

UCLA

Department of Statistics Papers

Title

Resampling Methods for the Extrema of Certain Conditional Sample Functions

Permalink

<https://escholarship.org/uc/item/5np8s1q7>

Author

Zhenwei Zhou

Publication Date

2011-10-25

RESAMPLING METHODS FOR THE EXTREMA OF CERTAIN CONDITIONAL SAMPLE FUNCTIONS

ZHENWEI ZHOU

ABSTRACT. In this paper some resampling methods for the extrema of certain conditional functions, including delta method along with jackknifing and bootstrapping, are examined, and the properties of the resulted statistics are discussed.

1. INTRODUCTION

Let's motivate our study from the following model. Consider a bundle of n parallel filaments of equal length whose individual strengths are denoted by X_1, \dots, X_n , where X_1, \dots, X_n are assumed to be i.i.d. random variables. If we assume that the force of a free load on the bundle is distributed equally on each filament and the strength of an individual filament is independent of the number of filaments in a bundle, then the minimum load B_n beyond which all the filaments of the bundle give away is defined to be the *strength of the bundle*.

Let $X_{n1} \leq \dots \leq X_{nn}$ be the n ordered values of X_1, \dots, X_n , if a bundle breaks under a load L , then the inequalities

$$nX_{n1} \leq L, (n-1)X_{n2} \leq L, \dots, X_{nn} \leq L$$

are simultaneously satisfied. Consequently, the bundle strength can be represented as:

$$B_n = \max\{nX_{n1}, (n-1)X_{n2}, \dots, X_{nn}\}.$$

Note that if $F_n(x)$ is the empirical distribution function for X_1, \dots, X_n , then

$$B_n^* \equiv n^{-1}B_n = \sup_{x \geq 0} x[1 - F_n(x)].$$

Daniels (1945) investigated the probability distribution of B_n^* (B_n) and established its asymptotic normality. Later on, Sen et al. (1973), Phoenix and Taylor (1973), Smith et al.(1983), Harlow and Phoenix (1981), Daniels (1985, 1989), and others studied some extensions of Daniels' model.

1991 *Mathematics Subject Classification.* 62E20, 62E99.

Key words and phrases. bundle strength of filaments, normality, bias, asymptotic variance, jackknifing, bootstrapping, delta method, extrema.

Since the strength of a bundle of fibres depends on the materials, the individual length, diameter and so on, it is natural to take these kind of variations into account. Specifically, let $(X_1, \mathbf{Y}_1), \dots, (X_n, \mathbf{Y}_n)$ be iid as $(X, \mathbf{Y}) \in \mathcal{R}^{q+1}$, where $X \in \mathcal{R}^+$, $\mathbf{Y} \in \mathcal{R}^q$, and X_i ($i = 1, \dots, n$) represents the strength of the i -th fibre, \mathbf{Y}_i represents the corresponding individual variations, and $G(x|\mathbf{y})$ is the conditional distribution of X given $\mathbf{Y} = \mathbf{y}$. Instead of the statistics $\sup_x \Psi(x, F(x))$ considered as in Daniels model and Sen et al. (1973), we are interested in the functionals of the form

$$\theta(\mathbf{y}) = \sup_{x \in \mathcal{R}^+} \Psi(x, G(x|\mathbf{y})), \quad \mathbf{y} \in \mathcal{C} \text{ (compact)} \subset \mathcal{R}^q.$$

Remark:

when $\Psi(x, z) = x(1 - z)$, $0 \leq z \leq 1$ and $G(x|y) = G(x)$, then $\theta(\mathbf{y}) = \sup_x \{x(1 - G(x))\}$, which is the classical case.

Obviously $\theta(\mathbf{y})$ has a natural estimator

$$\hat{\theta}_n(\mathbf{y}) = \sup_{x \geq 0} \Psi(x, G_n(x|\mathbf{y})), \quad \mathbf{y} \in \mathcal{C},$$

where $G_n(x|\mathbf{y})$ is a estimator of conditional distribution $G(x|\mathbf{y})$ defined as the following. Suppose $\{(X_i, \mathbf{Y}_i), i \geq 1\}$ iid. as (X, \mathbf{Y}) . Given $\mathbf{Y} = \mathbf{y}$, we set $Z_i = \|\mathbf{Y}_i - \mathbf{y}\|$, $i = 1, \dots, n$, where $\|\cdot\|$ is a norm on \mathcal{R}^q (e.g., Euclidean). Let $0 \leq Z_{n1} \leq Z_{n2} \leq \dots \leq Z_{nn}$ be the order statistics corresponding to Z_1, Z_2, \dots, Z_n , and let $X_{n1}, X_{n2}, \dots, X_{nn}$ be the *induced order statistics*, i.e. $X_{ni} = X_j$ if $Z_{ni} = Z_j$ for $i, j = 1, \dots, n$. For every positive integer $k(\leq n)$, the k -NN empirical distribution estimator of $G(x|\mathbf{Y} = \mathbf{y})$ is defined by

$$G_{nk_n}(x|\mathbf{y}) = k_n^{-1} \sum_{i=1}^{k_n} I(X_{ni} \leq x), \quad x \in \mathcal{R}^1$$

