Bootstrap Estimation of Actual Significance Levels for Tests Based on Estimated Nuisance Parameters *

Qiwei Yao and Wenyang Zhang
Institute of Mathematics & Statistics
University of Kent at Canterbury
Canterbury, Kent CT2 7NF, UK

Howell Tong
Department of Statistics & Actuarial Science
University of Hong Kong
Pokfulam, Hong Kong

Abstract

Often for a non-regular parametric hypothesis, a tractable test statistic involves a nuisance parameter. A common practice is to replace the unknown nuisance parameter by its estimator. The validality of such a replacement can only be justified for an infinite sample in the sense that under appropriate conditions the *asymptotic* distribution of the statistic under the null hypothesis is unchanged when the nuisance parameter is replaced by its estimator (Crowder, 1998). We propose a bootstrap method to calibrate the error incurred in the significance level, for *finite* samples, due to the replacement. Further, we have proved that the bootstrap method provides the more accurate estimator for the unknown actual significance level than the nominal level. Simulations demonstrate the proposed methodology.

KEY WORDS: Bootstrap estimation, extreme value, hypothesis test, non-regular parametric family, nuisance parameter, significance level.

*This research was partially supported by grants from the BBSRC/EPSRC, UK, the RGC Hong Kong and the University of Hong Kong.

1 Introduction

For some hypothesis testing problems, relevant statistics with tractable sampling distributions often depend on some nuisance parameters. It is a common practice to replace the unknown nuisance parameters with their estimators; such a practice has been *asymptotically* justified by Randles (1982) and Pierce (1982) for regular parametric tests and by Crowder (1998) for some non-regular cases. However, to the best of our knowledge, little attention has been paid to the issue of quantifying the deviation from the nominal significance level due to the use of an estimated nuisance parameter based on a *finite* sample, which is practically the more relevant case. In this paper, we propose a bootstrap method to estimate the *actual* significance levels of tests involving estimated nuisance parameters. The finding is alarming; the actual significant level could be quite different from the nominal level. We have proved that the new estimator for the actual significance level is of higher order accuracy than the nominal level. We have conducted two simulation studies, in which the test statistics are functions of extreme sample values. Results of the studies lend further support to the above claim.

Our approach is related to (though not the same as) the bootstrap calibration method first proposed by Loh (1987) and subsequently refined in Loh (1988, 1991). Loh's idea is to improve a confidence set, based on an asymptotic approximation, by correcting its nominal level following a bootstrap argument. Its validality is proved by appealing to Edgeworth expansions. The same idea is readily applied to hypothesis tests. We argue that it is particularly pertinent to adopt bootstrap approach in the context of tests based on estimated nuisance parameters. However, we are unable to provide an analytic formula to correct the nominal level since we deal with non-regular cases in which Edgeworth expansions are not available.

Although our primary concern is on non-regular cases such as those studied by Crowder (1998), the methodology is directly applicable to regular cases, for which Theorem 1 in Section 2 below can be easily derived from readily available Edgeworth expansions (Hall, 1992). Since the test statistics concerned in this paper are not always asymptotically normal, our proof of Theorem 1 is to establish an analogue of Edgeworth expansion (up to the first two terms) for non-regular statistics. De Haan and Resnick (1996) derived such an analogue for extreme sample values in a more systematic way, which, unfortunately, does not apply to the more complex cases covered in this paper.

The paper is organised as follows. We introduce the new method and the main result in Section 2. Simulation results are reported in Section 3. All the technical arguments are relegated to the Appendix.

2 Methodology and main results

Let $Y_1, \cdots, Y_n$ be independent observations from a population which belongs to the parametric family $P(\cdot|\theta, \psi)$, where $(\theta, \psi) \in R^p \times R^q$ are parameters. We assume that P is known up to some unknown parameters and are interested in testing the hypothesis on parameter θ only, namely

$$H_0: \ \theta = \theta_0 \ \text{ vs } \ H_1: \ \theta \in \Theta, \tag{2.1}$$

where $\theta_0 \in \Theta$, Θ is a subsets of R^p, and ψ plays the role of a nuisance parameter.

