Note on Edgeworth Expansions and Asymptotic Refinements of Percentile t-Bootstrap Methods

Jungbin Hwang

Oct, 2019

1 Edgeworth Expansion for the Sample Mean

1.1 Review of Central Limit Theorem

Theorem 1 (Lévy's continuity) Let Z_n be a sequence of random variables. The sequence of corresponding characteristic function (ch.f) $\varphi_{Z_n}(t)$, which by definition are

$$\varphi_{Z_n}(t) = Ee^{itZ_n} \text{ for all } t \in \mathbb{R} \text{ and all } n \in \mathbb{N}.$$

If the sequence of characteristic functions converges to pointwise to some function $\varphi(\cdot)$, i.e. $\varphi_{Z_n}(t) \to \varphi(t)$ for all $t \in \mathbb{R}$. Then, $Z_n \xrightarrow{d} Z$ for some random variable Z if and only if $\varphi_Z(t) := \varphi(t)$ is continuous at $t = 0$.

***Lemma 1** If $E|X^k| < \infty$, then*

$$\left|\varphi_X(t) - \sum_{j=0}^{k} \frac{(it)^j E(X^j)}{j!}\right| \leq E\left[\min\left\{\frac{2|tX|^k}{k!}, \frac{|tX|^{k+1}}{(k+1)!}\right\}\right].$$

Proof. Expanding the function e^{itx} up to k times yields

$$\begin{aligned} e^{itx} &= 1 + (itx) + \left(\frac{itx}{2!}\right)^2 + ... + \left(\frac{itx}{k!}\right)^k \exp(it\alpha^* x) \\ &= \sum_{j=0}^{k} \left(\frac{itx}{j!}\right)^j + \left(\frac{itx}{k!}\right)^k (\exp(it\alpha^* x) - 1) \end{aligned} \quad (1)$$

for some $\alpha^* \in [0, 1]$. It is not difficult to check that the second term in the last equality is bounded by $2|tx|^k/k!$. Repeating the expansion of $\exp(itx)$ up to $k+1$ term also gives

$$e^{itx} = \sum_{j=0}^{k} \frac{(itx)^j}{j!} + \frac{(itx)^{k+1}}{(k+1)!} \exp(it\tilde{\alpha}x) \tag{2}$$

for some $\tilde{\alpha} \in [0,1]$. Here the second term on the right is bounded by $|tx|^{k+1}/(k+1)!$. Combining (1) and (2), we have

$$\left| e^{itx} - \sum_{j=0}^{k} \frac{(itx)^j}{j!} \right| \leq \min\left\{ \frac{2|tx|^k}{k!}, \frac{|tx|^{k+1}}{(k+1)!} \right\},$$

and the results follows by taking expectations on both sides and modulus inequality.

■

*Theorem 2 (**Lindberg-Lévy CLT**) If $\{X_i\}_{i=1}^n$ is an i.i.d sequence of random variables having mean $\mu$and variance σ^2, then*

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(X_i - \mu)}{\sigma} \xrightarrow{d} N(0, 1).$$

Proof. Without loss of generality (WLOG), we assume $\mu = 0$ and $\sigma^2 = 1$. From the i.i.d. assumption, the ch.f X_i's are all identical to $\varphi_X(t) = \varphi_{X_1}(t)$ for all t, and

$$\varphi_{Z_n}(t) = \varphi_X \left(\frac{t}{\sqrt{n}} \right)^n .$$

Applying Lemma 1 with $k=2$, we have

$$\begin{aligned}\left|\varphi_X\left(\frac{t}{\sqrt{n}}\right)-1+\frac{t^2}{2n}\right| &\leq E\left[\min\left\{\frac{t^2|X|^2}{n},\frac{t^3|X|^3}{6t^{3/2}}\right\}\right]\\ &\leq \frac{1}{n}E\left[\min\left\{t^2|X|^2,\frac{t^3|X|^3}{6n^{1/2}}\right\}\right]\end{aligned}$$

We claim that the rightside in the last equality is $o(n^{-1})$ by showing that

$$E\left[\min\left\{t^2|X|^2,\frac{t^3|X|^3}{6n^{1/2}}\right\}\right]\rightarrow 0 \text{ as } n\rightarrow\infty. \quad (3)$$

To see this, observe that on the one hand

$$\min\left\{t^2|X|^2,\frac{t^3|X|^3}{6n^{1/2}}\right\}\leq t^2|X|^2 \text{ a.s.}$$

and the bound on the inequality is L_1-integrable. Also,

$$\min\left\{t^2|X|^2,\frac{t^3|X|^3}{6n^{1/2}}\right\}\leq \frac{t^3|X|^3}{6n^{1/2}}\rightarrow 0 \text{ a.s. as } n\rightarrow\infty.$$

Therefore, we can apply the dominated convergence theorem to conclude (3), which makes it possible to expand $\phi_X(\lambda/(\sigma\sqrt{n}))$ as

$$\phi_X\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right),$$

and

$$\begin{aligned} \varphi_{Z_n}(t) &= \varphi_X\left(\frac{t}{\sqrt{n}}\right)^n = \left\{1 - \frac{t^2}{2n}\left(1 + o(1)\right)\right\}^n \\ &\to \exp\left(-\frac{t^2}{2}\right) = \phi_Z(t), \text{ as } n \to \infty, \end{aligned}$$

which is a ch.f of standard normal random variable. By Lévy's continuity theorem, we have the desired result. ■

Comments

- Let $G_n(\cdot)$ denote the cdf of Z_n. By CLT, for every $x \in \mathbb{R}$,

$$G_n(x) = \lim_{n\to\infty} \Phi(x),$$

  where $\Phi(x)$ is the cdf of the standard random variable Z.

