Journal Code:				Article ID					Dispatch 30.01.14	CE: Patrimonio, Cheryl F.
S	J	O	S	1	2	0	7	9	No. of Pages: 18	ME:

Scandinavian Journal of Statistics
doi: 10.1111/sjos.12079

Computing Critical Values of Exact Tests by Incorporating Monte Carlo Simulations Combined with Statistical Tables

ALBERT VEXLER
Department of Biostatistics, The University at Buffalo, State University of New York

YOUNG MIN KIM
Department of Biostatistics, The University at Buffalo, State University of New York

JIHNHEE YU
Department of Biostatistics, The University at Buffalo, State University of New York

NICOLE A. LAZAR
Department of Statistics, University of Georgia

ALAN D. HUTSON
Department of Biostatistics, The University at Buffalo, State University of New York

ABSTRACT. Various exact tests for statistical inference are available for powerful and accurate decision rules provided that corresponding critical values are tabulated or evaluated via Monte Carlo methods. This article introduces a novel hybrid method for computing p-values of exact tests by combining Monte Carlo simulations and statistical tables generated *a priori*. To use the data from Monte Carlo generations and tabulated critical values jointly, we employ kernel density estimation within Bayesian-type procedures. The p-values are linked to the posterior means of quantiles. In this framework, we present relevant information from the Monte Carlo experiments via likelihood-type functions, whereas tabulated critical values are used to reflect prior distributions. The local maximum likelihood technique is employed to compute functional forms of prior distributions from statistical tables. Empirical likelihood functions are proposed to replace parametric likelihood functions within the structure of the posterior mean calculations to provide a Bayesian-type procedure with a distribution-free set of assumptions. We derive the asymptotic properties of the proposed nonparametric posterior means of quantiles process. Using the theoretical propositions, we calculate the minimum number of needed Monte Carlo resamples for desired level of accuracy on the basis of distances between actual data characteristics (e.g. sample sizes) and characteristics of data used to present corresponding critical values in a table. The proposed approach makes practical applications of exact tests simple and rapid. Implementations of the proposed technique are easily carried out via the recently developed STATA and R statistical packages.

Key words: Bayesian point estimation, empirical likelihood, exact tests, local maximum likelihood, posterior expectation, quantile estimation

1. Introduction

Exact tests are well known to be simple, efficient and reliable statistical tools in a variety of applications. The statistical literature has extensively addressed many parametric, semi-parametric and nonparametric exact tests that have finite sample type I error control, for example, exact tests for quantiles (Serfling, 2002, p. 102), Fisher's exact test, the exact F-test, the Shapiro–Wilk test of normality, the Wilcoxon rank-sum test, Hall and Welsh's test for normality (Hall & Welsh, 1983) and the exact test for comparing multivariate distributions proposed by Rosenbaum (2005), to name just a few. Many of these tests are now found in standard statistical software packages, for example, SPSS. Exact tests are not only useful in the small sample size setting but also invaluable in the case of rare events in the large sample setting; for example, see Mudholkar & Hutson (1997) for the case of contingency tables and the inaccuracy of asymptotic methods when there is a large number of zero count cells. A major disadvantage is that, in general, exact tests are computationally intensive and may not be feasible for moderate to large sample sizes. In these instances, Monte Carlo (MC) methods may provide accurate and computationally feasible approximations. Commonly, tables of corresponding critical points are required for the exact tests' implementation. Because all possible critical values cannot be tabulated, two different approaches can be employed in practice to calculate the exact p-value. One method is related to interpolation/extrapolation of the relevant critical points on the basis of tabulated values. The use of tables with corresponding critical values is a standard method applied in various statistical software routines. Employing a table-based methodology for use within the testing algorithm is quick and efficient. However, the interpolation/ extrapolation method using table-based methods becomes much less reliable when real data characteristics (e.g. sample sizes) differ from those used to tabulate the critical values. Alternatively, as a second approach, one can conduct MC simulations to approximate the exact p-values. The approach based on MC simulations is simple and accurate but oftentimes computationally intensive.

In this article, we propose and examine a novel Bayesian hybrid method to compute p-values of exact tests on the basis of both statistical tables of related critical values and MC simulations. The key feature of the proposed method is that if actual data characteristics (e.g. sample sizes) are close to those used in the table of critical values, then relatively few MC simulations are needed to improve the accuracy of p-value calculations. When this is not the case, more MC simulations are necessary.

Consider the following simple illustration. Table 1 represents a part of 'percentiles of the χ^2 distribution' (Kutner *et al.*, 2005) at 29, 30, 40 and 50 degrees of freedom. Suppose that we are interested in the percentiles of the χ^2 distribution with 35 degrees of freedom and level of significance 0.05. Table 1 does not provide an exact percentile in this case. Interpolation between the 30-degree-of-freedom and 40-degree-of-freedom entries or an MC procedure based on randomly generated χ^2 values can accurately approximate the exact (missing and unknown) percentile. In another scenario, suppose we are interested in the case with 31 degrees of freedom and 0.05 level of significance. In this case, we could conduct an MC study with fewer replications than required in the previous case (35 degrees of freedom) because the desired percentile is nearer to the tabulated value at 30 degrees of freedom. The proposed Bayesian hybrid method calculates the percentile of interest by incorporating the information from Table 1 as a prior distribution to improve the accuracy and reduce the computational cost of the simulations.

In this article, we consider a Bayesian solution to the problem through the incorporation of information from MC simulations and tabulated critical values. Hence we consider the tables to

Table 1. *Percentiles of the χ^2 distribution (Kutner et al., 2005) where d.f. is degrees of freedom*

	Level of significance		
d.f.	0.025	0.050	0.100
29	14.26	16.05	17.71
30	14.95	16.79	18.49
40	22.16	24.43	26.51
50	37.48	40.48	43.19

represent prior information, whereas the likelihood part of the analysis is based on the output of the relevant MC simulations. Therefore, the two key issues that need to be addressed for the proposed hybrid method are as follows: (i) constructing the functional forms of the prior distribution based on of tabulated critical values; and (ii) developing nonparametric likelihood-type functions from information extracted from MC simulations.

In this work, we apply the local maximum likelihood (LML) technique (Tibshirani & Hastie, 1987; Fan & Gijbels, 1995, 1996; Fan *et al.*, 1995; Fan *et al.*, 1998) to obtain functional forms of prior distributions from tables of critical values. The LML methodology is usually applied to construct functional forms of dependency between variables. For example, in regression, the aim of the LML approach is to explore the association between dependent and independent variables when the regression functions are data driven instead of being limited to a certain presumed form (Fan & Gijbels, 1995). The flexibility and relative simplicity of the LML methodology makes it a practical solution to the first of our problems.

As for the second problem, that is, construction of a likelihood to summarize the data based on MC generations of exact-test-statistic values, we propose the nonparametric technique of empirical likelihood (EL; e.g. Owen, 1988, 1990, 2001; Lazar & Mykland, 1998; Yu *et al.*, 2010; Vexler & Gurevich, 2010; Vexler & Yu, 2011). It is well known that EL is a powerful nonparametric technique of statistical inference. Lazar (2003) further demonstrated that the EL functions can be used to construct distribution-free Bayesian posterior probabilities. Following the results of Lazar (2003), we develop the nonparametric posterior expectations of quantiles via EL in our proposed procedure for computing the p-values of exact tests.

Evaluation of the exact p-value in our context requires statistical inference for quantiles. We achieve this inference using the EL technique based on kernel densities. As shown by Chen & Hall (1993), the use of kernel densities to smooth EL significantly improves the performance of the EL ratio tests for quantiles, in terms of relative accuracy. Zhou & Jing (2003) proposed an alternative smoothed EL approach where the likelihood ratio has an explicit form based on M-estimators. Yu *et al.* (2011) also showed that the use of kernel densities significantly improves two-sample EL ratio tests for medians.

In order to implement our approach, we need to calculate the posterior expectation given the *a priori* information (tables of critical values) and the observed data (MC simulations). In the case of the classical Bayesian analysis, there are some instances where the integrals related to the posterior expectations may be evaluated analytically. More often than not, however, these calculations are intractable, and numerical methods are used to obtain a solution. Alternatively, this article provides an asymptotic expression for the proposed nonparametric posterior expectations, which utilizes the asymptotic Gaussian form of the posterior distribution of quantiles based on maximum likelihood estimation. Note that the proposed nonparametric posterior expectations are based on integrated EL functions. The EL functions themselves do not have closed analytical forms. Hence, we are required to use numerical methods. That is, computation of the posterior expectation in this instance is not trivial. In a similar manner to classical Bayesian inference (e.g. DasGupta, 2008; DiCiccio *et al.*, 1997; Kass & Raftery, 1995), the asymptotic results developed in this article have good utility relative to the posterior expectation calculation, thus making our methodology tractable. The asymptotic approach also provides a method for calculating the required number of samples needed to generate specific MC simulations in order to obtain accurate critical values for the proposed exact tests. The criterion for computing the minimum number of MC repetitions is based on the prior information (tables of pre-populated critical values).

This article is organized as follows. In Section 2, we describe the problem more fully and define our formal notation. We then define the nonparametric posterior expectation of the critical values based on EL with kernel functions. The asymptotic properties of the posterior expectation are shown, and the theoretical properties of the estimated critical values are evaluated asymptotically. Section 2 also presents the method to create functional forms of the prior distributions using the LML method based on statistical tables. Finally, we provide the procedure to calculate critical values of exact tests using our new method. Concluding remarks are presented in Section 3.

