International Journal of Statistics and Probability; Vol. 4, No. 3; 2015
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

# Generalized Likelihood Ratio Tests Based on The Asymptotic Variant of The Minimax Approach

Han Yu[1]

[1] Department of Mathematics, Computer Science and Information System, Northwest Missouri State University, USA

Correspondence: Han Yu, Department of Mathematics, Computer Science and Information System, Northwest Missouri State University, USA. E-mail: hanyu@nwmissouri.edu

Received: April 1, 2015 Accepted: May 28, 2015 Online Published: July 1, 2015

doi:10.5539/ijsp.v4n3p36 URL: http://dx.doi.org/10.5539/ijsp.v4n3p36

**Abstract**

Maximum likelihood ratio test statistics may not exist in general in nonparametric function estimation setting. In this paper a new class of generalized likelihood ratio (GLR) tests is proposed for nonparametric goodness-of-fit testing via the asymptotic variant of the minimax approach. The proposed nonparametric tests are developed to be asymptotically distribution-free based on latent variable representations. The nonparametric tests are ameliorated to be appropriately complex so that they are analytically tractable and numerically feasible. They are well applicable for the "adaptive" study of hypothesis testing problems of growing dimensions. To assess the proposed GLR tests, the asymptotic properties are derived. The procedure can be viewed as a novel nonparametric extension of the classical parametric likelihood ratio test as a guard against possible gross misspecification of the data-generating mechanism. Simulations of the proposed minimax-type GLR tests are investigated for the small sample size performance and show that the GLR tests have appealing small sample size properties.

**Keywords:** kernel, latent variable representations, generalized likelihood ratio, minimax approach

## 1. Introduction

Parametric models have the advantage of easy interpretation and efficient computation over nonparametric models. However, the parametric models are often suspected of being at the risk of misspecification in specific real applications. A lot of statistical problems may not be parametric in practice. For example, the objects of estimation or testing are *speech*, *images*, and *so on*. They can be treated as unknown infinite-dimensional parameters that belong to some specific functional set. As a guard against possible misspecification of the data-generating mechanism, nonparametric alternatives appear to avoid the misspecification problem by making statistical models rich enough to include essentially all relevant sampling distributions emerging in real world applications. Nonparametric goodness-of-fit tests deal with testing the adequacy of model complexity to data by adopting the nonparametric alternatives.

Nonparametric goodness-of-fit tests have a long history in statistics (D'Agostino and Stephens, 1986) and are increasingly commonly used in data mining, pattern recognition, machine learning and statistical analysis. There are many connections between nonparametric hypothesis testing and nonparametric estimation. However, the nonparametric hypothesis testing theory is essentially different from the estimation theory due to curse of dimensionality. There are a lot of new effects in nonparametric hypothesis testing theory compared with parametric hypothesis testing as well as nonparametric estimation (Ingster, 2002). In nonparametric goodness-of-fit tests, there are many discrepancy based approaches designed to solve the problems of testing against nonparametric alternatives. The classical approaches based on $L^2$ and $L^\infty$ are popular in nonparametric goodness-of-fit tests. They measure the difference between the estimators under null and alternative models and are the generalization of the Kolmogorov-Smirnov (KS) and Cramér-von Mises (CV) types of statistics. However, the approaches suffer some drawbacks. First of all, the null distribution of the test statistic is unknown and depends critically on nuisance parameters. Second, the choices of measures and weights can be arbitrary, which limits the applicability of the discrepancy based methods (Fan et al., 2001). Useful counterpart approaches are the maximum likelihood ratio approaches that are generally well applicable to most parametric hypothesis testing problems. The most fundamental property that significantly contributes to the success of the maximum likelihood ratio approaches is that their asymptotic distributions under null hypotheses are independent of nuisance parameters. The property can determine the null distribution of the likelihood ratio statistic via using either the asymptotic distribution or the Monte Carlo simulation by setting nuisance parameters at some estimated values. In addition, likelihood-based approaches generally lead to more efficient estimation of unknown model parameters and allow for the proper assessment of the uncertainty, which has a practical impact. The traditional maximum likelihood ratio tests are not naturally applicable to the problems with nonparametric models as alternatives in general. First of all, the nonparametric maximum likelihood estimate may usually not exist in a density function space that specifies the nonparametric density alternatives and thus the nonparametric maximum likelihood ratio tests are not applicable in general (Bahadur, 1958, Le Cam, 1990). Even if the nonparametric maximum likelihood estimate exists for the alternatives, it is awkward to select the same optimal smoothing parameter for both steps in the nonparametric estimation for the nonparametric alternatives and the testing against the nonparametric alternatives. Some likelihood ratio test procedures that are distribution-free under parametric alternatives may become dependent on nuisance parameters under nonparametric alternatives since infinite dimensional neighborhood is around a null hypothesis. To attenuate these difficulties arising from the nonparametric alternatives problems due to the curse of dimensionality, an approach based on the asymptotic variant of the minimax approach is proposed along the line of parametric likelihood ratio tests that possesses distribution-free property. It is a method that tests the adequacy of the model complexity to avoid overfitting in nonparametric setting. To aim at a unified principle for nonparametric alternatives problems from uncertainty perspective, the proposed approach replaces the maximum likelihood estimate under the nonparametric density alternatives by tractable least favorable nonparametric estimates allowing the flexibility of choosing the same optimal tuning parameters.

In the next section, the tests are introduced and formulated. Section 3 presents the main results on asymptotic rate of convergence. Simulation results on the small sample size performance of the GLR tests are contained in Section 4. All mathematical proofs of main results are collected in an appendix.

## 2. The Test Statistics

Assume $\boldsymbol{X} = (X_1, X_2, \cdots, X_n)$ is an independent identically distributed (i.i.d) random sample from the *probability density function* (pdf) $f(x)$ with *cumulative distribution function* (CDF) F. $X_{(i)}$, $i = 1, 2, \cdots, n$ are the order statistics of $\boldsymbol{X}$. Let $X_{(i)} = X_{(1)}$ if $i \leq 0$ and $X_{(i)} = X_{(n)}$ if $i \geq n+1$ thereafter for notational simplicity. The sample directly tells us much about the F via the *empirical cumulative distribution function* (ECDF) $\mathbb{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(X_i \leq x)$, where $\mathrm{I}(\cdot)$ is an indicator function. The *quantile function* (qf) $Q \equiv F^{-1}$ of the $F$ defined as

$$Q(u) := \inf\{x : F(x) \geq u\}, \quad 0 < u < 1,$$

is sometimes the object of more direct interest than the $F$ itself. The data $\boldsymbol{X}$ relate directly to $Q$ simply by taking the left-continuous inverse of $\mathbb{F}_n$, namely the usual *empirical quantile function* (eqf)

$$\mathbb{Q}_n(u) = \sum_{i=1}^{n} X_{(i)} \mathrm{I}_{(\frac{i-1}{n}, \frac{i}{n}]}(u), \quad 0 < u < 1. \tag{1}$$

Let $\mathscr{F}_0 = \{f_0(x, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta_0\}$ be a family of probability density functions with CDF $\mathrm{F}_0(x, \boldsymbol{\theta})$ that are measurable and absolutely continuous in $x$ for every $\boldsymbol{\theta}$ and continuous in $\boldsymbol{\theta}$ for every $x$, where $\boldsymbol{\Theta}_0$ is an open subset of a $d$-dimensional space $\mathbb{R}^d$. The problem of interest is to test the null hypothesis that the unknown density $f$ is in $\mathscr{F}_0$, i.e.

$$H_0: \quad f = f_0(x, \theta), \quad \text{for some } \boldsymbol{\theta} \in \Theta_0, \tag{2}$$

versus the alternative hypothesis

$$H_a: \quad f \neq f_0(x, \theta), \quad \text{for all } \boldsymbol{\theta} \in \Theta_0. \tag{3}$$

Our generalized likelihood ratio tests are proposed for the hypothesis via $\lambda_n = \ell_n(f) - \ell_n(f_0)$, where $\ell_n(\cdot)$ is a log-likelihood functional of a density function dependent on sample size $n$. A large value of $\lambda_n$ is an evidence against the null hypothesis $H_0$ since the alternative family of nonparametric models is far more likely to generate the observed data.

### *2.1 Nonparametric Likelihood Estimate under Nonparametric Density Alternatives*

On one hand, due to a lack of clear knowledge about the data-generating mechanism, we can only make very general assumptions and leave a large portion of the mechanism unspecified so that the distribution $f$ of the data is not specified by a finite number of parameters. Such nonparametric models are quite popular guard against possible gross misspecification of the data-generating mechanism especially when data can be collected adequately. Further, latent variables can be introduced to allow relatively more complex log-likelihood of $f$, i.e., $\ell_n(f)$ over observed data under the nonparametric alternatives, to be expressed in terms of more tractable nonparametric estimates of the $\ell_n(f)$ over the expanded space of observed and latent variables. The introduction of latent variables thus allows least favorable log-likelihood $\ell_n(f)$ to be constructed from simpler nonparametric estimates with the flexibility of choosing the same optimal tuning parameters. The latent variable representations for the log-likelihood by the introduction of the order statistics $U_{(1)}, \cdots, U_{(n)}$ are as follows:

$$\begin{aligned}\ell_n(f) &= \sum_{i=1}^{n} \log f(X_{(i)}) \\ &= -\sum_{i=1}^{n} \log \frac{1}{f(Q(U_{(i)}))} \\ &= -\sum_{i=1}^{n} \log q\left(U_{(i)}\right), \end{aligned} \tag{4}$$

where $q(\cdot) = (f \circ Q(\cdot))^{-1}$ that Turkey (1965) called the *sparsity function* and Parzen (1979) called the *quantile density function* (qdf). The quantile density function is of much practical relevance mainly because it appears as part of the asymptotic variance of empirical quantiles. Histogram-type estimators of the quantile density function have been suggested by Siddiqui (1960) and studied by Bloch and Gastwirth (1968), Bofinger (1975), Reiss (1978), Sheather and Maritz (1983), and Falk (1986). Falk (1984) showed that the kernel type estimate of $q$-quantile beats the sample $q$-quantile under suitable conditions. The kernel estimate of the quantile function $Q(\cdot)$ is

$$\widehat{Q}_{n,h_n}(t, \mathbb{Q}_n) := \int_0^1 \mathbb{Q}_n(u) K_{h_n}(t-u)du, \quad 0 < t < 1, \tag{5}$$

where $K_{h_n}(\cdot) = \frac{1}{h_n} K(\frac{\cdot}{h_n})$ with bandwith $h_n$. Then

$$\widehat{q}_{n,h_n}(t, \mathbb{Q}_n) := \frac{d}{dt}\widehat{Q}_{n,h_n}(t, \mathbb{Q}_n) = \frac{d}{dt}\int_0^1 \mathbb{Q}_n(u) K_{h_n}(t-u)\,du, \quad 0 < t < 1.$$

can be a kernel estimator of $q(\cdot)$. To make the estimator well defined, the kernels $K_{h_n}$ must satisfy certain differentiability conditions. The conditions detailed in the next section result in

$$\widehat{q}_{n,h_n}(t, \mathbb{Q}_n) = \int_0^1 \mathbb{Q}_n(u) k_{h_n}(t-u)du, \quad 0 < t < 1, \tag{6}$$

where the kernel $k(\cdot) = K'(\cdot)$ is the derivative of $K$ with $k_{h_n}(\cdot) = \frac{1}{h_n^2} k(\frac{\cdot}{h_n})$.

