Bandwidth Selection for Kernel Quantile Estimation

Ming-Yen Cheng[1] and Shan Sun[2]

Abstract

In this article, we summarize some quantile estimators and related bandwidth selection methods and give two new bandwidth selection methods. By four distributions: standard normal, exponential, double exponential and log normal we simulated the methods and compared their efficiencies to that of the empirical quantile. It turns out that kernel smoothed quantile estimators, with no matter which bandwidth selection method used, are more efficient than the empirical quantile estimator in most situations. And when sample size is relatively small, kernel smoothed estimators are especially more efficient than the empirical quantile estimator. However, no one method can beat any other methods for all distributions.

Keywords. Bandwidth, kernel, quantile, nonparametric smoothing.

Short Title. Quantile Estimation.

JEL subject classification: C14, C13.

[1] Department of Mathematics, National Taiwan University, Taipei 106, Taiwan. Email: cheng@math.ntu.edu.tw

[2]Department of Mathematics and Statistics, Texas Tech University, Lubbock, Texas 79409 - 1042, USA. Email: ssun@math.ttu.edu

1 Introduction

The estimation of population quantiles is of great interest when a parametric form for the underlying distribution is not available. In addition, quantiles often arise as the natural thing to estimate when the underlying distribution is skewed. Let $X_1, X_2, \cdots, X_n$ be an independent and identically distributed random sample drawn from an absolutely continuous distribution function F with density f. Let $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ denote the corresponding order statistics. The quantile function Q of the population is defined as $Q(p) = \inf\{x : F(x) \geq p\}, 0 < p < 1$. Note that Q is the left-continuous inverse of F. Denote, for each $0 < p < 1$, the pth quantile of F by ξ_p, that is, $\xi_p = Q(p)$.

A traditional nonparametric estimator of the distribution function is the empirical function $F_n(x)$, which is defined as

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,x]}(X_i)$$

where $I_A(x) = 1$ if $x \in A$ and 0 otherwise. Accordingly, a nonparametric estimator of ξ_p is the empirical quantile

$$Q_n(p) = \inf\{x : F_n(x) \geq p\} = X_{([np]+1)},$$

where $[np]$ denotes the integer part of np. Let $p_r = r/(n+1)$ and $q_r = 1 - p_r$. If we use $X_{(r)}$ to estimate the p_rth quantile, then the asymptotic bias and variance are

$$ABias\{X_{(r)}\} = \frac{p_r q_r Q''(p_r)}{2(n+2)} + \frac{p_r q_r}{(n+2)^2}\Big\{\frac{1}{3}(q_r - p_r)Q_r''' + \frac{1}{8}Q_r''''\Big\},$$

$$AVar\{X_{(r)}\} = \frac{p_r q_r}{(n+2)}Q_r'^2 + \frac{p_r q_r}{(n+2)^2}\Big\{2(q_r - p_r)Q_r'Q_r'' + p_r q_r(Q_r'Q_r''' + \frac{1}{2}Q_r'')\Big\}.$$

The asymptotic mean squared error of $X_{(r)}$ should be $AMSE\{X_{(r)}\} = ABias\{X_{(r)}\}^2 + AVar\{X_{(r)}\}$.

When F is continuous, it is more natural to use a smooth random function as an estimator of F since there is a substantial lack of efficiency, caused by the variability of individual order statistics. Indeed, the choice of F_n does not always lead to the best estimator of F (cf. Read (1972), who has shown that F_n is inadmissible with respect to the integrated square loss). Intuitively appealing and easily understood competitors to Q_n are the popular kernel quantile estimators, see Section 2.

Section **??** gives the asymptotic mean squared errors and asymptotically optimal bandwidths for two kernel smoothed quantile estimators. The optimal bandwidths depend on unknown quantities such as density derivatives and quantile derivatives. Kernel estimators and optimal bandwidths for these unknowns are addressed as well. In Section 4, we give four methods to select the bandwidths for the two kernel quantile estimators based on data. In Section 5 we implement these methods on four specific distributions and the results of the simulation. The Appendix gives some proofs.

2 Kernel smoothed quantile estimation

2.1 Inverse of kernel distribution function estimator

A popular kernel quantile estimator is based on the Nadaraya (1964) type estimator for F, defined as

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)$$

where

$$K_h(x) = \int_{-\infty}^{x} \frac{1}{h}k(\frac{t}{h})dt,$$

k is a kernel function satisfying $k \geq 0, \int_{-\infty}^{\infty} k(x)dx = 1$. Here $h = h_n > 0$ is called the smoothing parameter or bandwidth since it controls the amount of smoothness in the estimator for a given sample of size n. We make the assumption that $h \to 0$ as $n \to \infty$. The corresponding estimator of the quantile function $Q = F^{-1}$ is then defined by

$$\hat{Q}_n(p) = \inf\{x : \hat{F}_n(X) \geq p\},\ 0 < p < 1. \tag{1}$$

Nadaraya (1964) showed under some assumptions for k, f and $h, \hat{Q}_n(p)$ (appropriately normalized) has an asymptotic standard normal distribution. Another notable property of $\hat{Q}_n(p)$, namely the almost sure consistency, was obtained by Yamato (1973). Ralescu and Sun (1992) obtained the necessary and sufficient conditions for the asymptotic normality of $\hat{Q}_n(p)$. Azzalini (1981) and an unpublished report used heuristic arguments based on second order approximations and performed some numerical comparisons of $\hat{Q}_n(p)$ with the classical sample quantile for estimating the 95th quantile of the Gamma (1) distribution. These studies indicated a considerable amount of empirical evidence to support the superiority of $\hat{Q}_n(p)$ for a variety of smooth distribution functions.

Azzalini (1981) considered second order property of $\hat{F}_n$ under the following assumptions: (i) $h \to 0$ as $n \to \infty$; (ii) the kernel has a finite support, that is, $k(t) = 0$ if $|t| > t_0$ for some positive t_0; (iii) the density f is continuous in the interval $(x - t_0 h, x + t_0 h)$; and (iv) $f'(x)$ exists. He pointed out that the asymptotic optimal bandwidth for $\hat{F}$ is of the form

$$h_{opt} = \left(\frac{u}{4vn}\right)^{\frac{1}{3}} \tag{2}$$

where

$$u = f(x)\Big\{t_0 - \int_{-t_0}^{t_0} K^2(t)dt)\Big\},\ v = \Big\{\frac{1}{2}f'(x)\int_{-t_0}^{t_0} t^2 k(t)dt\Big\}^2.$$

Also, Azzalini (1981) suggested, without offering a proof, that (2) is again the asymptotically optimal choice of h for $\hat{Q}_n(p)$. We state the result in the following theorem and the proof of the theorem can be found in Shankar (1998).

We make the following assumptions:

Assumption A

(1) f is differentiable with a bounded derivative f';

(2) f' is continuous in the neighborhood of ξ_p and $f'(\xi_p) \neq 0$;

(3) $\int_{-\infty}^{\infty} xk(x)dx = 0$ and $\int_{-\infty}^{\infty} x^2 k(x)dx < \infty$.