Section S1 of the Supporting information demonstrates and examines the applicability of the proposed approach in practice. Note also that Tanajian *et al.* (2013) and Shepherd *et al.* (2013) have employed the proposed method in the STATA and R software packages, respectively, to apply new statistical testing strategies in practice. The STATA package provides nonparametric tests for symmetry of data distributions as well as comparing K-sample distributions. Recognizing the fact that recent statistical software packages do not sufficiently address K-sample nonparametric comparisons of data distributions, Tanajian *et al.* (2013) proposed a new STATA command VX_DBEL to execute exact tests on the basis of K-samples. To calculate p-values of the exact tests, the STATA command uses the following methods: (i) a classical technique based on MC p-value evaluations; (ii) an interpolation technique based on tabulated critical values; and (iii) a new hybrid technique that combines (i) and (ii). The second and third cutting edge methods are based on the methodology presented in this paper. The R package provides a function dbELnorm to be used for joint assessment of normality of K-independent samples with varying means and standard deviations. The function provides the test statistic and associated p-values. The p-values can be calculated by incorporating MC simulations combined with statistical tables in a manner similar to the technique considered in this paper.

2. Statements and methods

In this section, we formally state the problem related to the procedure development in order to compute critical values of exact tests using MC simulations and statistical tables. We begin with the following example. Suppose, for simplicity, that we have a two sample exact test with critical values $q_{n,m}^{1-\alpha}$, where n, m are the respective sample sizes and α is the level of significance. Let the values of $q_{n,m}^{1-\alpha}$ be tabulated for $n \in N$, $m \in M$ and $\alpha \in A$, where N and M are sets of integer numbers and A is a set of real numbers from 0 to 1. That is, we have the table defined as $\left\{q_{n,m}^{1-\alpha}; n \in N, m \in M, \alpha \in A\right\}$. We are interested in obtaining a value of $q_{n_0,m_0}^{1-\alpha_0} \notin \left\{q_{n,m}^{1-\alpha}; n \in N, m \in M, \alpha \in A\right\}$. There are two ways to evaluate $q_{n_0,m_0}^{1-\alpha_0}$: (i) a technique based on interpolation/extrapolation using the table values $\left\{q_{n,m}^{1-\alpha}; n \in N, m \in M, \alpha \in A\right\}$; and (ii) an approach based on a MC study for generating values of the test statistic. In this paper, we use the table $\left\{q_{n,m}^{1-\alpha}; n \in N, m \in M, \alpha \in A\right\}$ to construct a prior distribution for $q_{n_0,m_0}^{1-\alpha_0}$ that will be combined through Bayes rule with a nonparametric likelihood function based on MC generated values of the test statistic. The value of $q_{n_0,m_0}^{1-\alpha_0}$, the $(1-\alpha_0) \times 100\%$ quantile, is then estimated from the posterior distribution that results. The next section presents the nonparametric Bayesian approach to estimate $q_{n_0,m_0}^{1-\alpha_0}$ using the posterior expectation of quantiles based on the smoothed EL method. Section 2.2 introduces the LML technique to derive a prior distribution of quantiles based on related statistical tables. In Section 2.3, the final procedure of the hybrid method is provided to be applied to calculate critical values of exact tests efficiently.

2.1. *Bayesian empirical likelihood evaluation of quantiles*

Smoothed empirical likelihood for quantiles. In this section, we denote the distribution-free posterior expectation of quantiles based on independent identically distributed observations $X_1, \ldots, X_t$. Let X_j $(j = 1, \ldots, t)$ denote a realization of the test statistic value at the j-th MC iteration when the MC simulations provide in total t generated values of the test statistic. The distribution function, F, of X is unknown. We focus on the evaluation of the $(1-\alpha) \times 100\%$ quantile, $\Pr\left(X < q_0^{1-\alpha}\right) = 1-\alpha$, $\alpha \in (0,1)$. To provide nonparametric statistical inference regarding $q_0^{1-\alpha}$, we apply the nonparametric likelihood approach proposed by Chen & Hall (1993). Towards this end, denote $K(\cdot)$ to be a smoothed version of the indicator function F_0 defined by $F_0(u) = 1$, for $u \geq 0$, 0 otherwise. Let $k(\cdot)$ be a τ-th order kernel density function, commonly used in nonparametric density estimation (e.g. Silverman, 1986; Härdle, 1990). Then, the function $K(\cdot)$ can be represented in the form

$$K(x) = \int_{u<x} k(u)du, \tag{1}$$

where $k(\cdot)$ satisfies

$$\int u^j k(u)du = \begin{cases} 1 \;\; if \;\; j = 0, \\ 0 \;\; if \;\; 1 \leq j \leq \tau - 1, \\ \gamma \;\; if \;\; j = \tau, \end{cases}$$

for any integer $\tau \geq 2$ and a constant $\gamma \neq 0$. We use the notation $K_h(\cdot) = K(\cdot/h)$ and $k_h(\cdot) = h^{-1}k(\cdot/h)$ as the kernel distribution and kernel density representations, respectively, with bandwidth h satisfying the following conditions: $h \to 0$ and $th \to \infty$ as $t \to \infty$.

Write the smoothed EL function for quantiles as

$$EL(q) = \sup\left[\prod_{j=1}^{t} p_j \,\middle|\, p_j \geq 0, \sum_{j=1}^{t} p_j = 1 \;\; and \;\; \sum_{j=1}^{t} p_j \left\{K_h\left(q - X_j\right) - (1-\alpha)\right\} = 0\right]. \tag{2}$$

The method of Lagrange multipliers then can be used to find the values of $p_1, \ldots, p_t$ in (2) yielding the result:

$$p_j = \left(t + \lambda\left\{K_h\left(q - X_j\right) - (1-\alpha)\right\}\right)^{-1}, \tag{3}$$

where λ, the Lagrange multiplier, is a root of

$$\sum_{j=1}^{t} \frac{K_h(q - X_j)}{t + \lambda\left\{K_h(q - X_j) - (1-\alpha)\right\}} = 1 - \alpha. \tag{4}$$

Chen & Hall (1993) showed that this smoothed EL for quantiles improves the coverage accuracy of confidence intervals from order $t^{-1/2}$ to order t^{-1}.

Bayesian approach based on the smoothed empirical likelihood for quantiles. We assume that, in addition to the sample, $X_1, \ldots, X_t$, we have prior information regarding the quantile $q_0^{1-\alpha}$ in the form of a density function denoted as $\pi(q)$. Without loss of generality and for simplicity of notation, we also assume that the form of the density function of X_1 is known to be $f(x|q)$. Classical Bayesian methodology yields the posterior probability of the quantile $q_0^{1-\alpha}$ as

$$P_P(q) = \frac{\prod_{j=1}^{t} f\left(X_j|q\right)\ \pi(q)}{\int_{-\infty}^{\infty} \prod_{j=1}^{t} f(X_j|q)\ \pi\ (q)\ dq}. \tag{5}$$

Thus, the posterior expectation of a given quantile is

$$E(q|X_1, \ldots, X_t) = \int_{-\infty}^{\infty} qP_p(q)\ dq = \frac{\int_{-\infty}^{\infty} q \prod_{j=1}^{t} f(X_j|q)\ \pi(q)dq}{\int_{-\infty}^{\infty} \prod_{j=1}^{t} f(X_j|q)\ \pi(q)dq}. \tag{6}$$

In this work, we replace in (5) with a nonparametric counterpart. Lazar (2003) proposed and validated the Bayesian EL method as an alternative to the classical parametric setting. This suggests the nonparametric form of (5)

$$P_{NP}(q) = \frac{EL(q)\ \pi(q)}{\int_{X_{(1)}}^{X_{(t)}} EL(q)\ \pi(q)\ dq}\quad , \tag{7}$$

where $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(t)}$ are the order statistics based on $X_1, \ldots, X_t$ and the function $EL(q)$ is defined at (2), the kernel density-smoothed EL. It follows that the nonparametric posterior expectation of the (1-α)th quantile is given by

$$\hat{q}_{NP}^{1-\alpha} = \int\limits_{X_{(1)}}^{X_{(t)}} qP_{NP}(q)\ dq = \frac{\int_{X_{(1)}}^{X_{(t)}} qEL(q)\ \pi(q)\ dq}{\int_{X_{(1)}}^{X_{(t)}} EL(q)\ \pi(q)\ dq}. \tag{8}$$

Note that if, for example, for some q, we have $\max_{j=1,\ldots,t}\{K_h(q - X_j) - (1-\alpha)\} < 0$, then there are no $p_1, \ldots, p_t$ summing to 1 for which $\sum_{j=1}^{t} p_j\{K_h(q - X_j) - (1-\alpha)\} = 0$. If this happens, we define $EL(q) = 0$ (Owen, 2001), and hence, by convention, and without loss of generality, we propose the support of the integrals in (7) and (8) to be restricted to $[X_{(1)}, X_{(t)}]$.

