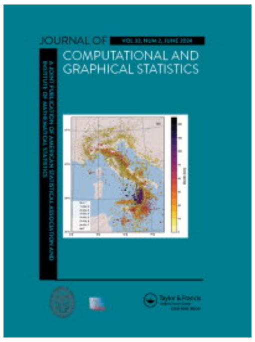


Taylor & Francis
Taylor & Francis Group

Journal of Computational and Graphical Statistics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ucgs20

Universal Inference Meets Random Projections: A Scalable Test for Log-Concavity

Robin Dunn, Aditya Gangrade, Larry Wasserman & Aaditya Ramdas

To cite this article: Robin Dunn, Aditya Gangrade, Larry Wasserman & Aaditya Ramdas (31 May 2024): Universal Inference Meets Random Projections: A Scalable Test for Log-Concavity, Journal of Computational and Graphical Statistics, DOI: 10.1080/10618600.2024.2347338

To link to this article: https://doi.org/10.1080/10618600.2024.2347338

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2024, VOL. 00, NO. 0, 1–13
https://doi.org/10.1080/10618600.2024.2347338

Universal Inference Meets Random Projections: A Scalable Test for Log-Concavity

Robin Dunn[a], Aditya Gangrade[b,c], Larry Wasserman[d], and Aaditya Ramdas[d]

[a]Advanced Methodology and Data Science, Novartis Pharmaceuticals Corporation, East Hanover, NJ; [b]Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI; [c]Department of Electrical & Computer Engineering, Boston University, Boston, MA; [d]Department of Statistics & Data Science and Machine Learning Department, Carnegie Mellon University, Pittsburgh, PA

ABSTRACT
Shape constraints yield flexible middle grounds between fully nonparametric and fully parametric approaches to modeling distributions of data. The specific assumption of log-concavity is motivated by applications across economics, survival modeling, and reliability theory. However, there do not currently exist valid tests for whether the underlying density of given data is log-concave. The recent universal inference methodology provides a valid test. The universal test relies on maximum likelihood estimation (MLE), and efficient methods already exist for finding the log-concave MLE. This yields the first test of log-concavity that is provably valid in finite samples in any dimension, for which we also establish asymptotic consistency results. Empirically, we find that a random projections approach that converts the *d*-dimensional testing problem into many one-dimensional problems can yield high power, leading to a simple procedure that is statistically and computationally efficient.

ARTICLE HISTORY
Received May 2023
Accepted April 2024

KEYWORDS
Density estimation; Finite-sample validity; Hypothesis testing; Shape constraints

1. Introduction

Statisticians frequently use density estimation to understand the underlying structure of their data. To perform nonparametric density estimation on a sample, it is common for researchers to incorporate shape constraints (Carroll, Delaigle, and Hall 2011; Koenker and Mizera 2018). Log-concavity is one popular choice of shape constraint; a density f is called log-concave if it has the form $f = e^g$ for some concave function g. This class of densities encompasses many common families, such as the normal, uniform (over a compact domain), exponential, logistic, and extreme value densities (Bagnoli and Bergstrom 2005, Table 1). Furthermore, specifying that the density is log-concave poses a middle ground between fully nonparametric density estimation and use of a parametric density family. As noted in Cule, Samworth, and Stewart (2010), log-concave density estimation does not require the choice of a bandwidth, whereas kernel density estimation in d dimensions requires a $d \times d$ bandwidth matrix.

Log-concave densities have multiple appealing properties; An (1997) describes several. For example, log-concave densities are unimodal, they have at most exponentially decaying tails (i.e., $f(x) = O(\exp(-c\|x\|))$ for some $c > 0$), and all moments of the density exist. Log-concave densities are also closed under convolution, meaning that if X and Y are independent random variables from log-concave densities, then the density of $X+Y$ is log-concave as well. A unimodal density f is strongly unimodal if the convolution of f with any unimodal density g is unimodal. Proposition 2 of An (1997) states that a density f is log-concave if and only if f is strongly unimodal.

In addition, log-concave densities have applications in many domains. Bagnoli and Bergstrom (2005) describe applications of log-concavity across economics, reliability theory, and survival modeling. (The latter two appear to use similar methods in the different domains of engineering and medicine, respectively.) Suppose a survival density function f is defined on (a, b) and has a survival function (or reliability function) $\bar{F}(x) = \int_x^b f(t)dt$. If f is log-concave, then its survival function is log-concave as well. The failure rate associated with f is $r(x) = f(x)/\bar{F}(x) = -\bar{F}'(x)/\bar{F}(x)$. Corollary 2 of Bagnoli and Bergstrom (2005) states that if f is log-concave on (a, b), then the failure rate $r(x)$ is monotone increasing on (a, b). Proposition 12 of An (1997) states that if a survival function $\bar{F}(x)$ is log-concave, then for any pair of nonnegative numbers x_1, x_2, the survival function satisfies $\bar{F}(x_1 + x_2) \leq \bar{F}(x_1)\bar{F}(x_2)$. This property is called the new-is-better-than-used property; it implies that the probability that a new unit will survive for time x_1 is greater than or equal to the probability that at time x_2, an existing unit will survive an additional time x_1.

Given the favorable properties of log-concave densities and their applications across fields, it is important to be able to test the log-concavity assumption. Previous researchers have considered this question as well. Cule, Samworth, and Stewart (2010) develop a permutation test based on simulating from the log-concave MLE and computing the proportion of original and simulated observations in spherical regions. Chen and Samworth (2013) construct an approach similar to the permutation test, using a test statistic based on covariance matrices. Hazelton (2011) develops a kernel bandwidth test, where the test statistic

CONTACT Robin Dunn robin.dunn@novartis.com Advanced Methodology and Data Science, Novartis Pharmaceuticals Corporation, East Hanover, NJ.

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JCGS.

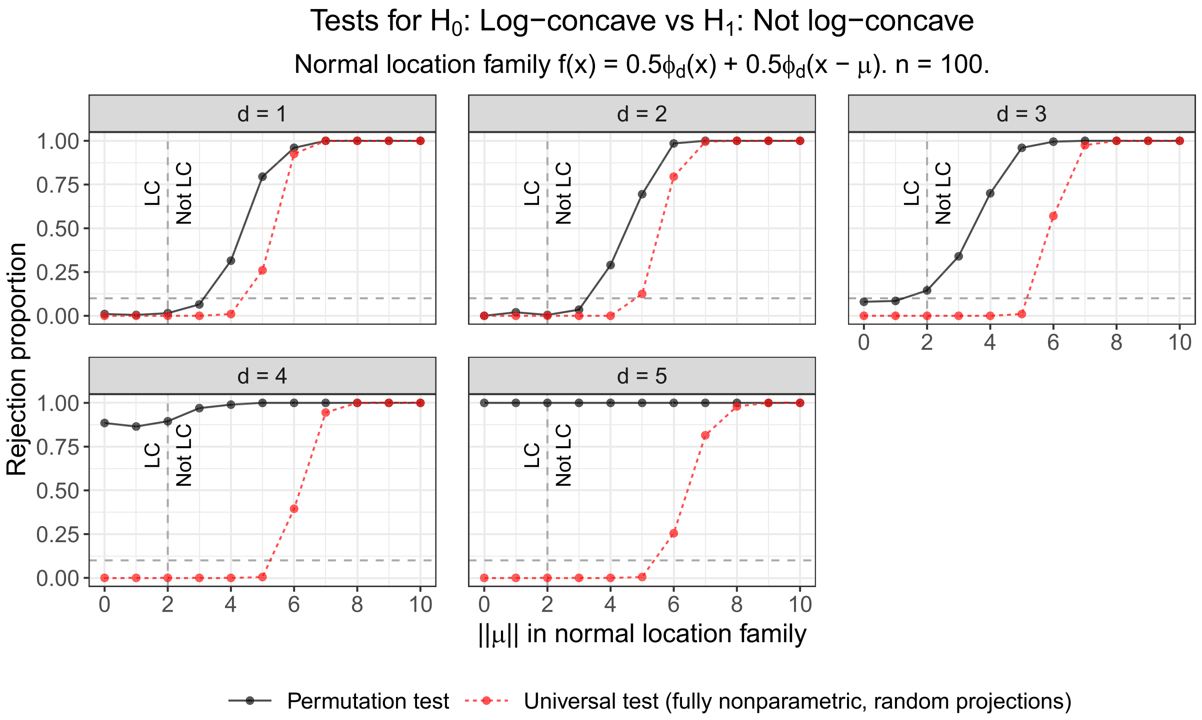


Figure 1. Rejection proportions for tests of $H_0 : f^*$ is log-concave versus $H_1 : f^*$ is not log-concave. The permutation test from Cule, Samworth, and Stewart (2010) is valid or approximately valid for $d \le 3$, but it is not valid for $d \ge 4$. Our test that combines random projections and universal inference (Algorithm 3) is provably valid for all n and d while having high power.

is the smallest kernel bandwidth that produces a log-concave density. Carroll, Delaigle, and Hall (2011) construct a metric for the necessary amount of modification to the weights of a kernel density estimator to satisfy the shape constraint of log-concavity, and they use the bootstrap for calibration. While these approaches exhibit reasonable empirical performance in some settings, none of the aforementioned papers have proofs of validity (or asymptotic validity) for their proposed methods. As one exception, An (1997) uses asymptotically normal test statistics to test implications of log-concavity (e.g., increasing hazard rate) in the univariate, nonnegative setting. A general valid test for log-concavity has proved elusive.

As an example, Figure 1 shows the performance of the permutation test of log-concavity from Cule, Samworth, and Stewart (2010) and a random projection variant of our universal test. Section 3.1 explains the details of the permutation test, and Algorithm 3 explains this universal test. Section 4 provides more extensive simulations. If ϕ_d represents the $N(0, I_d)$ density and $\gamma \in (0, 1)$, then $f(x) = \gamma\phi_d(x)+(1-\gamma)\phi_d(x-\mu)$ is log-concave only when $\|\mu\| \le 2$. We simulate the permutation test in the $\gamma = 1/2$ setting for $1 \le d \le 5$ and $n = 100$, testing the null hypothesis that the true underlying density f^* is log-concave. We set $\mu = (\mu_1, 0, \ldots, 0)$, so that $\|\mu\| = |\mu_1|$. We use a significance level of $\alpha = 0.10$. Each point represents the proportion of times we reject H_0 over 200 simulations. Figure 1 shows that the permutation test is valid at $d = 1$ and $d = 2$ and approximately valid at $d = 3$. Alternatively, at $d = 4$ and $d = 5$, this test rejects H_0 at proportions much higher than α, even when the underlying density is log-concave ($\|\mu\| \le 2$). In contrast, the universal test is valid in all dimensions. Furthermore, this universal test has high power for reasonable $\|\mu\|$ even as we increase d.

