SINGH'S THEOREM IN THE LATTICE CASE

Michael WOODROOFE * and Myoungshic JHUN **

Department of Statistics, The University of Michigan, 1444 Mason Hall, Ann Arbor, MI 48109–1027, USA

Abstract The asymptotic behavior of the parametric bootstrap estimator of the sampling distribution of a maximum likelihood estimator is investigated in a simple lattice case, integer valued random variables whose distributions form an exponential family. The exptected value of the bootstrap estimator is compared with an Edgeworth expansion, less the continuity correction

AMS (1980) Subject Classification 62E20.

Keywords bootstrap estimator, Edgeworth expansions, fractional part, maximum likelihood estimator, very weak expansions.

1. The main result

Singh's (1981) theorem asserts that the bootstrap estimator of the distribution of a standardized sum of n i.i.d. non-lattice random variables with three moments differs from the true value by a term of smaller order of magnitude than $1/\sqrt{n}$ w.p. 1 as $n \to \infty$. The purpose of this note is to investigate the lattice case. It is shown that the expected value of the bootstrap estimator differs from the very weak expansions of Stein (1985) and Woodroofe (1986) by a term of order $1/\sqrt{n}$, the coefficient of which is very small in many examples.

To isolate the issues, the result is presented in a simple context. Thus, let X_1, X_2, ... denote i.i.d. integer valued random variables with common discrete density (mass function) of the form

$$f_\omega(x) = h(x) \exp[\omega x - \psi(\omega)], \quad x \in \mathbf{Z}, \ \omega \in \Omega,$$

where $\mathbf{Z}$ denotes the integers, h is a non-negative function and ω denotes an unknown parameter

with values in an open interval Ω. It is assumed throughout the greatest common divisor of $x - y$ for which $h(x) > 0$ and $h(y) > 0$ is one.

the mean and variance of X are

$$\theta = \psi'(\omega) \quad \text{and} \quad \sigma^2(\theta) = \psi''(\omega),$$

where $'$ denotes differentiation. Regarding σ^2 as a function of θ here is justified, since θ and ω are increasing functions of each other. Let $\bar{X}_n$ denote the sample mean; $\bar{X}_n = S_n/n$, where $S_n = X_1 + \cdots + X_n$, $n \geq 1$. Then $\bar{X}_n$ is the maximum likelihood estimator of θ, whenever it is in $\psi'(\Omega)$. For purposes of inference, the distributions of the (approximately) pivotal quantities

$$Z_n = \frac{\sqrt{n}}{\sigma}\left(\bar{X}_n - \theta\right) = \frac{S_n - n\theta}{\sigma\sqrt{n}}, \quad n \geq 1, \qquad (1)$$

are of interest. The distribution function of Z_n is denoted by

$$G_n(\omega, t) = P_\omega\{Z_n \leq t\},$$

where $t \in \mathbf{R}$, $\omega \in \Omega$, $n \geq 1$. It depends on the unknown parameter ω as well as n.

To estimate these distributions, Efron (1979) has suggested the (parametric) bootstrap estimator

$$\hat{G}_n(t) = G_n(\hat{\omega}_n, t), \quad t \in \mathbf{R}, \ n \geq 1,$$

where

$$\hat{\omega}_n = \hat{\omega}_n(X_1, \ldots, X_n)$$

denotes the maximum likelihood estimator of ω (the solution to the equation $\psi'(\omega) = \overline{X}_n$, if one exists). For present purposes, it is convenient to allow $\hat{\omega}_n$, $n \geqslant 1$, to denote any strongly consistent sequence of estimators for which

$$\lim_n \left\{ P_\omega \{ |n\psi'(\hat{\omega}_n) - S_n| \geqslant \varepsilon \} \right.$$

$$\left. + \sqrt{n} \, P_\omega \{ |\hat{\omega}_n - \omega| \geqslant \varepsilon \} \right\} = 0 \qquad (2)$$

for all $\varepsilon > 0$ and $\omega \in \Omega$. Such a sequence may be constructed by letting $\hat{\omega}_n$ denote the maximum likelihood estimator for a restricted model in which Ω is replaced by a compact $\Omega_n \subset \Omega$ and Ω_n increases to Ω as $n \to \infty$.

For the related case in which the distribution of X_1 is non-lattice, (a simple variation on) Singh's (1981) Theorem shows that $\hat{G}_n$ and $G(\omega, \cdot)$ differ by $o(1/\sqrt{n})$ w.p. 1 (P_ω) for all ω. This is not true in the lattice case.