To investigate the asymptotic properties of the nonparametric posterior expectation of quantiles at (8), we assume that the data distribution function, $F(\cdot)$, is three times differentiable in a neighbourhood of $q_0^{1-\alpha}$ and that the applied kernel function is twice differentiable in a neighbourhood of $q_0^{1-\alpha}$. Let us also define $z^{(r)}(q) = d^r z(q)/dq^r$ for a smooth function $z(\cdot)$ and $r = 3, 4, \ldots$, and $\bar{k}_h(q) = t^{-1}\sum_{j=1}^{t} k_h(q - X_j)$, $\bar{k}'_h(q) = (th^2)^{-1}\sum_{j=1}^{t} k'(h^{-1}(q - X_j))$ and $\bar{K}_{s,h}(q) = t^{-1}\sum_{j=1}^{t}\{K_h(q - X_j) - (1-\alpha)\}^s$ for $s = 1, 2, 3$. Then we have the following:

Proposition 1. *Let $X_1, \ldots, X_t$ be a random sample from a density function $f(\cdot)$. Suppose that the prior density is defined as $\pi(q) = \exp\left(-(q-\mu_0)^2(2\sigma_0^2)^{-1}\right)\Big/\sqrt{2\pi\sigma_0^2}$. Then, the asymptotic approximation to the nonparametric posterior expectation of the (1-α)-th quantile is given as*

$$\hat{q}_{NP}^{1-\alpha} = \frac{\hat{q}_M^{1-\alpha} + \frac{\mu_0}{\sigma_0^2 B_t^2(\hat{q}_M^{1-\alpha})\,t}}{1 + \frac{1}{\sigma_0^2 B_t^2(\hat{q}_M^{1-\alpha})\,t}} + o_p\left(\frac{1}{t}\right) \textit{as } t \to \infty,$$

where $\hat{q}_M^{1-\alpha}$ satisfies $1 - \alpha = t^{-1}\sum_{j=1}^{t} K_h\left(\hat{q}_M^{1-\alpha} - X_j\right)$ and $B_t^2\left(\hat{q}_M^{1-\alpha}\right) = \left\{\bar{k}_h\left(\hat{q}_M^{1-\alpha}\right)\right\}^2\Big/\bar{K}_{2,h}\left(\hat{q}_M^{1-\alpha}\right)$.

By proposition 1, one can show that the asymptotic distribution of the estimator (8) is

$$\sqrt{t}\,V_t^{-1/2}\left(\hat{q}_{NP}^{1-\alpha} - q_0^{1-\alpha}\right)\left(1 + \frac{1}{\sigma_0^2 B_t^2\left(\hat{q}_M^{1-\alpha}\right)\,t}\right) - \frac{\sqrt{t}\,V_t^{-1/2}\left(\mu_0 - q_0^{1-\alpha}\right)}{\sigma_0^2 B_t^2\left(\hat{q}_M^{1-\alpha}\right)\,t} \xrightarrow{d} N(0,1) \text{ as } t \to \infty,$$

where $V_t = tVar\left(\hat{q}_M^{1-\alpha}\right)$.

The next proposition provides a method to control the accuracy of the proposed p-values' evaluations. To formulate the following result, we assume that $F(x)$ is an absolutely continuous cumulative distribution function with corresponding density function $f(x)$ and

$F''(x) = f'(x)$. We also assume that $F^{(5)}\left(q_0^{1-\alpha}\right) = f^{(4)}\left(q_0^{1-\alpha}\right)$ exists. The application of the Taylor theorem then implies

$$\begin{aligned} 1 - F\left(\hat{q}_{NP}^{1-\alpha}\right) = \alpha - \left(\hat{q}_{NP}^{1-\alpha} - q_0^{1-\alpha}\right) f\left(q_0^{1-\alpha}\right) - \frac{1}{2}\left(\hat{q}_{NP}^{1-\alpha} - q_0^{1-\alpha}\right)^2 f'\left(q_0^{1-\alpha}\right) \\ - \frac{1}{6}\left(\hat{q}_{NP}^{1-\alpha} - q_0^{1-\alpha}\right)^3 f''\left(q_0^{1-\alpha}\right) + \dots . \end{aligned} \tag{9}$$

In this case, controlling values of $\tilde{\Delta}(t) = \left(\hat{q}_{NP}^{1-\alpha} - q_0^{1-\alpha}\right) f\left(q_0^{1-\alpha}\right) + 0.5\left(\hat{q}_{NP}^{1-\alpha} - q_0^{1-\alpha}\right)^2 f'\left(q_0^{1-\alpha}\right)$, we can monitor the accuracy of the p-value evaluations. Here, $\tilde{\Delta}(t)$ may be thought of as a *risk function*. For example, the choice of t such that $\left|E\tilde{\Delta}(t)\right| < \upsilon$ (e.g. $\upsilon = 0.001$) provides the presumed accuracy of the $q_0^{1-\alpha}$-estimation. This leads to the following:

Proposition 2. *The asymptotic representation of the expectation of the estimated probability at (9) is*

$$E\left\{1 - F\left(\hat{q}_{NP}^{1-\alpha}\right)\right\} = \alpha - E\tilde{\Delta}(t) + o(t^{-1}),$$

where

$$\begin{aligned} E\tilde{\Delta}(t) = & -\frac{f\left(q_0^{1-\alpha}\right)}{\left(1 + \frac{\alpha(1-\alpha)}{\sigma_0^2 f^2\left(q_0^{1-\alpha}\right)t}\right)} \left\{ -\frac{d_2 f'\left(q_0^{1-\alpha}\right)h^2}{2f\left(q_0^{1-\alpha}\right)} - \frac{d_4 f^{(3)}\left(q_0^{1-\alpha}\right)h^4}{24 f\left(q_0^{1-\alpha}\right)} \right. \\ & \left. + \frac{(d_2)^2 f'\left(q_0^{1-\alpha}\right) f''\left(q_0^{1-\alpha}\right)h^4}{4f^2\left(q_0^{1-\alpha}\right)} + \frac{(-1+2\alpha) f\left(q_0^{1-\alpha}\right)}{t} \right\} \\ & - \frac{f\left(q_0^{1-\alpha}\right)}{\left(1 + \frac{\alpha(1-\alpha)}{\sigma_0^2 f^2\left(q_0^{1-\alpha}\right)t}\right)} \left\{ -\frac{\alpha(1-\alpha) f'\left(q_0^{1-\alpha}\right)}{2f^3\left(q_0^{1-\alpha}\right)t} + \frac{\alpha(1-\alpha)\left(\mu_0 - q_0^{1-\alpha}\right)}{\sigma_0^2 f^2\left(q_0^{1-\alpha}\right)t} \right\} \\ & - \frac{f'\left(q_0^{1-\alpha}\right)}{2\left(1 + \frac{\alpha(1-\alpha)}{\sigma_0^2 f^2\left(q_0^{1-\alpha}\right)t}\right)^2} \left\{ \frac{\alpha(1-\alpha)}{f^2\left(q_0^{1-\alpha}\right)t} \right\}, \end{aligned}$$

and $d_\nu = \int u^\nu k(u)du, \nu = 2, 4$.

We outline the proofs of propositions 1 and 2 in the Appendix.

Remark 1. Proposition 1 illustrates that the asymptotic result has a remainder term of order $o(t^{-1})$. By virtue of proposition 1, we can easily show that $\hat{q}_{NP}^{1-\alpha} = \hat{q}_M^{1-\alpha} + \left(\mu_0 - \hat{q}_M^{1-\alpha}\right)\left(\sigma_0^2 B_t^2\left(\hat{q}_M^{1-\alpha}\right) t\right)^{-1} + o_p(t^{-1})$. This indicates that we may asymptotically detect effects of the prior on the estimated quantile (8) with relative accuracy, because the term dependent on the prior information vanishes as t^{-1}. One can further show that if the kernel method is not used, the EL function could still be represented through the empirical distribution functions, for example, $EL(q) \approx -t(1-\alpha-F_t(q))^2/(2\alpha(1-\alpha))$, where $F_t(q) = t^{-1}\sum_{j=1}^{t} I(X_j \le q)$ for an indicator function, $I(\cdot)$. In this case, because of lack of smoothness of the estimator, the technique used to prove proposition 1 cannot be used because F_t is a step function. Instead, we can apply the classic Bahadur asymptotic results (e.g. Serfling, 2002, p. 93) ; these provide remainder terms calculated to be of order $t^{-3/4}$. This is not enough evidence to assess the asymptotic impact of the prior information in the evaluation of the distribution-free posterior expectation.

Remark 2. If X is discrete or has defined discontinuities, we can reformulate the propositions below using the smoothing transformation given by $Z_j = X_j + \eta\varepsilon_j$, where $\varepsilon_j \sim N(0,1)$, $j = 1, 2, \ldots, t$, and $\eta \to 0$ as $t \to \infty$.

Remark 3. We suggest to choose a bandwidth of $h = 0.2t^{-1/6}$. In the context of the quantile evaluations, this choice of bandwidth provides reasonable performance for various underlying distributions in our extensive MC study. In addition, the choice of bandwidth satisfies the asymptotic conditions on η required in Chen & Hall (1993) (see also Yu *et al.*, 2011, for details).