To develop a test for log-concavity with validity guarantees, we consider the universal likelihood ratio test (LRT) introduced in Wasserman, Ramdas, and Balakrishnan (2020). This approach provides valid hypothesis tests in any setting in which we can maximize (or upper bound) the null likelihood. Importantly, validity holds in finite samples and without regularity conditions on the class of models. Thus, it holds even in high-dimensional settings without assumptions.

Suppose $\mathcal{F}_d$ is a (potentially nonparametric) class of densities in d dimensions. The universal LRT allows us to test hypotheses of the form $H_0 : f^* \in \mathcal{F}_d$ versus $H_1 : f^* \notin \mathcal{F}_d$. In this article, $\mathcal{F}_d$ will represent the class of all log-concave densities in d dimensions.

Assume we have n iid observations $Y_1, \ldots, Y_n$ with some true density f^*. To implement the split universal LRT, we randomly partition the indices from 1 to n, denoted as $[n]$, into $\mathcal{D}_0$ and $\mathcal{D}_1$. (Our simulations assume $|\mathcal{D}_0| = |\mathcal{D}_1| = n/2$, but any split proportion is valid.) Using the data indexed by $\mathcal{D}_1$, we fit *any* density $\widehat{f}_1$ of our choice, such as a kernel density estimator. The likelihood function evaluated on a density f over the data indexed by $\mathcal{D}_0$ is denoted $\mathcal{L}_0(f) = \prod_{i\in\mathcal{D}_0} f(Y_i)$. Using the data indexed by $\mathcal{D}_0$, we fit $\widehat{f}_0 = \arg\max_{f\in\mathcal{F}_d} \mathcal{L}_0(f)$, which is the null maximum likelihood estimator (MLE). The split LRT statistic is

$$T_n(f) = \mathcal{L}_0(\widehat{f}_1)/\mathcal{L}_0(f).$$

The test rejects if $T_n(\widehat{f}_0) \ge 1/\alpha$.

Theorem 1. (Wasserman, Ramdas, and Balakrishnan 2020) $T_n(\widehat{f}_0)$ is an e-value, meaning that it has expectation at most one under the null. Hence, $1/T_n(\widehat{f}_0)$ is a valid p-value, and rejecting the null when $T_n(\widehat{f}_0) \geq 1/\alpha$ is a valid level-α test. That is, under $H_0 : f^* \in \mathcal{F}_d$,

$$\mathbb{P}(T_n(\widehat{f}_0) \geq 1/\alpha) \leq \alpha.$$

Wasserman, Ramdas, and Balakrishnan (2020) prove Theorem 1, but Appendix A.1 contains a proof for completeness. It is also possible to invert this test, yielding a confidence set for f^*, but for nonparametric classes $\mathcal{F}_d$, these are not in closed form and are hard to compute numerically, so we do not pursue this direction further. Nevertheless, as long as we are able to construct $\widehat{f}_0$ (or actually simply calculate or upper bound its likelihood), it is possible to perform the nonparametric hypothesis test described in Theorem 1.

Prior to the universal LRT developed by Wasserman, Ramdas, and Balakrishnan (2020), there was no hypothesis test for $H_0 : f^*$ is log-concave versus $H_1 : f^*$ is not log-concave with finite sample validity, or even asymptotic validity. Since it is possible to compute the log-concave MLE on any sample of size $n \geq d + 1$, the universal LRT described above provides a valid test as long as $|\mathcal{D}_0| \geq d + 1$. The randomization in the splitting above can be entirely removed—without affecting the validity guarantee—at the expense of more computation. Wasserman, Ramdas, and Balakrishnan (2020) show that one can repeatedly compute $T_n(\widehat{f}_0)$ under independent random splits and average all the test statistics; since each has expectation at most one under the null, so does their average. It follows that the test based on averaging over multiple splits still has finite sample validity.

Section 2 reviews critical work on the construction and convergence of log-concave MLE densities. Section 3 describes the permutation test from Cule, Samworth, and Stewart (2010) and proposes several universal tests for log-concavity. The log-concave MLE does suffer from a curse of dimensionality, both computationally and statistically. Hence, our most important contribution is a scalable method using random projections to reduce the multivariate problem into many univariate testing problems, where the log-concave MLE is easy to compute. (This relies on the fact that if a density is log-concave then every projection is also log-concave.) Section 4 compares these tests through a simulation study. Section 5 explains a theoretical result about the power of the universal LRT for tests of log-concavity. All proofs and several additional simulations are available in the appendices. Code to reproduce all analyses is available at *https://github.com/RobinMDunn/LogConcaveUniv*.

2. Finding the Log-Concave MLE

Suppose we observe an iid sample $X_1, \ldots, X_n \in \mathbb{R}^d$ from a d-dimensional density f^*, where $n \geq d + 1$. Recall that $\mathcal{F}_d$ is the class of log-concave densities in d dimensions. The log-concave MLE is $\widehat{f}_n = \underset{f \in \mathcal{F}_d}{\arg\max} \sum_{i=1}^n \log\{f(X_i)\}$. Theorem 1 of Cule, Samworth, and Stewart (2010) states that with probability 1, $\widehat{f}_n$ exists and is unique. Importantly, this does not require $f^* \in \mathcal{F}_d$.

The construction of $\widehat{f}_n$ relies on the concept of a tent function $\bar{h}_y : \mathbb{R}^d \to \mathbb{R}$. For a given vector $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ and given the sample $X_1, \ldots, X_n$, the tent function $\bar{h}_y$ is the smallest concave function that satisfies $\bar{h}_y(X_i) \geq y_i$ for $i = 1, \ldots, n$. Let C_n be the convex hull of the observations $X_1, \ldots, X_n$. Consider the objective function

$$\sigma(y_1, \ldots, y_n) = -\frac{1}{n}\sum_{i=1}^n y_i + \int_{C_n} \exp\{\bar{h}_y(x)\}dx.$$

Theorem 2 of Cule, Samworth, and Stewart (2010) states that σ is a convex function, and it has a unique minimum at the value $y^* \in \mathbb{R}^n$ that satisfies $\log(\widehat{f}_n) = \bar{h}_{y^*}$.

Thus, to find the tent function that defines the log-concave MLE, we need to minimize σ over $y \in \mathbb{R}^n$. σ is not differentiable, but Shor's algorithm (Shor 2012) uses a subgradient method to optimize convex, non-differentiable functions. This method is guaranteed to converge, but convergence can be slow. Shor's r-algorithm involves some computational speed-ups over Shor's algorithm, and Cule, Samworth, and Stewart (2010) use this algorithm in their implementation. Shor's r-algorithm is not guaranteed to converge, but Cule, Samworth, and Stewart (2010) state that they agree with Kappel and Kuntsevich (2000) that the algorithm is "robust, efficient, and accurate." The `LogConcDEAD` package for log-concave density estimation in arbitrary dimensions implements this method (Cule, Gramacy, and Samworth 2009).

Alternatively, the `logcondens` package implements an active set approach to solve for the log-concave MLE in one dimension (Dümbgen and Rufibach 2011). This algorithm is based on solving for a vector that satisfies a set of active constraints and then using the tent function structure to compute the log-concave density associated with that vector. See Section 3.2 of Dümbgen, Hüsler, and Rufibach (2007) for more details.

Figure 2 shows the true f^* and log-concave MLE $(\widehat{f}_n)$ densities of several samples from two-component Gaussian mixtures. The underlying density is $f^*(x) = 0.5\phi_d(x) + 0.5\phi_d(x - \mu)$. Again, this density is log-concave if and only if $\|\mu\| \leq 2$. (We develop this example further in Section 4.) In the $n = 5000$ and $d = 1$ setting, we simulate samples $X_1, \ldots, X_n \sim f^*$ and compute the log-concave MLE $\widehat{f}_n$ on each random sample. These simulations use both the `LogConcDEAD` and `logcondens` packages to fit $\widehat{f}_n$. `logcondens` only works in one dimension but is much faster than `LogConcDEAD`. The two packages produce densities with similar appearances. Furthermore, we include values of $n^{-1}\sum_{i=1}^n \log(f^*(x_i))$ on the true density plots and $n^{-1}\sum_{i=1}^n \log(\widehat{f}_n(x_i))$ on the log-concave MLE plots. The log-likelihood is approximately the same for the two density estimation methods.

In the first two rows of Figure 2, the true density is log-concave and in this case, we see that $n^{-1}\sum_{i=1}^n \log(\widehat{f}_n(x_i))$ is approximately equal to $n^{-1}\sum_{i=1}^n \log(f^*(x_i))$. When $\|\mu\| = 4$, the underlying density is not log-concave. The log-concave MLE at $\|\mu\| = 4$ and $n = 5000$ seems to have normal tails, but it is nearly uniform in the middle.

Appendix B.1 contains additional plots of the true densities and log-concave MLE densities when $n = 50$ and $d = 1$, $n = 50$ and $d = 2$, and $n = 500$ and $d = 2$. In the smaller sample $d = 1$ setting, we still observe agreement between `LogConcDEAD` and `logcondens`. When $d = 2$ and the true density is log-concave, the log-concave MLE is closer to the true density at

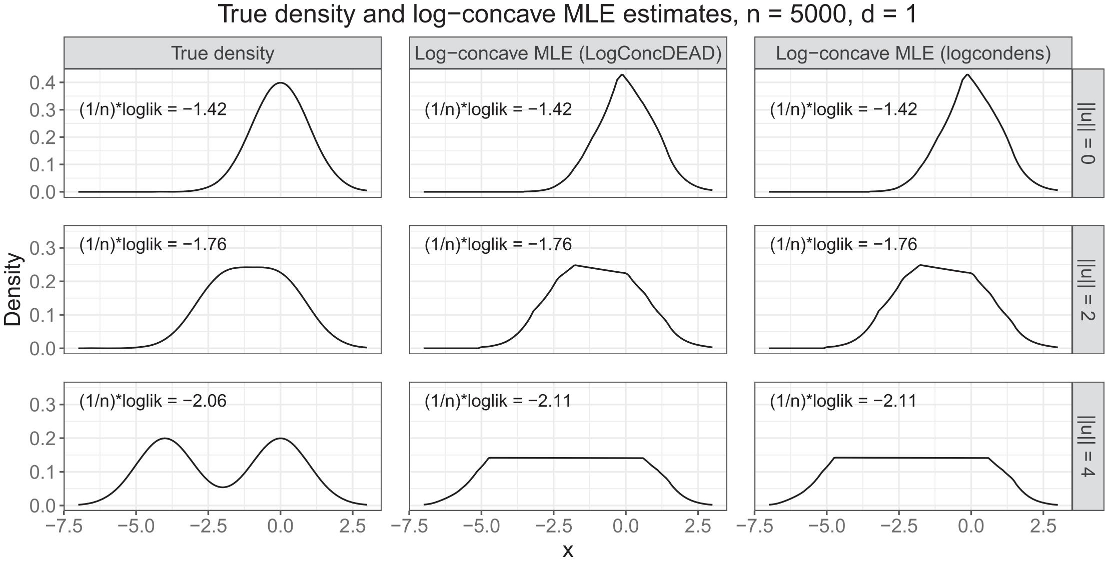


Figure 2. Densities from fitting log-concave MLE on $n = 5000$ observations. The true density is the Normal mixture $f^*(x) = 0.5\phi_1(x) + 0.5\phi_1(x - \mu)$. In all settings, the `LogConcDEAD` and `logcondens` packages return similar results. In the $\|\mu\| = 0$ and $\|\mu\| = 2$ log-concave settings, the log-concave MLE is close to the true density. In the $\|\mu\| = 4$ non-log-concave setting, the log-concave densities appear to have normal tails and uniform centers.

larger n. Alternatively, when $d = 2$ and the true density is not log-concave, the log-concave MLE density again appears to be uniform in the center.