where $I(A)$ stands for the indicator function of the set A . Under certain conditions, Zhou(1994) Obtained that

$$k_n^{\frac{1}{2}}[\theta_n(y) - \theta(y)] = \Psi_{01}(x^0, G(x^0|y))k_n^{\frac{1}{2}}[G_{nk_n}(x^0|y) - G(x^0|y)] + o_p(1) \quad (1.1)$$

where $k_n = [n^{4/5}]$, $\theta(y) = \sup_x \Psi(x, G(x|y))$, $\theta_n(y) = \max_i \Psi(X_{ni}, G_{nk_n}(X_{ni}|y))$, $\Psi_{01}(\cdot, \cdot)$ is the partial derivative of $\Psi(x, y)$ w.r.t. y , and x^0 is the point at which $\sup_x \Psi(x, G(x|y))$ attains it unique maximum. From (1.1) we have

$$k_n^{\frac{1}{2}}[\theta_n(y) - \theta(y)] \rightarrow N(\mu, \gamma^2) \quad (1.2)$$

where

$$\mu(y) = \frac{1}{24} \{G_{yy}(x^0|y)f(y) + 2f'(y)G_y(x^0|y)\} / f^3(y) \quad (1.3)$$

$$\gamma^2(y) = \Psi_{01}^2(x^0, G(x^0|y))G(x^0|y)(1 - G(x^0|y)) \quad (1.4)$$

here $f(y)$ is the marginal distribution of Y , G_y G_{yy} are the first and second derivatives of $G(x|y)$ w.r.t. y . Therefore, one may conclude that the typical bias term for $\theta_n(y)$ is of the order $k_n^{-\frac{1}{2}}$.

In the following, we first discuss the bias reduction for $\theta_n(y)$, then, probe the bootstrap and jackknife estimator $\hat{\gamma}^2$ of γ^2 .

2. ORDER OF THE ASYMPTOTIC BIAS

Jackknifing serves a dual purpose of reducing the possible bias of a statistics and providing an estimator of its mean square error. For the classical jackknifing, it consists of identifying the n subsamples of size $n-1$ by deleting one observation at a time from the base sample, and incorporating them in the formulation of the pseudovariables. Using those pseudovariables, one can construct a jackknifed estimator (for the statistics) as well as an estimator of its asymptotic mean square error. In our case, since estimating $\theta(y)$ only involves k_n observations (in the way defined as before), we will take those k_n observations as the base sample (both for jackknifing and bootstrapping). As mentioned before, the order of the bias term of $\theta_n(y)$ is of $O(k_n^{-\frac{1}{2}})$, the classical jackknifing may not work in the sense that it does not reduce the order of the bias term. To see this, suppose

$$E(\theta_n(y)) = \theta(y) + a(G, y)k_n^{-\frac{1}{2}} + o(k_n^{-\frac{1}{2}}) \quad (2.1)$$

where $a(G, y)$ is a constant independent of k_n (depending on conditional distribution $G(x|y)$ and y). Then

$$E\{k_n\theta_n(y) - (k_n - 1)\theta_{n-1}(y)\} = \theta + \frac{1}{2}a(G, y)k_n^{-\frac{1}{2}} + o(k_n^{-\frac{1}{2}}),$$

the order of the bias term remains the same, though its contribution is discounted by the factor $\frac{1}{2}$. Therefore, there may not be enough incentive in using the classical jackknife for bias reduction, although its utility in estimating the asymptotic variance may still remain intact. We'll discuss it more later on.

In order to reduce the order of the bias term, one may need some alternative methods other than the classical jackknife. One of them coming into mind might be so-called delete- d jackknifing method, and this d usually depends on n (k_n) and is increasing in n . Suppose that each time d observations are deleted from the base sample (we will denote those d

observations by $\tilde{\sim}^i = (i_1, \dots, i_d)$ and denote the corresponding estimate as $\theta_{style_{\tilde{\sim}^i}}$. Let

$$\hat{\theta}_{(\cdot)} = \frac{1}{\binom{k_n}{d}} \sum_{\tilde{\sim}^i} \theta_{\tilde{\sim}^i}. \quad (2.2)$$

The jackknife estimator $\tilde{\theta}_n(y)$ is defined as

$$\tilde{\theta}_n = \frac{k_n^{\frac{1}{2}} \theta_n - (k_n - d)^{\frac{1}{2}} \hat{\theta}_{(\cdot)}}{k_n^{\frac{1}{2}} - (k_n - d)^{\frac{1}{2}}}. \quad (2.3)$$

If (2.1) still is in force, then

$$E(\tilde{\theta}_n) = \theta + o\left(\frac{1}{k_n^{\frac{1}{2}} - (k_n - d)^{\frac{1}{2}}}\right) \quad (2.4)$$

Note that

$$k_n^{\frac{1}{2}} - (k_n - d)^{\frac{1}{2}} = k_n^{\frac{1}{2}} [1 - (1 - \frac{d}{k_n})^{\frac{1}{2}}] \simeq \frac{1}{2} k_n^{-\frac{1}{2}} d \quad (2.5)$$