Suppose that we have a test statistic

$$T_n \equiv T_n(\psi) \equiv T_n(Y_1, \cdots, Y_n; \psi), \tag{2.2}$$

and its asymptotic distribution under the null hypothesis H_0, denoted by F, is known. We reject H_0 if $T_n(\psi) > F^{-1}(\alpha)$, where $F^{-1}(\alpha)$ denotes the upper-α point of distribution F, and $\alpha \in (0, 1)$ is the nominal significance level of the test. In practice, the nuisance parameter ψ is unknown and is estimated by an appropriate estimator

$$\hat{\psi} \equiv \hat{\psi}(Y_1, \cdots, Y_n). \tag{2.3}$$

Typically, the distribution of $T_n(\hat{\psi})$ is difficult to derive even asymptotically. A common practice is to reject H_0 when $T_n(\hat{\psi}) > F^{-1}(\alpha)$. Therefore the significance level of this modified test is

$$\alpha_n \equiv P\{ \ T_n(\hat{\psi}) > F^{-1}(\alpha) \mid \theta_0, \psi_0 \}, \tag{2.4}$$

which is unknown and which could be quite different from the nominal level α; see the examples in Section 3. In the above expression, ψ_0 denotes the (unknown) true value of ψ. Crowder (1998) proved that $|\alpha_n - \alpha| \to 0$ as $n \to \infty$ if $T_n(\cdot)$ is a reasonably smooth function and $\hat{\psi} \xrightarrow{P} \psi_0$ sufficiently fast.

To calibrate the discrepancy between α_n and α, we adopt a bootstrap scheme: draw independent samples $Y_1^*, \cdots, Y_n^*$ from distribution $P(\cdot|\theta_0, \hat{\psi})$, let $\hat{\psi}^* = \hat{\psi}(Y_1^*, \cdots, Y_n^*)$ (see (2.3) above), and define

$$\hat{\alpha}_n = P^*\{ \ T_n(Y_1^*, \cdots, Y_n^*; \hat{\psi}^*) > \hat{F}^{-1}(\alpha)\}, \tag{2.5}$$

where P^* denotes the conditional probability measure of $\{Y_1^*, \cdots, Y_n^*\}$ given the sample $\{Y_1, \cdots, Y_n\}$, $\hat{F}^{-1}(\alpha)$ is the upper-α point of distribution function of $T_n(Y_1^*, \cdots, Y_n^*; \hat{\psi})$ under P^*, and T_n is given as in (2.2). Theorem 1 below ensures that $\hat{\alpha}_n$ is a better estimator for α_n than α. This means that we should regard the test as one with the significance level $\hat{\alpha}_n$ although it was originally constructed at the nominal level α. In practice, we estimate $\hat{\alpha}_n$ by the relative frequency of the event $\{T_n(Y_1^*, \cdots, Y_n^*; \hat{\psi}^*) > \hat{F}^{-1}(\alpha)\}$ in a repeated bootstrap sampling.

Theorem 1. Under conditions (C1) — (C3) listed in the Appendix, $|\hat{\alpha}_n - \alpha_n| = o_p(|\alpha - \alpha_n|)$ as $n \to \infty$.

3 Simulation

In this section, we demonstrate the proposed bootstrap method through the same examples as those studied by Crowder (1998). We repeat the simulation 100 times for each of the three different sample sizes $n = 50$, $n = 100$ and $n = 200$ and for each example. We always repeat the bootstrap sampling 10,000 times. We always set the nominal level $\alpha = 0.05$, and estimate the actual significance level α_n through a simulation with 10,000 replications. We measure the improvement in the accuracy of the bootstrap estimator by the ratio

$$R_n = |\hat{\alpha}_n - \alpha_n|/|\alpha - \alpha_n|.$$

Obviously, $\hat{\alpha}_n$ is closer to α_n than α when $R_n < 1$. The smaller is the value of R_n, the larger is the improvement.