Cule, Samworth, and Stewart (2010) formalize the convergence of $\widehat{f}_n$. Let $D_{\mathrm{KL}}(g\|f)$ be the Kullback-Leibler (KL) divergence of g from f. Define $f^{\mathrm{LC}} = \underset{f \in \mathcal{F}_d}{\arg\min}\ D_{\mathrm{KL}}(f^*\|f)$ as the log-concave projection of f^* onto the set of all log-concave densities $\mathcal{F}_d$ (Samworth 2018; Barber and Samworth 2021). In the simplest case, if $f^* \in \mathcal{F}_d$, then $f^{\mathrm{LC}} = f^*$. Regardless of whether $f^* \in \mathcal{F}_d$, suppose f^* satisfies the following conditions: $\int_{\mathbb{R}^d} \|x\| f^*(x)dx < \infty$, $\int_{\mathbb{R}^d} f^* \log_+(f^*) < \infty$ (where $\log_+(x) = \max\{\log(x), 0\}$), and the support of f^* contains an open set. By Lemma 1 of Cule and Samworth (2010), there exists some $a_0 > 0$ and $b_0 \in \mathbb{R}$ such that $f^{\mathrm{LC}}(x) \le \exp(-a_0\|x\| + b_0)$ for any $x \in \mathbb{R}^d$. Theorem 3 of Cule, Samworth, and Stewart (2010) states that for any $a < a_0$,

$$\int_{\mathbb{R}^d} \exp(a\|x\|)|\widehat{f}_n(x) - f^{\mathrm{LC}}(x)|dx \to 0 \qquad \text{almost surely.}$$

This means that the integrated difference between $\widehat{f}_n$ and f^{LC} converges to 0 even when we multiply the tails by some exponential weight. Furthermore, Theorem 3 of Cule, Samworth, and Stewart (2010) states that if f^{LC} is continuous, then

$$\sup_{x \in \mathbb{R}^d} \left\{\exp(a\|x\|)|\widehat{f}_n(x) - f^{\mathrm{LC}}(x)|\right\} \to 0 \qquad \text{almost surely.}$$

In the case where $f^* \in \mathcal{F}_d$, it is possible to describe rates of convergence of the log-concave MLE in terms of the Hellinger distance. The squared Hellinger distance is

$$h^2(f, g) = \int_{\mathbb{R}^d} (f^{1/2} - g^{1/2})^2.$$

As stated in Chen, Mazumder, and Samworth (2021) and shown in Kim and Samworth (2016) and Kur, Dagan, and Rakhlin (2019), the rate of convergence of $\widehat{f}_n$ to f^* in squared Hellinger distance is

$$\sup_{f^* \in \mathcal{F}_d} \mathbb{E}[h^2(\widehat{f}_n - f^*)] \le K_d \cdot \begin{cases} n^{-4/5} & d = 1 \\ n^{-2/(d+1)} \log(n) & d \ge 2 \end{cases},$$

where $K_d > 0$ depends only on d.

3. Tests for Log-Concavity

We first describe a permutation test as developed in Cule, Samworth, and Stewart (2010), and then we propose several universal inference tests. The latter are guaranteed to control the Type I error at level α (theoretically and empirically), while the former is not always valid even in simulations, as already demonstrated in Figure 1.

3.1. Permutation Test (Cule, Samworth, and Stewart 2010)

Cule, Samworth, and Stewart (2010) describe a permutation test of the hypothesis $H_0 : f^* \in \mathcal{F}_d$ versus $H_1 : f^* \notin \mathcal{F}_d$. First, this test fits the log-concave MLE $\widehat{f}_n$ on $\mathcal{Y} = \{Y_1, \ldots, Y_n\}$. Then it draws another sample $\mathcal{Y}^* = \{Y_1^*, \ldots, Y_n^*\}$ from $\widehat{f}_n$. Next, it computes a test statistic based on the empirical distributions of $\mathcal{Y}$ and $\mathcal{Y}^*$. As the permutation step, the procedure repeatedly "shuffles the stars" to permute the observations in $\mathcal{Y} \cup \mathcal{Y}^*$ into two sets of size n, and it recomputes the test statistic on each permuted sample. We reject H_0 if the original test statistic exceeds the $1 - \alpha$ quantile of the test statistics computed from the permuted samples. We explain the permutation test in more detail in Appendix E.1.

Intuitively, this test assumes that if H_0 is true, $\mathcal{Y}$ and $\mathcal{Y}^*$ will be similar. Then the original test statistic will not be particularly large relative to the test statistics computed from the permuted samples. Alternatively, if H_0 is false, $\mathcal{Y}$ and $\mathcal{Y}^*$ will be dissimilar, and the converse will hold. This approach is not guaranteed to control the Type I error level. Figure 1 shows cases both where

Algorithm 1 For $H_0 : f^* \in \mathcal{F}_d$ versus $H_1 : f^* \notin \mathcal{F}_d$, compute the subsampling test statistic or run the test.

```
Input: $n$ iid $d$-dimensional observations $Y_1, \dots, Y_n$ from unknown density $f^*$,
    number of subsamples $B$, significance level $\alpha$, any density estimation approach.
Output: The subsampling test statistic $T_n$ or the test result.
for $b = 1, 2, \dots, B$ do
    Randomly partition $[n]$ into $\mathcal{D}_{0,b}$ and $\mathcal{D}_{1,b}$ such that $|\mathcal{D}_{0,b}| = |\mathcal{D}_{1,b}| = n/2$.
    Where $\mathcal{L}_{0,b}(f) = \prod_{i \in \mathcal{D}_{0,b}} f(Y_i)$, compute $\widehat{f}_{0,b} = \arg\max_{f \in \mathcal{F}_d} \mathcal{L}_{0,b}(f)$.
    Fit a density $\widehat{f}_{1,b}$ on $\{Y_i : i \in \mathcal{D}_{1,b}\}$, using the input density estimation approach.
    Compute $T_{n,b} = \mathcal{L}_{0,b}(\widehat{f}_{1,b}) / \mathcal{L}_{0,b}(\widehat{f}_{0,b})$.
    if $B^{-1} \sum_{j=1}^{b} T_{n,j} \geq 1/\alpha$ then return rejection of hypothesis.
return the subsampling test statistic $T_n = B^{-1} \sum_{j=1}^{B} T_{n,j}$.
```

the permutation test performs well and where the permutation test's false positive rate is much higher than α.

3.2. Universal Tests in d Dimensions

Alternatively, we can use universal approaches to test for log-concavity. Theorem 1 justifies the universal approach for testing whether $f^* \in \mathcal{F}_d$. Recall that the universal LRT provably controls the Type I error level in finite samples. To implement the universal test on a single subsample, we partition $[n]$ into $\mathcal{D}_0$ and $\mathcal{D}_1$. Let $\widehat{f}_0$ be the maximum likelihood log-concave density estimate fit on $\{Y_i : i \in \mathcal{D}_0\}$. Let $\widehat{f}_1$ be any density estimate fit on $\{Y_i : i \in \mathcal{D}_1\}$. The universal test rejects H_0 when

$$T_n = \prod_{i \in \mathcal{D}_0} \{\widehat{f}_1(Y_i) / \widehat{f}_0(Y_i)\} \geq 1/\alpha.$$

The universal test from Theorem 1 holds when T_n is replaced with an average of test statistics, each computed over random partitions of $[n]$. Algorithm 1 explains how to use subsampling to test $H_0 : f^* \in \mathcal{F}_d$ versus $H_1 : f^* \notin \mathcal{F}_d$. The jth random partition of $[n]$ produces a test statistic $T_{n,j}$. The subsampling approach rejects H_0 when $B^{-1} \sum_{j=1}^{B} T_{n,j} \geq 1/\alpha$. Note that each test statistic $T_{n,j}$ is nonnegative. In cases where we have sufficient evidence against H_0, it may be possible to reject H_0 at some iteration $b < B$. That is, for any b such that $1 \leq b < B$, $\sum_{j=b+1}^{B} T_{n,j} \geq 0$. If there is a value of $b < B$ such that $B^{-1} \sum_{j=1}^{b} T_{n,j} \geq 1/\alpha$, then it is guaranteed that $T_n = B^{-1} \sum_{j=1}^{B} T_{n,j} \geq 1/\alpha$. Algorithms 1–3 incorporate this fact by rejecting early if we have sufficient evidence against H_0.

Both `logcondens` ($d = 1$) and `LogConcDEAD` ($d \geq 1$) compute the log-concave MLE $\widehat{f}_0$. The choice of $\widehat{f}_1$, which can be any density, is flexible, and we explore several options.

Full Oracle The full oracle approach uses the true density f^* in the numerator. That is, in Algorithm 1, the input density estimation approach is to set $\widehat{f}_{1,b} = f^*$. This method is a helpful theoretical comparison, since it avoids the depletion in power that occurs when $\widehat{f}_{1,b}$ does not approximate f^* well. We would expect the power of this approach to exceed the power of any approach that estimates a numerator density on $\{Y_i : i \in \mathcal{D}_{1,b}\}$.

Partial Oracle The partial oracle approach uses a d-dimensional parametric MLE density estimate in the numerator. Suppose we know (or we guess) that the true density is parameterized by some unknown real-valued vector $\theta^* \in \mathbb{R}^p$ such that $f^* = f_{\theta^*}$. In Algorithm 1, the input density estimation approach is to set $\widehat{f}_{1,b} = f_{\widehat{\theta}_{1,b}}$, where $\widehat{\theta}_{1,b}$ is the MLE of θ over $\{Y_i : i \in \mathcal{D}_{1,b}\}$. If the true density is from the parametric family $(f_\theta : \theta \in \mathbb{R}^p)$, we would expect this method to have good power relative to other density estimation methods.

Fully Nonparametric The fully nonparametric method uses a d-dimensional kernel density estimate (KDE) in the numerator. In Algorithm 1, the input density estimation approach is to set $\widehat{f}_{1,b}$ to the kernel density estimate computed on $\mathcal{D}_{1,b}$. Kernel density estimation involves the choice of a bandwidth. The `ks` package (Duong 2021) in R can fit multidimensional KDEs and has several bandwidth computation procedures. These options include a plug-in bandwidth (Wand and Jones 1994; Duong and Hazelton 2003; Chacón and Duong 2010), a least squares cross-validated bandwidth (Rudemo 1982; Bowman 1984), and a smoothed cross-validation bandwidth (Jones, Marron, and Park 1991; Duong and Hazelton 2005). In the parametric density case, we would expect the fully nonparametric method to have lower power than the full oracle method and the partial oracle method. If we do not want to make assumptions about the true density, this may be a good choice.