The $q(U_{(i)})$ in (4) can be estimated by $\widehat{q}_{n,h_n}(U_{(i)}, \mathbb{Q}_n)$. Thus $\ell_n(f)$ in (4) can be estimated by

$$-\sum_{i=1}^{n} \log \widehat{q}_{n,h_n}(U_{(i)}, \mathbb{Q}_n). \tag{7}$$

On the other hand, both the nonparametric estimates of the quantile density function $\widehat{q}_{n,h_n}(\cdot, \mathbb{Q}_n)$ and the log likelihood function $-\log q(U_{(i)})$ are evaluated at the same training data set. To be a predictive quantity without modeling noise components of training data, the log-likelihood in (7) is evaluated at the grids of the means $EU_{(i)}$, $G_n$=$\{\frac{i}{n+1}, i = 1, \cdots, n\}$, in the interval [0, 1] given the sample size $n$. The grid $G_n$ then characterizes the $U_{(i)}$'s. The total number of points in the grid $G_n$ grows linearly with the sample size $n$. The statistic (7) thus can be ameliorated by setting $U_{(i)}$ as $\frac{i}{n+1}$ to be numerically efficient. With enough data, this comes arbitrarily close to the true log-likelihood function.

Let $a_n = \frac{n}{n+1}$, $\lceil x \rceil = \min\{k, k \geq x, k \in \mathbb{Z}\}$, the ceiling of a real number $x$ is the smallest integer $\geq x$. Calculations using integration by parts show that $\widehat{q}_{n,h_n}$ at $t = \frac{i}{n+1}$ is a sum of local weighted order statistics:

$$\widehat{q}_{n,h_n}(\tfrac{i}{n+1}, \mathbb{Q}_n) = \sum_{j \in \mathcal{J}_i} w_{h_n i j} X_{(j)},$$

where

$$w_{h_n ij} = K_{h_n}(\tfrac{i}{n+1} - \tfrac{j-1}{n}) - K_{h_n}(\tfrac{i}{n+1} - \tfrac{j}{n})$$

and $j \in \mathcal{J}_i = [\underline{m}_i, \overline{m}_i]$ with $\underline{m}_i = \lceil a_n i - nh_n + 1 \rceil$ and $\overline{m}_i = \lceil a_n i + nh_n \rceil$ for each $i$. Note that $w_{h_n ij}$ is a $\frac{i}{n+1}$-centered *local difference kernel* for $j$. In particular, if $\mathbb{Q}_n(\cdot) = \mathbb{U}_n(\cdot)$, the uniform empirical quantile function of $U_{(i)}$, then $\widehat{q}_{n,h_n}(\frac{i}{n+1}; \mathbb{U}_n) = \sum_{j\in\mathcal{J}_i} w_{h_n ij} U_{(j)}$. It is well known from Wald's Statistical Decision Theory that minimax problems correspond to the Bayesian problems for the least favorable priors. The proposed generalized log-likelihood $\widehat{\ell_n(f)}$ under nonparametric alternatives can thus be estimated nonparametrically with numerically efficiency in the sense of the least favorable priors,

$$\begin{aligned}\widehat{\ell_n(f)} &= -\sum_{i=1}^{n} \log \widehat{q}_{n,h_n}\left(\tfrac{i}{n+1}; \mathbb{Q}_n\right) \\ &= -\sum_{i=1}^{n} \log \left( \sum_{j\in\mathcal{J}_i} w_{h_n ij} X_{(j)} \right) \end{aligned} \tag{8}$$

*2.2 Parametric Likelihood Estimate under Parametric Null*

Under the null hypothesis $f = f_0(x, \boldsymbol{\theta}) \in \mathscr{F}_0$. $\boldsymbol{\theta}$ can be estimated by maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}_0}{argmax}\, L_n(\mathbf{X}, \boldsymbol{\theta})$, where $L_n(\mathbf{X}, \boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \log f_0(X_i, \boldsymbol{\theta})$, we denote

$$\hat{\ell}_n(f_0) := \sum_{i=1}^{n} \log f_0(X_{(i)}, \hat{\theta}_n) \tag{9}$$

*2.3 Generalized Likelihood Ratio Test*

The proposed minimax-type generalized likelihood ratio (GLR) test statistic $T_{n,h_n} = 2\hat{\lambda}_n$ with

$$\begin{aligned}\hat{\lambda}_n &= \widehat{\ell_n(f)} - \hat{\ell}_n(f_0) \\ &= -\sum_{i=1}^{n} \log \widehat{q}_{n,h_n}(\tfrac{i}{n+1}; \mathbb{Q}_n) - \sum_{i=1}^{n} \log f_0(X_{(i)}, \hat{\theta}_n) \\ &= -\sum_{i=1}^{n} \log \widehat{q}_{n,h_n}(\tfrac{i}{n+1}; \mathbb{U}_n) - \sum_{i=1}^{n} \log \frac{f(X_{(i)})\widehat{q}_{n,h_n}(\frac{i}{n+1}; \mathbb{Q}_n)}{\widehat{q}_{n,h_n}(\frac{i}{n+1}; \mathbb{U}_n)} + \sum_{i=1}^{n} \log \frac{f(X_{(i)})}{f_0(X_{(i)}, \hat{\theta}_n)} \\ &= \mathscr{U}_{n,h} + \mathscr{V}_{n,h} + \mathscr{W}_n(\hat{\theta}_n)\end{aligned}$$

where

$$\begin{aligned}\mathscr{U}_{n,h_n} &:= -\sum_{i=1}^{n} \log \widehat{q}_{n,h_n}(\tfrac{i}{n+1}; \mathbb{U}_n) \\ \mathscr{V}_{n,h_n} &:= -\sum_{i=1}^{n} \log \frac{f(X_{(i)})\widehat{q}_{n,h_n}(\frac{i}{n+1}; \mathbb{Q}_n)}{\widehat{q}_{n,h_n}(\frac{i}{n+1}; \mathbb{U}_n)}, \\ \mathscr{W}_n(\hat{\theta}_n) &:= \sum_{i=1}^{n} \log \frac{f(X_{(i)})}{f_0(X_{(i)}, \hat{\theta}_n)}.\end{aligned}$$

Our asymptotic analysis shows that the first term $\mathscr{U}_{n,h_n}$ under the certain conditions detailed in the next section dominates the asymptotic null distribution of $T_{n,h_n}$ while the other two terms can be asymptotically negligible. Since $\mathscr{U}_{n,h_n}$ is a function of uniform random variables that do not depend on the underlying pdf $f$, the asymptotic null distribution is therefore distribution-free.

## 3. Asymptotic Results

*3.1 Asymptotic Null Distribution*

Let $x_F := \sup\{x : F(x) = 0\}$ and $x^F := \inf\{x : F(x) = 1\}$, $-\infty \le x_F < x^F \le \infty$. We need regularity conditions on the kernel functions, smoothing parameters and density functions for our results. We assume that

(K.1) The $K(\cdot)$ is a pdf, symmetric about 0;

(K.2) $\|K\|_\infty < \infty$;

(K.3) $K(\cdot)$ is at least three times continuously differentiable on the compact support $(-1, 1)$.

(H.1) $h_n \downarrow 0$ and $nh_n \uparrow \infty$ as $n \to \infty$;

(H.2) $nh_n / \log n \to \infty$ as $n \to \infty$;

(H.3) $nh_n^{1+\frac{1}{2+\delta}} \to \infty$ for some $\delta > 0$ as $n \to \infty$ and $nh_n^{\frac{3}{2}} \log^2 n \to 0$ as $n \to \infty$.

(F.1) pdf $f$ with CDF $F(x)$ is strictly positive and continuously differentiable on $(x_F, x^F)$;

(F.2) $A = \lim_{x \downarrow x_F} f(x) < \infty$ and $B = \lim_{x \uparrow x^F} f(x) < \infty$; Either A=0 (B=0) or $f$ is nondecreasing (nonincreasing) on an interval to the right of $x_F$ (to the left of $x^F$).

(F.3) There exists a constant $\gamma > 0$ such that

$$\sup_{x_F < x < x^F} F(x)(1 - F(x)) \frac{| \dot{f}(x) |}{f^2(x)} \leq \gamma,$$

where $\dot{f}_0(x)$ denotes the first derivative with respect to $x$.

It is noteworthy that the inequality in (F.3) is equivalent to

(F.3′) $\sup_{0<u<1} |J(u)| u(1 - u) \leq \gamma$, where $J(u) := d \log q(u)/du$ is the score function.

Note that the term $\mathscr{W}_n(\hat{\theta}_n)$ doesn't depend on the smoothing parameter $h_n$. Under the null hypothesis $H_0$ it is simply the log-likelihood ratio statistic. Therefore, its order of magnitude would be $O_P(1)$ under the null hypothesis $H_0$. So we have

**Lemma 1.** *Assume kernel function $K(x)$ satisfies conditions (K.1-2-3); smoothing parameter $h_n$ satisfies conditions (H.1-2-3); and $f(x)$ satisfies conditions (F.1-2-3). Under the null hypothesis $H_0$, we have*

$$\sqrt{h_n} \mathscr{W}_n(\hat{\theta}_n) = o_P(1).$$

**Lemma 2.** *Assume kernel function $K(x)$ satisfies conditions (K.1-2-3); smoothing parameter $h_n$ satisfies conditions (H.1-2-3); and $f(x)$ satisfies conditions (F.1-2-3). Under the null hypothesis $H_0$, we have*

$$\sqrt{h_n} \mathscr{V}_{n,h_n} = o_P(1).$$

**Theorem 1.** *Assume kernel function $K(x)$ satisfies conditions (K.1-2-3); smoothing parameter $h_n$ satisfies conditions (H.1-2-3); and $f(x)$ satisfies condition (F.1-2-3). Under the null hypothesis $H_0$, we have*

$$\sqrt{h_n}(T_{n,h_n} - \mu_{n,h_n}(K)) \xrightarrow{\mathscr{L}} N(0, \sigma^2(K)), \quad as\ n \to \infty,$$

*where*

$$\mu_{n,h_n}(K) = \frac{1}{h_n} \|K\|_2^2 + o\left(\frac{1}{n^2 h_n^3}\right), \ \sigma^2(K) = \int_0^2 dz (2 \int_{-1+z}^1 K(x) K(x - z) dx)^2.$$

**Remark 1.** *The above result shows that the mean of $T_{n,h_n}(K)$ and the variance of $T_{n,h_n}(K)$ have the same rate of $h_n^{-1}$. The test statistic is considered as a $\chi^2$ type-test statistic in the sense.*

**Remark 2.** *The mean of the GLR statistic, whose leading term is $\frac{1}{h_n} \|K(x)\|_2^2$, is the counterpart of the degrees of freedom in $\chi^2$. It is equivalent to the difference of the effective number of parameters used under the null and alternative hypotheses. This can be explained as follows. Suppose that we partition the support of $Q(u)$ into equi-spaced intervals, each with length $h_n$. This results in roughly $\frac{1}{h_n} \|K(x)\|_2^2$ intervals. Since the local smoother uses overlapping intervals, the independence in $\chi^2$ is not satisfied and therefore the covariances result in the convolution factor making the effective number of parameters being slightly different from $\frac{1}{h_n} \|K(x)\|_2^2$.*