Example 1. Suppose $Y_1, \cdots, Y_n$ are independent observations from the distribution

$$F(y;\ \nu, \xi, \kappa) = \begin{cases} 1 - \exp\left\{\kappa^\nu - (\kappa + \xi y)^\nu\right\} & y > 0 \\ 0 & y \leq 0 \end{cases}, \quad \nu > 0,\ \kappa \geq 0,\ \xi > 0. \tag{3.1}$$

As in Crowder (1998) , we consider two tests, namely

$$H_0: \ \nu = 1 \ \text{ vs } \ H_1: \ \nu > 1, \tag{3.2}$$

and

$$H_0: \ \kappa = 0 \ \text{ vs } \ H_1: \ \kappa > 0. \tag{3.3}$$

Note that the associated likelihood-based methods are non-regular: $\kappa = 0$ is on the boundary of the parameter space, and κ disappears from the likelihood but not from the score function when $\nu = 1$ (Crowder 1990, 1998).

First, we consider the tests for hypotheses in (3.2), and let $\kappa = 2$. Under the null hypothesis H_0, the distribution (3.1) reduces to $1-\exp(-\xi y)$ with ξ playing the role of a nuisance parameter. The test statistic previously suggested in Crowder (1990, 1998) is $T_n(\xi) = \xi Y_{(n)} - \log n$, where $Y_{(n)} = \max\{Y_1,\ Y_2, \cdots,\ Y_n\}$. It is easy to see that under the null hypothesis the distribution of $T_n(\xi)$ is $\{1-n^{-1}\exp(-t)\}^n \to \exp\{-e^{-t}\}$; the maximum likelihood estimator of ξ is $\hat{\xi} = n/\sum_{i=1}^{n} Y_i$. Crowder (1998) shows that $T_n(\hat{\xi})$ shares the same asymptotic distribution as $T_n(\xi)$. Note that

$F(y;\ \nu, \xi, \kappa)$ is an increasing function of ν. Therefore $T_n(\xi)$ is stochastically decreasing with respect to ν when $\kappa > 1$. This leads to the reject region $\{T_n(\hat{\xi}) < z_\alpha\}$, where z_α is the lower α-point of the asymptotic distribution of $T_n(\xi)$.

In our simulation, we set the true value of ξ at 1. The simulation results with 100 replications and three different sample sizes, namely 50, 100 and 200, are presented in the third column of Table 1 and Fig.1(a). The simulated values for the actual significance α_n are quite different from the nominal level $\alpha = 0.05$, even with the sample size $n = 200$. The bootstrap estimator $\hat{\alpha}_n$ is always *much* closer to α_n than α. Fig.1(a) shows that the improvement of the knowledge on the actual significance level through bootstrap is quite dramatic, although the smaller is the sample size, the larger is the improvement.

Consider now the second setting (3.3). The distribution (3.1) under H_0 reduces to $F(y;\ \nu, \xi) = 1 - \exp\Big\{ - (\xi y)^\nu \Big\}$, $y > 0$. Now both ν and ξ are nuisance parameters. We use the same test statistic $T_n(\xi,\ \nu) = n(\xi Y_{(1)})^\nu$ as Crowder (1990, 1998), where $Y_{(1)} = \min\{Y_1,\ \cdots,\ Y_n\}$. Its asymptotic distribution under the null hypothesis is $1 - \exp(-t)$. It can be proved that the maximum likelihood estimator $(\hat{\xi}, \hat{\nu})$ is the solution of the equations

$$\begin{cases} n - \xi^\nu \sum\limits_{i=1}^{n} Y_i^\nu = 0 \\ n\nu^{-1} + \sum\limits_{i=1}^{n} \Big\{ \log Y_i - \xi^\nu Y_i^\nu \log Y_i \Big\} = 0. \end{cases}$$

Moreover, $T_n(\hat{\xi}, \hat{\nu})$ has the same asymptotic distribution as that of $T_n(\xi, \nu)$; see Crowder (1998).