3.3. Universal Tests with Dimension Reduction

Suppose we write each random variable $Y \in \mathbb{R}^d$ as $Y = (Y^{(1)}, \dots, Y^{(d)})$. As noted in An (1997), if the density of Y is log-concave, then the marginal densities of $Y^{(1)}, \dots, Y^{(d)}$ are all log-concave. In the converse direction, if the marginal densities of $Y^{(1)}, \dots, Y^{(d)}$ are all log-concave and $Y^{(1)}, \dots, Y^{(d)}$ are all independent, then the density of Y is log-concave. Proposition 1 of Cule, Samworth, and Stewart (2010) uses a result from Prékopa (1973) to deduce a more general result. We restate Proposition 1(a) in Theorem 2.

Theorem 2. (Proposition 1(a) of Cule, Samworth, and Stewart 2010) Suppose $Y \in \mathbb{R}^d$ is a random variable from a distribution having density f^* with respect to Lebesgue measure. Let V be a subspace of $\mathbb{R}^d$, and denote the orthogonal projection of y onto V as $P_V(y)$. If f^* is log-concave, then the marginal density of $P_V(Y)$ is log-concave and the conditional density $f^*_{Y|P_V(Y)}(\cdot \mid t)$ of Y given $P_V(Y) = t$ is log-concave for each t.

When considering how to test for log-concavity, An (1997) notes that univariate tests for log-concavity could be used in

Algorithm 2 For $H_0 : f^* \in \mathcal{F}_d$ versus $H_1 : f^* \notin \mathcal{F}_d$, compute the axis-aligned projection test statistics or run the test.

```
Input: n iid d-dimensional observations $Y_1, \ldots, Y_n$ from unknown density $f^*$,
    number of subsamples $B$, significance level $\alpha$.
Output: $d$ test statistics $T_n^{(k)}, k = 1, \ldots, d$, or the test result.
1: for $k = 1, 2, \ldots, d$ do
2:     for $b = 1, 2, \ldots, B$ do
3:         Randomly partition $[n]$ into $\mathcal{D}_{0,b}$ and $\mathcal{D}_{1,b}$ such that $|\mathcal{D}_{0,b}| = |\mathcal{D}_{1,b}| = n/2$.
4:         Estimate a one-dimensional density $\widehat{f}_{1,b,k}$ on $\{Y_i^{(k)} : i \in \mathcal{D}_{1,b}\}$.
5:         Estimate the log-concave MLE $\widehat{f}_{0,b,k}$ on $\{Y_i^{(k)} : i \in \mathcal{D}_{0,b}\}$.
6:         if $B^{-1} \sum_{j=1}^{b} \prod_{i \in \mathcal{D}_{0,j}} \{\widehat{f}_{1,j,k}(Y_i^{(k)}) / \widehat{f}_{0,j,k}(Y_i^{(k)})\} \geq d/\alpha$ then stop, reject null.
7:     Compute the test statistic $T_n^{(k)} = B^{-1} \sum_{b=1}^{B} \prod_{i \in \mathcal{D}_{0,b}} \{\widehat{f}_{1,b,k}(Y_i^{(k)}) / \widehat{f}_{0,b,k}(Y_i^{(k)})\}$
8: return the test statistics $T_n^{(k)}, k = 1, \ldots, d$.
```

the multivariate setting. For our purposes, we use Theorem 2's implication that if f^* is log-concave, then the one-dimensional projections of f^* are also log-concave. We develop new universal tests on these one-dimensional projections.

To reduce the data to one dimension, we take one of two approaches.

3.3.1. *Dimension Reduction Approach 1: Axis-aligned Projections*

We can represent any d-dimensional observation Y_i as $Y_i = (Y_i^{(1)}, Y_i^{(2)}, \ldots, Y_i^{(d)})$. Algorithm 2 describes an approach that computes a test statistic for each of the d dimensions.

We reject $H_0 : f^* \in \mathcal{F}_d$ if at least one of the d test statistics $T_n^{(1)}, \ldots, T_n^{(d)}$ exceeds d/α. Instead of checking this condition at the very end of the algorithm, we check this along the way, and we stop early to save computation if this condition is satisfied (line 6). This rejection rule has valid Type I error control because under H_0,

$$\mathbb{P}(\{T_n^{(1)} \geq d/\alpha\} \cup \{T_n^{(2)} \geq d/\alpha\} \cup \cdots \cup \{T_n^{(d)} \geq d/\alpha\})$$
$$\leq \sum_{k=1}^{d} \mathbb{P}(T_n^{(k)} \geq d/\alpha) \leq d(\alpha/d) = \alpha.$$

If we do not simply want an accept-reject decision but would instead like a real-valued measure of evidence, then we can note that $p_n := d \min_{k \in [d]} 1/T_n^{(k)}$ is a valid p-value. Indeed the above equation can be rewritten as the statement $\mathbb{P}(p_n \leq \alpha) \leq \alpha$, meaning that under the null, the distribution of p_n is stochastically larger than uniform.

To run the test, we must fit some one-dimensional density $\widehat{f}_{1,b,k}$ on $\{Y_i^{(k)} : i \in \mathcal{D}_{1,b}\}$. We consider two density estimation methods; the same applies to the next subsection. Thus, for the universal LRTs with dimension reduction, we consider four total combinations of two dimension reduction approaches and two density estimation methods.

Density Estimation Method 1: Partial Oracle This approach uses parametric knowledge about the true density. The numerator $\widehat{f}_{1,b,k}$ is the parametric MLE fit on $\mathcal{D}_{1,b}$.

Density Estimation Method 2: Fully Nonparametric This approach does not use any prior knowledge about the true density. Instead, we use kernel density estimation (e.g., `ks` package with plug-in bandwidth) to fit $\widehat{f}_{1,b,k}$.

3.3.2. *Dimension Reduction Approach 2: Random Projections*

We can also construct one-dimensional densities by projecting the data onto a vector drawn uniformly from the unit sphere. Algorithm 3 shows how to compute the random projection test statistic T_n. As discussed in Section 3.2, Theorem 1 justifies the validity of this approach. In short, each individual projection test statistic $T_{n,j}$ is an e-value, meaning that it has expectation of at most one under the null. Thus the average of $T_{n,j}$ values is also an e-value. Since each $T_{n,j}$ is nonnegative, if there is some $k < n_{\text{proj}}$ such that $(1/n_{\text{proj}}) \sum_{j=1}^{k} T_{n,j} \geq 1/\alpha$, then we can reject H_0 without computing all n_{proj} test statistics.

We expect random projections (with averaging) to work better when the deviations from log-concavity are "dense," meaning there is a small amount of evidence to be found scattered in different directions. In contrast, the axis-aligned projections (with Bonferroni) presented earlier are expected to work better when there is a strong signal along one or a few dimensions, with most dimensions carrying no evidence (meaning that the density is indeed log-concave along most axes).

4. Example: Testing Log-Concavity of Normal Mixture

We test the permutation approach and the universal approaches on a normal mixture distribution, which is log-concave only at certain parameter values. Naturally, when testing or fitting log-concave distributions in practice, one would eschew all parametric assumptions, so the restriction to normal mixtures is simply for a nice simulation example. See Appendix D for another such example over Beta densities.

Let ϕ_d be the $N(0, I_d)$ density. Cule, Samworth, and Stewart (2010) note a result that we state in Fact 1.

Fact 1. For $\gamma \in (0, 1)$, the normal location mixture $f(x) = \gamma\phi_d(x) + (1 - \gamma)\phi_d(x - \mu)$ is log-concave only when $\|\mu\| \leq 2$.

Cule, Samworth, and Stewart (2010) use the permutation test to test $H_0 : f^* \in \mathcal{F}_d$ versus $H_1 : f^* \notin \mathcal{F}_d$ in this setting. We explore the power and validity of both the permutation and the universal tests over $\gamma = 0.5$, varying $\|\mu\|$, and dimensions $d \in \{1, 2, 3, 4\}$.

Algorithm 3 For $H_0 : f^* \in \mathcal{F}_d$ versus $H_1 : f^* \notin \mathcal{F}_d$, compute the random projection test statistic or run the test.

```
Input: $n$ iid $d$-dimensional observations $Y_1, \dots, Y_n$ from unknown density $f^*$,
  number of subsamples $B$, significance level $\alpha$, number of random projections $n_{\text{proj}}$.
Output: The random projection test statistic $T_n$ or the test result.
for $k = 1, 2, \dots, n_{\text{proj}}$ do
    Draw a vector $V$ uniformly from the $d$-dimensional unit sphere. To obtain $V$,
      draw $X \sim N(0, I_d)$ and set $V = X/\|X\|$.
    Project each $Y$ observation onto $V$. The projection of $Y_i$ is $P_V(Y_i) = Y_i^T V$.
    for $b = 1, 2, \dots, B$ do
        Randomly partition $[n]$ into $\mathcal{D}_{0,b}$ and $\mathcal{D}_{1,b}$ such that $|\mathcal{D}_{0,b}| = |\mathcal{D}_{1,b}| = n/2$.
        Estimate a one-dimensional density $\widehat{f}_{1,b,k}$ on $\{P_V(Y_i) : i \in \mathcal{D}_{1,b}\}$.
        Estimate the log-concave MLE $\widehat{f}_{0,b,k}$ on $\{P_V(Y_i) : i \in \mathcal{D}_{0,b}\}$.
    Compute the test statistic $T_{n,k} = B^{-1} \sum_{b=1}^{B} \prod_{i \in \mathcal{D}_{0,b}} \{\widehat{f}_{1,b,k}(P_V(Y_i)) / \widehat{f}_{0,b,k}(P_V(Y_i))\}$.
    if $n_{\text{proj}}^{-1} \sum_{j=1}^{k} T_{n,j} \geq 1/\alpha$ then stop, reject null.
return the random projection test statistic $T_n = n_{\text{proj}}^{-1} \sum_{j=1}^{n_{\text{proj}}} T_{n,j}$.
```