Theorem 1. Under assumptions (1)-(3), the asymptotic mean squared error of $\hat{Q}(p)$ is

$$AMSE\{\hat{Q}(p)\} = \frac{p(1-p)}{nf(\xi_p)^2} + \frac{h^4}{4}\frac{f'(\xi_p)^2}{f(\xi_p)^2}\mu_2(k)^2 - \frac{h}{n}\frac{1}{f(\xi_p)}\psi(k)$$

and the asymptotically optimal choice of bandwidth for the smoothed empirical quantile function $\hat{Q}_n(p)$ is

$$h_{opt,1} = \left[\frac{f(\xi_p)\psi(k)}{n\{f'(\xi_p)\}^2\mu_2(k)^2}\right]^{\frac{1}{3}} \tag{3}$$

where $\mu_2(k) = \int_{-\infty}^{\infty} t^2 k(t)dt$ and $\psi(k) = 2\int yk(y)K(y)dy$. If we take k as the standard normal density, then $\int_{-\infty}^{\infty} tdK^2(t) = 1/\sqrt{\pi}$, $\mu_2(k) = 1$ and

$$h_{opt,1} = \left[\frac{f(\xi_p)}{\sqrt{\pi}n\{f'_{g_n^*}(\xi_p)\}^2}\right]^{\frac{1}{3}}.$$

2.2 Kernel smoothing the order statistics

Another type of smooth quantile estimator, provided by Yang (1985) and also traced to Parzen (1979), is

$$\tilde{Q}_n(p) = \sum_{i=1}^{n} X_{(i)} \int_{\frac{i-1}{n}}^{\frac{i}{n}} \frac{1}{h} k(\frac{p-x}{h})dx. \tag{4}$$

It is clear that when i/n is close to p, $\tilde{Q}_n(p)$ puts more weight on the order statistics $X_{(i)}$. The asymptotic normality and mean squared consistency of $\tilde{Q}_n(p)$ were provided by Yang (1985), while Falk (1984) showed that the asymptotic performance of $\tilde{Q}_n(p)$ is better than that of the empirical sample quantile $Q_n(p)$ in the sense of relative deficiency for appropriately chosen kernels and sufficiently smooth distribution functions.

Building on Faulk (1984), Sheater and Morron (1990) gave the asymptotic mean squared error (AMSE) of $\tilde{Q}_n(p)$ as follows if f is not symmetric or f is symmetric but $p \neq 0.5$:

$$AMSE\{\tilde{Q}_n(p)\} = \frac{p(1-p)}{n} q^2(p) + \frac{1}{4} h^4 q'(p)^2 \mu_2(k)^2 - \frac{h}{n} q^2(p)\psi(k) \tag{5}$$

where $q = Q'$ and $q' = Q''$. If $q = Q' > 0$ then

$$h_{opt,2} = \left\{\frac{Q'(p)^2\psi(k)}{nQ''(p)^2\mu_2(k)^2}\right\}^{\frac{1}{3}}. \tag{6}$$

Remark 2.1. *When F is symmetric and $p = 0.5$, then*

$$AMSE\{\tilde{Q}_n(p)\} = n^{-1}[q(0.5)]^2\{0.25 - 0.5h\psi(k) + n^{-1}h^{-1}R(k)\},$$

where $R(k) = \int k^2(x)dx$. In this case, there is no single optimal bandwidth minimizing the $AMSE$.

Remark 2.2. *If $q = 0$, we need higher order terms. The AMSE of $\tilde{Q}_n(p)$ can be shown as follows:*

$$AMSE\{\tilde{Q}_n(p)\} = (\frac{1}{4} - \frac{1}{n})h^4 Q''(q)^2\mu_2^2(k) + 2n^{-1}h^2 Q''(q)^2 \int (q - ht)tk(t)j(t)dt$$

where $j(t) = \int_{-\infty}^{t} xk(x)dx$. The proof is provided in the Appendix.

3 Density and quantile derivative estimation

The asymptotically optimal bandwidths $h_{opt,1}$ and $h_{opt,2}$ for $\hat{Q}_n(p))$ and $\tilde{Q}_n(p)$ depend on $f(\xi_p)$, $f'(\xi_p)$, $Q'(p)$ and $Q''(p)$. This section provides nonparametric estimators of these quantities and the asymptotically optimal bandwidths.

3.1 Density derivative estimation

From (3) we know that we need to estimate f'. A natural estimator of the rth derivative ($r \geq 1$) of f can be obtained by differentiating the estimator

$$\hat{f}_{g_n}(x) = \frac{d}{dx}\hat{F}_n(x) = \frac{d}{dx}\Big\{\frac{1}{n}\sum_{i=1}^{n} K_{g_n}(x - X_i)\Big\} = \frac{1}{n}\sum_{i=1}^{n} k_{g_n}(x - X_i) \tag{7}$$

of the density $f(x)$, giving

$$\hat{f}_{g_n}^{(r)}(x) = \frac{d^r}{dx^r}\frac{1}{ng_n}\sum_{i=1}^{n} k\big(\frac{x - X_i}{g_n}\big) = \frac{1}{ng_n^{r+1}}\sum_{i=1}^{n} k^{(r)}\big(\frac{x - X_i}{g_n}\big) \tag{8}$$

where g_n is the smoothing parameter (Wand and Jones, 1995). Therefore, the asymptotic mean squared error properties of $\hat{f}_{g_n}^{(r)}(x)$ can be derived straightforwardly to obtain (Wand and Jones, 1995)

$$AMSE\{\hat{f}_{g_n}^{(r)}(x)\} = \frac{1}{ng_n^{2r+1}}R(k^{(r)})f(x) + \frac{1}{4}g_n^4\{\mu_2(k)\}^2\big\{f^{(r+2)}(x)\big\}^2 \tag{9}$$

where $R(\eta) = \int \eta^2(x)dx$ for any square-integrable function η. It follows that the AMSE-optimal bandwidth for estimating $f^{(r)}(x)$ is of order $n^{-1/(2r+5)}$. The asymptotically optimal bandwidth for for $\hat{f}_{g_n}(x)$ is given by

$$g_n^* = \left\{\frac{R(k)f(x)}{n(\mu_2(k))^2 f''(x)^2}\right\}^{\frac{1}{5}}. \tag{10}$$

and the asymptotically optimal bandwidth for $\hat{f}'_{g_n}(x)$ is

$$g_n^{**} = \left\{ \frac{3R(k')f(x)}{n(\mu_2(k))^2 f'''(x)^2} \right\}^{\frac{1}{7}} . \tag{11}$$

When k is the standard Normal density,

$$g_n^{*} = \left\{ \frac{f(x)}{n\sqrt{\pi} f''(x)^2} \right\}^{\frac{1}{5}} , \qquad g_n^{**} = \left\{ \frac{3f(x)}{4n\sqrt{\pi} f'''(x)^2} \right\}^{\frac{1}{7}} .$$

3.2 Quantile derivative estimation

Next, we estimate $Q' = q$ and $Q'' = q'$ in the following ways. From (4), we know that the estimator of $Q' = q$ can be constructed as follows:

$$\begin{aligned} \tilde{q}(p) = \tilde{Q}'_n(p) \quad &= \textstyle\sum_{i=1}^n X_{(i)}[k_a(p - \frac{i-1}{n}) - k_a(p - \frac{i}{n})] \\ &= \textstyle\sum_{i=2}^n (X_{(i)} - X_{(i-1)})k_a(p - \frac{i-1}{n}) - X_{(n)}k_a(p-1) + X_{(1)}k_a(p). \end{aligned} \tag{12}$$

where $k_a(x) = \frac{1}{a}k(\frac{x}{a})$ and $a = a_n$ is the bandwidth for $\tilde{q}$. Jones (1992) derived that the asymptotic MSE of $\tilde{q}(p)$ is given as follows:

$$AMSE\{\tilde{q}(p)\} = \frac{a^4}{4} q''(p)^2 \mu_2(k)^2 + \frac{1}{na} q^2(p) \int k^2(y)dy \, . \tag{13}$$