Note that $F(y;\ \nu, \xi, \kappa)$ is an increasing function of κ. Therefore $T_n(\xi, \nu)$ is stochastically decreasing with respect to κ. Further, the rejection region is $T_n(\hat{\xi},\ \hat{\nu}) < z_\alpha$, where z_α is the lower-α point of the asymptotic distribution of $T_n(\xi, \nu)$.

In our simulation, we set $\xi = 1$, $\nu = 2$. The results are summarised in the fourth column of Table 1 and Fig.1(b). Now the asymptotic distribution provides a better approximation in the sense that the actual level α_n is much closer to α than that in the previous case. However, the improvement from using bootstrap is still significant; see Fig.1(b).

Example 2. Let $Y_1, \cdots, Y_n$ be the independent observations from the distribution

$$F(y;\ \psi, \xi, v) = \begin{cases} 1 - \Big[1 + \Big\{(y - \psi)/\xi\Big\}^v\Big]^{-1} & y \geq \psi \\ 0 & y < \psi \end{cases} \qquad v > 0,\ \xi > 0.$$

We consider the test of the following hypotheses on the parameter ξ:

$$H_0 :\ \xi = 1 \ \text{ vs } \ H_1 :\ \xi < 1. \tag{3.4}$$

Table 1: The simulated values for the actual significance level α_n and the means and standard deviation of the bootstrap estimator $\hat{\alpha}_n$. The nominal level $\alpha = 0.05$.

		Test for (3.2)	Test for (3.3)	Test for (3.4) $v = 1$	Test for (3.4) $v = 2$
n=50	α_n	0.0114	0.0401	0.0690	0.0408
	$E(\hat{\alpha}_n)$	0.0119	0.0413	0.0687	0.0408
	STD$(\hat{\alpha}_n)$	0.0010	0.0042	0.0026	0.0020
n=100	α_n	0.0240	0.0466	0.0656	0.0432
	$E(\hat{\alpha}_n)$	0.0230	0.0453	0.0627	0.0430
	STD$(\hat{\alpha}_n)$	0.0017	0.0019	0.0027	0.0024
n=200	α_n	0.0296	0.0479	0.0609	0.0432
	$E(\hat{\alpha}_n)$	0.0300	0.0462	0.0576	0.0427
	STD$(\hat{\alpha}_n)$	0.0026	0.0014	0.0034	0.0029

The parameter ψ plays the role of a nuisance parameter. We assume that the value of v is given to simplify the derivation. This problem is non-regular in the sense that the parameter ψ defines a boundary of the sample space; see Smith (1985) and Crowder (1998). Further, different values of v lead to different behaviour of the test. It is easy to see that the score test statistic is

$$T_n(\psi, v) = n^{-1/2} \sum_{i=1}^{n} h(Y_i;\ \psi, 1, v),$$

where $h(y;\ \psi, \xi, v) = (\frac{\partial}{\partial \xi}) \log\{(\frac{\partial}{\partial y}) F(y;\ \psi, \xi, v)\}$. We use $\hat{\psi} = Y_{(1)} \equiv \min\{Y_1, \cdots, Y_n\}$ to estimate the nuisance parameter.

First, we consider the case $v = 1$. It can be proved that $T_n(\psi, 1)$ is asymptotically normal with mean 0 and variance 1/3 under the null hypothesis, and $T_n(\hat{\psi}, 1)$ shares the same asymptotic distribution as $T_n(\psi, 1)$; see Crowder (1998). Further, we will reject the null hypothesis if $3^{1/2} T_n(\hat{\psi}, 1, 1) < z_\alpha$, where z_α is the lower-α point of the standard normal distribution. The simulation with $\psi = 0$ yields the results reported in the fifth column in Table 1. Now the nominal level under-estimates the actual significance level by 0.011 for $n = 200$ and by 0.019 for $n = 50$. The boxplots of simulated values of R_n are depicted in Fig.1(c), which demonstrates a substantial improvement by bootstrap.