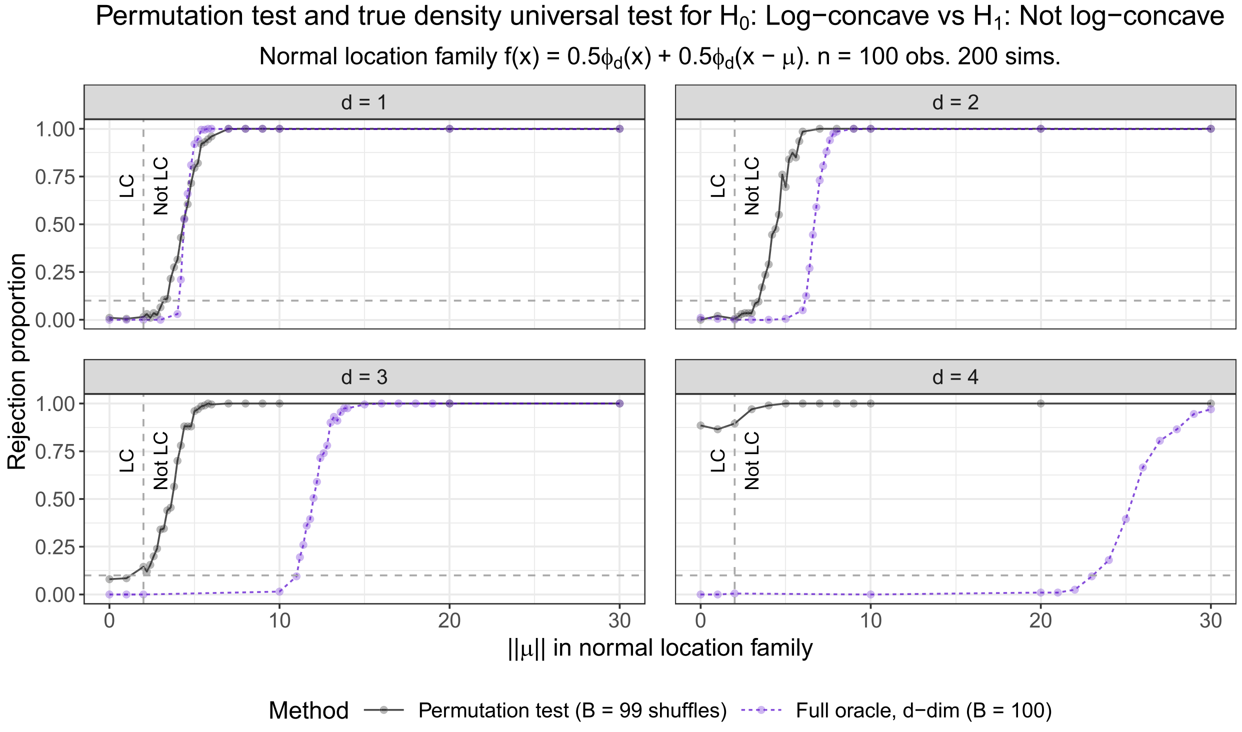


Figure 3. Rejection proportions for tests of $H_0 : f^*$ is log-concave versus $H_1 : f^*$ is not log-concave. When $d = 1$, the permutation test and the full oracle universal test have similar power. The full oracle universal approach remains valid in higher dimensions, but it has low power for moderate $\|\mu\|$ when $d \geq 3$.

4.1. Full Oracle (in d Dimensions) has Inadequate Power

We compare the permutation test from Section 3.1 to the full oracle universal test, which uses the true density in the numerator, from Section 3.2. Figure 1 showed that the permutation test is not valid in this setting for $d \geq 4$, $n = 100$, and $B = 99$. In Appendix B.2, we show that the permutation test's rejection proportion is similar if we increase B to $B \in \{100, 200, 300, 400, 500\}$. In addition, we show that if we increase n to 250, the rejection proportion is still much higher than 0.10 for $\|\mu\| \leq 2$ at $d = 4$ and $d = 5$. In the same appendix, we also show that the discrete nature of the test statistic is not the reason for the test's conservativeness (e.g., $d = 1$) or anticonservativeness (e.g., $d = 5$).

To compare the permutation test to the full oracle universal test, we again set $\mu = (\mu_1, 0, \dots, 0)$. Figure 3 shows the power for $d \in \{1, 2, 3, 4\}$ on $n = 100$ observations. For the universal test, we subsample $B = 100$ times. Each point is the rejection proportion over 200 simulations. For some $\|\mu\|$ values in the $d = 1$ case, the full oracle test has higher power than the permutation test. For most $(d, \|\mu\|)$ combinations, though, the full oracle test has lower power than the permutation test. Unlike the permutation test, though, the universal test is provably valid for all d. In Figure 3, we see that as d increases, we need larger $\|\mu\|$ for the universal test to have power. More specifically, $\|\mu\|$ needs to grow exponentially with d to maintain a certain level of power. (See Figure 14 in Appendix B.3.)

In Appendix E.2, we briefly discuss the trace test from Section 3 of Chen and Samworth (2013) as an alternative to the permutation test. Similar to the permutation test, simulations suggest that the trace test is valid in the $d = 2$ setting, but it does not control Type I error in the general d-dimensional setting.

4.2. Superior Performance of Dimension Reduction Approaches

We have seen that the full oracle universal LRT requires $\|\mu\|$ to grow exponentially to maintain power as d increases. We turn to the dimension reduction universal LRT approaches, and we find that they produce higher power for smaller $\|\mu\|$ values.

We implement all four combinations of the two dimension reduction approaches (axis-aligned and random projections) and two density estimation methods (partial oracle and fully nonparametric). We compare them to three d-dimensional approaches: the permutation test, the full oracle test, and the partial oracle test. The full oracle d-dimensional approach uses the split LRT with the true density in the numerator and the d-dimensional log-concave MLE in the denominator. The partial oracle approaches use the split LRT with a two component Gaussian mixture in the numerator and the log-concave MLE in the denominator. We fit the Gaussian mixture using the EM algorithm, as implemented in the `mclust` package (Scrucca et al. 2016). The fully nonparametric approaches use a kernel density estimate in the numerator and the log-concave MLE in the denominator. We fit the kernel density estimate using the plug-in bandwidth in the `ks` package (Duong 2021).

Figures 4 and 5 compare the four dimension reduction approaches and the d-dimensional approaches. The six universal approaches subsample at $B = 100$. The random projection approaches set $n_{\text{proj}} = 100$. The permutation test uses $B = 99$ permutations to determine the significance level of the original test statistic. Both figures use the normal location model $f^*(x) = 0.5\phi_d(x) + 0.5\phi_d(x - \mu)$ as the underlying model. However, Figure 4 uses $\mu = -(\|\mu\|, 0, \ldots, 0)$, while Figure 5 uses $\mu = -(\|\mu\|d^{-1/2}, \|\mu\|d^{-1/2}, \ldots, \|\mu\|d^{-1/2})$. The axis-aligned projection method has higher power in the first setting, but the other methods do not have differences in power between the two settings.

Figures 4(a) and 5(a) compare all seven methods. There are several key takeaways. The universal approaches that fit one-dimensional densities (axis-aligned projections and random projections) have higher power than the universal approaches that fit d-dimensional densities. (When $d = 1$, the "Partial oracle, axis-aligned projections," "Partial oracle, d-dim," and "Partial oracle, random projections" approaches are the same, except that the final method uses Bn_{proj} subsamples rather than B subsamples.) As intuition for this behavior, even one projection with particularly strong evidence against log-concavity may provide sufficient evidence to reject log-concavity. See Appendix C for more discussion. The permutation test is not always valid, especially for $d \geq 4$.

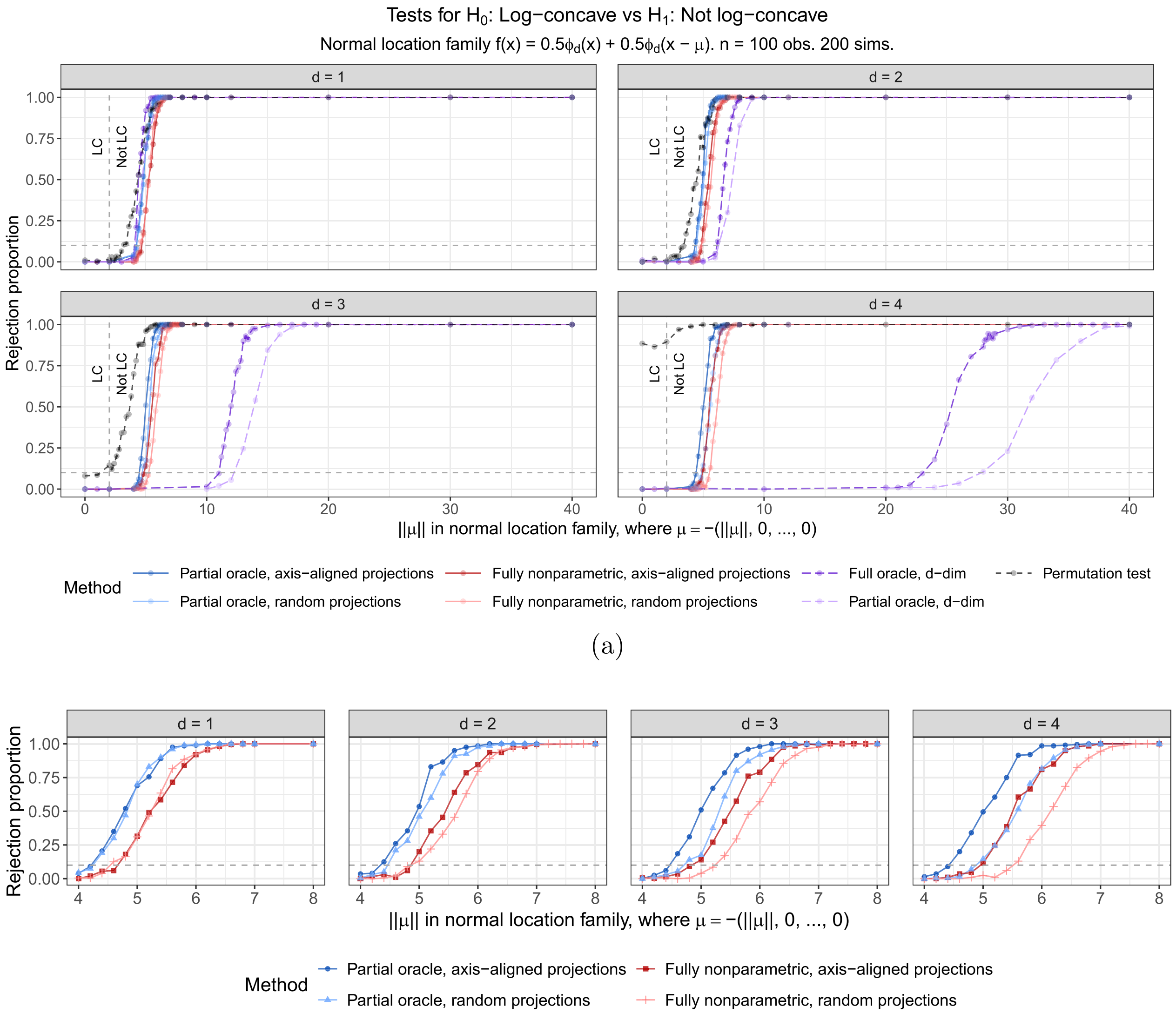
To compare the four universal approaches that fit one-dimensional densities, we consider Figures 4(b) and 5(b), which zoom in on a smaller range of $\|\mu\|$ values for those four methods. In both Figures 4(b) and 5(b), for a given dimension reduction approach (axis-aligned projections or random projections), the partial oracle approach has slightly higher power than the fully nonparametric approach. For a given density estimation approach, the dimension reduction approach with higher power changes based on the setting. When $\mu = -(\|\mu\|, 0, \ldots, 0)$ (Figure 4(b)), the axis-aligned projection approach has higher power than the random projections approach. This makes sense because a single dimension contains all of the signal. When $\mu = -(\|\mu\|d^{-1/2}, \|\mu\|d^{-1/2}, \ldots, \|\mu\|d^{-1/2})$ (Figure 5(b)), the random projections approach has higher power than the axis-aligned projection approach. This makes sense because all directions have some evidence against H_0, and there exist linear combinations of the coordinates that have higher power than any individual axis-aligned dimension.