Hence, in order to ensure that the bias term in (2.4) has the order of $o(k_n^{-\frac{1}{2}})$ under (2.1), d must be so chosen that has the order of k_n . If there exists an $\epsilon_0 > 0$, such that

$$\frac{d}{k_n} \geq \epsilon_0, \quad (2.6)$$

then, the jackknife estimator $\tilde{\theta}_n$ defined in (2.3) does have a bias term of higher order ($o(k_n^{-\frac{1}{2}})$) compared to original of $O(k_n^{-\frac{1}{2}})$, and looks like the bias being reduced. Unfortunately, in our setting, (2.1) is not true. This is due to the asymmetric property of $\theta_n(y, x_1, \dots, x_{k_n})$ w.r.t. x_1, \dots, x_{k_n} . To see this, from (1.1), we have

$$\theta_n = \theta + \Psi_{01}(x^0, G(x^0|y)) \frac{1}{k_n} \sum_i (I_{\{X_{ni} \leq x^0\}} - G(x^0|y)) + o_p(k_n^{-1/2}). \quad (2.7)$$

From this and the contiguity of measure q_{nk} to p_{nk} , we always have

$$\theta_{\tilde{\sim}^i} = \theta + \Psi_{01}(x^0, G(x^0|y)) \frac{1}{k_n} \sum_{i \text{ not in } \tilde{\sim}^i} (I_{\{X_{ni} \leq x^0\}} - G(x^0|y)) + o_p(k_n^{-1/2}). \quad (2.8)$$

Hence

$$\begin{aligned}
\hat{\theta}_{(\cdot)} &= \frac{1}{\binom{k_n}{d}} \sum_{\substack{-i \\ \sim i}} \theta_{-i} \\
&= \theta + \frac{1}{\binom{k_n}{d}} \frac{\binom{k_n-1}{d}}{k_n-d} \Psi_{01}(x^0, G(x^0|y)) \sum_i (I_{\{X_{ni} \leq x^0\}} - G(x^0|y)) \\
&\quad + o_p((k_n-d)^{-1/2}) \\
&= \theta + \frac{1}{k_n} \sum_i (I_{\{X_{ni} \leq x^0\}} - G(x^0|y)) + o_p((k_n-d)^{-1/2}), \tag{2.9}
\end{aligned}$$

which has exactly the same leading bias term as in θ_n . Therefore, the jackknifing method may not work in our setting (no matter delete-1 or delete- d). We are going to propose another method to carry out the task of reducing the bias.

Let $k_{na} = [an^{4/5}]$, $k_{nb} = [bn^{4/5}]$, $0 < a < b < \infty$, and θ_{na} , θ_{nb} be the corresponding estimators obtained. From (2.7) (replacing k_n by k_{na}), we have

$$E(\theta_{na}) = \theta + \Psi_{01} \beta a^2 n^{-2/5} + o(n^{-2/5}). \tag{2.10}$$

where

$$\begin{aligned}
\Psi_{01} &= \Psi_{01}(x^0, G(x^0|y)) \\
\beta &= \frac{1}{24} \{G_{yy}(x^0|y)f(y) + 2f'(y)G_y(x^0|y)\} / f^3(y)
\end{aligned}$$

Define

$$\theta_{nab} = \frac{b^2 \theta_{na} - a^2 \theta_{nb}}{b^2 - a^2}, \tag{2.11}$$

then from (2.10)

$$E(\theta_{nab}) = \theta + o(n^{-2/5}). \tag{2.12}$$

Such a defined estimator really reduces the bias of the original estimator θ_n . Its bias is negligible compared with its variance. However, the possible penalty for doing this is that

θ_{nab} may have bigger variance compared with θ_{na} or θ_{nb} . Note that

$$\begin{aligned}
Var(\theta_{na}) &= \frac{\gamma^2}{k_{na}} + o(k_{na}^{-1}) \\
Cov(\theta_{na}, \theta_{nb}) &= E\{[\theta_{na} - E(\theta_{na})][\frac{k_{na}}{k_{nb}}(\theta_{na} - E(\theta_{na})) \\
&\quad + \frac{1}{k_{nb}}(\sum_{i=k_{na}+1}^{k_{nb}} (I_{\{X_{ni} \leq x^0\}} - G) - E \sum_{i=k_{na}+1}^{k_{nb}} (I_{\{X_{ni} \leq x^0\}} - G))]\} + o(n^{-4/5}) \\
&= \frac{k_{na}}{k_{nb}} Var(\theta_{na}) + o(n^{-4/5}) \\
&= \frac{\gamma^2}{k_{nb}} + o(n^{-4/5}). \tag{2.13}
\end{aligned}$$

Hence,

$$\begin{aligned}
Var(\theta_{nab}) &= \frac{1}{(b^2 - a^2)^2} [b^4 Var(\theta_{na}) + a^4 Var(\theta_{nb}) - 2a^2 b^2 Cov(\theta_{na}, \theta_{nb})] \\
&= \frac{\gamma^2}{(b^2 - a^2)^2} [\frac{b^4}{a} + \frac{a^4}{b} - 2a^2 b] n^{-4/5} + o(n^{-4/5}). \tag{2.14}
\end{aligned}$$