Finally, we consider the case with $v = 2$. It can be proved that under the null hypothesis, the asymptotic distributions of $T_n(\hat{\psi}, 2)$ and $T_n(\psi, 2)$ are no longer the same; see Crowder (1998). This case is beyond the scope of Theorem 1. (In fact, the condition $\sigma_n \lambda_n \to 0$ is no longer fulfilled now; see condition (C2) in the Appendix.) However, our simulation suggests that the bootstrap method can still provide more accurate estimates for the actual significance level than the nominal value of α; see the last column Table 1 and Fig.1(d).

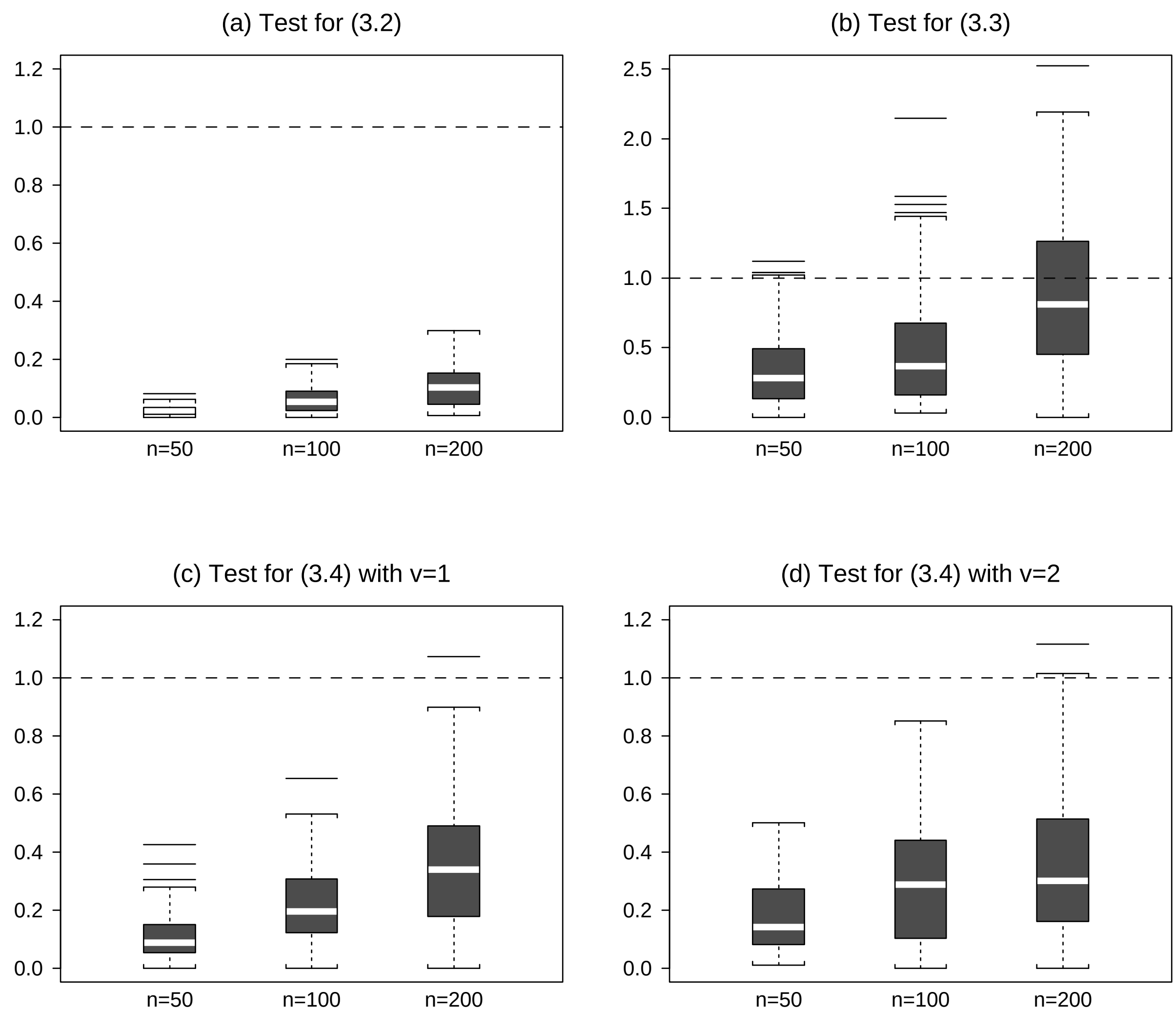

Figure 1: The box-plots of ratios $R_n = |\hat{\alpha}_n - \alpha_n|/|\alpha - \alpha_n|$ for Examples 1 and 2. Bootstrap provides a better estimator for α whenever $R_n < 1$.

Appendix

Conditions

(C1) There exists a sequence of positive numbers $\{\sigma_n\}$ for which $(\hat{\psi} - \psi)/\sigma_n = O_P(1)$ under the distribution $P(\cdot|\theta_0, \psi)$.

(C2) There exist sequences of positive numbers $\{\lambda_n\}$ and $\{\delta_n\}$ such that under the distribution $P(\cdot|\theta_0, \psi)$

$$\{\dot{T}_n(\psi) - \tau_n(\psi)\}/\lambda_n \xrightarrow{P} 0, \quad \ddot{T}_n(\psi)/\delta_n \xrightarrow{P} c,$$

and $\tau_n = O(\lambda_n)$, $\delta_n \sigma_n = o(\lambda_n)$, and $\sigma_n \lambda_n \to 0$, where

$$\dot{T}_n(\psi) = \partial T_n(\psi)/\partial \psi, \quad \tau_n(\psi) = E\{\dot{T}_n(\psi)|\theta_0, \psi\}, \quad \ddot{T}_n(\psi) = \partial^2 T_n(\psi)/\partial \psi \partial \psi^T.$$