2.2. *Derivation of the functional form of prior information via the local maximum likelihood estimation based on statistical tables*

In this section, we use the table $\{q^{1-\alpha}_{n,m}; n \in N, m \in M, \alpha \in A\}$ to define a prior distribution on $q^{1-\alpha_0}_{n_0,m_0}$. Consider the model

$$q^{1-\alpha}_{n,m} = \mu(\alpha, n, m) + \sigma\varepsilon^{1-\alpha}_{n,m}, \quad \varepsilon^{1-\alpha}_{n,m} \sim N(0,1), \tag{10}$$

where $\mu(\cdot)$ is an unknown function corresponding to location and $\sigma > 0$ is an unknown scale parameter. We estimate the function μ in order to provide the prior location $\mu(\alpha_0, n_0, m_0)$ for $q^{1-\alpha_0}_{n_0,m_0}$, the parameter of interest. In order to estimate $\mu(\alpha_0, n_0, m_0)$ and σ, we propose directly applying the LML methodology (e.g. Fan *et al.*, 1998). The LML methodology is a nonparametric maximum likelihood technique based on a polynomial-type approximation of a parameterized function of interest, where the parameterized function is often unknown. The point of interest in our procedure is the point (α_0, n_0, m_0) for which the corresponding quantile is not tabulated in the table of critical values$\{q^{1-\alpha}_{n,m}; n \in N, m \in M, \alpha \in A\}$. In the LML approach, we approximate the parametric function via Taylor expansion as

$$\mu(\alpha, n, m) \cong \sum_{i=0}^{r_\alpha}\sum_{j=0}^{r_n}\sum_{k=0}^{r_m} \beta_{ijk}(\alpha-\alpha_0)^i (n-n_0)^j (m-m_0)^k, \quad \beta_{ijk} = \beta_{ikj}. \tag{11}$$

The restriction $\beta_{ijk} = \beta_{ikj}$ corresponds to a wide class of two sample hypothesis tests, including the applications considered in Section S1 of the Supporting information. This condition can be avoided in general evaluations. Equation (11) allows an approximation to the location parameter around the point of interest $\mu(\alpha_0, n_0, m_0)$. The ultimate parameter of interest is $\beta_{000} = \mu(\alpha_0, n_0, m_0)$ in the polynomial approximation, while incorporating all available data points into the parameter estimation. Let σ_0 be the standard deviation corresponding to (the estimator of) β_{000}. β_{ijk} and σ are estimated through maximizing the log likelihood given by

$$\begin{aligned} L(\boldsymbol{\beta}, \alpha_0, n_0, m_0) = &-\log\left(\sqrt{2\pi\sigma^2}\right) \sum_{n\in N, m\in M, \alpha\in A} \mathbf{k}(\alpha-\alpha_0, n-n_0, m-m_0) \\ &- \frac{1}{2\sigma^2} \sum_{n\in N, m\in M, \alpha\in A} \left(q^{1-\alpha}_{n,m} - \sum_{i=0}^{r_\alpha}\sum_{j=0}^{r_n}\sum_{k=0}^{r_m} \beta_{ijk}(\alpha-\alpha_0)^i (n-n_0)^j (m-m_0)^k \right)^2 \\ &\mathbf{k}(\alpha-\alpha_0, n-n_0, m-m_0), \end{aligned} \tag{12}$$

where $\boldsymbol{\beta}$ indicates the vector composed of β_{ijk} and $\mathbf{k}(\cdot)$ is a joint kernel function. We write $\mathbf{k}_{\alpha_0,n_0,m_0} = \mathbf{k}(\alpha-\alpha_0, n-n_0, m-m_0)$ in order to ease the notational burden of this formulation. The joint kernel function $\mathbf{k}_{\alpha_0,n_0,m_0}$ defines the contribution of tabulated data points to the log-likelihood (12), thus yielding a local kernel weighted log-likelihood function. The kernel function centred around the values of interest (α_0, n_0, m_0) provides a larger weight for the data points closer to the values of interest relative to those points more distant. Without compromising this concept, we define the joint kernel function as

$$\mathbf{k}_{\alpha_0,n_0,m_0} = k_{h_1}(\alpha-\alpha_0)k_{h_2}(n-n_0)k_{h_3}(m-m_0),$$

where $k_{h_i}(\cdot) = k(\cdot/h_i)/h_i$ is a univariate density function with $k_{h_i}(x) = k_{h_i}(-x)$ for all real x and bandwidth $h_i, h_i > 0$. Typically, the ranges of tabulated n and m are similar so that we let $h_2 = h_3$. The values of $r_\alpha,\ r_n,\ r_m$ and the bandwidth are selected on the basis of the bias and variance estimates of the nonparametric estimator (Fan *et al.*, 1998). As building blocks of the bias and variance estimation, we note that $\partial L(\boldsymbol{\beta},\alpha,n,m)/\partial\boldsymbol{\beta}$ is a vector whose elements corresponding to $\beta_{i'j'k'}$ are given as

$$\frac{1}{\sigma^2}\sum_{n\in N,m\in M,\alpha\in A}\left[\left\{q_{n,m}^{1-\alpha}-\sum_{i=0}^{r_\alpha}\sum_{j=0}^{r_n}\sum_{k=0}^{r_m}\beta_{ijk}(\alpha-\alpha_0)^i(n-n_0)^j(m-m_0)^k\right\}\right.$$
$$\left.(\alpha-\alpha_0)^{i'}(n-n_0)^{j'}(m-m_0)^{k'}\right]\mathbf{k}_{\alpha_0,n_0,m_0}, \tag{13}$$

and $\partial^2 L(\boldsymbol{\beta};\alpha,n,m)\big/\partial\boldsymbol{\beta}^2$ is a matrix whose elements are given as

$$\frac{\partial^2 L(\boldsymbol{\beta},\alpha,n,m)}{\partial\beta_{i'j'k'}\partial\beta_{i^*j^*k^*}}, \tag{14}$$

with diagonal elements corresponding to $\beta_{i'j'k'}$, $-\sigma^{-2}\sum_{n\in N,m\in M,\alpha\in A}(\alpha-\alpha_0)^{2i'}(n-n_0)^{2j'}(m-m_0)^{2k'}\mathbf{k}_{\alpha_0,n_0,m_0}$ and off-diagonal elements corresponding to $\beta_{i'j'k'}$ and $\beta_{i^*j^*k^*}$, $-\sigma^{-2}\sum_{n\in N,m\in M,\alpha\in A}(\alpha-\alpha_0)^{i'+i^*}(n-n_0)^{j'+j^*}(m-m_0)^{k'+k^*}k_{\alpha_0,n_0,m_0}$. Analogous to the univariate approach of Fan *et al.* (1998), the bias for the vector of estimators $\hat{\boldsymbol{\beta}}$ can be estimated by

$$\left\{\frac{\partial^2 L(\boldsymbol{\beta},\alpha,n,m)}{\partial\boldsymbol{\beta}^2}\right\}^{-1}\frac{\partial L(\boldsymbol{\beta},\alpha,n,m)}{\partial\boldsymbol{\beta}}\Bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \tag{15}$$

given values of r_α, r_n and r_m. If we set $(\alpha,n,m) = (\alpha_0,n_0,m_0)$, then on the basis of the results in (13) and (14), the bias defined at (15) becomes 0 for each component. Thus the bias is not of concern when we obtain the likelihood function estimates at the value of interest, namely (α_0,n_0,m_0). This result follows immediately because the Taylor expansion is carried out around the point (α_0,n_0,m_0) so that the first term of (11) always corresponds to the true value of β_{000} regardless of the order of the expansion. Hence there is no bias involved in the estimation of β_{000}.

Analogous to Fan *et al.* (1998), the variance estimators $\hat{\boldsymbol{\beta}}$ are obtained in the form of

$$\operatorname{var}(\hat{\boldsymbol{\beta}},\alpha,n,m) = \left\{\frac{\partial^2 L(\boldsymbol{\beta},\alpha,n,m)}{\partial\boldsymbol{\beta}^2}\right\}^{-1}\operatorname{var}\left\{\frac{\partial L(\boldsymbol{\beta},\alpha,n,m)}{\partial\boldsymbol{\beta}}\right\}\left\{\frac{\partial^2 L(\boldsymbol{\beta},\alpha,n,m)}{\partial\boldsymbol{\beta}^2}\right\}^{-1}\Bigg|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \tag{16}$$

In (16), because elements of the vector, $\operatorname{var}\{\partial L(\boldsymbol{\beta},\alpha,n,m)/\partial\boldsymbol{\beta}\}$ corresponding to $\beta_{i'j'k'}$, have the form

$$\frac{1}{\sigma^2}\sum_{n\in N,m\in M,\alpha\in A} k^2_{\alpha_0,n_0,m_0}\operatorname{var}\left\{q^{1-\alpha}_{n,m}-\sum_{i=0}^{r_\alpha}\sum_{j=0}^{r_n}\sum_{k=0}^{r_m}\beta_{ijk}(\alpha-\alpha_0)^i(n-n_0)^j(m-m_0)^k\right\}$$
$$(\alpha-\alpha_0)^{2i'}(n-n_0)^{2j'}(m-m_0)^{2k'}, \tag{17}$$

the only term not to be 0 is the element for β_{000}, when evaluated at $(\alpha,n,m)=(\alpha_0,n_0,m_0)$. Thus, on the basis of (14), (15) and (16), (17) reduces to

$$\operatorname{var}(\hat\beta_{000},\alpha_0,n_0,m_0)=\left.\frac{\sum\limits_{n\in N,m\in M,\alpha\in A}\mathbf{k}^2_{\alpha_0,n_0,m_0}\operatorname{var}\left\{q^{1-\alpha}_{n,m}-\beta_{000}\right\}}{\left(\sum\limits_{n\in N,m\in M,\alpha\in A}\mathbf{k}_{\alpha_0,n_0,m_0}\right)^2}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}. \tag{18}$$

Thus, we have a prior for $q_0^{1-\alpha}=q^{1-\alpha_0}_{n_0,m_0}\sim N\left(\mu_0,\sigma_0^2\right)$, where $\mu_0=\beta_{000}\equiv\mu(\alpha_0,n_0,m_0)$ and $\sigma_0^2=\operatorname{var}\left(\hat\beta_{000}\right)$. The values of r_α, r_n and r_m are obtained to minimize $\sigma_0^2=\operatorname{var}\left(\hat\beta_{000}\right)$. We employ the optimal bandwidth h_i on the basis of minimum variance (Simonoff, 1998, p. 105). In practice, the values of r_α, r_n and r_m also can be assumed to be fixed, say, to be 2 or 3.

