_{j=1}^{m} \left\{ e^{-\lambda_0 \nu_j e^{i\psi}} e^{\varepsilon^2 \theta} - e^{-\lambda_0 \nu_j e^{i\psi}} + 1 \right\} = \prod_{j=1}^{m} \left\{ e^{-\lambda_0 \nu_j e^{i\psi}} (\varepsilon^2 \theta + O(\varepsilon^4)) + 1 \right\}$$

$$= \exp \left\{ \sum_{j=1}^{m} \log \left[e^{-\lambda_0 \nu_j e^{i\psi}} (\varepsilon^2 \theta + O(\varepsilon^4)) + 1 \right] \right\}$$

Substituting $\lambda_0 = n(1 + O(\varepsilon^2))$ and $\nu_j = 1/m + O(\varepsilon/m)$, since $n = O(m)$, this is

$$= \exp \left\{ \sum_{j=1}^{m} \log \left[e^{-\frac{n}{m} e^{i\psi}(1+O(\varepsilon))} (\varepsilon^2 \theta + O(\varepsilon^4)) + 1 \right] \right\}$$

$$= \exp \left\{ m \left[e^{-\frac{n}{m} e^{i\psi}(1+O(\varepsilon))} (\varepsilon^2 \theta + O(\varepsilon^4)) \right] \right\}$$

$$= \exp \left\{ m e^{-\frac{n}{m} e^{i\psi}} \varepsilon^2 \theta (1 + O(\varepsilon^2)) + O(n\varepsilon^3) \right\}$$

∎

Lemma 78 *Suppose Assumption 5 holds. For $\lambda_0 = n(1 + e^{-n/m}\varepsilon^2\theta)$, $\nu_j = \frac{1}{m} + O(\frac{\varepsilon}{m})$, and*

$$f(k) = \exp(\varepsilon^2\theta\mathbb{1}_{\{k=0\}})$$

and

$$G(\psi) = -in\psi + \lambda_0 e^{i\psi} + \sum_{j=1}^{m} \log\left\{e^{-\lambda_0\nu_j e^{i\psi}}e^{\varepsilon^2\theta} - e^{\lambda_0\nu_j e^{i\psi}} + 1\right\}$$

we have

$$G''(\psi) = -ne^{i\psi} + O(n\varepsilon^2)$$

Proof First, note that for $c > 0$, since $\mathbb{1}_{\{k+c=0\}} = 0$ for every $k \geq 0$, $f(k+c) = 1$, for every $k \geq 0$. So,

$$\sum_{k=0}^{\infty}\left\{\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}e^{-\lambda_0\nu_j e^{i\psi}}f(k+c)\right\} = 1$$

By Lemma 76, we have that

$$\sum_{k=0}^{\infty}\left\{\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}e^{-\lambda_0\nu_j e^{i\psi}}f(k)\right\} = 1 + O(\varepsilon^2)$$

So, we have

$$\left(\sum_{k=0}^{\infty}\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}e^{-\lambda_0\nu_j e^{i\psi}}f(k+1)\right)^2$$
$$-\left(\sum_{k=0}^{\infty}\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}e^{-\lambda_0\nu_j e^{i\psi}}f(k)\right)\left(\sum_{k=0}^{\infty}\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}e^{-\lambda_0\nu_j e^{i\psi}}f(k+2)\right) = O(\varepsilon^2)$$

and

$$\left(\sum_{k=0}^{\infty}\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}e^{-\lambda_0\nu_j e^{i\psi}}f(k)\right)\left(\sum_{k=0}^{\infty}\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}e^{-\lambda_0\nu_j e^{i\psi}}f(k+2)\right) = 1 + O(\varepsilon^2)$$

and

$$\left(\sum_{k=0}^{\infty}\frac{(\lambda_0\nu_j e^{i\psi})^k}{k!}e^{-\lambda_0\nu_j e^{i\psi}}f(k)\right)^2 = 1 + O(\varepsilon^2)$$

So, by Lemma 51, we have that

$$G''(\psi) = \sum_{j=1}^{m}\frac{(\lambda_0\nu_j e^{i\psi})^2 O(\varepsilon^2) - (\lambda_0\nu_j e^{i\psi})(1 + O(\varepsilon^2))}{1 + O(\varepsilon^2)}$$

$$= \sum_{j=1}^{m}\left\{O(\frac{n^2\varepsilon^2}{m^2}) - (\lambda_0\nu_j e^{i\psi})(1 + O(\varepsilon^2))\right\}(1 + O(\varepsilon^2))$$

$$= -(\lambda_0 e^{i\psi})(1 + O(\varepsilon^2)) + O\left(\frac{n^2\varepsilon^2}{m}\right) = -ne^{i\psi} + O(n\varepsilon^2)$$

since $n = O(m)$ and $\lambda_0 = n(1 + O(\varepsilon^2))$. ∎