(C3) Let $G_n(x|y;\ \psi)$ be the conditional distribution of $\tau_n^T(\psi)(\hat{\psi} - \psi)/(\sigma_n \lambda_n)$ given $T_n(\psi) = y$ under $P(\cdot|\theta_0, \psi)$. Then, the limit functions

$$\phi_1(y, \psi) \equiv \lim_{n\to\infty} -(\sigma_n \lambda_n)^{-1} G_n\Big(-(\sigma_n \lambda_n)^{-1}|y;\ \psi\Big)$$

and

$$\phi_2(y, \psi) \equiv \lim_{n\to\infty} (\sigma_n \lambda_n)^{-1}\Big(1 - G_n\Big((\sigma_n \lambda_n)^{-1}|y;\ \psi\Big)\Big)$$

exist and are continuous with respect to ψ.

Condition (C1) is usually fulfilled with a good estimator (such as an MLS) with $\sigma_n \to 0$. Under condition (C2), the Statistic T_n is smooth. (C3) assumes the tails of the limit distribution of 'normalised' $\hat{\psi}$ decays fast enough and is continuous with ψ.

Proof of Theorem 1

Unless specified, $P(\cdot)$ stands for the population $P(\cdot|\theta_0, \psi)$ in the sequel. Let

$$\Delta_n = \Big(\dot{T}_n(\psi) - \tau_n\Big)^T + \frac{1}{2}(\hat{\psi} - \psi)^T \ddot{T}_n(\psi_1),$$

where ψ_1 lies between ψ and $\hat{\psi}$. It follows from a Taylor's expansion that $T_n(\hat{\psi}) = T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) + \Delta_n(\hat{\psi} - \psi)$. Therefore

$$\begin{aligned} P\Big\{T_n(\hat{\psi}) \le t\Big\} &= P\Big\{T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) + \Delta_n(\hat{\psi} - \psi) \le t\Big\} \\ &= P\Big\{T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) + \Delta_n(\hat{\psi} - \psi) \le t,\ \Delta_n(\hat{\psi} - \psi) \le 0\Big\} \\ &+ P\Big\{T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) + \Delta_n(\hat{\psi} - \psi) \le t,\ \Delta_n(\hat{\psi} - \psi) \ge 0\Big\}. \end{aligned}$$