If $t \in \mathbb{R}$, denote its integer part by $[t]$; and let $\langle t \rangle = t - [t]$, denote the fractional part of t.

Proposition. *If Ω_0 is any compact subinterval of Ω, then*

$$G_n(\omega, t)$$

$$= \Phi(t) + \frac{1}{\sqrt{n}} \phi(t) \left[\frac{\rho}{6}(1 - t^2) + R_n(\omega, t) \right]$$

$$+ o\left(\frac{1}{\sqrt{n}} \right),$$

uniformly in $t \in \mathbb{R}$ and $\omega \in \Omega_0$ as $n \to \infty$, where Φ denotes the standard normal distribution function, ϕ denotes the standard normal density,

$$\rho = \rho(\omega) = \psi''{}'(\omega)/\sigma^3,$$

and

$$R_n(\omega, t) = \frac{1}{\sigma} \left\{ \tfrac{1}{2} - \langle n\theta + \sqrt{n} \, \sigma t \rangle \right\}$$

for $t \in \mathbb{R}$, $\omega \in \Omega$, and $n \geqslant 1$.

Proof. For fixed ω, this is simply Theorem 1 (p. 213) of Gnedenko and Kolmogorov (1968). The uniformity follows from an examination of the proof.

Corollary 1. *For $t \in \mathbb{R}$, $\omega \in \Omega$, and $n \geqslant 1$, let*

$$G_n^*(\omega, t) = \Phi(t) + \frac{\rho}{6\sqrt{n}}(1 - t^2)\phi(t).$$

Then

$$\lim_n \int_\Omega \sqrt{n} \left[G_n(\omega, t) - G_n^*(\omega, t) \right] \xi(\omega) \, d\omega = 0$$

for all $t \in \mathbb{R}$ and all densities ξ with compact support in Ω.

Proof. It suffices to show that

$$\lim_n \int_\Omega \left\{ \tfrac{1}{2} - \langle n\theta + \sqrt{n} \, \sigma \rangle \right\} \xi \, d\omega = 0$$

for all integrable ξ; and since the term in $\{ \cdots \}$ is bounded, it suffices to show this when ξ is the indicator of an interval $(a, b) \subset \Omega$, by a standard approximation argument. See, for example, Billingsley (1979, p. 226). Let $v(\omega) = \psi'(\omega)$ and $v_n(\omega) = \psi'(\omega) + \sqrt{\psi''(\omega)}/\sqrt{n}$ for $\omega \in \Omega$ and $n \geqslant 1$. Then v_n converges to v and v_n' converges to v' uniformly on compacts as $n \to \infty$; and $v' = \psi''$ is continuous and positive. So, if Ω_0 is any compact subset of Ω, then the restrictions of v and v_n to Ω_0 have inverses w and w_n for which w_n and w_n' converge to w and w', pointwise and boundedly. If $(a, b]$ is any finite subinterval of Ω, then

$$\int_a^b \left\{ \tfrac{1}{2} - \langle n\theta + \sqrt{n} \, \sigma \rangle \right\} \xi \, d\omega$$

$$= \int_{\alpha_n}^{\beta_n} \left\{ \tfrac{1}{2} - \langle ny \rangle \right\} \frac{1}{w_n'(y)} \, dy,$$

where $\alpha_n = v_n(a)$ and $\beta_n = v_n(b)$ for $n \geqslant 1$. So, it suffices to show that the right side converges to zero as $n \to \infty$ for arbitrary choices of $a, b \in \Omega$. Since w_n' converges boundedly to w', it suffices to show this when w_n' is replaced by w'; by the standard approximation, it suffices to show it when $1/w'$ is replaced by the indicator of an interval; and the latter condition is clearly satisfied. $\square$

If $G(\omega, t)$ is regarded as the coverage probability of a confidence set at ω, then $\int_\Omega G_n(\omega, t) \xi(\omega) \, d\omega$ may be regarded as the long run relative frequency of coverage in many independent replications of the experiment, when ω is drawn from

the density ξ; so, $\int_\Omega G_n(\omega, t)\xi(\omega)\, d\omega$ may reasonably be called the average coverage probability at ξ. See Woodroofe (1986) for an elaboration of this point. The bootstrap estimator is compared to G_n^* below.