Simple algebra shows that $Var(\theta_{nab}) > Var(\theta_{na}) (Var(\theta_{nb}))$. But when the bias is took into account, in the sense of mean square error (mse), the estimator θ_{nab} may beat θ_{na} (θ_{nb}). Note that the asymptotic mse for θ_{na} is

$$\begin{aligned}
mse(\theta_{na}) &\simeq bias^2 + Var(\theta_{na}) \\
&= [\Psi_{01}^2 \beta^2 a^4 + \frac{\Psi_{01}^2 G(1-G)}{a}] n^{-4/5}
\end{aligned}$$

where $G = G(x^0|y)$, which is minimized at

$$a_0 = (\frac{G(1-G)}{4\beta^2})^{1/5}. \tag{2.15}$$

At a_0 the asymptotic mse of θ_{na} reaches its minimum

$$mse^0(\theta) = 5 \times 4^{-4/5} \Psi_{01}^2 \beta^{2/5} (G(1-G))^{4/5} n^{-4/5} \tag{2.16}$$

For such a a_0 , let $b = ta_0$ ($t > 1$), then the asymptotic mse of θ_{nab} can be expressed as (keep in mind that the bias of θ_{nab} is negligible)

$$\begin{aligned}
mse(\theta_{nab}) &\simeq \frac{\gamma^2}{a_0^4 (t^2 - 1)^2} [a_0^3 t^4 + \frac{a_0^3}{t} - 2a_0^2 t] n^{-4/5} \\
&= \frac{\gamma^2}{a_0 (t^2 - 1)^2} [t^4 + \frac{1}{t} - 2t]. \tag{2.17}
\end{aligned}$$

The natural question is: is there t (or b) such that the asymptotic mse of θ_{nab} is less or equal to that of θ_{na_0} ? The answer is yes. To see this, let

$$\begin{aligned}
 \psi(t) &= mse(\theta_{na_0b}) - mse(\theta_{na_0}) \\
 &\simeq \left\{ \frac{\gamma^2}{a_0(t^2-1)^2} \left[t^4 + \frac{1}{t} - 2t \right] - \Psi_{01}^2 \beta^2 a_0^4 \right\} n^{-4/5} \\
 &= \frac{\gamma^2}{4a_0} \left\{ \frac{4}{(t^2-1)^2} \left[t^4 + \frac{1}{t} - 2t \right] - 5 \right\} n^{-4/5} \\
 &= \frac{\gamma^2}{4a_0} \left\{ -\frac{t-1}{t(t^2-1)^2} [t^4 + t^3 - 9t^2 - t + 4] \right\} n^{-4/5} \tag{2.18}
 \end{aligned}$$

Note that $\psi(1.1) > 0$, but $\psi(2) < 0$, it means that there exists t (b) such that $\psi(t) < 0$, namely, the asymptotic mse of θ_{na_0b} is less than that of θ_{na_0} . Thus, in the sense of mse, θ_{nab} is a better estimator than θ_{na} (θ_{nb}) (one may show that $\psi(t)$ is decreasing in t).

In passing, we may remark that if one tends to use the classical bootstrap to reduce the bias, the situation may be no better than using the classical jackknife since the bootstrap sampling allows a possible duplication of some of the observations in the base sample (with a positive probability). The impact of ties arising could make the bias term even worse.

3. RESAMPLING SCHEMES FOR VARIANCE ESTIMATION

We know that

$$k_n^{\frac{1}{2}} [\theta_n - \theta] \rightarrow N(\mu, \gamma^2)$$

where μ and γ are defined as in (1.3) and (1.4) respectively. Here, the γ^2 is to be estimated. If one tends to use θ_{nab} , as we see in the previous section, it still requires the estimation of γ^2 . In the following, we discuss a few methods leading to this estimation.

A natural and naive estimator of γ^2 is

$$\hat{\gamma}_n^2 = \hat{\xi}_n^2 \hat{\pi}_{0n} (1 - \hat{\pi}_{0n}), \tag{3.1}$$

where $\hat{\pi}_{0n} = G_{nk_n}(X_{(nr)}|y)$, $\hat{\xi}_n^2 = \Psi_{01}^2(X_{(nr)}, G_{nk_n}(X_{(nr)}|y))$, and $X_{(nr)}$ is the point at which θ_n attains its maximum, i.e. $\theta_n = \max_i \Psi(X_{ni}, G(X_{ni}|y)) = \Psi(X_{(nr)}, G_{nk_n}(X_{(nr)}|y))$. This estimator is based on the classical delta method, and it is consistent under the conditions of theorem 3 (?). However, as one may notice that θ_n is highly nonlinear, the estimator $\hat{\gamma}_n$ may inherit the sensitivity of the delta method to basic nonlinearity of the functional. Hence, it may lead to significant bias. For this reason, we like to explore some alternative

methods such as jackknifing and bootstrapping in estimating $\hat{\gamma}_n$.