Minimizing (13) with respect to a, we obtain the asymptotically optimal bandwidth for $\tilde{q}(p)$ as

$$a^*_{opt} = \left\{ \frac{Q'(p)^2 \int k^2(y)dy}{nQ'''(p)^2 \mu_2(k)^2} \right\}^{\frac{1}{5}} . \tag{14}$$

To estimate $Q'' = q'$ in (6), note that

$$\tilde{Q}''_n(p) = \frac{d}{dp} \tilde{Q}'_n(p) = \frac{1}{a^2} \sum_{i=1}^n X_{(i)} \Big\{ k'\big(\frac{p - \frac{i-1}{n}}{a}\big) - k'\big(\frac{p - \frac{i}{n}}{a}\big) \Big\} . \tag{15}$$

Similarly, we obtain the asymptotically optimal bandwidth for $\tilde{Q}''_n(p)$ as

$$a^{**}_{opt} = \left\{ \frac{3 \int k'(x)^2 dx Q'(p)^2}{n\mu_2(k)^2 Q^{(4)}(p)^2} \right\}^{\frac{1}{7}} . \tag{16}$$

4 Bandwidth selection

In this section, we consider several data-based methods to find the asymptotically optimal bandwidths for the estimators $\hat{Q}_n(p)$ and $\tilde{Q}_n(p)$. Bandwidth plays a critical role in implementation of practical estimation. It determines the trade-off between the amount of smoothness obtained and closedness of the estimation to the true distribution. (see Wand and Jones)

4.1 Method 1. Approximate $h_{opt,1}$ for $\hat{Q}_n(p)$ using density derivative estimators

Note that the asymptotically optimal bandwidth $h_{opt,1}$ for $\hat{Q}_n(p)$, given in (3), involves $f(\xi_p)$ and $f'(\xi_p)$, which can be estimated by $\hat{f}_{g_n}(\hat{\xi}_p)$ and $\hat{f}'_{g_n}(\hat{\xi}_p)$ respectively. Here, $\hat{\xi}_p$ is the empirical p-th quantile $Q_n(p)$. Using g_n^* in (10) with $f(\hat{\xi}_p)$ and $f''(\hat{\xi}_p)$ replaced by their Normal(μ, σ^2) reference values, we obtain $\hat{f}_{g_n^*}(x)$. Using g_n^{**} in (11) with $f(\hat{\xi}_p)$ and $f'''(\hat{\xi}_p)$ replaced by their Normal(μ, σ^2) reference values, we obtain $\hat{f}'_{g_n^{**}}(x)$. Plugging this into (3), we have a data-based bandwidth

$$\hat{h}_{opt,1} = \left[\frac{\hat{f}_{g_n^*}(\hat{\xi}_p)\psi(k)}{n\left\{\hat{f}'_{g_n^{**}}(\hat{\xi}_p)\right\}^2 \mu_2(k)^2} \right]^{\frac{1}{3}} \tag{17}$$

for $\hat{Q}_n(p)$. If k is the standard normal density then

$$\hat{h}_{opt,1} = \left[\frac{\hat{f}_{g_n^*}(\hat{\xi}_p)}{n\sqrt{\pi}\left\{\hat{f}'_{g_n^{**}}(\hat{\xi}_p)\right\}^2} \right]^{\frac{1}{3}}. \tag{18}$$

Remark 4.1. *In the expression of the $h_{opt,1}$, we have the derivative of f in the denominator. If f' has zeros, then its estimates at these zeros are also very small. Hence the estimator $\hat{h}_{opt,1}$ of $h_{opt,1}$ at these zeros will be very unstable. For example, if f is standard normal, then $f' = -xf$ has a zero at $x = 0$, which corresponds to $p = 0.5$, and hence, when $p = 0.5$, the estimator $\hat{h}_{opt,1}$ is very unstable. Similarly, the first derivative of the double exponential density has a zero at $x = 0$ and the first derivative of the log normal density has a zero at $x = e^{-1}$.*

4.2 Method 2. Approximate $h_{opt,2}$ for $\tilde{Q}_n(p)$ using quantile derivative estimators

The asymptotically optimal bandwidth $h_{opt,2}$, given in (6), for $\tilde{Q}_n(p)$ involves the unknown quantities $Q'(p)$ and $Q''(p)$, which can be estimated by $\tilde{Q}'_n(p)$ and $\tilde{Q}''_n(p)$ in (12) and (15), respectively. The asymptotically optimal bandwidths a^*_{opt} and a^{**}_{opt}, given in (14) and (16), for $\tilde{Q}'_n(p)$ and $\tilde{Q}''_n(p)$ depend on $Q'(p)$, $Q'''(p)$ and $Q^{(4)}(p)$. We replace these unknowns by their Normal(μ, σ^2) reference values. Then, using $\tilde{Q}'_n(p)$ with $a = a^*_{opt}$ and $\tilde{Q}''_n(p)$ with $a = a^{**}_{opt}$, we have the following data-based bandwidth

$$\hat{h}_{opt,2} = \left\{ \frac{\tilde{Q}'_n(p)^2 \psi(k)}{n \tilde{Q}''_n(p)^2 \mu_2(k)^2} \right\}^{\frac{1}{3}} \tag{19}$$

for $\tilde{Q}_n(p)$.

4.3 Method 3. Approximate $h_{opt,1}$ for $\hat{Q}_n(p)$ using quantile derivative estimators

We introduce an alternative way of estimating $f(\xi_p)$ and $f'(\xi_p)$ in $h_{opt,1}$, see (3), which uses estimators of the quantile derivatives. Note that

$$Q'(p) = \frac{1}{f(F^{-1}(p))} = \frac{1}{f(Q(p))} = \frac{1}{f(\xi_p)} \tag{20}$$

$$Q''(p) = \frac{-f'(Q(p))}{f^3(Q(p))} = \frac{-f'(\xi_p)}{f^3(\xi_p)} \,. \tag{21}$$

Hence, (3) becomes

$$h_{opt,1} = \left\{ \frac{Q'_n(p)^5 \psi(k)}{n Q''_n(p)^2 \mu_2(k)^2} \right\}^{\frac{1}{3}} .$$

Similar to Method 2, first replace the unknowns in a^*_{opt} and a^{**}_{opt} by their Normal reference values, and then use $\tilde{Q}'_n(p)$ with $a = a^*_{opt}$ and $\tilde{Q}''_n(p)$ with $a = a^{**}_{opt}$ to get

$$\bar{h}_{opt,2} = \left\{ \frac{\tilde{Q}'_n(p)^5 \psi(k)}{n \tilde{Q}''_n(p)^2 \mu_2(k)^2} \right\}^{\frac{1}{3}} . \tag{22}$$

If we take k as the standard normal density, then

$$\bar{h}_{opt,2} = \left\{ \frac{\tilde{Q}'_n(p)^5}{n \sqrt{\pi} \tilde{Q}''_n(p)^2} \right\}^{\frac{1}{3}} .$$

4.4 Method 4. Approximate $h_{opt,2}$ for $\tilde{Q}_n(p)$ using density derivative estimators

From (20) and (21), we have

$$h_{opt,2} = \left\{ \frac{f(\xi_p)^4 \psi(k)}{n f'(\xi_p)^2 \mu_2(k)^2} \right\}^{\frac{1}{3}} . \tag{23}$$