**Remark 3.** *$\mu_{n,h_n}(K)$ and $\sigma^2(K)$ are independent of nuisance parameters. One does not have to theoretically derive the constant $\mu_{n,h}(K)$ and $\sigma^2(K)$ in order to use the GLR tests in applications, since as long as there is such a distribution-free property for the test statistic, one can simply simulate the null distributions by setting nuisance parameters under the null hypothesis at reasonable values or estimates.*

## 4. Simulation Study on Bandwidth Selection and Powers

Nonparametric methods are typically indexed by smoothing parameter (i.e. bandwidth) which controls the degree of complexity. The choice of bandwidth is therefore critical to implementation. In order to compute our test statistic for a given data set, one needs to specify the order of smoothing parameter $h$. Our asymptotic study suggests that $h$ should be chosen adaptively according to the sample size, for example, ranging from the order of $n^{-1}\log n$ to the order of $n^{-\frac{2}{3}}\log^{-\frac{4}{3}} n$ by conditions (H.2) and (H.3), and would ensure the distribution-free property and consistency of our test as far as the order of $h$ is concerned. In addition, the knowledge of distributional assumption in null hypothesis can be utilized to estimate the optimal bandwidth $h_n^{opt}$ in terms of power. Specifically, to test $H_0 : f_0(x,\theta) \in \mathscr{F}_0$, we choose $\hat{h}_n^{opt}$ as the estimate of $h_n^{opt}$ for the given sample size that minimizes the $\widehat{\ell_n(f)} - \hat{\ell}_n(f_0)$ with respect to $h$ under the null hypothesis. This is equivalent to minimize $\widehat{\ell_n(f)}$ because the observed log-likelihood $\hat{\ell}_n(f_0)$ does not depend on the bandwidth $h$. But this doesn't mean that the observed log-likelihood $\hat{\ell}_n(f_0)$ does not play any role in selecting $h$, i.e., the optimization is subject to $\widehat{\ell_n(f)} \geq \hat{\ell}_n(f_0)$. This motivates the following data-driven method of choosing smoothing parameter $h_n^{opt}$ in terms of log-likelihood:

$$\hat{h}_n^{opt} := \mathop{\arg\min}_{O(\frac{\log n}{n})<h<o(n^{-\frac{2}{3}}\log^{-\frac{4}{3}} n)} \left\{\widehat{\ell_n(f)} : \quad \widehat{\ell_n(f)} \geq \hat{\ell}_n(f_0)\right\}.$$

In practice a general guide for the choice of $h$ for a fixed and finite $n$ would be valuable since the distribution of our test statistic is dependent on the choice of $h$. In the simulation, we used the triweight kernel to consider the problem of testing the composite hypothesis of normality when both the mean and the variance were unspecified for sample size $n$ = 20, 30, 40, 50, 60, 80 and 100 at the level $\alpha = 0.05$. There is no close-form solution for the bandwidth. The choice of the bandwidth needs to be done by numerical optimization. The summary statistics of the bandwidth estimates $\hat{h}_n^{opt}$ by numerical optimization are presented in order of increasing sample size $n$ = 20, 30, 40, 50, 60, 80, 100 with 5000 replicates of null $N(0,1)$ in Table 1. The mean and the median of the estimates of $h$ decrease from roughly 0.3 to 0.2 as sample size increases from 20 to 100. We can see the means of the bandwidth estimates tend to be larger than the medians of the bandwidth estimates. This observation suggests that the bandwidth estimates at given sample size tend to be right skewed.

Table 1. Summary of $\hat{h}_n^{opt}$

| sample size | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|---|---|---|---|---|---|---|
| $n = 20$ | 0.0976 | 0.2476 | 0.2976 | 0.2928 | 0.3476 | 0.6477 |
| $n = 30$ | 0.0651 | 0.1984 | 0.2318 | 0.2478 | 0.2651 | 0.6572 |
| $n = 40$ | 0.0988 | 0.1988 | 0.2238 | 0.2380 | 0.2489 | 0.6488 |
| $n = 50$ | 0.0403 | 0.1408 | 0.1977 | 0.2183 | 0.2597 | 0.6357 |
| $n = 80$ | 0.0376 | 0.1376 | 0.1874 | 0.2122 | 0.2624 | 0.6122 |
| $n = 100$ | 0.0301 | 0.1401 | 0.1899 | 0.2173 | 0.2700 | 0.6382 |

To investigate the finite sample behavior of the proposed test statistic based on sample size and the associated estimated bandwidth, we generated 5000 replicate samples in order of increasing size $n$ = 20, 50 and 100 from the standard normal distribution. Both parametric likelihood estimates under null and nonparametric likelihood estimates under alternative were calculated for each of these samples. The kernel density estimate of generalized likelihood ratio test statistic $T_{\hat{h}}$ (solid curve) was plotted with superimposed normal fit (i.e., the normal density curve with the sample mean and variance) shown by red dashed line as reference in Figures 1, 2 and 3 for each of these samples. Figures 1 and 2 illustrate the departure of the GLR test statistic $T_{\hat{h}}$ from normality. There is a skewness in the direction of a slightly heavy right tail. The height of the peak is higher than the parametric fit. For $n = 100$ the asymptotic normality holds to a remarkable degree shown by Figure 3. Figures 1, 2 and 3 demonstrate that the generalized likelihood ratio behaves closer and closer to a normal distribution as sample size $n$ is larger and larger.

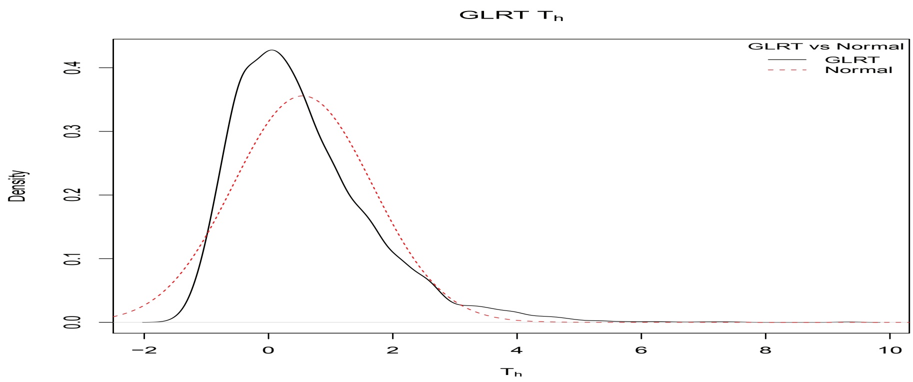


Figure 1. Distribution of GLRT $T_h$ (n=20)

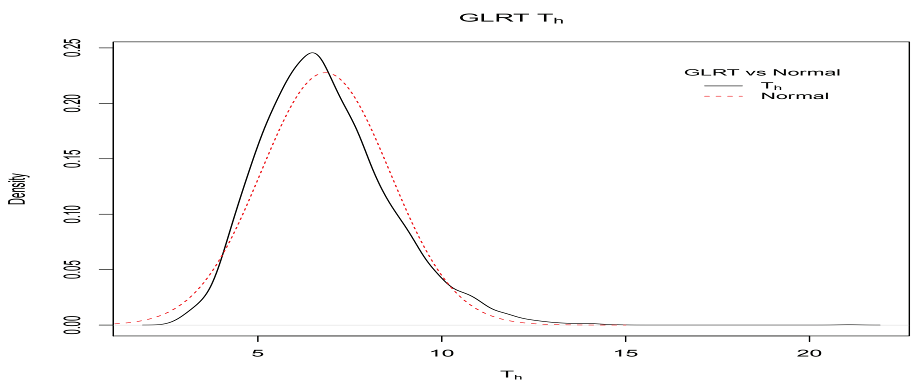


Figure 2. Distribution of GLRT $T_h$ (n=50)

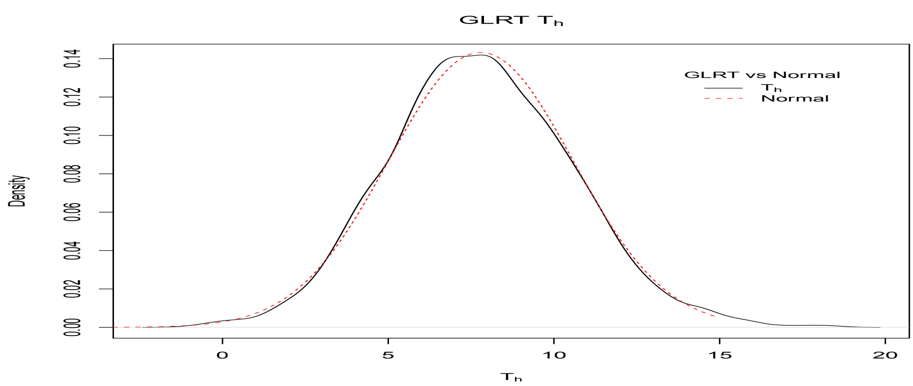


Figure 3. Distribution of GLRT $T_h$ (n=100)

To assess power, we considered the problem of testing the composite hypothesis of normality when both the mean and the variance are unspecified against ten alternatives that were used in previous studies of nonparametric tests for sample size $n = 20$, $n = 50$ and $n = 100$ at the level $\alpha = 0.05$ using the triweight kernel. Specifically, these alternative distributions are standard exponential denoted by Exp(1), gamma distribution with shape parameter $p = 2$ and scale parameter $\lambda = 1$ denoted by Gamma(2, 1), Uniform distribution on (0, 1) denoted by U(0, 1), Beta distribution with parameters 2 and 1 denoted by Beta(2, 1), Beta distribution with parameters 2 and 6 denoted by Beta(2, 6), Laplace distribution with density function given by

$$f(x, \theta) = \frac{1}{2\phi} exp(-|x - \mu|/\phi)$$

where $\theta := (\mu, \phi) = (0, 0.25)$ denoted by Laplace(0, 0.25)), log-normal with density function given by

$$f(x, \theta) = \frac{1}{x\tau\sqrt{2\pi}} exp(-\frac{1}{2\tau^2}(\log x - \nu)^2)$$

where $\theta := (\nu, \tau) = (2, 0.25)$ denoted by Lognormal(2, 0.25). The last six alternatives in Table 2 were added to present various shapes of densities similar to a normal density. To determine critical values of $T_{h_n}$, we generalized 5000 replicate samples of size 20, 50 and 100 respectively from the standard normal distribution. For each sample, $T_{h_n}$ was calculated using triweight kernel on the regular grid of $h$ ranging in order from .05 to .45 by 0.05 and the corresponding estimated $\hat{h}_n$. Note that the $\widehat{\ell_n(f)}$ might be misleading for small value of $h$ for small sample size $n = 20$ and $n = 50$. This arises when there is data rounding and $nh$ is too small to be close to 0. The $(1 - \alpha)th$ quantiles of $T_{h_n}$ were then estimated. Once these critical values had been determined, the powers of the test were estimated by simulations, i.e., for each alternative and each $h_n$, 5000 samples of size 20, size 50 and size 100 were generated from the corresponding alternative distribution and the powers were thus estimated. These Monte Carlo power estimates are given in Tables 2-3-4. It is not surprising to see poor performance with the pretty low powers for some alternatives at bandwidth $h$ close to no-smoothing points.