Delete-d jackknife method. As one has noticed, the usual deleting one jackknife method does not work well in reducing the bias in our setting. As such a delete- d jackknife method is incorporated for the bias reducing purpose. Based on that, an estimator for γ^2 is proposed as the following:

$$\hat{\gamma}_{J(d)}^2 = \frac{k_n(k_n - d)}{dN} \sum_{\sim i} (\theta_{\sim i} - \theta_n)^2 \quad (3.2)$$

where $N = \binom{k_n}{d}$ and $\sum_{\sim i}$ is the summation over all possible choice of (i_1, \dots, i_d) . Note that from theorem () we have

$$\theta_n = \theta + \frac{1}{k_n} \sum_i \phi(X_{ni}) + R_{k_n} \quad (3.3)$$

where

$$\begin{aligned} \phi(X_{ni}) &= \Psi_{01}(X_{(nr)}, G(X_{(nr)}|y))(I_{\{X_{ni} \leq X_{(nr)}\}} - G(X_{(nr)}|y)) \\ R_{k_n} &= o_p(k_n^{-\frac{1}{2}}), \end{aligned} \quad (3.4)$$

here X_{ni} ($i = 1, \dots, k_n$) are the induced order statistics w.r.t. y . From Sen (1993a), we know that the joint measure q_{nk} of X_{ni} ($i = 1, \dots, k_n$) is contiguous to the joint measure p_{nk} of X_i ($i = 1, \dots, k_n$) which are i.i.d. with distribution function $G(x|y)$. Under p_{nk} , $\phi(X_i)$ has mean zero and positive variance γ^2 . From (1.2), we have

$$Var(\theta_n) = \frac{\gamma^2}{k_n} + o(k_n^{-1}). \quad (3.5)$$

From (3.4) and (3.5), by virtue of lemma 1 of Shao and Wu (1989), we may reach that under p_{nk} ,

$$E(R_{k_n})^2 = o(k_n^{-1}). \quad (3.6)$$

If d is so chosen that there exists $\epsilon_0 > 0$ such that

$$\frac{d}{k_n} \geq \epsilon_0, \quad (3.7)$$

then from (3.6) (3.7) and theorem 1 of Shao and Wu (1989), we may conclude under p_{nk}

$$\hat{\gamma}_{J(d)}^2 = \gamma^2 + o_p(1). \quad (3.8)$$

From this together with the contiguity of q_{nk} to p_{nk} we actually proved (under q_{nk})

$$\hat{\gamma}_{J(d)}^2 \rightarrow \gamma^2 \text{ in probability,}$$

namely, $\hat{\gamma}_{J(d)}^2$ is a (weakly) consistent estimator of γ^2 .

Delete-1 jackknife method. Although we proved the consistency of $\hat{\gamma}_{J(d)}^2$ for the delete-d (under (3.7)) jackknife. It is very computationaly extensive. As a matter of fact, in our setting the delete-1 jackknife estimator $\gamma_{J(1)}^2$ is consistent. Note that in this case $\gamma_{J(1)}^2$ has the form

$$\gamma_{J(1)}^2 = (k_n - 1) \sum_i (\theta_{-i} - \theta_n)^2 \quad (3.9)$$

where θ_{-i} ($i = 1, \dots, k_n$) are the estimators constructed each time by deleting observation X_{ni} from the base sample. Also, let $G_{n-1}^i(x|y)$ be the corresponding conditional distribution estimator constructed from this subsample, then

$$\max_i \sup_x |G_{n-1}^i(x|y) - G_{nk_n}(x|y)| = k_n^{-1} \quad (3.10)$$

Further, let X_{n-1}^i be the point at which θ_{-i} reaches maximum, i.e.

$$\theta_{-i} = \max_{j \neq i} \Psi(X_{nj}, G_{n-1}^i(X_{nj}|y)) = \Psi(X_{n-1}^i, G_{n-1}^i(X_{n-1}^i|y)),$$

and $\pi_{n-1}^i = G_{n-1}^i(X_{n-1}^i|y)$. Then

$$\max_i |X_{n-1}^i - x_0| \rightarrow 0 \text{ and } \max_i |\pi_{n-1}^i - \pi_n| \rightarrow 0 \text{ in probability} \quad (3.11)$$

Let's define

$$X_{n-1}^* : G_{nk_n}(X_{n-1}^*|y) = \max_i G_{nk_n}(X_{n-1}^i|y) \quad (3.12)$$

$$X_{n-1}^{**} : G_{nk_n}(X_{n-1}^{**}|y) = \min_i G_{nk_n}(X_{n-1}^i|y) \quad (3.13)$$

Since $X_{n-1}^i \rightarrow x^0$ uniformly in i , it means

$$G_{nk_n}(X_{n-1}^*|y) - G_{nk_n}(X_{n-1}^{**}|y) \rightarrow 0 \quad (3.14)$$

Note that

$$\begin{aligned} & \Psi(X_{(nr)}, G_{n-1}^i(X_{(nr)}|y)) - \Psi(X_{(nr)}, G_{nk_n}^i(X_{(nr)}|y)) \leq \theta_{-i} - \theta_n \\ & \leq \Psi(X_{n-1}^i, G_{n-1}^i(X_{n-1}^i|y)) - \Psi(X_{n-1}^i, G_{nk_n}^i(X_{n-1}^i|y)). \end{aligned} \quad (3.15)$$