Then, plugin the estimators of $f(\xi_p)$ and $f'(\xi_p)$ in Method 1, see (17), to obtain

$$\bar{h}_{opt,2} = \left\{ \frac{\hat{f}_{g_n^*}(\hat{\xi}_p)^4 \psi(k)}{n \hat{f}'_{g_n^{**}}(\hat{\xi}_p)^2 \mu_2(k)^2} \right\}^{\frac{1}{3}} . \tag{24}$$

When k is standard normal density, $\bar{h}_{opt,2}$ becomes

$$\bar{h}_{opt,2} = \left\{ \frac{\hat{f}_{g_n^*}(\hat{\xi}_p)^4}{n\sqrt{\pi} \hat{f}'_{g_n^{**}}(\hat{\xi}_p)^2} \right\}^{\frac{1}{3}} .$$

5 Numerical Performance

We implement the methods in Section 4. Four distributions are selected: Exponential, Double Exponential, Lognormal and standard Normal. We shall use the standard normal density as the kernel k, i.e. $k(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$. Then $k'(x) = -xk(x)$. and we can find

$$\mu_2(k) = \int x^2 k(x) dx = 1,$$

$$\psi(k) = 2 \int \{k(x)[\int_{-\infty}^{x} k(t)dt]\} dx = \frac{1}{\sqrt{\pi}},$$

$$R(k) = \int k^2(x) dx = \frac{1}{2\sqrt{\pi}},$$

$$R(k') = \int \{k'(x)\}^2 dx = \int x^2 k^2(x) dx = \frac{1}{4\sqrt{\pi}}.$$

5.1 True values

In the following we compute the asymptotically optimal bandwidths and the $AMSE$s for the four distributions. First, we have the relationship between $Q(p)$ and $f(\xi_p)$ as following

$$Q'(p) = \frac{1}{f(\xi_p)},\ Q''(p) = -\frac{f'(\xi_p)}{f(\xi_p)^3},\ Q'''(p) = \frac{3f'(\xi_p)^2 - f(\xi_p)f''(\xi_p)}{f(\xi_p)^5},$$

$$Q^{(4)}(p) = \frac{10f(\xi_p)f'(\xi_p)f''(\xi_p) - f(\xi_p)^2 f'''(\xi_p) - 15f'(\xi_p)^3}{f(\xi_p)^7}.$$

Using the above results, the asymptotic *mse* of $\tilde{Q}_n(p)$ is

$$AMSE\{\tilde{Q}_n(p)\} = \frac{p(1-p)}{nf(\xi_p)^2} + \frac{h^4 f'(\xi_p)^2}{4f(\xi_p)^6} - \frac{h}{n\sqrt{\pi}f(\xi_p)^2}\,.$$

Also we have

$$a^* = \left\{\frac{f(\xi_p)^8}{2n\sqrt{\pi}(3f'(\xi_p)^2 - f(\xi_p)f''(\xi_p))^2}\right\}^{\frac{1}{5}},$$

$$a^{**} = \left\{\frac{3f(\xi_p)^{12}}{4n\sqrt{\pi}(10f(\xi_p)f'(\xi_p)f''(\xi_p) - f(\xi_p)^2 f'''(\xi_p) - 15f'(\xi_p)^3)^2}\right\}^{\frac{1}{7}}.$$

Case 1. *f is the standard normal density.* We have

$$f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\ f'(x) = (-x)f(x),\ f''(x) = (x^2-1)f(x),\ f'''(x) = (3x - x^3)f(x).$$

Hence, with $x = \xi_p$,

$$g_n^* = \left\{\frac{\sqrt{2}\exp{(x^2/2)}}{n(x^2-1)^2}\right\}^{\frac{1}{5}},\ g_n^{**} = \left\{\frac{3\sqrt{2}\exp{(x^2/2)}}{4n(3x-x^3)^2}\right\}^{\frac{1}{7}},$$

$$AMSE\{\hat{Q}_n(p)\} = \frac{2\pi p(1-p)}{n}e^{x^2} - \frac{\sqrt{2}h}{n}e^{\frac{x^2}{2}} + \frac{h^4}{4}x^2,$$

$$a^* = \frac{1}{\sqrt{2\pi}}\left[\frac{e^{-2x^2}}{\sqrt{2}n(2x^2+1)^2}\right]^{\frac{1}{5}},\ a^{**} = \frac{1}{\sqrt{2\pi}}\left\{\frac{3e^{-3x^2}}{2\sqrt{2}n(6x^3+7x)^2}\right\}^{\frac{1}{7}},$$

$$AMSE\{\tilde{Q}_n(p)\} = \frac{2\pi p(1-p)}{n}e^{x^2} + \pi^2 h^4 x^2 e^{2x^2} - \frac{2\sqrt{\pi}h}{n}e^{x^2}.$$

Case 2. *f is the density of Exponential*(1). We have

$$f(x) = e^{-x} = -f'(x) = f''(x) = -f'''(x).$$

Hence

$$g_n^* = \left\{\frac{\exp(x)}{n\sqrt{\pi}}\right\}^{\frac{1}{5}},\ g_n^{**} = \left\{\frac{3\exp(x)}{4n\sqrt{\pi}}\right\}^{\frac{1}{7}},$$

$$AMSE\{\hat{Q}_n(p)\} = \frac{p(1-p)}{n}e^{2x} - \frac{h}{n\sqrt{\pi}}e^{x} + \frac{h^4}{4},$$

$$a^* = \left\{\frac{e^{-4x}}{8\sqrt{\pi}n}\right\}^{\frac{1}{5}},\ a^{**} = \left\{\frac{3e^{-6x}}{144\sqrt{\pi}n}\right\}^{\frac{1}{7}},$$

$$AMSE\{\tilde{Q}_n(p)\} = \frac{p(1-p)}{n}e^{2x} + \frac{h^4}{4}e^{4x} - \frac{h}{n\sqrt{\pi}}e^{2x}$$

Case 3. *f is the density of lognormal.* We have

$$f(x) = \frac{1}{\sqrt{2\pi}x}e^{-\frac{\log^2 x}{2}},$$

$$f'(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\log^2 x}{2}}\left(-\frac{1}{x^2} - \frac{1}{x^2}\log x\right) = -\frac{f(x)}{x}(1+\log x),$$

$$f''(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\log^2 x}{2}}\frac{1}{x^3}(1+3\log x+\log^2 x) = \frac{f(x)}{x^2}(1+3\log x+\log^2 x),$$

$$f'''(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{\log^2 x}{2}}\frac{1}{x^4}(-8\log x-6\log^2 x-\log^3 x) = -\frac{f(x)}{x^3}(8\log x+6\log^2 x+\log^3 x).$$

Hence

$$g_n^* = \left\{\frac{x^4}{n\sqrt{\pi}(1+3\log x+\log^2 x)^2 f(x)}\right\}^{\frac{1}{5}},\ g_n^{**} = \left\{\frac{3x^6}{4n\sqrt{\pi}(8\log x+6\log^2 x+\log^3 x)^2 f(x)}\right\}^{\frac{1}{7}},$$

$$AMSE\{\hat{Q}_n(p)\} = \frac{2\pi p(1-p)}{n}x^2e^{\log^2 x} - \frac{\sqrt{2}h}{n}xe^{\frac{\log^2 x}{2}} + \frac{h^4}{4}\frac{(1+\log x)^2}{x^2},$$

$$a^* = \frac{1}{\sqrt{2\pi}}\left\{\frac{e^{-2\log^2 x}}{n\sqrt{2}(2+3\log x+2\log^2 x)^2}\right\}^{\frac{1}{5}},$$

$$a^{**} = \frac{1}{\sqrt{2\pi}}\left\{\frac{3e^{-3\log^2 x}}{2n\sqrt{2}(5+13\log x+11\log^2 x+6\log^3 x)^2}\right\}^{\frac{1}{7}},$$

$$AMSE\{\tilde{Q}_n(p)\} = \pi^2h^4(1+\log x)^2e^{2\log^2 x} + 2\pi x^2e^{\log^2 x}\left\{p(1-p) - \frac{h}{\sqrt{\pi}}\right\}.$$