Theorem 1. *For all* $t \in \mathbb{R}$ *and all* $\omega \in \Omega$ *for which* $\rho(\omega) \neq 0$,

$$\lim_n \sqrt{n}\left[\hat{G}_n(t) - G_n^*(\omega, t)\right]$$
$$- \frac{1}{\sigma}\phi(T)\left\{\tfrac{1}{2} - \langle\langle t\sqrt{n}\,\sigma\rangle + \tfrac{1}{2}\rho\sigma t Z_n\rangle\right\} = 0$$

in P_ω-*probability, where* Z_n *is as in* (1).

Proof. Fix values of $t \in \mathbb{R}$ and $\omega \in \Omega$. Then, by the Proposition,

$$\sqrt{n}\left[\hat{G}_n(t) - G_n^*(\omega, t)\right]$$
$$= \tfrac{1}{6}(\hat{\rho}_n - \rho)(1 - t^2)\phi(t)$$
$$+ \phi(t)R_n(\hat{\omega}_n, t) + o(1)$$

w.p. 1 (P_ω) as $n \to \infty$, where $\hat{\rho}_n = \rho(\hat{\omega}_n)$, for all $n \geqslant 1$. Clearly

$$\lim_n \left|\tfrac{1}{6}(\hat{\rho}_n - \rho)(1 - t^2)\phi(t)\right| = 0$$

in P_ω-probability as $n \to \infty$, by the assumed consistency of $\hat{\omega}_n$, $n \geqslant 1$. Next, since S_n is an integer.

$$R_n(\hat{\omega}_n, t) = \hat{\sigma}_n^{-1}\left\{\tfrac{1}{2} - \langle\langle t\sqrt{n}\,\sigma\rangle + t\sqrt{n}\,(\hat{\sigma}_n - \sigma)\right.$$
$$\left. + n\hat{\theta}_n - S_n\rangle\right\}$$

for all n, where $\hat{\theta}_n = \psi'(\hat{\omega}_n)$ and $\hat{\sigma}_n = \sigma(\hat{\theta}_n) = \sqrt{\psi''(\hat{\omega}_n)}$. By (2), a simple Taylor series approximation, and the law of the iterated logarithm

$$t\sqrt{n}\,(\hat{\sigma}_n - \sigma) + n\hat{\theta}_n - S_n = \tfrac{1}{2}\rho\sigma t Z_n + o_p(1)$$

in P_ω-probability as $n \to \infty$. So, since the fractional part is continuous, except for a set of content zero,

$$R_n(\hat{\omega}_n, t) = \sigma^{-1}\left\{\tfrac{1}{2} - \langle\langle t\sqrt{n}\,\sigma\rangle + \tfrac{1}{2}\rho\sigma t Z_n\rangle\right\}$$
$$+ o_p(1) \qquad (3)$$

under P_ω as $n \to \infty$. See the lemma below with

$$U_n = \langle t\sqrt{n}\,\sigma\rangle + t\sqrt{n}\,(\hat{\sigma}_n - \sigma) + n\hat{\theta}_n - S_n$$

and

$$V_n = \langle t\sqrt{n}\,\sigma\rangle + \tfrac{1}{2}\rho\sigma t Z_n.$$

The theorem follows by collecting terms. □

Corollary 2. *For* $r > 0$ *and* $0 \leqslant m < 1$, *let*

$$e(m, r) = \int_\mathbb{R}\left[\tfrac{1}{2} - \langle m + rs\rangle\right]\phi(s)\, ds.$$

Then

$$E_\omega\left\{\hat{G}_n(t) - G_n^*(\omega, t)\right\}$$
$$= \frac{1}{\sigma\sqrt{n}}\phi(t)e(\langle t\sqrt{n}\,\sigma\rangle, \tfrac{1}{2}\rho\sigma t) + o\left(\frac{1}{\sqrt{n}}\right)$$

as $n \to \infty$ *for all* $t \in \mathbb{R}$ *and all* $\omega \in \Omega$ *for which* $\rho(\omega) \neq 0$.