On the other hand

$$\begin{aligned} & G_{n-1}^i(x|y) - G_{nk_n}(x|y) \\ & = (k_n - 1)^{-1} \{k_n G_{nk_n}(x|y) - I_{\{X_{ni} \leq x\}}\} - G_{nk_n}(x|y) \\ & = (k_n - 1)^{-1} \{G_{nk_n}(x|y) - I_{\{X_{ni} \leq x\}}\} \end{aligned} \quad (3.16)$$

By (3.11) and (3.16), the right hand side of (3.15) can be written as

$$\Psi_{01}(X_{(nr)}, G_{nk_n}(X_{(nr)}|y)[(k_n - 1)(G_{nk_n}(X_{n-1}^i|y) - I_{\{X_{ni} \leq X_{n-1}^i\}})] + o_p(k_n^{-1}) \quad (3.17)$$

But

$$G_{nk_n}(X_{n-1}^i|y) - I_{\{X_{ni} \leq X_{n-1}^i\}} = G_{nk_n}(X_{(nr)}|y) - I_{\{X_{ni} \leq X_{(nr)}\}} + R_{ni} \quad (3.18)$$

where

$$R_{ni} = [G_{nk_n}(X_{n-1}^i|y) - G_{nk_n}(X_{(nr)}|y)] - [I_{\{X_{ni} \leq X_{n-1}^i\}} - I_{\{X_{ni} \leq X_{(nr)}\}}]. \quad (3.19)$$

By (3.19) and (3.14), we have

$$\begin{aligned} & \frac{1}{k_n} \sum_i R_{ni}^2 \\ &= \frac{1}{k_n} \sum_i \{[G_{nk_n}(X_{n-1}^i|y) - G_{nk_n}(X_{(nr)}|y)] - [I_{\{X_{ni} \leq X_{n-1}^i\}} - I_{\{X_{ni} \leq X_{(nr)}\}}]\}^2 \\ &\leq 2 \frac{1}{k_n} \sum_i \{[G_{nk_n}(X_{n-1}^i|y) - G_{nk_n}(X_{(nr)}|y)]^2 + [I_{\{X_{ni} \leq X_{n-1}^i\}} - I_{\{X_{ni} \leq X_{(nr)}\}}]^2\} \\ &\leq 2 \frac{1}{k_n} \sum_i \{[G_{nk_n}(X_{n-1}^i|y) - G_{nk_n}(X_{(nr)}|y)]^2 \\ &\quad + 2[G_{nk_n}(X_{n-1}^*|y) + G_{nk_n}(X_{(nr)}|y) - 2G_{nk_n}(\min(X_{n-1}^{**}, X_{(nr)}))|y)]\}. \end{aligned} \quad (3.20)$$

Since $X_{n-1}^i \rightarrow x^0$ uniformly in i , and $X_{(nr)} \rightarrow x^0$ in probability, it follows that the first term in (3.20) is of $o_p(1)$. Likewise, X_{n-1}^* , X_{n-1}^{**} and $X_{(nr)}$ all tend to x^0 . It ensures the second term in (3.20) also is of $o_p(1)$, and it follows

$$\frac{1}{k_n} \sum_i R_{ni}^2 = o_p(1). \quad (3.21)$$

Therefore

$$\begin{aligned} & \gamma_{J(1)}^2 \quad (3.22) \\ &= (k_n - 1) \Psi_{01}^2(X_{(nr)}, G_{nk_n}(X_{(nr)}|y)) [(k_n - 1)^{-1} \sum_i (G_{nk_n}(X_{(nr)}|y) - I_{\{X_{ni} \leq X_{(nr)}\}})]^2 \\ &\quad + o_p(1) \\ &= (k_n - 1)^{-1} \Psi_{01}^2(X_{(nr)}, G_{nk_n}(X_{(nr)}|y)) [k_n G_{nk_n}(X_{(nr)}|y) - k_n G_{nk_n}^2(X_{(nr)}|y)] + o_p(1) \\ &= k_n (k_n - 1)^{-1} \Psi_{01}^2(X_{(nr)}, G_{nk_n}(X_{(nr)}|y)) G_{nk_n}(X_{(nr)}|y) (1 - G_{nk_n}(X_{(nr)}|y)) + o_p(1) \\ &= k_n (k_n - 1)^{-1} \hat{\gamma}_n^2 + o_p(1) \end{aligned} \quad (3.23)$$

where $\hat{\gamma}_n^2$ is defined in (3.1), which is a strong consistent estimator of γ^2 . Hence, $\gamma_{J(1)}^2$ is a (weakly) consistent estimator of γ^2 , i.e.