Case 4. *f is the density of double exponential.*

We have $f(x) = \frac{1}{2}e^{-|x|} = f''(x)$ except at $x = 0$ and

$$f'(x) = \begin{cases} -\frac{1}{2}e^{-x} & x > 0 \\ \frac{1}{2}e^{x} & x < 0 \end{cases} = -\frac{1}{2}sign(x)e^{-|x|} = f'''(x).$$

Hence

$$g_n^{**} = \left\{ \frac{2e^{|x|}}{n\sqrt{\pi}} \right\}^{\frac{1}{5}}, \; g_n^{**} = \left\{ \frac{3e^{|x|}}{2n\sqrt{\pi}} \right\}^{\frac{1}{7}}$$

$$AMSE\{\hat{Q}_n(p)\} = \frac{4p(1-p)}{n} e^{2|x|} - \frac{2h}{n\sqrt{\pi}} e^{|x|} + \frac{h^4}{4},$$

$$a^* = \left\{ \frac{e^{-4|x|}}{2^7 n\sqrt{\pi}} \right\}^{\frac{1}{5}}, \; a^{**} = \left\{ \frac{e^{-6|x|}}{2^{10} 3n\sqrt{\pi}} \right\}^{\frac{1}{7}},$$

$$AMSE\{\tilde{Q}_n(p)\} = 4h^4 e^{4|x|} + \frac{4p(1-p)}{n} e^{2|x|} - \frac{4h}{n\sqrt{\pi}} e^{2|x|}.$$

5.2 Simulation results

We sampled from the four distributions of size 50, 100, 500, and 1000, and computed the bandwidths and AMSE's at values of p from 0.05 to 0.95 with step size 0.05. However, by remark 2.1.1, we omitted $p = 0.5$ 0.5 for normal and double exponential distributions and $p = 0.35$ for lognormal. We repeated the computation for 100 times.

In the first several times of simulations, we obtained some extremely large or small bandwidths, which certainly resulted in extremely large asymptotic MSE. Hence we adopted the strategy in Sheather and Marron (1990) to adjust too small or large bandwidths. For example, in method 1, we forced $\hat{f}'(\xi_p)^{-2}$ to be in the interval [0.05, 1.5] as follows: if it is not in the interval, we replace it by the closest endpoint of the interval. Simulation results are displayed by figures. In the figures, plotted against p is the relative efficiency, i.e. the ratio of the AMSE of the different methods to the AMSE of the empirical quantile. Figures 1–4 summarize performance of different methods with the same sample size for the four distributions. Figures 5–8 show performance of one method with different sample sizes.

From Figures 1-4 we can see that the solid line, which corresponds to sample size $n = 50$, is almost the lowest in each plot. This is because when sample size is small, the empirical quantile has a relatively bigger MSE. Hence the kernel estimators are relatively more efficient.

Generally speaking, the four methods did a better job than empirical quantiles. For example, in Figure 6, we can see that when $n = 50$ only method 2 gave an efficiency more than 1 with p values between 0.75 and 0.95. Efficiency of all other methods are under 1 with all p values. But, unfortunately, no method works better than all the other methods for all distributions and all sample sizes. In Figure 8, for example, Method 2 sometimes works better than the others, but sometimes worse than the others. From this Figure it seems that Method 1 is always more efficient than Method 3. But if we look at Figure 6, Method 3 is more efficient than Method 1 for many p values in each sample size. We can also see from Figures 5-8 that plots of Method 1 (2) are similar to plots of Method 3 (4). This is not casual because we use the same formula to compute their asymptotic MSEs. From Figures 1–4, we observe that another common behavior for Method 2 and Method 4 is that they performance badly near the boundaries, i.e. when p is close to 0 or 1.

In a word, the kernel quantile estimators, wit no matter which bandwidth selection method, are more efficient than the empirical quantile estimator in most situations. And when sample size n is relatively small, say n=50, they are significantly more efficient than the empirical quantile estimator. But no one single method is most efficient in any situations.

References

[1] Azzalini, A. (1981). A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika* 68, 326-328.

[2] Faulk, M. (1984). Relative deficiency of kernel type estimators of quantiles, *Ann. Stat.*, 12, 261-268.

[3] Jones, M. C. (1992). Estimating densities, quantile, quantile densities and density quantiles. *Ann. Inst. Statist. Math.*, 44, 721-727

[4] Nadaraya, E. A. (1964). Some new estimates for distribution functions. *Theory Probab. Appl.*, 9, 497-500.

[5] Parzen, E. (1979). Nonparametric statistical data modeling. *J. Amer. Stat. Assoc.*, 74, 105-131.

[6] Read, R.R. (1972). The asymptotic inadmissibility of the sample distribution function. *Ann. Math. Statist.*, 43, 89-95.

[7] Ralescu, S. S. and Sun, S. (1993). Necessary and sufficient conditions for the asymptotic normality of perturbed sample quantiles. *J. Statist. Plann. Inference*, 35 55-64.

[8] Shankar, B. (1998). An optimal choice of bandwidth for perturbed sample quantiles, master thesis.

[9] Sheather, S. J. and Marron, J. S. (1990). Kernel quantile estimators. *J. Amer. Statist. Assoc.*, 85, 410-416.

[10] Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*. Chapman and Hall, London.

[11] Yamato, H. (1973). Uniform convergence of an estimator of a distribution function. *Bull. Math. Statist.*, 15, 69–78.

[12] Yang, S. S. (1985). A smooth nonparametric estimation of a quantile function. *J. Amer. Stat. Assoc.*, 80, 1004-1011.

Appendix

We now provide the proof for $AMSE$ in Remark 2.2. Here we follow the notation of Faulk (1984). Since $F^{-1\prime}(q) = Q'(q) = 0$, we have

$$
\begin{aligned}
Var\{\tilde{Q}_n(p)\} &= n^{-1}\int_0^1\{\int k(x)(q-\alpha_n x - 1_{(0,q-\alpha_n x)}(y))F^{-1\prime}(q-\alpha_n x)dx\}^2dy \\
&= n^{-1}\int_0^1\{\int k(x)(q-\alpha_n x - 1_{(0,q-\alpha_n x)}(y))[F^{-1\prime}(q)-\alpha_n xF^{-1\prime\prime}(q)+O(\alpha_n^2)]dx\}^2dy \\
&= n^{-1}\int_0^1\{\int k(x)(q-\alpha_n x - 1_{(0,q-\alpha_n x)}(y))(-\alpha_n xF^{-1\prime\prime}(q))dx\}^2dy + O(n^{-1}\alpha_n^2) \\
&= b\int_0^1\{\int k(x)(q-\alpha_n x - 1_{(0,q-\alpha_n x)}(y))xdx\}^2dy + O(n^{-1}\alpha_n^2) \\
&= b\int_0^1\{q\int xk(x)dx - \alpha_n\int x^2k(x)dx - \int xk(x)1_{(0,q-\alpha_n x)}(y))dx\}^2dy + O(n^{-1}\alpha_n^2) \\
&= b\int_0^1[\alpha_n\mu_2(k) + \int xk(x)1_{(0,q-\alpha_n x)}(y))dx]^2dy + O(n^{-1}\alpha_n^2) \\
&= b\int_0^1\{\alpha_n^2\mu_2^2(k) + 2\alpha_n\mu_2(k)\int xk(x)1_{(0,q-\alpha_n x)}(y))dx \\
&\quad +[\int xk(x)1_{(0,q-\alpha_n x)}(y))dx]^2\}dy + O(n^{-1}\alpha_n^2) \\
&= b\alpha_n^2\mu_2^2(k) + 2c\alpha_n\mu_2(k)\int_0^1\int xk(x)1_{(0,q-\alpha_n x)}(y))dxdy \\
&\quad +b\int_0^1[\int xk(x)1_{(0,q-\alpha_n x)}(y))dx]^2\}dy + O(n^{-1}\alpha_n^2) \\
&\triangleq b\alpha_n^2\mu_2^2(k) + 2b\alpha_n\mu_2(k)S_1 + bS_2 + O(n^{-1}\alpha_n^2)
\end{aligned}
$$