Table 2. Power Estimates for Various Choices of $h$ and Alternatives ($n = 20$, replicate= 5000, $\alpha = 0.05$)

| Alternative | h=.05 | h=.10 | h=.15 | h=.20 | h=.25 | h=.30 | h=.35 | h=.40 | h=.45 | $\hat{h}_n$ |
|---|---|---|---|---|---|---|---|---|---|---|
| Exp(1) | 0.0206 | 0.2524 | 0.7030 | 0.8142 | 0.8452 | 0.8594 | 0.8570 | 0.8576 | 0.8518 | 0.8508 |
| Gamma(2, 1) | 0.0000 | 0.0208 | 0.2484 | 0.4018 | 0.4602 | 0.4828 | 0.4790 | 0.4792 | 0.4764 | 0.4670 |
| U(0, 1) | 0.0000 | 0.0004 | 0.0312 | 0.1362 | 0.2576 | 0.3316 | 0.3688 | 0.3966 | 0.4170 | 0.3210 |
| Beta(2, 1) | 0.0000 | 0.0000 | 0.0004 | 0.0164 | 0.1322 | 0.2654 | 0.3320 | 0.3826 | 0.4124 | 0.2616 |
| Beta(2, 6) | 0.0000 | 0.0010 | 0.0478 | 0.1286 | 0.1702 | 0.1996 | 0.2074 | 0.2148 | 0.2180 | 0.1860 |
| Laplace(0, 0.25) | 0.0322 | 0.1532 | 0.2104 | 0.1756 | 0.1360 | 0.1072 | 0.0788 | 0.0640 | 0.0538 | 0.1052 |
| Lognormal(2, 0.25) | 0.0000 | 0.0000 | 0.0050 | 0.0364 | 0.0716 | 0.0948 | 0.1010 | 0.1072 | 0.1102 | 0.0778 |
| t(3) | 0.0310 | 0.2058 | 0.2634 | 0.2340 | 0.2024 | 0.1692 | 0.1442 | 0.1246 | 0.1108 | 0.1686 |
| t(5) | 0.0146 | 0.1006 | 0.1390 | 0.1192 | 0.0980 | 0.0846 | 0.0706 | 0.0654 | 0.0598 | 0.0828 |
| Weilbull(2, 0.5) | 0.0000 | 0.0000 | 0.0250 | 0.0718 | 0.1034 | 0.1234 | 0.1292 | 0.1322 | 0.1338 | 0.1092 |

Table 3. Power Estimates for Various Choices of $h$ and Alternatives ($n = 50$, replicate=5000, $\alpha = 0.05$)

| Alternative | h=.05 | h=.10 | h=.15 | h=.20 | h=.25 | h=.30 | h=.35 | h=.40 | h=.45 | h=.50 | $\hat{h}_n$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Exp(1) | 0.0108 | 0.9970 | 0.9986 | 0.9994 | 0.9990 | 0.9990 | 0.9990 | 0.9978 | 0.9968 | 0.9954 | 0.9992 |
| Gamma(2, 1) | 0.0000 | 0.7652 | 0.8630 | 0.9220 | 0.9174 | 0.9144 | 0.8942 | 0.8714 | 0.8330 | 0.7992 | 0.9200 |
| U(0, 1) | 0.0000 | 0.3626 | 0.8372 | 0.9266 | 0.9486 | 0.9638 | 0.9694 | 0.9724 | 0.9738 | 0.9746 | 0.9248 |
| Beta(2, 1) | 0.0000 | 0.0252 | 0.7940 | 0.8994 | 0.9370 | 0.9428 | 0.9438 | 0.9412 | 0.9386 | 0.9348 | 0.9058 |
| Beta(2, 6) | 0.0000 | 0.2166 | 0.3922 | 0.5594 | 0.5680 | 0.5786 | 0.5656 | 0.5486 | 0.5230 | 0.5032 | 0.5528 |
| Laplace(0, 0.25) | 0.0256 | 0.4128 | 0.3040 | 0.2278 | 0.1302 | 0.0666 | 0.0324 | 0.0190 | 0.0116 | 0.0084 | 0.2280 |
| Lognormal(2, 0.25) | 0.0000 | 0.0302 | 0.0978 | 0.2314 | 0.2156 | 0.2204 | 0.1946 | 0.1780 | 0.1564 | 0.1454 | 0.2242 |
| t(3) | 0.0266 | 0.4792 | 0.3868 | 0.3162 | 0.2288 | 0.1556 | 0.1042 | 0.0682 | 0.0468 | 0.0326 | 0.3156 |
| t(5) | 0.0130 | 0.2148 | 0.1534 | 0.1134 | 0.0690 | 0.0470 | 0.0312 | 0.0236 | 0.0176 | 0.0136 | 0.1136 |
| Weilbull(2, 0.5) | 0.0000 | 0.0916 | 0.1974 | 0.3210 | 0.3276 | 0.3394 | 0.3306 | 0.3186 | 0.2992 | 0.2822 | 0.3144 |

Table 4. Power Estimates for Various Choices of $h$ and Alternatives ( $n = 100$, replicate=5000, $\alpha = 0.05$ )

| Alternative | h=.05 | h=.10 | h=.15 | h=.20 | h=.25 | h=.30 | h=.35 | h=.40 | h=.45 | h=.50 | $\hat{h}_n$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Exp(1) | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9996 | 1.0000 |
| Gamma(2, 1) | 0.8404 | 0.9972 | 0.9982 | 0.9976 | 0.9952 | 0.9900 | 0.9816 | 0.9692 | 0.9444 | 0.9150 | 0.9980 |
| U(0, 1) | 0.1562 | 0.9980 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Beta(2, 1) | 0.0000 | 0.9932 | 0.9996 | 0.9996 | 0.9998 | 1.0000 | 0.9998 | 0.9996 | 0.9996 | 0.9994 | 0.9996 |
| Beta(2, 6) | 0.1050 | 0.8556 | 0.9138 | 0.9132 | 0.8974 | 0.8784 | 0.8524 | 0.8254 | 0.7988 | 0.7708 | 0.9144 |
| Laplace(0. 0.25) | 0.6974 | 0.6000 | 0.4374 | 0.2446 | 0.1004 | 0.0324 | 0.0094 | 0.0032 | 0.0024 | 0.0020 | 0.2888 |
| Lognormal(2, 0.25) | 0.0006 | 0.3398 | 0.4376 | 0.4026 | 0.3494 | 0.2952 | 0.2426 | 0.2078 | 0.1764 | 0.1578 | 0.4164 |
| t(3) | 0.7606 | 0.6786 | 0.5154 | 0.3248 | 0.1788 | 0.0914 | 0.0498 | 0.0270 | 0.0140 | 0.0086 | 0.3630 |
| t(5) | 0.3672 | 0.2526 | 0.1464 | 0.0708 | 0.0282 | 0.0134 | 0.0054 | 0.0034 | 0.0028 | 0.0016 | 0.0812 |
| Weilbull(2, 0.5) | 0.0184 | 0.5692 | 0.6724 | 0.6718 | 0.6416 | 0.6042 | 0.5658 | 0.5332 | 0.4996 | 0.4736 | 0.6770 |

These power simulations show that for a fixed $n$, there does not exist an $h$ which is optimal uniformly for all alternatives considered. This makes sense in situations when the tests have to take into account departures from the null hypothesis over all directions in nonparametric density alternatives, since alternatives are vague and the choice of the bandwidth $h$ is designed to guard against all nonparametric density alternatives, and it is natural not to expect that the chosen $\hat{h}_n$ would beat all other choices of $h$ in terms of power. Our simulation results are very encouraging. From Tables 2-3-4, we can see that the powers for $\hat{h}_n$ are far greater than or as close as the median powers for all choices of $h$ for sample sizes $n = 20$, 50 and 100. These results suggest that the data-driven method of choosing $h$ is a very promising procedure to overcome the dependence problem of the power of the test on $h$. These results also suggest one possible way of choosing the optimal $h$ in situations where one has in mind a priori knowledge on a particular alternative being tested against, i.e., if a specific alternative is of special interest then the best way of choosing $h$ would be to choose $h$ that yields the highest power in the direction of this alternative for the given sample size and level $\alpha$. Furthermore, these results suggest another way to improve power against all nonparametric density alternatives or a particular alternative in mind is to increase the sample size by its own nature in nonparametric setting. This was shown in our simulations that all the powers increases as sample size increases. To see if it was actually the case, we repeated the simulation study with sample size $n$ increased to 200. The results are presented in Table 5 for h=0.01 to 0.50, which support the fact that the asymptotics of the theory presented here takes some time to "take effect".

Table 5. Power Estimates for Various Choices of $h$ and Alternatives ( $n = 200$, h=0.01 to 0.50, replicate=5000, $\alpha = 0.05$ )

| Alternative | h=.01 | h=.02 | h=.03 | h=0.04 | h=.05 | h=.10 | h=.15 | h=.20 | h=.25 | h=.30 | h=.35 | h=.40 | h=.45 | h=.50 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Exp(1) | 0.0040 | 0.9986 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Gamma(2, 1) | 0.0000 | 0.1000 | 0.9832 | 0.9996 | 1.0000 | 1.0000 | 0.9998 | 0.9998 | 0.9998 | 0.9992 | 0.9974 | 0.9924 | 0.9814 | 0.9642 |
| U(0, 1) | 0.0000 | 0.0000 | 0.3184 | 0.9986 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Beta(2, 1) | 0.0000 | 0.0000 | 0.0000 | 0.7154 | 0.9996 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| Beta(2, 6) | 0.0000 | 0.0000 | 0.1402 | 0.8194 | 0.9764 | 0.9980 | 0.9970 | 0.9938 | 0.9898 | 0.9834 | 0.9708 | 0.9618 | 0.9534 | 0.9404 |
| Laplace(0. 0.25) | 0.0194 | 0.8936 | 0.9292 | 0.9144 | 0.9082 | 0.7778 | 0.4742 | 0.1850 | 0.0470 | 0.0090 | 0.0014 | 0.0002 | 0.0000 | 0.0000 |
| Lognormal(2, 0.25) | 0.0000 | 0.0000 | 0.0000 | 0.0494 | 0.3224 | 0.7142 | 0.6246 | 0.5088 | 0.4132 | 0.3318 | 0.2706 | 0.2236 | 0.1870 | 0.1654 |
| t(3) | 0.0124 | 0.9272 | 0.9536 | 0.9466 | 0.9408 | 0.8182 | 0.5286 | 0.2356 | 0.0816 | 0.0328 | 0.0158 | 0.0088 | 0.0038 | 0.0020 |
| t(5) | 0.0052 | 0.5228 | 0.6030 | 0.5640 | 0.5334 | 0.2830 | 0.0784 | 0.0182 | 0.0050 | 0.0020 | 0.0014 | 0.0006 | 0.0006 | 0.0006 |
| Weilbull(2, 0.5) | 0.0000 | 0.0000 | 0.0078 | 0.3576 | 0.7454 | 0.9472 | 0.9210 | 0.8748 | 0.8348 | 0.7898 | 0.7542 | 0.7186 | 0.6858 | 0.6574 |

## 5. Appendix: Proofs of Main Results

This appendix presents the proofs of the lemmas and the theorem in section 3. To keep notation simple, we suppress the dependence of $h_n$ on the sample size $n$ to $h$. Then the weight $w_{h_n ij}$ is conveniently abbreviated as $w_{hij}$ and $\widehat{q}_{n,h_n}$ as $\widehat{q}_{n,h}$.