4.3. Time Benchmarking

Table 1 displays the average run times of the seven methods that we consider. Each cell corresponds to an average (in seconds) over 10 simulations at $(\|\mu\| = 0, \|\mu\| = 5, \|\mu\| = 10)$. Except for the restriction on $\|\mu\|$, these simulations use the same parameters as Figure 4. We arrange the methods in rough order from longest to shortest run times in $d = 1$. The random projection methods have some of the longest run times at $\|\mu\| = 0$. If the random projection methods do not have sufficient evidence to reject H_0 early, they will construct Bn_{proj} test statistics. Each of those Bn_{proj} test statistics requires fitting a one-dimensional log-concave MLE and estimating a partial oracle or fully nonparametric numerator density. The permutation test is faster than the random projection tests in the $\|\mu\| = 0$ setting, but it does not stop early for larger $\|\mu\|$. The axis projection and d-dimensional universal approaches have similar run times for $d \leq 2$. (In fact, for $d = 1$ the partial oracle axis projection and partial oracle d-dimensional methods are the same.) The axis projection methods compute a maximum of Bd test statistics, and the d-dimensional methods compute a maximum of B test statistics. Since the d-dimensional universal approaches repeatedly fit d-dimensional log-concave densities, they are the most computationally expensive approaches for large d.

4.4. Counterpoint: Non-Log-Concave Density with Log-Concave Marginals

In the previous normal location mixture example, the projection methods have power because the projection distributions are not all log-concave. While Theorem 2 states that log-concavity of a density implies log-concavity of the lower dimensional projections, the converse is not guaranteed to hold. As an example, suppose $f^* : \mathbb{R}^2 \to \mathbb{R}$ is the normal mixture density given by $f^*(x) = (2 \cdot 2\pi)^{-1}(\exp(-\|x\|^2/2) + \sigma^{-2}\exp(-\|x\|^2/2\sigma^2))$, where $\sigma = \sqrt{3}$. This density is not log-concave, but all of its one-dimensional projections are log-concave. (See Appendix A.2.) Figure 6 shows power results from simulations of several full-dimensional and projection approaches to test $H_0 : f^* \in \mathcal{F}_d$ versus $H_1 : f^* \notin \mathcal{F}_d$ across varying n. The d-dimensional



(a)

(b)

Figure 4. Power of tests of H_0 : f^* is log-concave versus H_1 : f^* is not log-concave. μ vector for second component is $\mu = -(\|\mu\|, 0, \ldots, 0)$. (a) The projection-based universal tests decrease the gap in power between the permutation test and the d-dimensional universal tests for $d \geq 2$. The power of the projection tests only exhibits a moderate curse of dimensionality. (b) Partial oracle numerators have higher power than fully nonparametric numerators. Since all signal against H_0 is in the first component of μ, the axis-aligned projection tests have higher power than the random projection tests within each choice of numerator.

approaches have some power to detect H_1, but even at $n = 1600$ the estimated power is only about 0.13. The projection methods do not have power since all projections are log-concave. Hence, the projection methods will not always have higher power than the d-dimensional approaches, but in this example even the d-dimensional approaches do not have high power. If one is not sure whether the projected or full-dimensional test will have higher power, one can simply run both, average the resulting test statistics, and threshold the average at $1/\alpha$. Since the average of e-values is an e-value, such a test is still valid, and the test is consistent if either of the original tests is consistent.

5. Theoretical Power of Log-Concave Universal Tests

Our simulations have shown that the universal LRTs can have high power to test $H_0 : f^* \in \mathcal{F}_d$ versus $H_1 : f^* \notin \mathcal{F}_d$. We complement this with a proof of the consistency of this test. For the rest of this section, think of $f^* \notin \mathcal{F}_d$.

First, we review and introduce some notation. Let $\widehat{f}_{1,n}$ be an estimate of f^*, fit on $\mathcal{D}_{1,n}$. Let $\widehat{f}_{0,n}$ be the log-concave MLE of f^* fit on $\mathcal{D}_{0,n}$, that is, $\widehat{f}_{0,n} = \underset{f \in \mathcal{F}_d}{\arg\max} \sum_{i \in \mathcal{D}_{0,n}} \log(f(Y_i))$. The universal test statistic is

$$T_n = \prod_{i \in \mathcal{D}_{0,n}} \frac{\widehat{f}_{1,n}(Y_i)}{\widehat{f}_{0,n}(Y_i)}, \tag{1}$$

and we reject H_0 if $T_n \geq 1/\alpha$.

Let f^{LC} denote the log-concave projection of f^*, that is, $f^{\text{LC}} = \underset{f \in \mathcal{F}_d}{\arg\min}\, D_{\text{KL}}(f^* \| f)$, where for densities p, q, $D_{\text{KL}}(p\|q)$ is the KL divergence $D_{\text{KL}}(p\|q) := \int p(x) \log(p(x)/q(x))\, dx$. We further define the Hellinger divergence $h(p, q) := \|\sqrt{p/2} - \sqrt{q/2}\|_2$. h is well-defined for nonnegative functions (and not only densities) and is a metric on such functions.

Finally, we define some set notation: $\text{supp}(f^*) := \{x : f^*(x) > 0\}$ denotes the support of the measure induced by f^*,

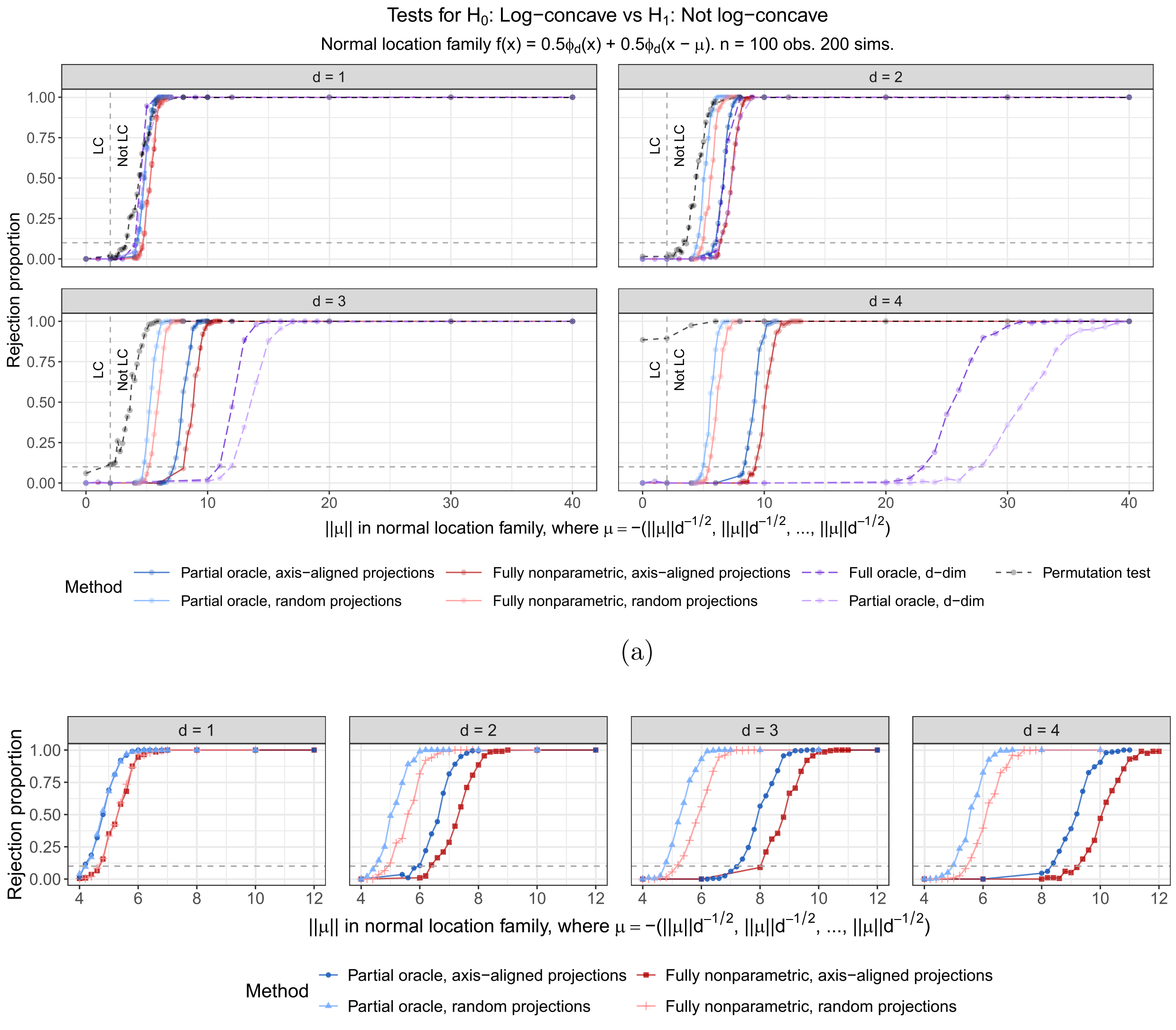


Figure 5. Power of tests of $H_0 : f^*$ is log-concave versus $H_1 : f^*$ is not log-concave. μ vector for second component is $\mu = -(\|\mu\|d^{-1/2}, \|\mu\|d^{-1/2}, \ldots, \|\mu\|d^{-1/2})$. (a) In this second choice of μ structure, the projection-based universal tests also decrease the gap in power between the permutation test and the d-dimensional universal tests for $d \geq 2$. The power of the projection tests only exhibits a moderate curse of dimensionality. (b) Since the signal against H_0 is equally distributed across the components of μ, the random projection tests have higher power than the axis-aligned projection tests.

and for a measurable set $S \subset \mathbb{R}^d$, $\text{int}(S)$ and ∂S, respectively, denote its topological interior and boundary.

5.1. Assumptions

We reiterate that the validity of the universal LRT only requires the assumption of iid data. However, for the universal LRT to be powerful (when $f^* \notin \mathcal{F}_d$), we need a few relatively mild conditions.

Assumption 1. (Regularity of f^*) Let $P^*(\mathrm{d}x) := f^*(x)\mathrm{d}x$. We assume that if $X \sim f^*$, then $\mathbb{E}[\|X\|] < \infty$, $\mathbb{E}[\max(\log f^*(X), 0)] < \infty$, $\text{int}(\text{supp}(f^*)) \neq \varnothing$, and that for every hyperplane H, $P^*(H) < 1$.