2.3. *The procedure to calculate critical values of exact tests incorporating Monte Carlo simulations and statistical tables*

In this section, we provide the algorithm for executing the proposed method in practice. The procedure is based on the following steps:

(a) Obtain the prior distribution with the parameters $\left(\mu_0,\sigma_0^2\right)$ using the LML method on the basis of tabulated critical values. This prior has the form defined in Section 2.2.
(b) Generate a learning sample (e.g. defining $t=200$) of the test statistic values under the corresponding null hypothesis using MC simulations.
(c) Using the learning sample, estimate $f\left(q_0^{1-\alpha}\right)$ $f'\left(q_0^{1-\alpha}\right)$ $f''\left(q_0^{1-\alpha}\right)$ and $f^{(3)}\left(q_0^{1-\alpha}\right)$ to present $E\tilde\Delta(t)$ of proposition 2 as a function of t.
(d) Compare $\left|E\tilde\Delta(t)\right|$ with a presumed threshold (e.g. $\upsilon=0.001$) to compute an appropriate value of t_0 when $\left|E\tilde\Delta(t_0)\right|<\upsilon$.
(e) Run the MC simulations to obtain t_0 values of the exact test statistic.
(f) Use (8) or its asymptotic form from proposition 1 to obtain the estimator of the quantile of interest, evaluating the critical values.

Remark 4. The proposed estimator given at (8) or its asymptotic form from proposition 1 can be used to directly evaluate the critical values given a fixed number of MC simulations (e.g. $t_0=10,000$). In this case, proposition 2 provides the method from which we can measure the accuracy of the given estimation procedure.

Remark 5. Classical hypothesis testing methodology offers to preselect the significance level of the test by defining the corresponding critical value of the test statistic. The test statistic is then compared with the critical value with the possible decisions of rejecting the null hypothesis H_0 or not rejecting H_0. An alternative approach to hypothesis testing is based on computing the

p-value (e.g. Gibbons & Chakraborti, 2011). Generally speaking, the p-value is a value of a uniformly distributed random variable that can measure the smallest significance level at which the null hypothesis would be rejected conditionally on the observed data. The procedure we have developed can be easily modified for evaluating p-values. For example, using proposition 1, we can define p as a root of

$$T_0 = \left(\hat{q}_M^{1-p} + \frac{\mu_0}{\sigma_0^2 B_t^2 \left(\hat{q}_M^{1-p} \right) t} \right) \left(1 + \frac{1}{\sigma_0^2 B_t^2 \left(\hat{q}_M^{1-p} \right) t} \right)^{-1}, \qquad (19)$$

where T_0 represents a value of the test statistic computed conditional on the observed data. That is, $\hat{q}_M^{1-p}$ can be computed as a function of T_0, and then by virtue of the definition$1 - p = t^{-1} \sum_{j=1}^{t} K_h \left(\hat{q}_M^{1-p} - X_j \right)$, we obtain a value of p that gives an approximation to the true underlying p-value. Similarly, but slightly more complex, one can use (8) to obtain a p-value approximation. Several other procedures to examine the p-value, on the basis of the proposed method, are possible. Algorithms employed to develop the software packages of Tanajian *et al.* (2013) and Shepherd *et al.* (2013), for example, consist of the following steps: (1): (a), (b) and (c) as mentioned earlier with $q_0^{1-\alpha} = T_0$; (2) compute the value of p using (19); and (3) taking $\alpha = p$, compare $|E\tilde{\Delta}(t)|$ with a presumed threshold υ. If $|E\tilde{\Delta}(t)| \le \upsilon$, we approximate the p-value using p. If $|E\tilde{\Delta}(t)| > \upsilon$, then we generate more test statistic values under the corresponding null hypothesis using MC simulations, thus obtaining $t = t + d$, where, for example, $d = 200$, and go to Step (1.c).

3. Concluding remarks

In this paper, we defined and examined the nonparametric posterior expectation of quantiles within the general setting of hypothesis testing. The goal of the proposed approach was to develop a general framework for evaluating the p-values of a variety of exact tests, incorporating the relevant MC simulations and statistical tables in a Bayesian framework as observed and prior information, respectively. We used the smoothed EL function (Chen & Hall, 1993) to encapsulate information from the MC experiments, whereas tabulated critical values were used to compute prior distributions. The new method improves the schemes used to calculate critical values of exact tests. We showed two asymptotic propositions to conduct exact tests using asymptotic critical points and provided the rule to select the appropriate number of MC simulations needed for a desired degree of accuracy. The asymptotic result we developed is useful in practice for evaluating critical values of interest. The MC simulation studies demonstrated that the proposed method yields very accurate estimators for the critical values of exact tests. It should be noted that the proposed method can also be applied to power calculations (Cohen, 1988). The proposed method is also very flexible in terms of the variety of applications. For example, Tanajian *et al.* (2013) employed the main idea of this paper to develop STATA statistical software packages for testing symmetry and for K-sample comparisons. In a subsequent paper, we plan to address the use of the bootstrap and asymptotic Wilks-type propositions to achieve accurate estimation of critical values. Further studies are needed to evaluate the approach in other contexts. We hope that this paper will stimulate future theoretical and applied research on this topic.

Appendix: The proofs of propositions 1 and 2

To prove proposition 1, we calculate the integral $\int EL(q)\pi(q)dq$ in the neighbourhood of $\hat{q}_M^{1-\alpha}$ for large t, where $\hat{q}_M^{1-\alpha}$ is a maximum of the smoothed EL function, $EL(q)$. The approach for evaluating the integral is based on the following stages. The Taylor theorem can be applied to represent approximately the function $L(\lambda)$ at (A.1), which is the first derivative of $EL(q)$ with respect to the Lagrange multiplier λ. The obtained Taylor approximation of$L(\lambda)$ at (A.5) can be used in the evaluation of λ on the basis of the observed data points. This form can yield values of the integrals of interest. Thus, we begin with a lemma about the existence of the maximum value of the smoothed EL function in the convex hull of $\{X_1, \dots, X_t\}$. The lemma shows that the behaviour of the EL is similar to that of parametric likelihoods with respect to increasing/decreasing of the likelihoods and their maximum occurrences.

Lemma 1. *The smooth EL function, $EL(q)$, defined at (2) monotonically decreases if $q > \hat{q}_M^{1-\alpha}$, and monotonically increases if $q < \hat{q}_M^{1-\alpha}$, where the maximum of $EL(q)$ occurs at $q = \hat{q}_M^{1-\alpha}$ that satisfies $1-\alpha = t^{-1}\sum_{j=1}^{t} K_h\left(\hat{q}_M^{1-\alpha} - X_j\right)$.*

Proof. By virtue of (2) and (3), the derivative of the log $EL(q)$ is

$$\begin{aligned}
\frac{\partial}{\partial q}\log EL(q) &= \frac{\partial}{\partial q}\log\prod_{j=1}^{t}\frac{1}{t+\lambda\{K_h(q-X_j)-(1-\alpha)\}}\\
&= -\sum_{j=1}^{t}\frac{\left(\frac{\partial}{\partial q}\lambda\right)\{K_h(q-X_j)-(1-\alpha)\}+\lambda k_h(q-X_j)}{t+\lambda\{K_h(q-X_j)-(1-\alpha)\}}\\
&= -\sum_{j=1}^{t}\frac{\lambda k_h(q-X_j)}{t+\lambda\{K_h(q-X_j)-(1-\alpha)\}},
\end{aligned}$$

this follows from the fact that λ is a root of

$$\sum_{j=1}^{t}\frac{K_h(q-X_j)-(1-\alpha)}{t+\lambda\{K_h(q-X_j)-(1-\alpha)\}} = 0.$$

Thus, the sign of $\partial\log EL(q)/\partial q$ depends on the sign of $\lambda(q)$. In order to obtain the sign of $\lambda(q)$, we define the function

$$L(\lambda) \equiv \sum_{j=1}^{t}\frac{K_h(q-X_j)-(1-\alpha)}{t+\lambda\{K_h(q-X_j)-(1-\alpha)\}}. \qquad \text{(A.1)}$$

Note that $\partial L(\lambda)/\partial\lambda < 0$, and hence $L(\lambda)$ is decreasing with respect to λ. Because $\hat{q}_M^{1-\alpha}$ satisfies $t^{-1}\sum_{j=1}^{t} K_h\left(\hat{q}_M^{1-\alpha} - X_j\right) - (1-\alpha) = 0$, λ at the point $q = \hat{q}_M^{1-\alpha}$ is 0, that is, $\lambda\left(\hat{q}_M^{1-\alpha}\right) = 0$. In this case, p_j, $j = 1, 2, \dots, t$, in (8) are just t^{-1}, and the constraint, $\sum_{j=1}^{t} p_j\{K_h(q-X_j)-(1-\alpha)\} = 0$, is automatically satisfied.