### *5.1 Proof of Lemma 2*

Let $\{U_i\}_{i=1}^n$ be independent observations from the uniform distribution $U(0,1)$ and put $U^*_{(i)}$ equal to 0, $U_{(i)}$, 1 according to $i = 0$, $1 \le i \le n$, $i = n+1$. The following Lemma 3 is quoted for convenience from Rao and Zhao (1997) and its proof is omitted here.

**Lemma 3.** *Assume that assumption* $\frac{nh}{\log n} \to \infty$ *holds, then*

$$T_n \stackrel{def}{=} max^* \left| \left( (n+1) \frac{(U^*_{(j)} - U^*_{(i)})}{(j-i)} \right) - 1 \right| \xrightarrow{a.s.} 0,$$

*where* $max^*$ *is taken for all* $(i, j)$ *with* $0 \leq i < j \leq n+1$ *and* $j - i \geq Cnh$.

Let $\mathscr{D}_{inh} := \frac{f(X_{(i)})\widehat{q}_{n,h_n}(\frac{i}{n+1};\mathbb{Q}_n)}{\widehat{q}_{n,h_n}(\frac{i}{n+1};\mathbb{U}_n)}$, then

$$\begin{aligned} -\mathscr{V}_{n,h} &= \sum_{i\in\mathcal{I}_1} \log \mathscr{D}_{inh} + \sum_{i\in\mathcal{I}_2} \log \mathscr{D}_{inh} + \sum_{i\in\mathcal{I}_3} \log \mathscr{D}_{inh} \\ &= \mathscr{V}^{(1)}_{nh} + \mathscr{V}^{(2)}_{nh} + \mathscr{V}^{(3)}_{nh}, \end{aligned}$$

where we decompose the index set for $i$ into three segments $\mathcal{I}_1 := \{k \in \mathbb{N}; \quad 1 \leq k \leq nh\}$, $\mathcal{I}_2 := \{k \in \mathbb{N}; \quad nh < k < n - nh\}$ and $\mathcal{I}_3 := \{k \in \mathbb{N}; \quad n - nh \leq k \leq n\}$. We shall demonstrate the proof for $\mathcal{I}_2$.

**Lemma 4.** *Under assumptions H.2 and F.3, there exists* $0 < \alpha \leq 1$ *such that, for n large,*

$$\alpha \leq \frac{f(Q(U_{(i)}))}{f(Q(p))} \leq \alpha^{-1},$$

*almost surely for all* $i \in \mathcal{I}_2$ *and* $p \in (U_{(\underline{m}_i)}, U_{(\overline{m}_i)})$.

*Proof.* Under assumptions H.2 and F.3, Lemma 1 of Csörgő and Révész (1978) implies that

$$\frac{f(Q(u_1))}{f(Q(u_2))} \leq \varDelta^{\gamma}(u_1, u_2),$$

for every pair $u_1, u_2 \in (0, 1)$ with $\gamma$ as in F.3. and $\varDelta(u_1, u_2) := \frac{u_1 \vee u_2(1-u_1 \wedge u_2)}{u_1 \wedge u_2(1-u_1 \vee u_2)}$. It follows from the symmetry of $\varDelta(u_1, u_2)$ that

$$\varDelta^{-\gamma}(U_{(i)}, p) \leq \frac{f(Q(U_{(i)}))}{f(Q(p))} \leq \varDelta^{\gamma}(U_{(i)}, p).$$

If $U_{(\underline{m}_i)} \leq p \leq U_{(i)}$, then

$$\varDelta^{-\gamma}(U_{(i)}, p) \geq \left[1 - \frac{U_{(i)} - U_{(\underline{m}_i)}}{1 - U_{(\underline{m}_i)}}\right]^{\gamma} \left[1 - \frac{U_{(i)} - U_{(\underline{m}_i)}}{U_{(i)}}\right]^{\gamma}.$$

By Lemma 3, with probability one, for $n$ large, we have for all $i \in \mathcal{I}_2$ that

$$0 \leq \frac{U_{(i)} - U_{(\underline{m}_i)}}{1 - U_{(\underline{m}_i)}} \leq \frac{U_{(i)} - U_{(\underline{m}_i)}}{1 - U_{(i)}} \leq \frac{1}{2},$$

and

$$0 \leq \frac{U_{(i)} - U_{(\underline{m}_i)}}{U_{(i)}} \leq \frac{1}{2}.$$

Hence $\varDelta^{-\gamma}(U_{(i)}, p) \geq \left(\frac{1}{4}\right)^{\gamma}$.
If $U_{(i)} \leq p \leq U_{(\overline{m}_i)}$, then

$$\varDelta^{-\gamma}(U_{(i)}, p) \geq \left[1 - \frac{U_{(\overline{m}_i)} - U_{(i)}}{1 - U_{(i)}}\right]^{\gamma} \left[1 - \frac{U_{(\overline{m}_i)} - U_{(i)}}{U_{(\overline{m}_i)}}\right]^{\gamma}.$$

Again by using the same arguments given above, we have, with probability one, for $n$ large,

$$0 \leq \frac{U_{(\overline{m}_i)} - U_{(i)}}{1 - U_{(i)}} \leq \frac{1}{2} \qquad \text{and} \qquad 0 \leq \frac{U_{(\overline{m}_i)} - U_{(i)}}{U_{(\overline{m}_i)}} \leq \frac{1}{2},$$

for all $i \in \mathcal{I}_2$, which proves that $\varDelta^{-\gamma}(U_{(i)}, p) \geq \left(\frac{1}{4}\right)^{\gamma}$.

So almost surely for all $i \in \mathcal{I}_2$ and $U_{(\underline{m}_i)} \leq p \leq U_{(\overline{m}_i)}$, we have $\varDelta^{-\gamma}(U_{(i)}, p) \geq \left(\frac{1}{4}\right)^{\gamma}$. The lemma follows by observing $\varDelta^{\gamma}(U_{(i)}, p) \leq (4)^{\gamma}$, almost surely for all $i \in \mathcal{I}_2$. □

**Lemma 5.** *If* $\frac{nh}{\log n} \to \infty$, *then*

$$\sqrt{h}\left|\mathscr{V}_{n,h}^{(2)}\right| \leq C\sqrt{h}\sum_{i\in\mathcal{I}_2}|\mathscr{D}_{inh}-1|.$$

*Proof.* Using mean value theorem,

$$\begin{aligned}\mathscr{D}_{inh} &= \frac{f(X_{(i)})\sum\limits_{j\in\mathcal{J}_i} w_{hij}X_{(j)}}{\sum\limits_{j\in\mathcal{J}_i} w_{hij}U_{(j)}}\\ &= \frac{f(Q(U_{(i)}))}{f(Q(U_{ih}^*))},\end{aligned}$$

where $U_{(\underline{m}_i)} < U_{ih}^* < U_{(\overline{m}_i)}$.

It follows from lemma 4 that there exists a positive constant $\alpha \leq 1$ such that for $n$ large, $\alpha \leq \mathscr{D}_{inh} \leq \alpha^{-1}$ almost surely for all $i \in \mathcal{I}_2$. This fact together with Taylor's theorem implies that for all $i \in \mathcal{I}_2$

$$\left|\log \mathscr{D}_{inh}\right| \leq C\left|\mathscr{D}_{inh}-1\right|,$$

and hence

$$\begin{aligned}\sqrt{h}\left|\mathscr{V}_{nh}^{(2)}\right| &= \sqrt{h}\left|\sum_{i\in\mathcal{I}_2}\log\mathscr{D}_{inh}\right|\\ &\leq C\sqrt{h}\sum_{i\in\mathcal{I}_2}|\mathscr{D}_{inh}-1|.\end{aligned}$$

□

**Lemma 6.** *If* $n^2h^3\log^4 n \to 0 \quad as\ n \to \infty$, *then*

$$\sqrt{h}\sum_{i\in\mathcal{I}_2}|\mathscr{D}_{inh}-1| = o_P(1).$$

*Proof.* Since

$$\begin{aligned}\mathscr{D}_{inh} &= \frac{f(X_{(i)})\sum\limits_{j\in\mathcal{J}_i} w_{hij}X_{(j)}}{\sum\limits_{j\in\mathcal{J}_i} w_{hij}U_{(j)}}\\ &= \frac{f(Q(U_{(i)}))\sum\limits_{j\in\mathcal{J}_i} w_{hij}Q(U_{(j)})}{\sum\limits_{j\in\mathcal{J}_i} w_{hij}U_{(j)}}.\end{aligned}$$

Applying Taylor's theorem with an integral form of the remainder, we have

$$\begin{aligned}&\sum_{j\in\mathcal{J}_i} w_{hij}Q(U_{(j)})\\ =&\sum_{j\in\mathcal{J}_i} w_{hij}\left[Q(U_{(i)}) + q(U_{(i)})(U_{(j)}-U_{(i)}) + \int_{U_{(i)}}^{U_{(j)}}(U_{(j)}-p)\frac{dq(p)}{dp}dp\right]\\ =&q(U_{(i)})\sum_{j\in\mathcal{J}_i} w_{hij}U_{(j)} + \sum_{j\in\mathcal{J}_i} w_{hij}\int_{U_{(i)}}^{U_{(j)}}(U_{(j)}-p)\frac{dq(p)}{dp}dp.\end{aligned}$$

So

$$f(Q(U_{(i)})\sum_{j\in\mathcal{J}_i} w_{hij}Q(U_{(j)}) - \sum_{j\in\mathcal{J}_i} w_{hij}U_{(j)} = \sum_{j\in\mathcal{J}_i} w_{hij}\int_{U_{(i)}}^{U_{(j)}} (U_{(j)} - p)f(Q(U_{(i)})\frac{dq(p)}{dp}dp.$$

We have

$$\begin{aligned}
\mathscr{D}_{inh} - 1 &= \frac{\sum\limits_{j\in\mathcal{J}_i} w_{hij}\int_{U_{(i)}}^{U_{(j)}}(U_{(j)} - p)f(Q(U_{(i)}))\frac{dq(p)}{dp}dp}{\sum\limits_{j\in\mathcal{J}_i} w_{hij}U_{(j)}} \\
&= \frac{\sum\limits_{j\in\mathcal{J}_i} w_{hij}(U_{(j)} - U_{(i)})\int_{U_{(i)}}^{U_{(j)}}\frac{U_{(j)}-p}{U_{(j)}-U_{(i)}}f(Q(U_{(i)}))\frac{dq(p)}{dp}dp}{\sum\limits_{j\in\mathcal{J}_i} w_{hij}U_{(j)}}.
\end{aligned}$$