Assumption 1 enforces standard conditions imposed in the log-concave estimation literature (Cule and Samworth 2010). In particular, the finiteness of $\mathbb{E}[\|X\|]$ and $\mathbb{E}[\max(\log f^*, 0)]$ yield the existence of the log-concave projection f^{LC} (defined earlier), and the remaining conditions impose weak regularity properties that ensure convergence of $\widehat{f}_{0,n}$ to f^{LC}.

For power, we fundamentally need our estimate $\widehat{f}_{1,n}$ to be close enough to the true (non-log-concave, under the alternative) distribution f^*, relative to $\mathcal{F}_d$. An assumption of KL-consistency, that is, assuming $D_{\text{KL}}(f^* \| \widehat{f}_{1,n}) \to 0$ as $n \to \infty$, would certainly suffice. The following condition is weaker, though, and it only requires $D_{\text{KL}}(f^* \| \widehat{f}_{1,n})$ to get smaller than a critical Hellinger distance of f^* from log-concavity.

Assumption 2. (Estimability of f^*) We assume that $\widehat{f}_{1,n}$ is a good estimator of f^*, in the sense that if we use $n/2$ iid draws from f^* to construct $\widehat{f}_{1,n}$, then for all $\theta > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{D_{\text{KL}}(f^* \| \widehat{f}_{1,n})}{h^2(f^*, f^{\text{LC}})} < \frac{1}{200}, \int f^*(x) \log^2(f^*(x)/\widehat{f}_{1,n}(x))\,\mathrm{d}x < \theta n \right) = 1.$$

Table 1. Average run time (in seconds) of log-concave tests at ($\|\mu\| = 0$, $\|\mu\| = 5$, $\|\mu\| = 10$), using the same parameters as Figure 4.

Method	$d = 1$	$d = 2$	$d = 3$	$d = 4$
Partial oracle, random projections	(150, 50, 0.94)	(160, 110, 1.7)	(150, 140, 3.2)	(160, 140, 3.7)
Fully NP, random projections	(120, 71, 0.67)	(110, 100, 1.3)	(120, 110, 2.6)	(120, 120, 4.2)
Permutation test	(51, 50, 51)	(52, 52, 52)	(54, 54, 53)	(61, 62, 62)
Partial oracle, axis projections	(8, 3.6, 0.1)	(16, 5.7, 0.19)	(24, 12, 0.26)	(32, 25, 0.34)
Fully NP, axis projections	(7.6, 5.8, 0.098)	(15, 11, 0.17)	(23, 19, 0.25)	(30, 23, 0.33)
Partial oracle, d-dim	(6.6, 3.9, 0.093)	(21, 22, 0.68)	(71, 74, 75)	(300, 330, 350)
Full oracle, d-dim	(6.1, 1.2, 0.073)	(20, 21, 0.3)	(70, 72, 74)	(300, 340, 340)

NOTE: The universal methods run faster when there is more evidence against H_0, which allows for early rejection. It is faster to compute log-concave MLEs in one dimension (e.g., projection methods) than in general d dimensions.

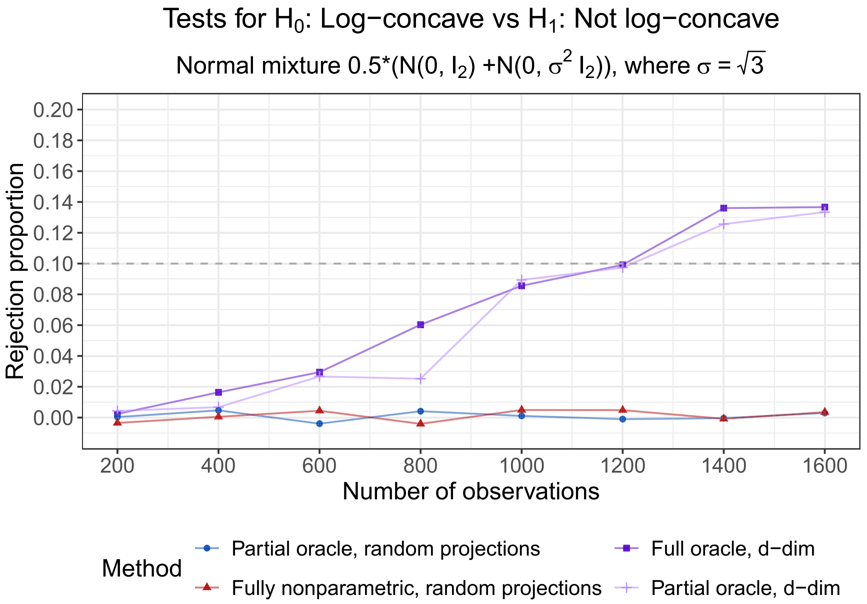


Figure 6. Power of the universal test for log-concavity in a setting where the two-dimensional density is not log-concave but all one-dimensional projections are log-concave. The estimated power is the rejection proportion over 200 simulations at $\alpha = 0.10$ for a given value of n. Values of rejection proportions are jittered by at most 0.005 for plotting. Full-dimensional approaches have power above α at $n \in \{1400, 1600\}$. Projection approaches never reject log-concavity.

Assumption 2 is satisfied if $\widehat{f}_{1,n}$ estimates f^* in a KL divergence sense better than f^{LC} approximates f^* in a Hellinger sense and if the variance of the log-ratio $\log(f^*/\widehat{f}_{1,n})$ does not grow too fast. Estimation in KL divergence is largely driven by the tail behavior of f^*; the $\widehat{f}_{1,n}$ estimation procedure needs to ensure that $\widehat{f}_{1,n}$ does not underestimate the tails of f^*. In many settings, we will have $D_{\mathrm{KL}}(f^* \| \widehat{f}_{1,n}) \to 0$, thus satisfying the assumption. However, we do not require $D_{\mathrm{KL}}(f^* \| \widehat{f}_{1,n})$ to go to 0—it is enough for the divergence to get smaller than $h^2(f^*, f^{\mathrm{LC}})/200$. While this criterion is still stringent enough to be practically relevant, it makes the theoretically favorable point that the universal LRT for log-concavity does not require consistent density estimation of f^* in a strong (KL) sense. In our argument, the role of Assumption 2 is to control how large $f^*/\widehat{f}_{1,n}$ can get in a manner similar to Theorems 3 and 4 of Wong and Shen (1995).

5.2. *Result and Proof Sketch*

We are now in a position to state our main result. We provide a proof sketch, leaving the details to Appendix A.3.

Theorem 3. Suppose Assumptions 1 and 2 hold. Then the universal likelihood ratio test for log-concavity with test statistic (1) is consistent. That is, $\lim_{n\to\infty} \mathbb{P}_{H_1}(T_n \geq 1/\alpha) = 1$.

Proof sketch. We assume throughout that H_1 is true, meaning that f^* is not log-concave. For brevity, we use $\mathbb{P}$ to denote $\mathbb{P}_{H_1}$. We begin by decomposing T_n into

$$T_n = \underbrace{\prod_{i \in \mathcal{D}_{0,n}} \frac{\widehat{f}_{1,n}(Y_i)}{f^*(Y_i)}}_{=:1/R_n} \cdot \underbrace{\prod_{i \in \mathcal{D}_{0,n}} \frac{f^*(Y_i)}{\widehat{f}_{0,n}(Y_i)}}_{=:S_n} = S_n/R_n.$$

Let $\varepsilon := h(f^*, f^{\mathrm{LC}})$. Further, suppose $n \geq 100\log(1/\alpha)/\varepsilon^2$. Now observe that

$$\{T_n < 1/\alpha\} \subset \{R_n > \exp(n\varepsilon^2/100)\} \cup \{S_n < \exp(n\varepsilon^2/50)\},$$

since outside this union, $S_n/R_n \geq \exp(n\varepsilon^2/100) \geq 1/\alpha$. Thus, it suffices to argue that

$$\mathbb{P}\left(R_n > \exp(n\varepsilon^2/100)\right) + \mathbb{P}\left(S_n < \exp(n\varepsilon^2/50)\right) \to 0.$$

The two assumptions contribute to bounding these terms. In particular, Assumption 1 implies that S_n is big, while Assumption 2 implies that R_n is small.

S_n **is big**. Observe that since $h(f^*, f^{\mathrm{LC}}) > 0$, the likelihood ratio $\prod_{i \in \mathcal{D}_{0,n}} f^*(Y_i)/f^{\mathrm{LC}}(Y_i)$ tends to be exponentially large with high probability. We can use Markov's inequality and the properties of Hellinger distance to show that for $\xi > 0$ (and particularly for large ξ),

$$\mathbb{P}\left(\prod_{i\in\mathcal{D}_{0,n}} f^*(Y_i)/f^{\mathrm{LC}}(Y_i) < \xi\right) \le \sqrt{\xi}\,\mathbb{E}_{Y\sim f^*}\left[\sqrt{f^{\mathrm{LC}}(Y)/f^*(Y)}\right]^{n/2}$$
$$= \sqrt{\xi}\,(1 - h^2(f^*, f^{\mathrm{LC}}))^{n/2}.$$

We would thus expect that the same holds for the ratio of interest $\prod_{i\in\mathcal{D}_{0,n}} f^*(Y_i)/\widehat{f}_{0,n}(Y_i)$. The regularity conditions from Assumption 1 enable this, by establishing that $\widehat{f}_{0,n} \to f^{\mathrm{LC}}$ in the strong sense that for large n, $\widehat{f}_{0,n}$ lies in a small bracket containing f^{LC}. This pointwise control allows us to use results from empirical process theory to show that for large enough n, $\prod_{i\in\mathcal{D}_{0,n}} f^*(Y_i)/\widehat{f}_{0,n}(Y_i)$ grows at an exponential rate similar to $\prod_{i\in\mathcal{D}_{0,n}} f^*(Y_i)/f^{\mathrm{LC}}(Y_i)$.

R_n **is small**. The smallness of R_n relies on two facts: (1) $\widehat{f}_{1,n}$ is fit on $\{Y_i : i \in \mathcal{D}_{1,n}\}$ and evaluated on an independent dataset $\{Y_i : i \in \mathcal{D}_{0,n}\}$, and (2) under Assumption 2, $\widehat{f}_{1,n}$ approximates f^* in a strong sense with high probability.