Consider the two situations:

i) If $q < \hat{q}_M^{1-\alpha}$. If $\lambda(q) = 0$, then

$$L(\lambda(q)=0) = \frac{1}{t}\sum_{j=1}^{t}\{K_h(q-X_j)-(1-\alpha)\} \le \frac{1}{t}\sum_{j=1}^{t}\left\{K_h\left(\hat{q}_M^{1-\alpha}-X_j\right)-(1-\alpha)\right\} = 0,$$

because K_h is a smooth function belonging to $(0,1)$.

We are interested in λ_0 such that $L(\lambda_0) = 0$. So, we have $\lambda_0 < 0$, for $q < \hat{q}_M^{1-\alpha}$.
FA.1 The details of this case are depicted in Fig. A.1, taking the function, $L(\lambda)$, into account which decreases when λ increases. Therefore, because $\partial\log EL(q)/\partial q > 0$, $\log EL(q)$ monotonically increases for $q < \hat{q}_M^{1-\alpha}$.

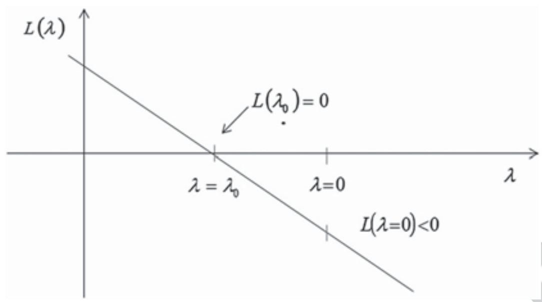

Fig. A.1. The function $L(\lambda)$ is plotted against λ for the case of $q < \hat{q}_M^{1-\alpha}$ to evaluation of the sign of λ_0 such that $L(\lambda_0) = 0$.

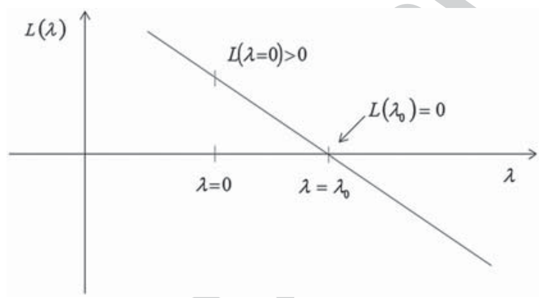

Fig. A.2. The function $L(\lambda)$ is plotted against λ for the case of $q > \hat{q}_M^{1-\alpha}$ to evaluation of the sign of λ_0 such that $L(\lambda_0) = 0$.

ii) If $q > \hat{q}_M^{1-\alpha}$. Likewise, we have

$$L\{\lambda(q)\} = \frac{1}{t}\sum_{j=1}^{t}\{K_h(q - X_j) - (1-\alpha)\} \geq \frac{1}{t}\sum_{j=1}^{t}\left\{K_h\left(\hat{q}_M^{1-\alpha} - X_j\right) - (1-\alpha)\right\} = 0.$$

We have $\lambda_0 > 0$, for $q > \hat{q}_M^{1-\alpha}$, shown in Fig. A.2. $L(\lambda)$ decreases as λ decreases. **FA.2**
In a similar manner to the case, $q < \hat{q}_M^{1-\alpha}$, because $\partial \log EL(q)/\partial q < 0$, $\log EL(q)$ monotonically decreases for $q > \hat{q}_M^{1-\alpha}$. The proof of lemma 1 is complete.

□

To simplify the proof of proposition 1, let us define $\bar{k}_h(q) = \sum_{j=1}^{t} k_h(q - X_j)/t$ and $\bar{K}_{s,h}(q) = \sum_{j=1}^{t}\left\{K_h\left(q - X_j\right) - (1-\alpha)\right\}^s/t$ for $s = 1, 2, 3$.

Proof of proposition 1. To prove proposition 1, we first note that

$$\int_{X_{(1)}}^{X_{(t)}} EL(q)\pi(q)dq = \int_{X_{(1)}}^{\hat{q}_M^{1-\alpha}-\varphi_t/\sqrt{t}} EL(q)\pi(q)dq + \int_{\hat{q}_M^{1-\alpha}-\varphi_t/\sqrt{t}}^{\hat{q}_M^{1-\alpha}+\varphi_t/\sqrt{t}} EL(q)\pi(q)dq$$
$$+ \int_{\hat{q}_M^{1-\alpha}+\varphi_t/\sqrt{t}}^{X_{(t)}} EL(q)\pi(q)dq, \qquad \text{(A.2)}$$

given $\varphi_t = t^{\beta}$, $0 < \beta < 1/2$. By lemma 1, we have

$$\begin{aligned}\int_{X_{(1)}}^{\hat{q}_M^{1-\alpha}-\varphi_t/\sqrt{t}} \exp[\log EL(q)]\pi(q)dq &\leq \exp\left[\log EL\left(\hat{q}_M^{1-\alpha} - \varphi_t\Big/\sqrt{t}\right)\right]\int_{-\infty}^{\infty}\pi(q)dq \\ &= \exp\left[\log EL\left(\hat{q}_M^{1-\alpha} - \varphi_t\Big/\sqrt{t}\right)\right].\end{aligned} \tag{A.3}$$

In order to show that the upper bound at (A.3) vanishes to 0, we consider the function $L(\lambda)$ defined in (A.1). This function can be rewritten as

$$L(\lambda) = \bar{K}_{1,h}(q) - \frac{\lambda}{t^2}\sum_{j=1}^{t}\frac{\{K(h^{-1}(q-X_j)) - (1-\alpha)\}^2}{1 + \frac{\lambda}{t}\{K(h^{-1}(q-X_j)) - (1-\alpha)\}}. \tag{A.4}$$

Define $\lambda_0 = t/g_t$ with $g_t = t^{\gamma}$. Then, defining $\gamma \in (0, 1/2)$, one can show that

$$\sqrt{t}L(\lambda_0) = \sqrt{t}\bar{K}_{1,h}(q) - \frac{\sqrt{t}}{g_t}\sum_{j=1}^{t}\frac{\{K(h^{-1}(q-X_j)) - (1-\alpha)\}^2}{1 + \frac{1}{g_t}\{K(h^{-1}(q-X_j)) - (1-\alpha)\}} \to -\infty \text{ and}$$

$$\sqrt{t}L(-\lambda_0) = \sqrt{t}\bar{K}_{1,h}(q) + \frac{\sqrt{t}}{g_t}\sum_{j=1}^{t}\frac{\{K(h^{-1}(q-X_j)) - (1-\alpha)\}^2}{1 - \frac{1}{g_t}\{K(h^{-1}(q-X_j)) - (1-\alpha)\}} \to \infty, \text{ as } t \to \infty,$$

because $\sqrt{t}\bar{K}_{1,h}(q) = O_p(1)$ (Chen & Hall, 1993). Thus, we obtain the order $\lambda = O(t/g_t)$, where λ is the solution to $L(\lambda) = 0$. In this case, because $\bar{K}_{4,h}(q) = O_p(1)$ by Chen & Hall (1993), the Taylor's expansion of the function $L(\lambda)$ with respect to λ around 0 provides the approximation

$$L(\lambda) = \bar{K}_{1,h}(q) - \frac{\lambda}{t}\bar{K}_{2,h}(q) + \frac{\lambda^2}{t^2}\bar{K}_{3,h}(q) + O_p\left(\frac{\lambda^3}{t^3}\right). \tag{A.5}$$

Solving (A.5) for λ gives

$$\lambda = \frac{t\bar{K}_{1,h}(q)}{\bar{K}_{2,h}(q)} + O_p\left(\frac{t}{g_t^2}\right), \tag{A.6}$$

which gives $g_t = o(h^{-1})$; that is, if $h = o(t^{-1/6})$, then $\gamma \in (0, 1/6)$. Because $\bar{K}_{3,h}(q) = O_p(1)$ by Chen & Hall (1993), the Taylor expansion of the function $\log EL(q)$ when λ is considered around 0, and (A.6) implies that

$$\log EL(q) = -\lambda\bar{K}_{1,h}(q) + \frac{\lambda^2}{2t}\bar{K}_{2,h}(q) + O_p\left(\frac{\lambda^3}{t^2}\right) = -\frac{t\{\bar{K}_{1,h}(q)\}^2}{2\bar{K}_{2,h}(q)} + O_p\left(\frac{\lambda^3}{t^2}\right). \tag{A.7}$$