Using Lemma 4 and $w_{hij}(U_{(j)} - U_{(i)}) \geq 0$ for all $j \in \mathcal{J}_i$, we have

$$\begin{aligned}
&\sqrt{h}\sum_{i\in\mathcal{I}_2}|\mathscr{D}_{inh} - 1| \\
&\leq \sqrt{h}\sum_{i\in\mathcal{I}_2}\frac{\sum\limits_{j\in\mathcal{J}_i} w_{hij}(U_{(j)} - U_{(i)})\left|\int_{U_{(i)}}^{U_{(j)}}\frac{U_{(j)}-p}{U_{(j)}-U_{(i)}}f(Q(U_{(i)}))\frac{dq(p)}{dp}dp\right|}{\sum\limits_{j\in\mathcal{J}_i} w_{hij}U_{(j)}} \\
&\leq \sqrt{h}\sum_{i\in\mathcal{I}_2}\frac{\sum\limits_{j\in\mathcal{J}_i} w_{hij}(U_{(j)} - U_{(i)})\left(\int_{U_{(i)}}^{U_{(\overline{m}_i)}} Cf(Q(p))|\frac{dq(p)}{dp}|dp + \int_{U_{(\underline{m}_i)}}^{U_{(i)}} Cf(Q(p))|\frac{dq(p)}{dp}|dp\right)}{\sum\limits_{j\in\mathcal{J}_i} w_{hij}U_{(j)}} \\
&= C\sqrt{h}\sum_{i\in\mathcal{I}_2}\frac{\left(\int_{U_{(i)}}^{U_{(\overline{m}_i)}}|J(p)|dp + \int_{U_{(\underline{m}_i)}}^{U_{(i)}}|J(p)|dp\right)\sum\limits_{j\in\mathcal{J}_i} w_{hij}(U_{(j)} - U_{(i)})}{\sum\limits_{j\in\mathcal{J}_i} w_{hij}U_{(j)}} \\
&= C\sqrt{h}\sum_{i\in\mathcal{I}_2}\left(\int_{U_{(i)}}^{U_{(\overline{m}_i)}}|J(p)|dp + \int_{U_{(\underline{m}_i)}}^{U_{(i)}}|J(p)|dp\right).
\end{aligned}$$

Let $g_{r,s}^{(n)}(x, y)$ denote the joint density of the order statistic $U_{(r)}$ and $U_{(s)}$ of a random sample of size n from the uniform $U(0, 1)$ distribution, where $1 \leq r < s \leq n$, and let $b_{u,v}(x)$ denote the beta density function with parameters $u$ and $v$, then

$$\begin{aligned}
\mathbb{E}\int_{U_{(\underline{m}_i)}}^{U_{(i)}}|J(p)|dp &\leq \iint\limits_{0<x_1\leq x_2<1}\frac{g_{(\underline{m}_i),i}^{(n)}(x_1, x_2)}{x_1(1 - x_2)}\int_{x_1}^{x_2} p(1 - p)|J(p)|dpdx_1dx_2 \\
&\leq \iint\limits_{0<x_1\leq x_2<1}(x_2 - x_1)\frac{g_{(\underline{m}_i),i}^{(n)}(x_1, x_2)}{x_1(1 - x_2)}G_0(x_1)dx_1dx_2 \\
&= \frac{a_n i - \underline{m}_i}{nh}\iint\limits_{0<x_1\leq x_2<1}\frac{nh\cdot n}{(\underline{m}_i - 1)(n - i)}g_{(\underline{m}_i-1),i}^{(n-1)}(x_1, x_2)G_0(x_1)dx_1dx_2 \\
&= \frac{a_n i - \underline{m}_i}{nh}\cdot\frac{n^2h}{(\underline{m}_i - 1)(n - i)}\int_0^1 G_0(x_1)b_{(\underline{m}_i-1),n-\underline{m}_i+1}(x_1)dx_1 \\
&\leq C\frac{n^2h}{(\underline{m}_i - 1)(n - i)}(R_{n-1} - R_{\underline{m}_i-2} + 1),
\end{aligned}$$

where $G_0(x) = \sup_{1>y\geq x} \frac{1}{y-x}\int_x^y p(1-p)|J(p)|dp$. The last equality follows from the fact that $g^{(n-1)}_{(\underline{m}_i-1),i}(x_1, x_2)$ is the joint density of the uniform order statistics $U_{(\underline{m}_i-1)}$ and $U_{(i)}$ of the sample size $(n-1)$.

Noting that $\frac{n^2 h}{(\underline{m}_i-1)(n-i)} = \frac{n^2 h}{n-i+\underline{m}_i-1}(\frac{1}{\underline{m}_i-1} + \frac{1}{n-i})$, we have

$$\sqrt{h}\sum_{i\in\mathcal{I}_2} E\int_{U_{(\underline{m}_i)}}^{U_{(i)}} |J(p)|dp = O((n^2h^3\log^4 n)^{\frac{1}{2}}) + O((n^2h^3\log^2 n)^{\frac{1}{2}})$$
$$= O((n^2h^3\log^4 n)^{\frac{1}{2}}).$$

By the same argument, it can be shown that

$$\sqrt{h}\sum_{i\in\mathcal{I}_2} E\int_{U_{(i)}}^{U_{(\overline{m}_i)}} |J(p)|dp = O((n^2h^3\log^4 n)^{\frac{1}{2}}).$$

Hence, by assumption $n^2h^3\log^4 n \longrightarrow 0 \quad as\; n \longrightarrow \infty$, we have

$$\sqrt{h}\sum_{i\in\mathcal{I}_2}|\mathscr{D}_{inh} - 1| = o_P(1).$$

□

Lemma 2 is proved by combining Lemmas 5 and 6 with an analogous lengthy arguments as the above proofs establishing that the parts corresponding to $\mathcal{I}_1$ and $\mathcal{I}_3$ with less than $nh$ terms are asymptotically negligible.

*5.2 Proof of Theorem 1*

Put $Z_k = \log(1/U_k), k = 0, 1, \cdots, S_0 = 0, S_j = \sum_{k=1}^j Z_k, k = 1, 2, \cdots,$

$$\tilde{\mathbb{U}}_n(y) = \frac{S_k}{S_{n+1}}, \qquad \text{if} \quad \frac{k-1}{n} < y \leq \frac{k}{n}, \qquad k = 1, 2, \cdots, n.$$

Then $Z_k$ are independent exponential random variables with mean value one and, for each $n$,

$$\{\mathbb{U}_n(y); \quad 0 \leq y \leq 1\} \overset{D}{=} \{\tilde{\mathbb{U}}_n(y); \quad 0 \leq y \leq 1\}.$$

A Taylor expansion shows that

$$\mathscr{U}_{n,h} \overset{D}{=} -\sum_{i=1}^n \log\widehat{q}_{n,h}(\tfrac{i}{n+1}; \tilde{\mathbb{U}}_n)$$
$$=: -\sum_{i=1}^n \varDelta_i + \frac{1}{2}\sum_{i=1}^n \varDelta_i^2 + \tilde{R}_{nh},$$

where $\varDelta_i = \widehat{q}_{n,h}(\frac{i}{n+1}; \tilde{\mathbb{U}}_n) - 1$ and $\tilde{R}_{nh} = -\frac{1}{3}\sum_{i=1}^n \frac{\varDelta_i^3}{1+\theta_i\varDelta_i^3}, \quad 0 < \theta_i < 1$. We define the centered standard exponential variables as $\xi_k = Z_k - 1,\ 1 \leq k \leq N$ with $N = n+1$. Let $T_{i,n} := \frac{1}{N}\sum_{j\in\mathcal{J}_i}(w_{hij}\xi_{\cdot j})$ where $\xi_{\cdot j} := \sum_{k=1}^{j}\xi_k$, $\overline{Y}_n := \frac{1}{N}\sum_{i=1}^N \xi_i$, $C_{nhik} := K(\frac{a_n i-k+1}{nh}) - K(\frac{a_n i-\lfloor a_n i+nh\rfloor}{nh})$ and $D_{i,n} := \frac{1}{N}\sum_{j\in\mathcal{J}_i} jw_{hij}$. Then

$$\mathscr{U}_{n,h} := -\left(\sum_{i\in\mathcal{I}_1} + \sum_{i\in\mathcal{I}_2} + \sum_{i\in\mathcal{I}_3}\right)\log\widehat{q}_{n,h_n}(\tfrac{i}{n+1}; \mathbb{U}_n)$$
$$= \mathscr{U}^{(1)}_{nh} + \mathscr{U}^{(2)}_{nh} + \mathscr{U}^{(3)}_{nh},$$

**Lemma 7.** *Assume kernel function $K(x)$ satisfies conditions (K.1-2-3); smoothing parameter h satisfies conditions (H.1-2-3); and $f(x)$ satisfies conditions (F.1-2-3). Under null hypothesis $H_0$, we have*

$$2\sqrt{h}\left(-\sum_{i=1}^{n}\Delta_i+\frac{1}{2}\sum_{i=1}^{n}\Delta_i^2+\tilde{R}_{nh}\right)=\sqrt{h}\sum_{i=1}^{n}T_{i,n}^2+o_P(1).$$

*Proof.* We can write $2\sqrt{h}\left(-\sum_{i\in\mathcal{I}_2}\Delta_i+\frac{1}{2}\sum_{i\in\mathcal{I}_2}\Delta_i^2+\tilde{R}_{nh}\right)=\sqrt{h}\sum_{i\in\mathcal{I}_2}T_{i,n}^2+\sum_{i=1}^{11}B_{i,n}+\sqrt{h}\tilde{R}_{nh}$,
where

$$\begin{aligned}
B_{1,n} &= \sum_{i\in\mathcal{I}_2}\sqrt{h}(\tfrac{1}{2}D_{i,n}^2-2D_{i,n}+\tfrac{3}{2}),\\
B_{2,n} &= \sum_{i\in\mathcal{I}_2}\sqrt{h}(D_{i,n}-1)T_{i,n},\\
B_{3,n} &= \sum_{i\in\mathcal{I}_2}\sqrt{h}(D_{i,n}-D_{i,n}^2)\overline{Y}_n,\\
B_{4,n} &= \sum_{i\in\mathcal{I}_2}\sqrt{h}(D_{i,n}-1)\overline{Y}_n,\\
B_{5,n} &= \sum_{i\in\mathcal{I}_2}\sqrt{h}(\overline{Y}_n-T_{i,n}),\\
B_{6,n} &= \sum_{i\in\mathcal{I}_2}\sqrt{h}(2-2D_{i,n})T_{i,n}\overline{Y}_n,\\
B_{7,n} &= \sum_{i\in\mathcal{I}_2}\sqrt{h}\left(\tfrac{3}{2}D_{i,n}^2-2D_{i,n}\right)\overline{Y}_n^2,\\
B_{8,n} &= \sum_{i\in\mathcal{I}_2}\sqrt{h}(-T_{i,n}^2)\overline{Y}_n,\\
B_{9,n} &= \sum_{i\in\mathcal{I}_2}\sqrt{h}(3D_{i,n}-2)T_{i,n}\overline{Y}_n^2,\\
B_{10,n} &= \sum_{i\in\mathcal{I}_2}\sqrt{h}T_{i,n}^2\overline{Y}_n^2,\\
B_{11,n} &= \sqrt{h}\sum_{i\in\mathcal{I}_2}\left[(D_{i,n}^2+2D_{i,n}T_{i,n}+T_{i,n}^2)O_P\left(n^{-\frac{3}{2}}\right)+(D_{i,n}+T_{i,n})O_P\left(n^{-\frac{3}{2}}\right)\right].
\end{aligned}$$

Using the facts $D_{i,n}=1+O(\frac{1}{nh})+O(\frac{1}{n^{1+\epsilon}h^{2+\epsilon}})$ for any $\epsilon>0$ and $\overline{Y}_n=O_p(\frac{1}{\sqrt{n}})$, elementary calculation shows that $B_{i,n}=o_P(1), i=1,2,\cdots,11$ and $\sqrt{h}\tilde{R}_{nh}=o_p(1)$.