$$\gamma_{J(1)}^2 \rightarrow \gamma^2 \text{ a.s.} \quad (3.24)$$

We may remark that if the delete-1 jackknife γ^2 estimator is defined as

$$\begin{aligned} \tilde{\gamma}_{J(1)}^2 &= (k_n - 1)^{-1} \sum_i (\theta_{-i} - \tilde{\theta}_n)^2 \\ &= (k_n - 1) \sum_i (\theta_{-i} - \theta_{(\cdot)})^2, \end{aligned} \quad (3.25)$$

then $\tilde{\gamma}_{J(1)}^2$ still is also a consistent estimator of γ^2 . Where θ_{-i} ($i = 1, \dots, k_n$) are defined as before, $\theta_{(\cdot)} = \frac{1}{k_n} \sum_i \theta_{-i}$, and $\tilde{\theta}_n = k_n - (k_n - 1)\theta_{(\cdot)}$. To verify the asymptotic equivalence of $\tilde{\gamma}_{J(1)}^2$ in (3.25) to $\gamma_{J(1)}^2$ in (3.9), it is sufficient to prove that

$$|\theta_{(\cdot)} - \theta_n| = o_p(k_n^{-1}) \quad (3.26)$$

Toward this, note (3.15), from (3.16) the left side of (3.15) can be written as

$$\Psi_{01}(X_{(nr)}, G_{nk_n}(X_{(nr)}|y))\{(k_n - 1)[G_{nk_n}(X_{(nr)}|y) - I_{\{X_{ni} \leq X_{(nr)}\}}]\} + o_p(k_n^{-1}), \quad (3.27)$$

uniformly in $i = 1, \dots, k_n$. Similarly, the right side of (3.15) can be expressed as

$$\Psi_{01}(X_{n-1}^i, G_{n-1}^i(X_{n-1}^i|y))\{(k_n - 1)[G_{n-1}(X_{n-1}^i|y) - I_{\{X_{ni} \leq X_{n-1}^i\}}]\} + o_p(k_n^{-1}) \quad (3.28)$$

Since $X_{n-1}^i \rightarrow x^0$ and $G_{n-1}^i(x|y) \rightarrow G(x|y)$ uniformly in i , the first term in both (3.27) and (3.28) can be replaced by $\Psi(x^0, G(x^0|y))$, and the remainder terms can be absorbed in $o(k_n^{-1})$. Namely, (3.27) and (3.28) are equivalent, respectively, to

$$\Psi_{01}(x^0, G(x^0|y))\{(k_n - 1)[G_{nk_n}(X_{(nr)}|y) - I_{\{X_{ni} \leq X_{(nr)}\}}]\} + o_p(k_n^{-1}) \quad (3.29)$$

$$\Psi_{01}(x^0, G(x^0|y))\{(k_n - 1)[G_{n-1}(X_{n-1}^i|y) - I_{\{X_{ni} \leq X_{n-1}^i\}}]\} + o_p(k_n^{-1}). \quad (3.30)$$

From (3.15), (3.29) and the fact that $\frac{1}{k_n} \sum_i [G_{nk_n}(X_{(nr)}|y) - I_{\{X_{ni} \leq X_{(nr)}\}}] = 0$, we have $\frac{1}{k_n} \sum_i (\theta_{-i} - \theta_n) \geq o_p(k_n^{-1})$, namely

$$\theta_{(\cdot)} - \theta_n \geq o_p(k_n^{-1}) \quad (3.31)$$

To obtain the opposite inequality, Note the definitions of X_{n-1}^* and X_{n-1}^{**} in (3.12) and (3.13), we have

$$\begin{aligned} G_{nk_n}(X_{n-1}^{**}|y) - G_{nk_n}(X_{n-1}^*|y) &\leq \frac{1}{k_n} \sum_i [G_{n-1}(X_{n-1}^i|y) - I_{\{X_{ni} \leq X_{n-1}^i\}}] \\ &\leq G_{nk_n}(X_{n-1}^*|y) - G_{nk_n}(X_{n-1}^{**}|y). \end{aligned} \quad (3.32)$$

From (3.14), both sides of (3.32) tend to zero. This, together with (3.15) and (3.30) leads to $\frac{1}{k_n} \sum_i (\theta_{-i} - \theta_n) \leq o_p(k_n^{-1})$, i.e.

$$\theta_{(\cdot)} - \theta_n \leq o_p(k_n^{-1}) \quad (3.33)$$

Combining (3.31) and (3.33), we finish the proof of the remark.

Bootstrap estimate of γ^2 . Still we assume the base sample is $(X_{n1}, \dots, X_{nk_n})$, each time a sample of size k_n is drawn from this base sample. Let X_{ni}^* ($i = 1, \dots, k_n$) be k_n (conditionally) i.i.d. observations drawn from the base sample, $G_{nk_n}^*(x|y)$ be the corresponding empirical distribution, and define

$$\theta_n^* = \max_i \Psi(X_{ni}^*, G_{nk_n}^*(X_{ni}^*|y)). \quad (3.34)$$

We draw M such (conditionally) independent bootstrap samples from the base sample $(X_{n1}, \dots, X_{nk_n})$, and denote the corresponding estimators by $\theta_{n1}, \dots, \theta_{nM}$. Then, let

$$\hat{\gamma}_B^2 = k_n M^{-1} \sum_i (\theta_{ni}^* - \theta_n)^2 \quad (3.35)$$

be the bootstrap estimator of γ^2 of the asymptotic variance of $k_n^{\frac{1}{2}}(\theta_n - \theta)$.