By the Taylor expansion of the function (A.7) around $\hat{q}_M^{1-\alpha}$ at $q = \hat{q}_M^{1-\alpha} - \varphi_t/\sqrt{t}$, $\bar{k}_h\left(\hat{q}_M^{1-\alpha}\right) = O_p(1)$ and $\bar{K}_{2,h}\left(\hat{q}_M^{1-\alpha}\right) = O_p(1)$ (Chen & Hall, 1993), we obtain

$$\begin{aligned}\log EL\left(\hat{q}_M^{1-\alpha} - \varphi_t\Big/\sqrt{t}\right) &= -\frac{t\left\{\bar{k}_h\left(\hat{q}_M^{1-\alpha}\right)\varphi_t/\sqrt{t}\right\}^2}{2\bar{K}_{2,h}\left(\hat{q}_M^{1-\alpha}\right)} + O_p\left(\frac{\varphi_t^3}{t^{1/2}}\right) \\ &= -C\varphi_t^2 + O_p\left(\frac{\varphi_t^3}{t^{1/2}}\right) \to -\infty \text{ as } t \to \infty,\end{aligned} \tag{A.8}$$

for some constant $C > 0$. Therefore, by virtue of (A.3) and (A.6), we conclude that

$$\int_{X_{(1)}}^{\hat{q}_M^{1-\alpha}-\varphi_t/\sqrt{t}} EL(q)\pi(q)dq = o_p\left(\frac{1}{t}\right). \tag{A.9}$$

Similarly, we also obtain

$$\int_{\hat{q}_M^{1-\alpha}+\varphi_t/\sqrt{t}}^{X_{(t)}} EL(q)\pi(q)dq = o_p\left(\frac{1}{t}\right). \tag{A.10}$$

Equations (A.9) and (A.10) imply that (A.2) can be rewritten as

$$\int_{X_{(1)}}^{X_{(t)}} EL(q)\pi(q)dq = \int_{\hat{q}_M^{1-\alpha}-\varphi_t/\sqrt{t}}^{\hat{q}_M^{1-\alpha}+\varphi_t/\sqrt{t}} EL(q)\pi(q)dq + o_p\left(\frac{1}{t}\right). \tag{A.11}$$

Through the result at (A.7), the right hand side of (A.11) can be written as

$$\int_{\hat{q}_M^{1-\alpha}-\varphi_t/\sqrt{t}}^{\hat{q}_M^{1-\alpha}+\varphi_t/\sqrt{t}} EL(q)\pi(q)dq + o_p\left(\frac{1}{t}\right) = \int_{\hat{q}_M^{1-\alpha}-\varphi_t/\sqrt{t}}^{\hat{q}_M^{1-\alpha}+\varphi_t/\sqrt{t}} \exp\left[-\frac{t\left\{\bar{K}_{1,2}(q)\right\}^2}{2\bar{K}_{2,h}(q)}\right]\pi(q)dq + o_p\left(\frac{1}{t}\right). \tag{A.12}$$

Hence, we approximate (A.12) as the function

$$\int_{-\infty}^{\infty} \exp\left[-\frac{t\left\{\bar{K}_{1,2}(q)\right\}^2}{2\bar{K}_{2,h}(q)}\right]\pi(q)dq + o_p\left(\frac{1}{t}\right).$$

By the assumption of $\pi(q) = \exp\left\{-(q-\mu_0)^2/2\sigma_0^2\right\}/\sqrt{2\pi\sigma_0^2}$ and the Taylor expansion of the function $\bar{K}_{1.h}(q)/\sqrt{\bar{K}_{2,h}(q)}$ with respect to q around $\hat{q}_M^{1-\alpha}$, we have

$$\begin{aligned}
\exp\left[-\frac{t\{\bar{K}_{1,2}(q)\}^2}{2\bar{K}_{2,h}(q)} - \frac{(q-\mu_0)^2}{2\sigma_0^2}\right] &= \exp\left[-\frac{t\left\{\bar{k}_h\left(\hat{q}_M^{1-\alpha}\right)\right\}^2}{2\bar{K}_{2,h}\left(\hat{q}_M^{1-\alpha}\right)}\left(q-\hat{q}_M^{1-\alpha}\right)^2 - \frac{(q-\mu_0)^2}{2\sigma_0^2}\right] + o_p\left(\frac{1}{t}\right)\\
&= \exp\left[\frac{1}{2}\left\{tB_t^2\left(\hat{q}_M^{1-\alpha}\right)+\frac{1}{\sigma_0^2}\right\}\left\{q-\left(tB_t^2\left(\hat{q}_M^{1-\alpha}\right)+\frac{1}{\sigma_0^2}\right)^{-1}\right.\right.\\
&\qquad \left.\left.\times\left(tB_t^2\left(\hat{q}_M^{1-\alpha}\right)\hat{q}_M^{1-\alpha}+\frac{\mu_0}{\sigma_0^2}\right)\right\}^2\right] + o_p\left(\frac{1}{t}\right),
\end{aligned}$$

where $B_t^2\left(\hat{q}_M^{1-\alpha}\right) = \left\{\bar{k}_h\left(\hat{q}_M^{1-\alpha}\right)\right\}^2 \Big/ \bar{K}_{2,h}\left(\hat{q}_M^{1-\alpha}\right)$. Thus, the asymptotic estimator of the nonparametric posterior expectation of the quantiles is

$$\hat{q}_{NP}^{1-\alpha} = \frac{\hat{q}_M^{1-\alpha} + \frac{\mu_0}{\sigma_0^2 B_t^2\left(\hat{q}_M^{1-\alpha}\right)t}}{1 + \frac{1}{\sigma_0^2 B_t^2\left(\hat{q}_M^{1-\alpha}\right)t}} + o_p\left(\frac{1}{t}\right). \tag{A.13}$$

This completes the proof. □

To prove proposition 2, we start with the fact that $\hat{q}_M^{1-\alpha}$ satisfies the relationship $1-\alpha = t^{-1}\sum_{j=1}^{t} K_h\left(\hat{q}_M^{1-\alpha} - X_j\right)$. It follows that the Taylor expansion of the function around $q_0^{1-\alpha}$ provides the approximation F2

$$\hat{q}_M^{1-\alpha} = q_0^{1-\alpha} - \frac{\bar{K}_{1,h}\left(q_0^{1-\alpha}\right)}{\bar{k}_h\left(q_0^{1-\alpha}\right)} - \frac{\bar{k}_h'\left(q_0^{1-\alpha}\right)}{2\bar{k}_h\left(q_0^{1-\alpha}\right)}\left(\hat{q}_M^{1-\alpha} - q_0^{1-\alpha}\right)^2 + o_p\left(\frac{1}{t}\right). \tag{A.14}$$

Note that we have covariance expression

$$\begin{aligned}\operatorname{cov}\left(\bar{k}_h\left(q_0^{1-\alpha}\right), \bar{K}_{1,h}\left(q_0^{1-\alpha}\right)\right) &= \tfrac{1}{t} cov\left(\tfrac{1}{k}k\left(\tfrac{q_0^{1-\alpha}-X_1}{h}\right), K\left(\tfrac{q_0^{1-\alpha}-X_1}{h}\right) - (1-\alpha)\right) \\ &= \tfrac{f\left(q_0^{1-\alpha}\right)(-1+2\alpha)}{t} + o\left(\tfrac{1}{t}\right).\end{aligned} \tag{A.15}$$

Proof of proposition 2. Because $\left(\hat{q}_M^{1-\alpha} - q_0^{1-\alpha}\right)^3 = o_P\left(t^{-1}\right)$ from (A.14) and $\left\{\bar{K}_{1,h}\left(q_0^{1-\alpha}\right)\right\}^3 = O_p\left(h^2/t\right)$ and (9), we obtain the following equation as

$$\begin{aligned}1 - F\left(\hat{q}_{NP}^{1-\alpha}\right) &= \alpha - \left(\hat{q}_{NP}^{1-\alpha} - q_0^{1-\alpha}\right) f\left(q_0^{1-\alpha}\right) - \frac{\left(\hat{q}_{NP}^{1-\alpha} - q_0^{1-\alpha}\right)^2 f'\left(q_0^{1-\alpha}\right)}{2} \\ &\quad + o_p\left(\frac{1}{t}\right) \equiv \alpha - \tilde{\Delta}(t) + o_p\left(\frac{1}{t}\right).\end{aligned} \tag{A.16}$$

Equations (A.14) and (A.16) jointly imply that

$$\tilde{\Delta}(t) = \left(\frac{\left(q_M^{1-\alpha} - q_0^{1-\alpha}\right) + \frac{\left(\mu_0 - q_0^{1-\alpha}\right)}{\sigma_0^2 B_t^2\left(\hat{q}_M^{1-\alpha}\right) t}}{1 + \frac{1}{\sigma_0^2 B_t^2\left(\hat{q}_M^{1-\alpha}\right) t}}\right) f\left(q^{1-\alpha}\right) \frac{\left(\left(q_M^{1-\alpha} - q_0^{1-\alpha}\right) + \frac{\left(\mu_0 - q_0^{1-\alpha}\right)}{\sigma_0^2 B_t^2\left(\hat{q}_M^{1-\alpha}\right) t}\right)^2 f'\left(q^{1-\alpha}\right)}{2\left(1 + \frac{1}{\sigma_0^2 B_t^2\left(\hat{q}_M^{1-\alpha}\right) t}\right)^2}. \tag{A.17}$$