where $b = n^{-1}\alpha_n^2 F^{-1\prime\prime}(q)^2$. But

$$\begin{aligned} S_1 &= \textstyle\int_0^1 \int xk(x)1_{(0,q-\alpha_n x)}(y))dxdy \\ &= \textstyle\int xk(x)\int_0^1 1_{(0,q-\alpha_n x)}(y))dydx \\ &= \textstyle\int xk(x)(q-\alpha_n x)dx \\ &= \textstyle q\int xk(x)dx - \alpha_n \int x^2 k(x)dx \\ &= -\alpha_n \mu_2(k) \end{aligned}$$

and

$$\begin{aligned} S_2 &= \textstyle\int_0^1 [\int xk(x)1_{(0,q-\alpha_n x)}(y))dx]^2\}dy \\ &= \textstyle\int_0^1 [\int_{\frac{q-1}{\alpha_n}}^{\frac{q-y}{\alpha_n}} xk(x)dx]^2 dy \\ &= \textstyle\{y[\int_{\frac{q-1}{\alpha_n}}^{\frac{q-y}{\alpha_n}} xk(x)dx]^2\}|_0^1 - \int_0^1 yd\{[\int_{\frac{q-1}{\alpha_n}}^{\frac{q-y}{\alpha_n}} xk(x)dx]^2\} \\ &= \textstyle -2\int_0^1 \{y[\int_{\frac{q-1}{\alpha_n}}^{\frac{q-y}{\alpha_n}} xk(x)dx]\frac{q-y}{\alpha_n}k(\frac{q-y}{\alpha_n})(-\frac{1}{\alpha_n})\}dy \\ &= \textstyle \frac{2}{\alpha_n}\int_0^1 \{y\frac{q-y}{\alpha_n}k(\frac{q-y}{\alpha_n})[\int_{\frac{q-1}{\alpha_n}}^{\frac{q-y}{\alpha_n}} xk(x)dx]\}dy \\ &= \textstyle \frac{2}{\alpha_n}\int_{\frac{q}{\alpha_n}}^{\frac{q-1}{\alpha_n}} \{(q-\alpha_n t)tk(t)[\int_{\frac{q-1}{\alpha_n}}^{t} xk(x)dx]\}d(-\alpha_n t) \\ &= \textstyle 2\int_{\frac{q-1}{\alpha_n}}^{\frac{q}{\alpha_n}} \{(q-\alpha_n t)tk(t)[\int_{\frac{q-1}{\alpha_n}}^{t} xk(x)dx]\}dt \\ &= \textstyle 2\int_{\frac{q-1}{\alpha_n}}^{\frac{q}{\alpha_n}} (q-\alpha_n t)tk(t)j(t)dt \end{aligned}$$

where $j(t) \triangleq \int_{-c}^{t} xk(x)dx$ and c is such that k is finitely supported in $[-c, c]$. Then

$$Var\{\tilde{Q}_n(p)\} = -n^{-1}\alpha_n^2 F^{-1\prime\prime}(q)^2\alpha_n^2\mu_2^2(k) + 2n^{-1}\alpha_n^2 F^{-1\prime\prime}(q)^2 \int (q-\alpha_n t)tk(t)j(t)dt + O(n^{-1}\alpha_n^2).$$

If we replace α_n by h and $F^{-1\prime\prime}(q)$ by $Q''(q)$, then

$$Var\{\tilde{Q}_n(p)\} = -n^{-1}h^4 Q''(q)^2\mu_2^2(k) + 2n^{-1}h^2 Q''(q)^2 \int (q-ht)tk(t)j(t)dt + O(n^{-1}h^2).$$

But the bias of $\tilde{Q}_n(p)$ is

$$bias = \frac{1}{2}h^2\mu_2(k)Q''(q) + O(h^2) + O(n^{-1}).$$

Hence the MSE of $\tilde{Q}_n(p)$ is

$$\begin{aligned} MSE\{\tilde{Q}_n(p)\} &= \tfrac{h^4}{4}\mu_2^2(k)Q''(q)^2 + O(h^4) + O(n^{-1}h^2) - n^{-1}h^4Q''(q)^2\mu_2^2(k) \\ &\quad +2n^{-1}h^2Q''(q)^2\textstyle\int(q-ht)tk(t)j(t)dt + O(n^{-1}h^2). \end{aligned}$$

That is

$$AMSE\{\tilde{Q}_n(p)\} = (\frac{1}{4} - \frac{1}{n})h^4Q''(q)^2\mu_2^2(k) + 2n^{-1}h^2Q''(q)^2\int(q-ht)tk(t)j(t)dt.$$

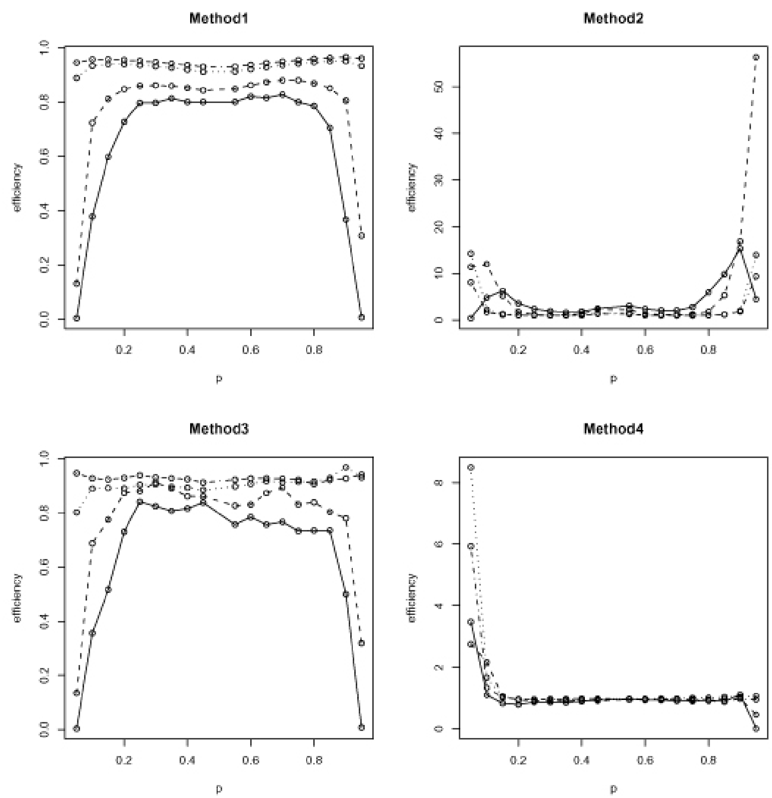

Figure 1: **Efficiency under double exponential.** Different panels correspond to different methods.