Suppose

$$\theta_n^* = \max_i \Psi(X_{ni}^*, G_{nk_n}^*(X_{ni}^*|y)) = \Psi(\hat{X}_{0n}^*, G_{nk_n}^*(\hat{X}_{0n}^*|y))$$

Note that

$$\begin{aligned} & \Psi(X_{(nr)}, G_{nk_n}^*(X_{(nr)}|y)) - \Psi(X_{(nr)}, G_{nk_n}(X_{(nr)}|y)) \leq \theta_n^* - \theta_n \\ & \leq \Psi(\hat{X}_{0n}^*, G_{nk_n}^*(\hat{X}_{0n}^*|y)) - \Psi(\hat{X}_{0n}^*, G_{nk_n}(\hat{X}_{0n}^*|y)). \end{aligned} \quad (3.36)$$

Also

$$k_n^{\frac{1}{2}} \|G_{nk_n}^* - G_{nk_n}\| = O_p(1), \quad (3.37)$$

holds. To see this, by the Kolmogrove inequality, under p_{nk} we have

$$\begin{aligned} & P(\max k_n^{\frac{1}{2}} \left| \frac{1}{k_n} \sum (I_{\{X_{ni}^* \leq x\}} - I_{\{X_{ni} \leq x\}}) \right| > c) \\ & \leq \frac{1}{c^2 k_n} \sum (G_{nk_n}(x|y) + I_{\{X_{ni} \leq x\}} - 2G_{nk_n}(x|y)I_{\{X_{ni} \leq x\}}) \\ & = \frac{2}{c^2} G_{nk_n}(x|y)(1 - G_{nk_n}(x|y)) \\ & \rightarrow 0 \text{ as } c \rightarrow \infty. \end{aligned} \quad (3.38)$$

Hence, under q_{nk} , (3.37) is true. Further,

$$|X_{0n}^* - X_{(nr)}| \rightarrow 0 \text{ and } |G_{nk_n}^*(X_{0n}^*|y) - G_{nk_n}(X_{(nr)}|y)| \rightarrow 0 \text{ in probability} \quad (3.39)$$

From (3.36), (3.37) and (3.38), we have

$$\begin{aligned} & \Psi_{01}(X_{(nr)}, G_{nk_n}(X_{(nr)}|y))[G_{nk_n}^*(X_{(nr)}|y) - G_{nk_n}(X_{(nr)}|y)] + o_p(k_n^{-\frac{1}{2}}) \\ & \leq \theta_n^* - \theta \\ & \leq \Psi_{01}(\hat{X}_{0n}^*, G_{nk_n}(\hat{X}_{0n}^*|y))[G_{nk_n}^*(\hat{X}_{0n}^*|y) - G_{nk_n}(\hat{X}_{0n}^*|y) + o_p(k_n^{-\frac{1}{2}})]. \end{aligned} \quad (3.40)$$

Combining (3.39) and (3.40) we may conclude

$$\begin{aligned} & k_n^{\frac{1}{2}}(\theta_n^* - \theta_n) - \Psi_{01}(X_{(nr)}, G_{nk_n}(X_{(nr)}|y))k_n^{\frac{1}{2}}[G_{nk_n}^*(X_{(nr)}|y) - G_{nk_n}(X_{(nr)}|y)] \\ & \rightarrow 0 \text{ in probability.} \end{aligned} \quad (3.41)$$

Since $\theta_{n1}^*, \dots, \theta_n^*$ are i.i.d. (as θ_n), from (3.41) we may conclude

$$\hat{\gamma}_B^2 - \Psi_{01}(X_{(nr)}, G_{nk_n}(X_{(nr)}|y))k_n^{\frac{1}{2}}G_{nk_n}(X_{(nr)}|y)[1 - G_{nk_n}(X_{(nr)}|y)] \rightarrow 0 \text{ in probability,} \quad (3.42)$$

which finishes the proof of the consistency of $\hat{\gamma}_B^2$.

REFERENCES

- [1] Bhattacharya, P. K. (1974), Convergence of sample paths of normalized sums of induced order statistics. *Ann. Stat.*, **2**,1034-1039.
- [2] Bhattacharya, P. K. and Gangopadhyay, A. K. (1990), Kernel and nearest neighbor estimation of a conditional quantile. *Ann. Stat.*, **18**, 1400-1415.
- [3] Daniels, H.E. (1945) The statistical theory of the strength of bundles of threads. *Proc. Roy. Soc. London, Ser. A*, bf 183,405-435.
- [4] Efron, Bradley (1982), The jackknife, the Bootstrap and other resampling plans. *CBMS-NSF lecture note*, vol. 38.
- [5] Sen, P.K. (1981) Sequential nonparametrics. *Wiley*
- [6] Sen, P.K. (1993a) Regression quantiles in nonparametric regression. *Presented at the Third Eugene Lukacs Symposium at Bowling Green State Univ., Bowling Green, Ohio.*
- [7] Sen, P.K. (1993b) Resampling methods for the extrema of certain sample functions. (Unpublished manuscript)
- [8] Sen, P.K., Bhattacharyya, B.B. and Suh, M.W. (1973) Limiting behavior of the extremum of certain sample functions. *Ann. Prob.*, **1**,297-311.
- [9] Zhou, Z.W. (1994) Limiting behavior of the extrema of certain conditional sample functions. (Unpublished manuscript)

DEPARTMENT OF STATISTICS, UNIVERSITY OF CALIFORNIA AT LOS ANGELES, LOS ANGELES, LA 90095-1554