Let $\mathbf{J}_1$ and $\mathbf{J}_2$ denote the two terms on the RHS of the above expression respectively. Then,

$$\begin{aligned} \mathbf{J}_1 &= P\Big\{T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) \le t,\ \Delta_n(\hat{\psi} - \psi) \le 0\Big\} \\ &+ P\Big\{t \le T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) \le t - \Delta_n(\hat{\psi} - \psi),\ \Delta_n(\hat{\psi} - \psi) \le 0\Big\}, \end{aligned}$$

and the second term on the RHS of the above expression is equal to

$$\begin{aligned} & P\left\{t \le T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) \le t - \Delta_n \sigma_n(\hat{\psi} - \psi)\sigma_n^{-1},\ \Delta_n(\hat{\psi} - \psi) \le 0\right\} \\ = \ & P\left\{t \le T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) \le t - \Big(\dot{T}_n(\psi) - \tau_n\Big)^T \sigma_n(\hat{\psi} - \psi)\sigma_n^{-1},\ \Delta_n(\hat{\psi} - \psi) \le 0\right\} \\ & + O(\delta_n \sigma_n^2) \ = \ o(\sigma_n \lambda_n). \end{aligned}$$

Thus, we have that

$$\mathbf{J}_1 = P\Big\{T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) \le t,\ \Delta_n(\hat{\psi} - \psi) \le 0\Big\} + o(\sigma_n\lambda_n).$$

The similar arguments lead to

$$\mathbf{J}_2 = P\Big[T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) \le t,\ \Delta_n(\hat{\psi} - \psi) \ge 0\Big] + o(\sigma_n\lambda_n).$$

Combing the above two equations, we have that

$$P\Big\{T_n(\hat{\psi}) \le t\Big\} = P\Big\{T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) \le t\Big\} + o(\sigma_n\lambda_n). \tag{A.1}$$

On the other hand, the following decompositions are obvious.

$$\begin{aligned} & P\Big\{T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) \le t\Big\} \\ = \ & P\Big\{T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) \le t,\ \tau_n^T(\hat{\psi} - \psi) \le 0\Big\} + P\Big\{T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) \le t,\ \tau_n^T(\hat{\psi} - \psi) \ge 0\Big\} \\ \triangleq \ & \mathbf{m}_1 + \mathbf{m}_2, \end{aligned}$$

$$\begin{aligned} \mathbf{m}_1 &= P\Big\{T_n(\psi) \le t,\ \tau_n^T(\hat{\psi} - \psi) \le 0\Big\} + P\Big\{t \le T_n(\psi) \le t - \sigma_n\tau_n^T(\hat{\psi} - \psi)/\sigma_n,\ \tau_n^T(\hat{\psi} - \psi) \le 0\Big\} \\ &\triangleq \mathbf{m}_{1,1} + \mathbf{m}_{1,2}, \end{aligned}$$

$$\begin{aligned} \mathbf{m}_2 &= P\Big\{T_n(\psi) \le t,\ \tau_n^T(\hat{\psi} - \psi) \ge 0\Big\} - P\Big\{t - \sigma_n\tau_n^T(\hat{\psi} - \psi)/\sigma_n \le T_n(\psi) \le t,\ \tau_n^T(\hat{\psi} - \psi) \ge 0\Big\} \\ &\triangleq \mathbf{m}_{2,1} - \mathbf{m}_{2,2}. \end{aligned}$$