Proof. If $\omega \in \Omega$, then there is a compact $\Omega_0 \subset \Omega$, for which $P_\omega\{\hat{\omega}_n \notin \Omega_0\} = o(1/\sqrt{n})$ as $n \to \infty$, by (2). It is clear that

$$(\hat{\rho}_n - \rho)I\{\hat{\omega}_n \in \Omega_0\}, \quad n \geqslant 1,$$

and

$$(1/\hat{\sigma}_n)R_n(\hat{\omega}_n, t)I\{\hat{\omega}_n \in \Omega_0\}, \quad n \geqslant 1,$$

are uniformly integrable random variables. It follows that

$$\lim_n E_\omega\left\{(\hat{\rho}_n - \rho)I\{\hat{\omega}_n \in \Omega_0\}\right\} = 0.$$

$$\lim_n E_\omega\left\{(1/\hat{\sigma}_n)R_n(\hat{\omega}_n, t)I\{\hat{\omega}_n \in \Omega_0\}\right\}$$
$$\cdot - \frac{1}{\sigma}\phi(t)e(\langle t\sqrt{n}\,\sigma\rangle, \tfrac{1}{2}\rho\sigma t) = 0.$$

(For the latter assertion, it is convenient to consider subsequences along which $\langle t\sqrt{n}\,\sigma\rangle$ is convergent.) the corollary follows easily. □

Remark. The function e decreases quickly as r increeases. To see why, write

$$\tfrac{1}{2} - \langle t\rangle = \sum_{k=1}^\infty \frac{1}{k\pi}\sin(2\pi kt), \quad t \in \mathbb{R}.$$

Then

$$e(m, r) = \sum_{k=1}^\infty \frac{1}{k\pi}\int_\mathbb{R}\sin(2\pi kt)\frac{1}{r}\phi\left(\frac{t - m}{r}\right)\, dt$$
$$= \sum_{k=1}^\infty \frac{1}{k\pi}\mathrm{Im}\left\{\exp(2\pi imk - 2r^2\pi^2k^2)\right\}$$
$$= \sum_{k=1}^\infty \frac{1}{k\pi}\sin(2\pi mk)\, e^{-2r^2\pi^2k^2}$$

and for $0 \leqslant m < 1$ and $r > 0$. For example, $\sup_m |e(m, \frac{1}{2})| < 0.01$, $\sup_m |e(m, 1)| < 10^{-9}$.

In cases where $\sup_m |e(m, \rho\sigma/2)|$ is small, the theorem suggests that bootstrap approximations should be close to the very weak expansions cited above. This is a weaker property than that asserted in Singh's Theorem, but provides some comfort.

Example. *If X_1 has the Poisson distribution with unknown mean $\theta > 0$, and if $t = 2$ (as in an upper 97.5% confidence bound), then $r = \rho\sigma t/2 = 1$ for all ω, so that the coefficient to $1/\sqrt{n}$ is less than 10^{-9}.*

The following lemma completes the justification of (3).

Lemma. *Let U_n, $n \geqslant 1$, and V_n, $n \geqslant 1$, denote random variables for which*:

(i) *the distrubitons of V_n, $n \geqslant 1$, are tight*;

(ii) *all weak limit points of the distributions of V_n, $n \geqslant 1$, assign measure zero to $\mathbb{Z}$*; *and*

(iii) $U_n - V_n \rightarrow 0$ *in probability as $n \rightarrow \infty$.*
Then

$$\langle U_n \rangle - \langle V_n \rangle \rightarrow 0 \quad \text{in probability}.$$

Proof. Let $\mathbb{Z}_\delta$ denote the set of x whose distance from $\mathbb{Z}$ is at most δ. Then $\mathbb{Z}_\delta$ is closed, and $\mathbb{Z}_\delta$ decreases to $\mathbb{Z}$ as δ decreases to zero. So,

$$\lim_{\sigma \rightarrow 0} \lim_n \sup P\{V_n \in \mathbb{Z}_\delta\} = 0,$$

by Theorem 2.1 of Billingsely (1968, pp. 11–12) and conditions (i) and (ii) above. Thus, given $\varepsilon > 0$, there are n_0 and $\delta < \varepsilon$ for which $P\{V_n \in \mathbb{Z}_{2\delta}\} \leqslant \varepsilon/2$ for all $n \geqslant n_0$. So,

$$P\{ |\langle U_n \rangle - \langle V_n \rangle| \geqslant \varepsilon \}$$
$$\leqslant P\{ |U_n - V_n| \geqslant \delta \} + P\{ V_n \in \mathbb{Z}_{2\delta} \}.$$

for all $n \geqslant n_0$; and this is less than ε for all sufficiently large n.