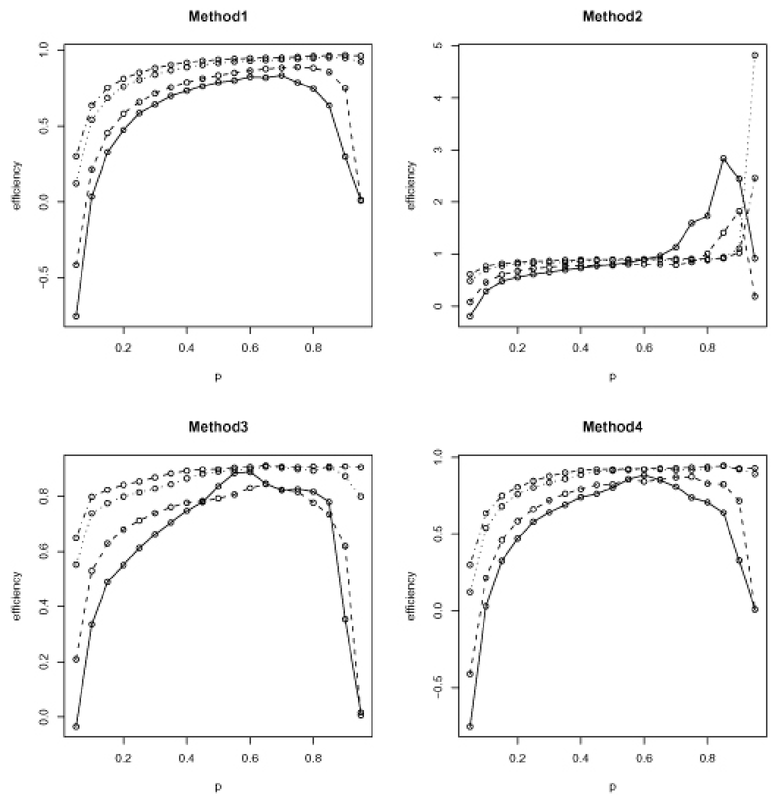

Figure 2: **Efficiency under exponential.** Different panels correspond to different methods.

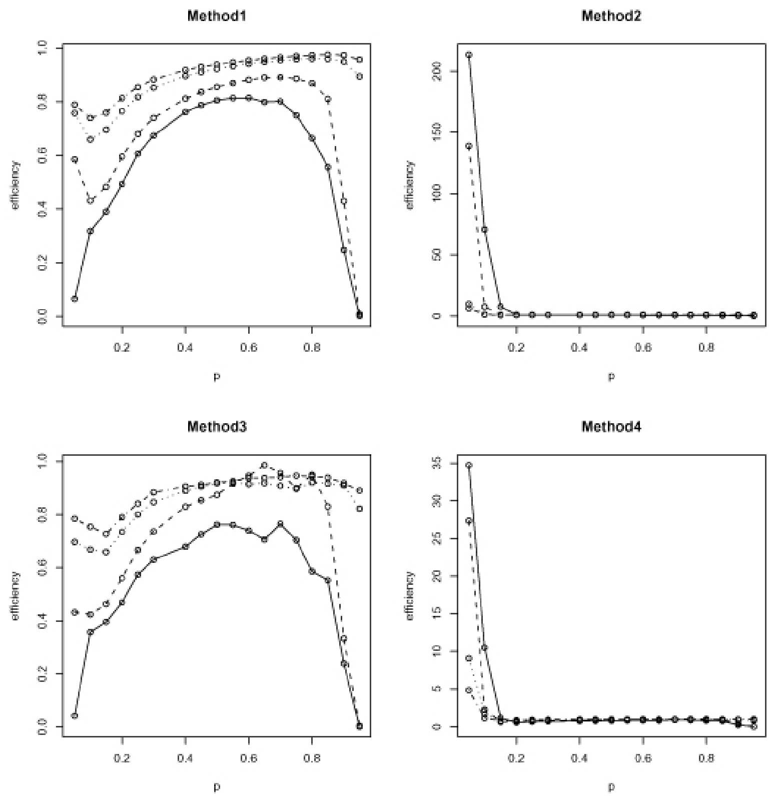

Figure 3: **Efficiency under Log Normal.** Different panels correspond to different methods.

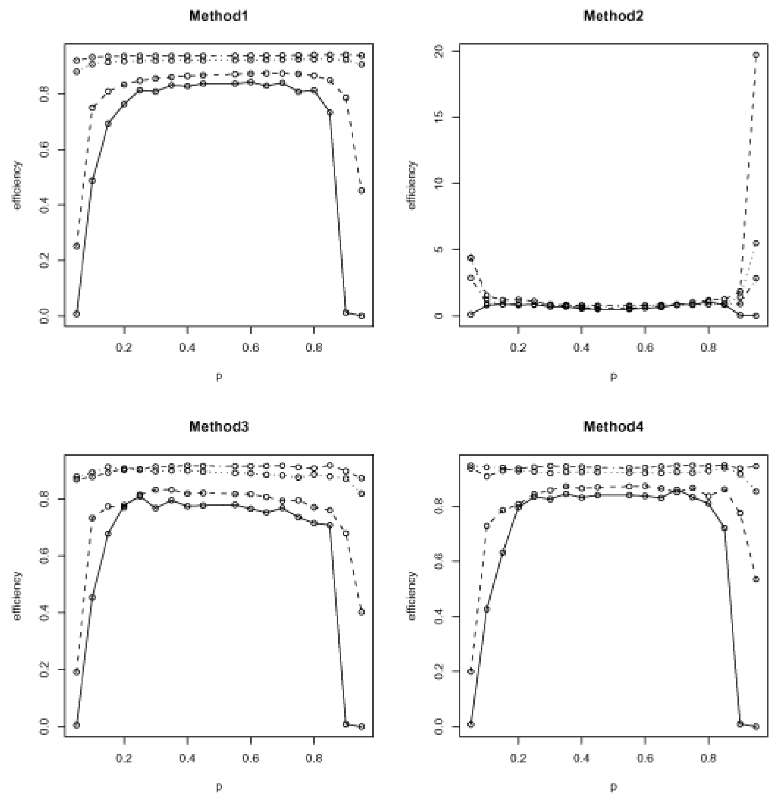

Figure 4: **Efficiency under standard Normal.** Different panels correspond to different methods.