□

**Lemma 8.** *Assume kernel function $K(x)$ satisfies conditions (K.1-2-3); smoothing parameter h satisfies conditions (H.1-2-3); and $f(x)$ satisfies conditions (F.1-2-3). Then the mean of $\sum_{i=1}^{n}T_{i,n}^2$ is*

$$\mathbb{E}\left(\sum_{i=1}^{n}T_{i,n}^2\right)=\frac{1}{h}\int_{-1}^{1}K^2(x)dx+o\left(\frac{1}{n^2h^3}\right).$$

*Proof.*

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^{n}T_{i,n}^2\right) &= \mathbb{E}\left(\tfrac{1}{N^2h^2}\sum_{i=1}^{n}\sum_{a_ni-nh\le k\le a_ni+nh}C_{nhik}^2\xi_k^2\right)\\
&= \tfrac{1}{N^2h^2}\sum_{i=1}^{n}\sum_{a_ni-nh\le k\le a_ni+nh}C_{nhik}^2\\
&= \tfrac{1}{h}\int_{-1}^{1}K^2(x)dx+o\left(\tfrac{1}{n^2h^3}\right).
\end{aligned}$$

□

**Lemma 9.** *Assume kernel function $K(x)$ satisfies conditions (K.1-2-3); smoothing parameter $h$ satisfies conditions (H.1-2-3); and $f(x)$ satisfies conditions (F.1-2-3). Under null hypothesis $H_0$, we have*

$$\sqrt{h}(T_{n,h}-\mu_{n,h}(K))) = \sqrt{h}\left(\sum_{i\in\mathcal{I}_2} T_{i,n}^2 - \mu_{n,h}(K)\right) + o_P(1),$$

*where* $\mu_{n,h}(K)=\frac{1}{h}\int_{-1}^{1} K^2(x)dx + o\left(\frac{1}{n^2h^3}\right)$.

*Proof.*

$$\begin{aligned}
&\sqrt{h}(T_{n,h}-\mu_{n,h}(K))\\
&= \sqrt{h}\left(2\mathscr{U}_{nh}-\mu_{n,h}(K)\right) + 2\sqrt{h}\mathscr{V}_{nh} + 2\sqrt{h}\mathscr{W}_n(\hat{\theta})\\
&\overset{d}{=} 2\sqrt{h}\left[-\sum_{i\in\mathcal{I}_2}\varDelta_i + \frac{1}{2}\sum_{i\in\mathcal{I}_2}\varDelta_i^2 - \mu_{n,h}(K)\right] + \tilde{R}_{nh} + o_P(1)\\
&= \sqrt{h}\left(\sum_{i\in\mathcal{I}_2} T_{i,n}^2 - \mu_{n,h}(K)\right) + o_P(1).
\end{aligned}$$

□

Hence the limiting behavior of $\sqrt{h}(T_{n,h}-\mu_{n,h}(K))$ is determined by $\sqrt{h}\left(\sum_{i\in\mathcal{I}_2} T_{i,n}^2 - \mu_{n,h}(K)\right)$. Since $T_{i,n}^2$ are not independent, it naturally comes to mind that the martingale approach might be used for the asymptotic results for $T_{n,h}$.

**Lemma 10.** *Assume kernel function $K(x)$ satisfies conditions (K.1-2-3); smoothing parameter $h$ satisfies conditions (H.1-2-3); and $f(x)$ satisfies conditions (F.1-2-3). We have*

$$\begin{aligned}
&\sqrt{h}\sum_{i\in\mathcal{I}_2}(T_{i,n}^2-\mu_{n,h}(K))\\
&= \sum_{(a_n+1)nh<l\le a_n n-(a_n-1)nh} \sqrt{h}W_l + o_P(1).
\end{aligned}$$

*where $W_l$ is a martingale difference sequence w. r. t. $\mathscr{F}_{l-1} = \sigma(\xi_1,\xi_2,\ldots,\xi_{l-1})$.*

*Proof.*

$$\begin{aligned}
&\sum_{i\in\mathcal{I}_2}(T_{i,n}^2-\mu_{n,h}(K))\\
&= \sum_{i\in\mathcal{I}_2}\left[\left(\tfrac{1}{N}\sum_{j\in\mathcal{J}_i} w_{hij}\xi_{\cdot j}\right)^2 - \mu_{n,h}(K)\right]\\
&= \sum_{i\in\mathcal{I}_2}\left[\left(\tfrac{1}{Nh}\sum_{a_ni-nh\le k\le a_ni+nh} C_{nhik}\xi_k\right)^2 - \mu_{n,h}(K)\right]\\
&= \sum_{i\in\mathcal{I}_2}\left[\left(\tfrac{1}{N^2h^2}\sum_{a_ni-nh\le k\le a_ni+nh} C_{nhik}^2\xi_k^2 - \mu_{n,h}(K)\right) + \tfrac{2}{N^2h^2}\sum\sum_{a_ni-nh\le k<l\le a_ni+nh} C_{nhik}C_{nhil}\xi_k\xi_l\right]\\
&= \sum_{i\in\mathcal{I}_2}\left(\tfrac{1}{N^2h^2}\sum_{a_ni-nh\le k\le a_ni+nh} C_{nhik}^2\xi_k^2 - \mu_{n,h}(K)\right)\\
&+ \tfrac{2}{N^2h^2}\sum\sum_{0<k<l\le a_nn-(a_n-1)nh}\left(\sum_{\frac{1}{a_n}(l-nh)\vee nh\le i\le\frac{1}{a_n}(k+nh)\wedge(n-nh)} C_{nhik}C_{nhil}\right)\xi_k\xi_l\\
&:= A_I + A_{II}.
\end{aligned}$$

$$\begin{aligned}
A_{II} &= \tfrac{2}{N^2h^2} \sum_{(a_n+1)nh<l<a_nn-(a_n-1)nh} \sum_{0<k<l} \left( \sum_{\frac{1}{a_n}(l-nh)\le i\le \frac{1}{a_n}(k+nh)\wedge(n-nh)} C_{nhik}C_{nhil} \right) \xi_k\xi_l \\
&+ \tfrac{2}{N^2h^2} \sum_{1<l\le(a_n+1)nh} \sum_{0<k<l} \left( \sum_{nh\le i\le \frac{1}{a_n}(k+nh)} C_{nhik}C_{nhil} \right) \xi_k\xi_l \\
&= \tfrac{2}{N^2h^2} \sum_{(a_n+1)nh<l\le a_nn-(a_n-1)nh} \sum_{0<k\le(l-1)\wedge(a_nn-(a_n+1)nh)} d_{nhkl}\xi_k\xi_l \\
&+ \tfrac{2}{N^2h^2} \sum_{a_nn-(a_n+1)nh<l<a_nn-(a_n-1)nh} \sum_{a_nn-(a_n+1)nh<k<l} \left( \sum_{\frac{1}{a_n}(l-nh)\le i\le(n-nh)} C_{nhik}C_{nhil} \right) \xi_k\xi_l \\
&+ \tfrac{2}{N^2h^2} \sum_{1<l\le(a_n+1)nh} \sum_{0<k<l} \left( \sum_{nh\le i\le \frac{1}{a_n}(k+nh)} C_{nhik}C_{nhil} \right) \xi_k\xi_l \\
&:= \sum_{(a_n+1)nh<l\le a_nn-(a_n-1)nh} W_l + r_{nh},
\end{aligned}$$

where

$$\begin{aligned}
d_{nhkl} &:= \sum_{\frac{1}{a_n}(l-nh)\le i\le \frac{1}{a_n}(k+nh)} C_{nhik}C_{nhil} \\
W_l &:= \tfrac{2}{N^2h^2} \sum_{0<k\le(l-1)\wedge[a_nn-(a_n+1)nh]} d_{nhkl}\xi_k\xi_l \\
&= \tfrac{2}{N^2h^2} \left( \sum_{(l-2nh)<k\le(l-1)\wedge[a_nn-(a_n+1)nh]} d_{nhkl}\xi_k \right) \xi_l.
\end{aligned}$$

Then $W_l$ is a martingale difference sequence w. r. t. $\mathscr{F}_{l-1} = \sigma(\xi_1, \xi_2, \ldots, \xi_{l-1})$. Thus $\sum_{(a_n+1)nh\le l\le a_nn-(a_n-1)nh} W_l$ is a martingale sequence with respect to $\mathscr{F}_n = \sigma(\xi_1, \xi_2, \ldots, \xi_n)$. It can be shown that $A_I = o_P(\frac{1}{\sqrt{h}})$ and $r_{nh} = o_P(\frac{1}{\sqrt{h}})$. Hence we have

$$\sqrt{h} \sum_{i\in\mathcal{I}_2} (T_{i,n}^2 - \mu_{n,h}(K)) = \sum_{(a_n+1)nh<l\le a_nn-(a_n-1)nh} \sqrt{h}W_l + o_P(1).$$

So the lemma follows. □

The conditional variance is an intrinsic measure of time for a martingale. For many purposes the time taken for a martingale to cross a level is best represented through its conditional variance rather than the number of increments up to the crossing (see Hall, 1980, p. 54). The conditional variance is given as follows:

$$\begin{aligned}
V_n^2 &:= \sum_{(a_n+1)nh<l\le a_nn-(a_n-1)nh} \mathbb{E}\left[ \left(\sqrt{h}W_l\right)^2 |\xi_1, \xi_2, \ldots, \xi_{l-1} \right] \\
&= \mathbb{E}\left[ \left( \tfrac{2\sqrt{h}}{N^2h^2} \sum_{0<k\le(l-1)\wedge\lfloor a_nn-(a_n+1)nh\rfloor} d_{nhkl}\xi_k\xi_l \right)^2 \;\middle|\; \xi_1, \xi_2, \ldots, \xi_{l-1} \right] \\
&= \left( \tfrac{2\sqrt{h}}{N^2h^2} \sum_{0<k\le(l-1)\wedge\lfloor a_nn-(a_n+1)nh\rfloor} d_{nhkl}\xi_k \right)^2 \mathbb{E}\left( \xi_l^2 \;\middle|\; \xi_1, \xi_2, \ldots, \xi_{l-1} \right) \\
&= \sum_{(a_n+1)nh<l\le a_nn-(a_n-1)nh} \left( \tfrac{2\sqrt{h}}{N^2h^2} \sum_{0<k\le(l-1)\wedge\lfloor a_nn-(a_n+1)nh\rfloor} d_{nhkl}\xi_k \right)^2.
\end{aligned}$$