2. A Studentized version

In this section, suppose that $\hat{\omega}_n$ satisfies the more stringent condition

$$\lim_n \sqrt{n}\, P_\omega\{ \hat{\theta}_n \neq \bar{X}_n \} = 0, \quad \forall \omega. \tag{2'}$$

Let

$$T_n = (\sqrt{n}/\hat{\sigma}_n)(\bar{X}_n - \theta)$$

and

$$H_n(\omega, t) = P_\omega\{ T_n \leqslant t \}$$

for all $t \in \mathbb{R}$, $\omega \in \Omega$, and $n \geqslant 1$. Then an expansion for H_n may be obtained as follows. For fixed ω, let $h(x) = h_\omega(x) = (x - \theta)/\sigma(x)$, for $x \in \psi'(\Omega)$. Then it is easily seen that h is increasing near $x = \theta$, so that the restriction of h to a suitable neighborhood of θ has an inverse function $g = g_\omega = h_\omega^{-1}$. Moreover, it is also easily seen that $g(0) = \theta$, $g'(0) = \sigma$ and $g''(0) = \rho\sigma$. For fixed $t \in \mathbb{R}$ and $\omega \in \Omega$, let

$$t_n = t_n(\omega, t) = \frac{\sqrt{n}}{\sigma}\left[g\left(\frac{t}{\sqrt{n}} \right) - g(0) \right].$$

Then, $t_n = t + \rho t^2/2\sqrt{n} + o(1/\sqrt{n})$ as $n \rightarrow \infty$; and, by (2') and Chebyshev's Inequality,

$$H_n(\omega, t)$$
$$= G_n(\omega, t_n) + o(1/\sqrt{n})$$
$$= \Phi(t) + \frac{1}{\sqrt{n}}\phi(t)\left[\frac{\rho}{6}(1 + 2t^2) + R_n(\omega, t_n) \right]$$
$$+ o\left(\frac{1}{\sqrt{n}} \right) \tag{4}$$

for each t and ω. In fact, (4) holds uniformly on compacts with respect to $t \in \mathbb{R}$ and $\omega \in \Omega$. Let

$$\hat{H}_n(t) = h_n(\hat{\omega}_n, t)$$

and

$$H_n^*(\omega, t) = \Phi(t) + \frac{\rho}{6\sqrt{n}}(1 + 2t^2)\phi(t)$$

for $t \in \mathbb{R}$, $\omega \in \Omega$, and $n \geqslant 1$.

Theorem 2. *With the notation of the previous paragraph,*

$$\lim_n \sqrt{n}\left[\hat{H}_n(t) - H_n^*(\omega, t) \right]$$
$$- \frac{1}{\sigma}\phi(t)\left\{ \frac{1}{2} - \langle\langle \sqrt{n}\, \rho\sigma t^2 \rangle + \frac{1}{2}\rho\sigma t Z_n \rangle \right\} = 0$$

In P_ω-probability for each $t \in \mathbb{R}$ and $\omega \in \Omega$ for which $\rho(\omega) \neq 0$.

Proof. The proof is similar to that of Theorem 1, but slightly more complicated, since $\hat{t}_n := t_n(\hat{\omega}_n, t)$ depends on n and $X_1, \ldots, X_n$. In the analysis, one finds that (with probability approaching one),

$$\langle \sqrt{n}\, \hat{\sigma}_n \hat{t}_n \rangle = \langle V_n + o_p(1) \rangle,$$

where

$$V_n = \langle \sqrt{n}\, \sigma t + \tfrac{1}{2}\rho\sigma t^2 \rangle + \tfrac{1}{2}t\rho\sigma Z_n, \quad n \geqslant 1.$$

Here V_n, $n \geqslant 1$, satisfies the conditions of the lemma; and the argument proceeds as above.

Acknowledgement

Thanks to Bob Keener for helpful discussions.

References

Billingsley, P. (1968), *Convergence of Probability Measures* (Wiley, New York).

Billingsley, P. (1979), *Probability and Measure* (Wiley, New York).

Efron, B (1979), Bootstrap methods. Another look at the jackknife, *Ann Statist* 7, 1–26

Gnedenko, B and A. Kolmogorov (1968), *Limit theorems for sums of independent random variables* (Addison-Wesley, Reading, MA)

Singh, K (1981), On the asymptotic accuracy of Efron's bootstrap, *Ann Statist* 9, 1187–1195.

Stein, C. (1985), On the coverage probability of confidence sets based on a prior distribution, *Banach Center Publications* **16**, 485–514

Woodroofe, M. (1986), Very weak expansions for sequential confidence levels, *Ann Statist* **14**, 1049–1067