Because $B_t^2\left(q_0^{1-\alpha}\right) = f^2\left(q_0^{1-\alpha}\right)/\alpha(1-a) + o_P(1)$ and $\hat{q}_M^{1-\alpha} = q_0^{1-\alpha} + o_P(1)$, (A.17) can be rewritten as

$$\tilde{\Delta}(t) = \left(\frac{\left(q_M^{1-\alpha} - q_0^{1-\alpha}\right) + \frac{\alpha(1-\alpha)\left(\mu_0 - q_0^{1-\alpha}\right)}{\sigma_0^2 f^2\left(q_0^{1-\alpha}\right) t}}{1 + \frac{\alpha(1-\alpha)}{\sigma_0^2 f^2\left(q_0^{1-\alpha}\right) t}}\right) f\left(q^{1-\alpha}\right) \frac{\left(q_M^{1-\alpha} - q_0^{1-\alpha}\right)^2 f'\left(q^{1-\alpha}\right)}{2\left(1 + \frac{\alpha(1-\alpha)}{\sigma_0^2 f^2\left(q_0^{1-\alpha}\right) t}\right)^2} + o_p\left(\frac{1}{t}\right). \tag{A.18}$$

We take the expectation of the function $\tilde{\Delta}(t)$ in (A.18) using (A.14). Then, we obtain

$$\begin{aligned}E\tilde{\Delta}(t) &= \left(\frac{E\left(q_M^{1-\alpha} - q_0^{1-\alpha}\right) + \frac{\alpha(1-\alpha)\left(\mu_0 - q_0^{1-\alpha}\right)}{\sigma_0^2 f^2\left(q_0^{1-\alpha}\right) t}}{1 + \frac{\alpha(1-\alpha)}{\sigma_0^2 f^2\left(q_0^{1-\alpha}\right) t}}\right) f\left(q^{1-\alpha}\right) \frac{E\left(q_M^{1-\alpha} - q_0^{1-\alpha}\right)^2 f'\left(q^{1-\alpha}\right)}{2\left(1 + \frac{\alpha(1-\alpha)}{\sigma_0^2 f^2\left(q_0^{1-\alpha}\right) t}\right)^2} + o\left(\frac{1}{t}\right) \\ &= \left(\frac{E\left(-\frac{\bar{K}_{1,h}\left(q_0^{1-\alpha}\right)}{\bar{k}_h\left(q_0^{1-\alpha}\right)}\right) - \frac{\alpha(1-\alpha) f'\left(q_0^{1-\alpha}\right)}{2f^3\left(q_0^{1-\alpha}\right)t} + \frac{\alpha(1-\alpha)\left(\mu_0 - q_0^{1-\alpha}\right)}{\sigma_0^2 f^2\left(q_0^{1-\alpha}\right) t}}{1 + \frac{\alpha(1-\alpha)}{\sigma_0^2 f^2\left(q_0^{1-\alpha}\right) t}}\right) \\ &\quad \times f\left(q^{1-\alpha}\right) \frac{\frac{\alpha(1-\alpha) f'\left(q^{1-\alpha}\right)}{f^2\left(q_0^{1-\alpha}\right) t}}{2\left(1 + \frac{\alpha(1-\alpha)}{\sigma_0^2 f^2\left(q_0^{1-\alpha}\right) t}\right)^2} + o\left(\frac{1}{t}\right).\end{aligned} \tag{A.19}$$

To compute $E\left(\bar{K}_{1,h}\left(q_0^{1-\alpha}\right)/\bar{k}_h\left(q_0^{1-\alpha}\right)\right)$, we use the Taylor expansion in two variables of $f(x,y) = y/x$ given as

$$\begin{aligned} f(x,y) = \frac{y}{x} &\approx \frac{Ey}{Ex} - \frac{Ey}{(Ex)^2}(x-Ex) + \frac{1}{(Ex)^2}(y-Ey) \\ &+ \frac{3Ey}{(Ex)^3}(x-Ex)^2 - \frac{1}{(Ex)^2}(x-Ex)(y-Ey). \end{aligned} \tag{A.20}$$

Equations (A.15) and (A.20) imply that

$$\begin{aligned} E\left(\frac{\bar{K}_{1,h}\left(q_0^{1-\alpha}\right)}{\bar{k}_h\left(q_0^{1-\alpha}\right)}\right) &= \frac{\sigma_k^2 f'\left(q_0^{1-\alpha}\right)h^2}{2f\left(q_0^{1-\alpha}\right)} + \frac{\sigma_k^4 f^{(3)}\left(q_0^{1-\alpha}\right)h^4}{24f\left(q_0^{1-\alpha}\right)} \\ &- \frac{\left(\sigma_k^2\right)^2 f'\left(q_0^{1-\alpha}\right) f''\left(q_0^{1-\alpha}\right)h^4}{4f^2\left(q_0^{1-\alpha}\right)} - \frac{f\left(q_0^{1-\alpha}\right)(-1+2\alpha)}{t} + o\left(\frac{1}{t}\right). \end{aligned}$$

The proof of proposition 2 is complete. □

Acknowledgements

This research is supported by the NIH grant 1R03DE020851 - 01A1 (the National Institute of Dental and Craniofacial Research). The authors are grateful to the editor and the referee for suggestions that led to a substantial improvement in this paper.

References

Chen, S. X. & Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Stat.* **21**, 1166–1181.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, Lawrence Erlbaumn Associates, New Jersey.

DasGupta, A. (2008). *Asymptotic theory of statistics and probability*, Springer, New York.

DiCiccio, T. J., Kass, R. E., Raftery, A. & Wasserman, L. (1997). Computing Bayes factors by combining simulation and asymptotic approximations. *J. Am. Stat. Assoc.* **92**, 903–915.

Fan, J., Farmen, M. & Gijbels, I. (1998). Local maximum likelihood estimation and inference. *J. R. Stat. Soc. Ser. B* **60**, 591–608.

Fan, J. & Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. R. Stat. Soc. Ser. B* **57**, 371–394.

Fan, J. & Gijbels, I. (1996). *Local polynomial modelling and its applications*, Chapman and Hall, London.

Fan, J., Heckman, N. E. & Wand, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Am. Stat. Assoc.* **90**, 141–150.

Gibbons, J. D. & Chakraborti, S. (2011). *Nonparametric statistical inference*, Chapman and Hall, London.

Hall, P. & Welsh, A. H. (1983). A test for normality based on the empirical characteristic function. *Biometrika* **70**, 485–489.

Härdle, W. (1990). *Applied nonparametric regression*, Cambridge Univ. Press. AQ1

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *J. Am. Stat. Assoc.* **90**, 773–795.

Kutner, M. H., Nachtsheim, C. J., Neter, J. & Li, W. (2005). *Applied linear statistical methods*, (5^{th} edition)., McGraw Hill, New York.

Lazar, N. A. & Mykland, P. A. (1998). An evaluation of the power and conditionality properties of empirical likelihood. *Biometrika* **85**, 523–534.

Lazar, N. A. (2003). Bayesian empirical likelihood. *Biometrika* **90**, 319–326.

Mudholkar, G. S. & Hutson, A. D. (1997). Continuity corrected approximations for and "exact" inference with Pearson's X^2. *J. Stat. Plann. Inference* **59**, 61–78.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

Owen, A. B. (1990). Empirical likelihood ratio confidence regions. *Ann. Stat.* **19**, 237–249.

Owen, A. B. (2001). *Empirical likelihood*, Chapman and Hall/CRC, New York.

Rosenbaum, P. R. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *J. R. Stat. Soc. Ser. B* **67**, 515–530.

Serfling, R. J. (2002). *Approximation theorems of mathematical statistics*, Wiley, New York.

Shepherd, L. A., Tsai, W. -M., Vexler, A. & Miecznikowski, J. C. (2013). dbEmpLikeNorm: test
AQ2 for joint assessment of normality, R package. Available at: http://cran.r-project.org/web/packages/dbEmpLikeNorm/index.html.

Silverman, B. J. (1986). *Density estimation for statistics and data analysis*, Chapman and Hall, London.

Simonoff, J. S. (1998). *Smoothing methods in statistics*, Springer-Verlag, New York.

Tanajian, H., Vexler. A. & Hutson, A. D. (2013). Novel and efficient density based empirical likelihood pro-
AQ3 cedures for symmetry and K-sample comparisons: STATA package. Available at: http://sphhp.buffalo.edu/biostatistics/research-and-facilities/software/stata.html.

Tibshirani, R. & Hastie, T. J. (1987). Local likelihood estimation. *J. Am. Stat. Assoc.* **82**, 559–567.

Vexler, A. & Gurevich, G. (2010). Empirical likelihood ratios applied to goodness-of-fit tests based on sample entropy. *Comput. Stat. Data Anal.* **54**, 531–545.

Vexler, A. & Yu, J. (2011). Two-sample density-based empirical likelihood tests for incomplete data in applications to a pneumonia study. *Biometrical J.* **53**, 628–651.

Yu, J., Vexler, A. & Tian, L. (2010). Analyzing incomplete data subject to a threshold using empirical likelihood methods: an application to a pneumonia risk study in an ICU setting. *Biometrics* **66**, 123–130.

Yu, J., Vexler, A., Kim, S. & Hutson, A. D. (2011). Two-sample empirical likelihood ratio tests for medians application to biomarker evaluations. *Can. J. Stat.* **39**, 671–689.

Zhou, W. & Jing, B. -Y. (2003). Adjusted empirical likelihood method for quantiles. *Ann. Inst. Stat. Math.* **55**, 689–703.

Received December 2012, in final form January 2014

Albert Vexler, Department of Biostatistics, The University at Buffalo, State University of New York , NY 14214, USA.
E-mail: avexler@buffalo.edu

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's website.