**Lemma 11.** *Assume kernel function $K(x)$ satisfies conditions (K.1-2-3); smoothing parameter $h$ satisfies conditions (H.1-2-3); and $f(x)$ satisfies conditions (F.1-2-3). Then as $n \to \infty$,*

$$\mathbb{E}V_n^2 = \int_0^2 dz \left( \int_{-1+z}^{1} 2K(x)K(x-z)dx \right)^2 + O(h).$$

*Proof.*

$$\begin{aligned}
&(\tfrac{N^2h^2}{2\sqrt{h}})^2\, \mathbb{E}\, V_n^2 \\
=& \sum_{(a_n+1)nh \le l \le a_n n-(a_n-1)nh} \mathbb{E} \left( \sum_{0<k\le(l-1)\wedge\lfloor a_n-(a_n+1)\rfloor nh} d_{nhkl}\xi_k \right)^2 \\
=& \sum_{(a_n+1)nh \le l \le a_n n-(a_n-1)nh} \sum_{0<k\le(l-1)\wedge\lfloor a_n-(a_n+1)\rfloor nh} d_{nhkl}^2 \\
=& \sum_{(a_n+1)nh \le l \le a_n n-(a_n-1)nh} \sum_{0<k\le(l-1)\wedge\lfloor a_n-(a_n+1)\rfloor} \left( \sum_{l-nh\le a_n i\le k+nh} C_{nhik}C_{nhil} \right)^2 \\
=& \left( \sum_{(a_n+1)nh \le l \le a_n n-(a_n+1)nh} \sum_{0<k\le(l-1)} + \sum_{a_n n-(a_n+1)nh\le l\le a_n n-(a_n-1)nh} \sum_{0<k<a_n n-(a_n+1)nh} \right) \\
& \sum_{\substack{l-nh\le a_n i\le k+nh \\ l-nh\le a_n j\le k+nh}} \sum K(\frac{a_n i-k+1}{nh})K(\frac{a_n j-k+1}{nh})K(\frac{a_n i-l+1}{nh})K(\frac{a_n j-l+1}{nh}) + O(n^4h^4) \\
=& \left( \sum_{(a_n+1)nh \le l \le a_n n-(a_n+1)nh} \sum_{1<l-k\le 2nh} + \sum_{-(a_n+1)nh\le l-a_n n\le -(a_n-1)nh} \sum_{l-a_n n+(a_n+1)nh<l-k\le 2nh} \right) \\
& \sum_{\substack{l-k-nh\le a_n i-k\le nh \\ l-k-nh\le a_n j-k\le nh}} \sum K(\frac{a_n i-k+1}{nh})K(\frac{a_n j-k+1}{nh})K(\frac{a_n i-l+1}{nh})K(\frac{a_n j-l+1}{nh}) + O(n^4h^4) \\
=& n^4h^3 \int_0^2 dz \int_{z-1}^1 \int_{z-1}^1 K(x)K(y)K(x-z)K(y-z)dxdy + O(n^4h^4).
\end{aligned}$$

So

$$\mathbb{E}\, V_n^2 = \int_0^2 dz \left( \int_{-1+z}^{1} 2K(x)K(x-z)dx \right)^2 + O(h).$$

□

**Lemma 12.** *Assume kernel function $K(x)$ satisfies conditions (K.1-2-3); smoothing parameter $h$ satisfies conditions (H.1-2-3); and $f(x)$ satisfies conditions (F.1-2-3). Then as $n \longrightarrow 0$,*

$$\mathbb{V}\mathrm{ar}(V_n^2) = O(h).$$

*Proof.*

$$
\begin{aligned}
&\mathbb{V}\mathrm{ar}\left(\sum_{(a_n+1)nh\le l\le a_nn-(a_n-1)nh}\mathbb{E}((\sqrt{h}W_l)^2|\xi_1,\dots,\xi_{l-1})\right)\\
&=\mathbb{V}\mathrm{ar}\left(\sum_{(a_n+1)nh\le l\le a_nn-(a_n-1)nh}(\sum_{0<k<(l-1)\wedge\lfloor a_nn-(a_n+1)nh\rfloor}(\tfrac{2\sqrt{h}}{N^2h^2})^2d_{nhkl}\xi_k)^2\right)\\
&=\tfrac{16}{N^8h^6}\left[\mathbb{E}\left(\sum_{(a_n+1)nh\le l\le a_nn-(a_n-1)nh}\sum_{\substack{l-2nh<r<(l-1)\wedge\lfloor a_nn-(a_n+1)nh\rfloor\\ l-2nh<s<(l-1)\wedge\lfloor a_nn-(a_n+1)nh\rfloor}}\sum d_{nhrl}d_{nhsl}\xi_r\xi_s\right)^2\right.\\
&\quad\left.-\left(\mathbb{E}\sum_{(a_n+1)nh\le l\le a_nn-(a_n-1)nh}\sum_{\substack{l-2nh<r<(l-1)\wedge\lfloor a_nn-(a_n+1)nh\rfloor\\ l-2nh<s<(l-1)\wedge\lfloor a_nn-(a_n+1)nh\rfloor}}\sum d_{nhrl}d_{nhsl}\xi_r\xi_s\right)^2\right]\\
&=\tfrac{16}{N^8h^6}O(n^8h^7)\\
&=O(h).
\end{aligned}
$$

□

The following lemma follows immediately by lemma 11 and lemma 12.

**Lemma 13.** *Assume kernel function $K(x)$ satisfies conditions (K.1-2-3); smoothing parameter $h$ satisfies conditions (H.1-2-3); and $f(x)$ satisfies conditions (F.1-2-3). Then*

$$V_n^2\xrightarrow{P}\sigma^2(K).$$

**Lemma 14.** *Assume kernel function $K(x)$ satisfies conditions (K.1-2-3); smoothing parameter $h$ satisfies conditions (H.1-2-3); and $f(x)$ satisfies conditions (F.1-2-3). Then for any $\varepsilon>0$, we have*

$$\sum_{(a_n+1)nh\le l\le a_nn-(a_n-1)nh}\mathbb{E}\left[\left(\sqrt{h}W_l\right)^2\mathrm{I}_{(\sqrt{h}|W_l|>\varepsilon)}\ \middle|\ \xi_1,\xi_2,\dots,\xi_{l-1}\right]\xrightarrow{P}0.$$

*Proof.*

$$
\begin{aligned}
&\sum_{(a_n+1)nh\le l\le a_nn-(a_n-1)nh}\mathbb{E}(\sqrt{h}W_l)^4\\
&=\tfrac{16}{N^8h^6}\sum_{(a_n+1)nh\le l\le a_nn-(a_n-1)nh}\mathbb{E}\left(\sum_{l-2nh\le k\le(l-1)\wedge\lfloor a_nn-(a_n+1)nh\rfloor}d_{nhkl}\xi_k\xi_l\right)^4\\
&=\tfrac{16}{N^8h^6}\left(\sum_{(a_n+1)nh\le l\le a_nn-(a_n-1)nh}\ \sum\sum_{l-2nh\le r<s\le(l-1)\wedge\lfloor a_nn-(a_n+1)nh\rfloor}6d_{nhrl}^2d_{nhsl}^2\,\mathbb{E}\,\xi_l^4\right.\\
&\quad\left.+\sum_{(a_n+1)nh\le l\le a_nn-(a_n-1)nh}\ \sum_{l-2nh\le k\le(l-1)\wedge\lfloor a_nn-(a_n+1)nh\rfloor}d_{nhkl}^4(\mathbb{E}\,\xi_l^4)^2\right)\\
&=\tfrac{16}{N^8h^6}O(n^7h^6)\\
&=O(\tfrac{1}{N}).
\end{aligned}
$$

By Markov inequality, $\forall\varepsilon>0$,

$$\sum_{(a_n+1)nh\le l\le a_nn-(a_n-1)nh}\mathbb{E}\left[\left(\sqrt{h}W_l\right)^2\mathrm{I}_{(\sqrt{h}|W_l|>\varepsilon)}\middle|\xi_1,\dots,\xi_{l-1}\right]\xrightarrow{P}0.$$

□

Theorem 1 follows immediately from lemma 13 and lemma 14 (See Corollary 3.1 of Hall, 1980, p. 58).

## References


Bahadur, R. R. (1958). Examples of inconsistency of maximum likelihood estimates. *Sankhyā, 20*, 207-210.

Bloch, D., & Gastwirth, J. (1968). On a simple estimate of the reciprocal of the density function. *Ann. Math. Statist. 39*, 1083-1085. http://dx.doi.org/10.1214/aoms/1177698342

Bofinger, E. (1975). Estimation of a density function using order statistics. *Australian Journal of Statistics, 17*(1), 1-7. http://dx.doi.org/10.1111/j.1467-842X.1975.tb01366.x

Csörgő, M., & Révész, P. (1978). Strong approximations of the quantile process. *Ann. Statist., 6*, 882-894. http://dx.doi.org/10.1214/aos/1176344261

D'Agostino, R & Stephens, M. (1986). *Goodness-of-Fit Techniques*. Marcel Dekker, New York.

Falk, M. (1984). Relative deficiency of kernel type estimators of quantiles, *Ann. Statist., 12*, 261-268. http://dx.doi.org/10.1214/aos/1176346405

Falk, M. (1986). On the estimation of the quantile density function. *Statist. Probability Letter, 4*, 69-73. http://dx.doi.org/10.1016/0167-7152(86)90020-9

Fan, J., Zhang, C., & Zhang, J. (2001). Generalized likelihood ratio statistics and wilks phenomenon. *The annals of Statistics, 29* (1), 153-193. http://dx.doi.org/10.1214/aos/996986505

Hall, P., & Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press.

Ingster, Y. I., & Suslina, I. A. (2002). *Nonparametric Goodness-of-Fit Testing Under Gaussian Models*. Springer.

Le Cam, L. (1990) Maximum likelihoodan introduction. *ISI Review, 58*, 153-171.

Parzen, E. (1979). Nonparametric statistical modelling (with comments). *J. Amer. Statist. Assoc., 74*, 105-131. http://dx.doi.org/10.1080/01621459.1979.10481621

Rao, C. R., & Zhao, L. C. (1997). *A limiting distribution theorem in Festschrift for Lucien Le Cam*. Springer, Berlin, 325-335. http://dx.doi.org/10.1007/978-1-4612-1880-7_21

Reiss, R. D. (1978). Approximate distribution of the maximum deviation of histograms. *Metrika, 25*(1), 9-26. http://dx.doi.org/10.1007/BF02204347

Sheather, S. J., & Maritz, J. S. (1983). An estimate of the asymptotic standard error of the sample median. *Australian Journal of Statistics, 25*(1), 109-122. http://dx.doi.org/10.1111/j.1467-842X.1983.tb01203.x

Siddiqui, M. M. (1960). Distribution of quantitles in samples from a bivariate population. *Journal of Research of the National Bureau of Standards, Section B*, 64, 145-150. http://dx.doi.org/10.6028/jres.064B.017

Tukey, J. W. (1965) Which part of the sample contains the most information? *Proe. Nat. Acad. Sei. U.S.A., 53*, 127-134. http://dx.doi.org/10.1073/pnas.53.1.127


## Copyrights