It is easy to see that

$$P\Big\{T_n(\psi) + \tau_n^T(\hat{\psi} - \psi) \le t\Big\} = P\Big\{T_n(\psi) \le t\Big\} + \mathbf{m}_{1,2} - \mathbf{m}_{2,2},$$

$$P\Big\{T_n(\hat{\psi}) \le t\Big\} = P\Big\{T_n(\psi) \le t\Big\} + \mathbf{m}_{1,2} - \mathbf{m}_{2,2} + o(\sigma_n\lambda_n). \tag{A.2}$$

Moreover, it follows from condition (C3) that

$$\mathbf{m}_{1,2} = \sigma_n\lambda_n \int_t^{\infty} (t-y)^{-1}\phi_1(y,\psi)dF(y) + o(\sigma_n\lambda_n),$$

$$\mathbf{m}_{2,2} = \sigma_n\lambda_n \int_{-\infty}^{t} (t-y)^{-1}\phi_2(y,\psi)dF(y) + o(\sigma_n\lambda_n).$$

Thus, we have that

$$\begin{aligned} \alpha_n &= P\Big\{T_n(\psi) \le F^{-1}(\alpha)\Big\} + \sigma_n\lambda_n \Big(\int_t^{\infty} (t-y)^{-1}\phi_1(y,\psi)dF(y) \\ &- \int_{-\infty}^{t} (t-y)^{-1}\phi_2(y,\psi)dF(y)\Big) \ + \ o(\sigma_n\lambda_n), \end{aligned} \tag{A.3}$$

which leads to

$$\alpha_n - \alpha = O\left(\sigma_n\lambda_n + \left(P\left\{T_n(\psi) \le F^{-1}(\alpha)\right\} - \alpha\right)\right).$$

Note that $\hat{\alpha}_n$ can be seen as α_n if $\hat{\psi}$ is treated as the true value of ψ. Thus

$$\begin{aligned} \hat{\alpha}_n &= P\left\{T_n(\psi) \le F^{-1}(\alpha)\right\} + \sigma_n\lambda_n\left(\int_t^\infty (t-y)^{-1}\phi_1(y,\hat{\psi})dF(y)\right. \\ &- \left.\int_{-\infty}^t (t-y)^{-1}\phi_2(y,\hat{\psi})dF(y)\right) + o_P(\sigma_n\lambda_n). \end{aligned} \tag{A.4}$$

It follows from (A.3), (A.4) and conditions (C1) and (C3) that $\hat{\alpha}_n - \alpha = o_P(\sigma_n\lambda_n)$. Therefore, $\hat{\alpha}_n - \alpha_n = o_P(\alpha_n - \alpha)$.

References

Crowder, M. J. (1990). On some nonregular tests for a modified Weibull model. *Biometrika*, **77**, 499-506.

Crowder, M. J. (1996). Some tests based on extreme values for a parametric survival model. *J. Roy. Statist. Soc.* **B**, **58**, 417-424.

Crowder, M. J. (1998). Corrected P-values for tests based on estimated nuisance parameters. *manuscript*

De Haan, L. and Resnick, S. (1996). Second-order regular variation and rates of convergence in extreme-value theory, *Ann. Probab.*, **24**, 97-124.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, New York.

Loh, W. Y. (1987). Calibrating confidence coefficients, *J. Amer. Statist. Assoc.*, **82**, 155-162.

Loh, W. Y. (1988). Discussion of "Theoretical comparison of bootstrap confidence intervals" by P. Hall, *Ann. Statist.*, **16**, 972-976.

Loh, W. Y. (1991). Bootstrap calibration for confidence interval construction and selection, *Statist. Sinica*, **1**, 479-495.

Pierce, D. A. (1982). The asymptotic effect of substituting estimators for parameters in certain types of statistics, *Ann. Statist.*, **10**, 475-478.

Randles, R. H. (1982). On the asymptotic normality of statistics with estimated parameters, *Ann. Statist.*, **10**, 462-474.

Smith, R. L. (1985). Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, **72**, 67-90.