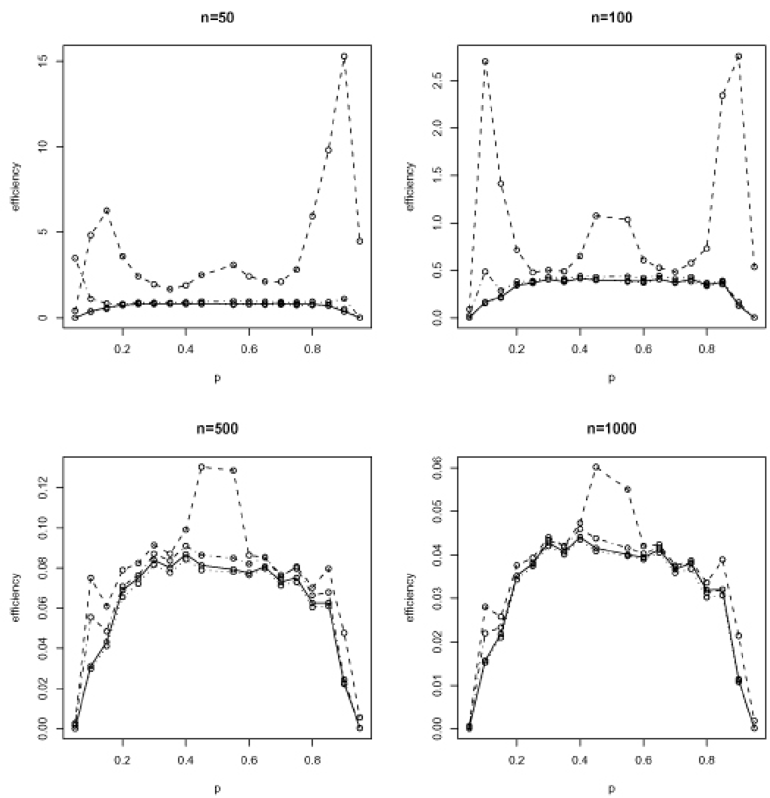

Figure 5: **Efficiency under double exponential.** Different panels correspond to different sample sizes.

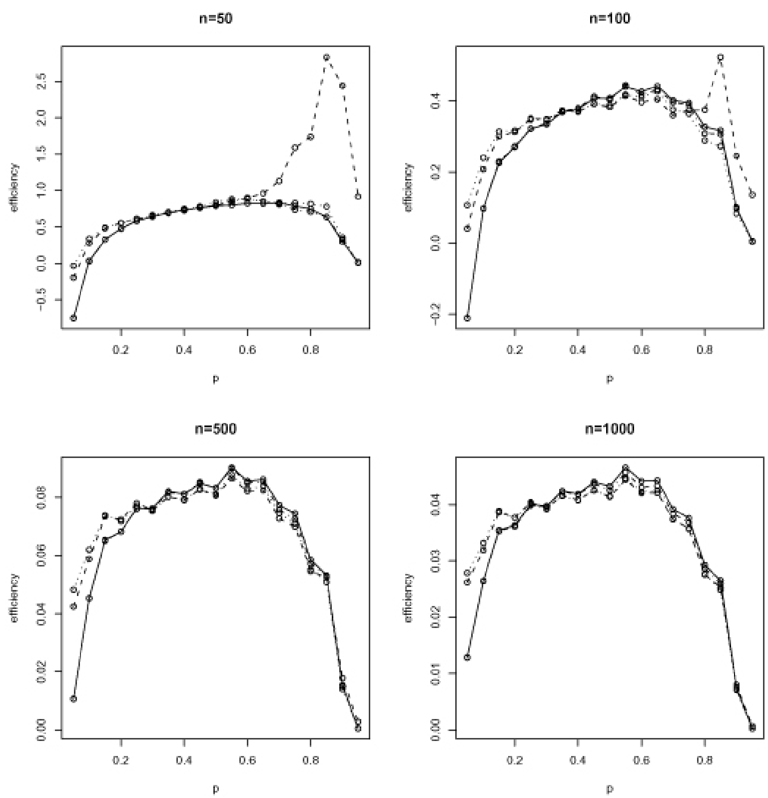

Figure 6: **Efficiency under exponential.** Different panels correspond to different sample sizes.

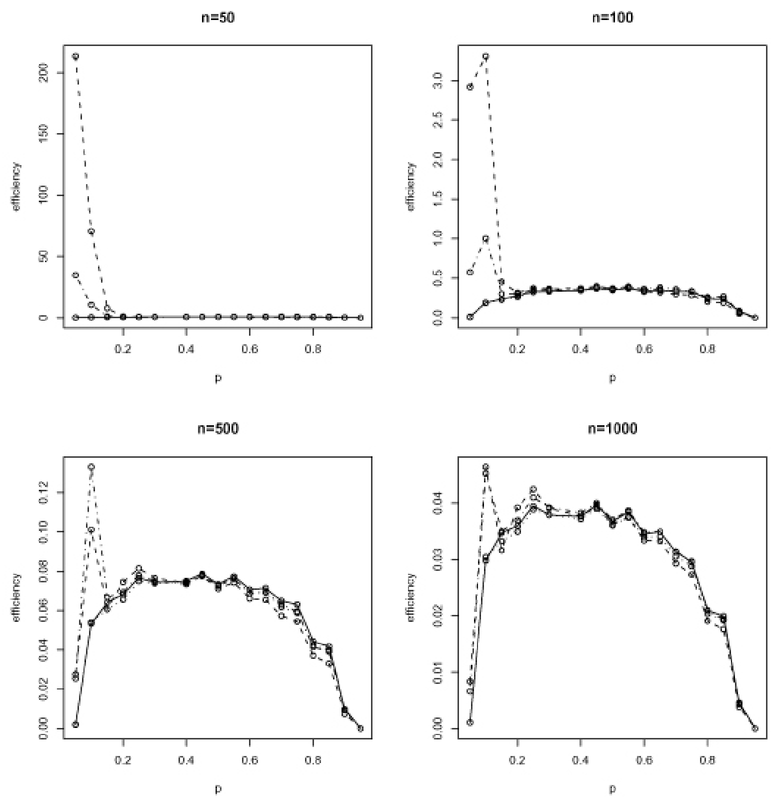

Figure 7: **Efficiency under Log Normal.** Different panels correspond to different sample sizes.

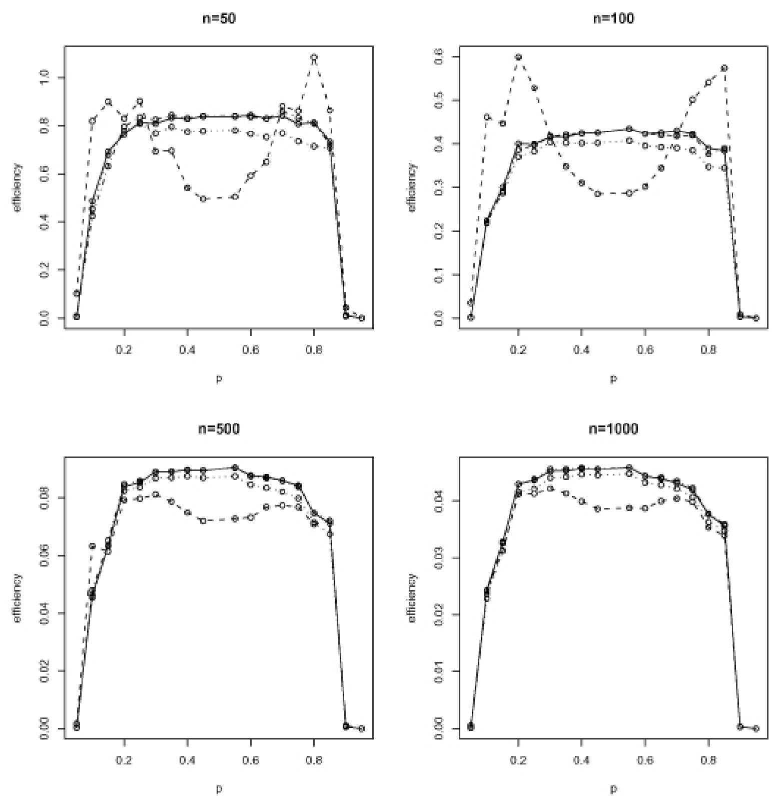

Figure 8: **Efficiency under standard Normal.** Different panels correspond to different sample sizes.