Concretely, we observe that since $\widehat{f}_{1,n}$ is determined given the data in $\mathcal{D}_{1,n}$, we may condition on $\mathcal{D}_{1,n}$ and study the tail behavior of R_n on $\mathcal{D}_{0,n}$. Further observing that $\mathbb{E}_{Y\sim f^*}[\log(f^*(Y)/\widehat{f}_{1,n}(Y))] = D_{KL}(f^*\|\widehat{f}_{1,n})$, applying Tchebycheff's inequality to $\log R_n$ yields

$$\mathbb{P}\left(R_n > \exp(n\varepsilon^2/100) \mid \{Y_i : i \in \mathcal{D}_{1,n}\}\right)$$
$$\le \frac{\int f^*(x)\log^2(f^*(x)/\widehat{f}_{1,n}(x))\mathrm{d}x}{n((\varepsilon^2/50 - D_{KL}(f^*\|\widehat{f}_{1,n}))_+)^2},$$

where $(z)_+ = \max(0, z)$. Assumption 2 lets us argue that with high probability over $\{Y_i : i \in \mathcal{D}_{1,n}\}$, this upper bound vanishes as $n \to \infty$.

6. Conclusion

We have implemented and evaluated several universal LRTs to test for log-concavity. These methods provide the first tests for log-concavity that are valid in finite samples under only the assumption that the data sample is iid. The tests include a full oracle (true density) approach, a partial oracle (parametric) approach, a fully nonparametric approach, and several LRTs that reduce the d-dimensional test to a set of one-dimensional tests. For reference, we compared these tests to a permutation test although that test is not guaranteed to be valid. In one dimension, the universal tests can have higher power than the permutation test. In higher dimensions, the permutation test may falsely reject H_0 at a rate much higher than α, but the universal tests are still valid in higher dimensions. As seen in the Gaussian mixture case, dimension reduction universal approaches can have notably stronger performance than the universal tests that work with d-dimensional densities.

Several open questions remain. Theorem 3 presented a set of conditions under which the universal LRT has power that converges to 1 as $n \to \infty$. As discussed, it may be possible to weaken some of these conditions. In addition, future work may seek to theoretically derive the power as a function of the dimension, number of observations, and signal strength. As shown in one example (Figure 14 of Appendix B), the signal may need to grow exponentially in d to maintain the same power.

Supplementary Materials

Appendix: The appendix contains proofs of all theoretical results (Appendix A), additional simulations and visualizations for the two-component normal mixture setting (Appendix B), discussions on the relative power of full-dimensional and projection tests (Appendix C), simulations to test log-concavity when data arise from a Beta distribution (Appendix D), and additional details on the permutation test and trace test for log-concavity (Appendix E). (pdf file)

R code: The R code to reproduce the simulations and figures is available at *https://github.com/RobinMDunn/LogConcaveUniv*.

Acknowledgments

This work made extensive use of the `R` statistical software (R Core Team 2021), as well as the `clustermq` (Schubert 2019), `data.table` (Dowle and Srinivasan 2021), `fitdistrplus` (Delignette-Muller and Dutang 2015), `kde1d` (Nagler and Vatter 2020), `ks` (Duong 2021), `LogConcDEAD` (Cule, Gramacy, and Samworth 2009), `logcondens` (Dümbgen and Rufibach 2011), `MASS` (Venables and Ripley 2002), `mclust` (Scrucca et al. 2016), `mvtnorm` (Genz et al. 2021; Genz and Bretz 2009), and `tidyverse` (Wickham et al. 2019) packages.

Disclosure Statement

The authors report that there are no competing interests to declare.

Funding

This work used the Extreme Science and Engineering Discovery Environment (XSEDE) (Towns et al. 2014), which is supported by National Science Foundation grant number ACI-1548562. Specifically, it used the Bridges system (Nystrom et al. 2015), which is supported by NSF award number ACI-1445606, at the Pittsburgh Supercomputing Center (PSC). RD is currently employed at Novartis Pharmaceuticals Corporation. This work was primarily conducted while RD was at Carnegie Mellon University. RD's research was supported by the National Science Foundation Graduate Research Fellowship Program under Grant Nos. DGE 1252522 and DGE 1745016. AR's research is supported by the National Science Foundation under Grant Nos. DMS (CAREER) 1945266 and DMS 2310718. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

ORCID

Robin Dunn http://orcid.org/0000-0003-1620-0413
Larry Wasserman http://orcid.org/0000-0003-1874-5601
Aaditya Ramdas http://orcid.org/0000-0003-0497-311X

References

An, M. Y. (1997), "Log-Concave Probability Distributions: Theory and Statistical Testing," *Duke University Dept of Economics Working Paper*. [1,2,5]

Bagnoli, M., and Bergstrom, T. (2005), "Log-Concave Probability and its Applications," *Economic Theory*, 26, 445–469. [1]

Barber, R. F., and Samworth, R. J. (2021), "Local Continuity of Log-Concave Projection, with Applications to Estimation under Model Misspecification," *Bernoulli*, 27, 2437–2472. [4]

Bowman, A. W. (1984), "An Alternative Method of Cross-Validation for the Smoothing of Density Estimates," *Biometrika*, 71, 353–360. [5]

Carroll, R. J., Delaigle, A., and Hall, P. (2011), "Testing and Estimating Shape-Constrained Nonparametric Density and Regression in the Presence of Measurement Error," *Journal of the American Statistical Association*, 106, 191–202. [1,2]

Chacón, J. E., and Duong, T. (2010), "Multivariate Plug-In Bandwidth Selection with Unconstrained Pilot Bandwidth Matrices," *Test*, 19, 375–398. [5]

Chen, W., Mazumder, R., and Samworth, R. (2021), "A New Computational Framework for Log-Concave Density Estimation," arXiv preprint arXiv:2105.11387. [4]

Chen, Y., and Samworth, R. J. (2013), "Smoothed Log-Concave Maximum Likelihood Estimation with Applications," *Statistica Sinica*, 23, 1373–1398. [1,8]

Cule, M., Gramacy, R., and Samworth, R. (2009), "LogConcDEAD: An R Package for Maximum Likelihood Estimation of a Multivariate Log-Concave Density," *Journal of Statistical Software*, 29, 1–20. [3,12]

Cule, M., and Samworth, R. (2010), "Theoretical Properties of the Log-Concave Maximum Likelihood Estimator of a Multidimensional Density," *Electronic Journal of Statistics*, 4, 254–270. [4,10]

Cule, M., Samworth, R., and Stewart, M. (2010), "Maximum Likelihood Estimation of a Multi-Dimensional Log-Concave Density," *Journal of the Royal Statistical Society*, Series B, 72, 545–607. [1,2,3,4,5,6]

Delignette-Muller, M. L., and Dutang, C. (2015), "fitdistrplus: An R Package for Fitting Distributions," *Journal of Statistical Software*, 64, 1–34. [12]

Dowle, M., and Srinivasan, A. (2021), *data.table: Extension of 'data.frame'*. R package version 1.14.2. [12]

Dümbgen, L., Hüsler, A., and Rufibach, K. (2007), "Active Set and EM Algorithms for Log-Concave Densities Based on Complete and Censored Data," arXiv preprint arXiv:0707.4643. [3]

Dümbgen, L., and Rufibach, K. (2011), "logcondens: Computations Related to Univariate Log-Concave Density Estimation," *Journal of Statistical Software*, 39, 1–28. [3,12]

Duong, T. (2021), *ks: Kernel Smoothing*. R package version 1.13.2. [5,8,12]

Duong, T., and Hazelton, M. L. (2003), "Plug-In Bandwidth Matrices for Bivariate Kernel Density Estimation," *Journal of Nonparametric Statistics*, 15, 17–30. [5]

——— (2005), "Cross-Validation Bandwidth Matrices for Multivariate Kernel Density Estimation," *Scandinavian Journal of Statistics*, 32, 485–506. [5]

Genz, A., and Bretz, F. (2009), *Computation of Multivariate Normal and t Probabilities*, Lecture Notes in Statistics, Heidelberg: Springer-Verlag. [12]

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2021), *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.1-3. [12]

Hazelton, M. L. (2011), "Assessing Log-Concavity of Multivariate Densities," *Statistics & Probability Letters*, 81, 121–125. [1]

Jones, M., Marron, J. S., and Park, B. U. (1991), "A Simple Root *n* Bandwidth Selector," *The Annals of Statistics*, 19, 1919–1932. [5]

Kappel, F., and Kuntsevich, A. V. (2000), "An Implementation of Shor's r-Algorithm," *Computational Optimization and Applications*, 15, 193–205. [3]

Kim, A. K., and Samworth, R. J. (2016), "Global Rates of Convergence in Log-Concave Density Estimation," *Annals of Statistics*, 44, 2756–2779. [4]

Koenker, R., and Mizera, I. (2018), "Shape Constrained Density Estimation via Penalized Rényi Divergence," *Statistical Science*, 33, 510–526. [1]

Kur, G., Dagan, Y., and Rakhlin, A. (2019), "Optimality of Maximum Likelihood for Log-Concave Density Estimation and Bounded Convex Regression," arXiv preprint arXiv:1903.05315. [4]

Nagler, T., and Vatter, T. (2020), *kde1d: Univariate Kernel Density Estimation*. R package version 1.13.2. [12]

Nystrom, N. A., Levine, M. J., Roskies, R. Z., and Scott, J. R. (2015), "Bridges: A Uniquely Flexible HPC Resource for New Communities and Data Analytics," in *Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure*, XSEDE '15, New York, NY, USA: Association for Computing Machinery. [12]

Prékopa, A. (1973), "On Logarithmic Concave Measures and Functions," *Acta Scientiarum Mathematicarum*, 34, 335–343. [5]

R Core Team. (2021), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. [12]

Rudemo, M. (1982), "Empirical Choice of Histograms and Kernel Density Estimators," *Scandinavian Journal of Statistics*, 9, 65–78. [5]

Samworth, R. J. (2018), "Recent Progress in Log-Concave Density Estimation," *Statistical Science*, 33, 493–509. [4]

Schubert, M. (2019), "clustermq Enables Efficient Parallelisation of Genomic Analyses," *Bioinformatics*, 35, 4493–4495. [12]

Scrucca, L., Fop, M., Murphy, T. B., and Raftery, A. E. (2016), "mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models," *The R Journal*, 8, 289–317. [8,12]

Shor, N. Z. (2012), *Minimization Methods for Non-Differentiable Functions* (Vol. 3), Berlin: Springer. [3]

Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G. D., Roskies, R., Scott, J. R., and Wilkins-Diehr, N. (2014), "XSEDE: Accelerating Scientific Discovery," *Computing in Science & Engineering*, 16, 62–74. [12]

Venables, W. N., and Ripley, B. D. (2002), *Modern Applied Statistics with S* (4th ed.), New York: Springer. [12]

Wand, M. P., and Jones, M. C. (1994), "Multivariate Plug-In Bandwidth Selection," *Computational Statistics*, 9, 97–116. [5]

Wasserman, L., Ramdas, A., and Balakrishnan, S. (2020), "Universal Inference," *Proceedings of the National Academy of Sciences*, 117, 16880–16890. [2,3]

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019), "Welcome to the tidyverse," *Journal of Open Source Software*, 4, 1686. [12]

Wong, W. H., and Shen, X. (1995), "Probability Inequalities for Likelihood Ratios and Convergence Rates of Sieve MLEs," *The Annals of Statistics*, 23, 339–362. [11]