- Thus, CLT justifies approximating the cdf of Z_n by that of Z for sufficiently large n, i.e.

$$G_n(x) = \Phi(x) + o(1).$$

- One way to measure, the term $o(1)$, the discrepancy between the actual distribution of sample statistics and asymptotic distribution is by higher-order expansions.

- Higher-order expansions of the distribution function are known as **Edgeworth expansions**.

Definition 1 (Cumulants) Let $K_X(t) = \log M_X(t)$ be a cumulant generating function (cgf) of random variable X which is the log of the moment generating function (mgf) $M_X(t)$. The r-th cumulant of the distribution X, κ_r, is the r-th derivative of $K(t)$, evaluated at $t = 0$, i.e.

$$\kappa_r = \frac{d^r}{dt^r} K(t) \bigg|_{t=0} = K^{(r)}(0).$$

- Since $M(0) = 1$, we see $K(0) = 0$. Expanding as a power series we obtain

$$K(t) = \sum_{r=1}^{\infty} \kappa_r \frac{t^r}{r!}.$$

- Define the r-th central moment as $\mu_r = E(X - E(X))^r$. Then, the first six cumulants are computed as follows:

$$\begin{aligned} \kappa_1 &= \mu_1,\ \kappa_2 = \mu_2,\ \kappa_3 = \mu_3; \\ \kappa_4 &= \mu_4 - 3\mu_2^2,\ \kappa_5 = \mu_5 - 10\mu_3\mu_2; \\ \kappa_6 &= \mu_6 - 15\mu_4\mu_2 - 10\mu_3^2 + 30\mu_2^3, \end{aligned} \tag{4}$$

  which are polynomial functions of the central moments.

Definition 2 (Hermite polynomial) Let $\phi(x)$ be the cdf and pdf of standard normal random variable, respectively, and

$$H_j(x) = (-1)^j \frac{\phi^{(j)}(x)}{\phi(x)}$$

is the j-th In particular,

$$H_2(x) = x^2 - 1, \; H_3(x) = x^3 - 3x; \; H_5(x) = x^5 - 10x^3 + 15x.$$

Assumption 1 $\{X_i\}$ is i.i.d random variables with $E(X_i^4) < \infty$.
Assumption 2 The characteristic function of X_i satisfies $\limsup_{|t|\to\infty} |\varphi_X(t)| < 1$.

- Assumption 2 is known as Cramer's condition - requires that the distribution of X_i to have an absolutely continuous component.

Theorem 3 (Edgeworth expansion) An Edgeworth expansion is a series representation for $G_n(x)$ expressed as powers of $n^{-1/2}$, i.e.

$$\begin{aligned} G_n(x) &= P(Z_n \leq x) \\ &= \Phi(x) - \frac{1}{\sqrt{n}}\left(\frac{\kappa_3}{6}H_2(x)\right)\phi(x) - \frac{1}{n}\left(\frac{\kappa_4}{24}H_3(x) + \frac{\kappa_3^2}{72}H_5(x)\right)\phi(x) + o\left(\frac{1}{n}\right). \end{aligned} \quad (5)$$

Before we prove the theorem, we introduce a useful technical result in Hermite polynomials.

Lemma 2 The Hermite polynomials satisfy

$$\begin{aligned}\frac{d}{dx}\left(H_j(x)\phi(x)\right) &= \frac{d}{dx}\Big((-1)^j\phi^{(j)}(x)\Big) \\ &= -(-1)^{j+1}\phi^{(j+1)}(x) \\ &= -H_{j+1}(x)\phi(x),\end{aligned} \quad (6)$$

and by the formula for normal mgf, the fact $H_0(x) = 1$, *and repeated integration by parts,*

$$\begin{aligned}\exp\left(\frac{t^2}{2}\right) &= \int_{-\infty}^{\infty} e^{tx}\phi(x)dx = t^{-j}\int_{-\infty}^{\infty} e^{tx}H_j(x)\phi(x)dx \\ \Rightarrow \exp\left(\frac{t^2}{2}\right)t^j &= \int_{-\infty}^{\infty} e^{tx}H_j(x)\phi(x)dx\end{aligned} \quad (7)$$

for $j = 0, 1, 2, \ldots$.

Proof. [of Theorem 3] We assume that the moment generating function $M_{Z_n}(t) = E[e^{tZ_n}]$ exists, and use it to derive (5). WLOG, assume $\mu = 0$ and $\sigma^2 = 1$. By definitions of the cumulants and (4),

$$K\left(\frac{t}{\sqrt{n}}\right) = \sum_{r=1}^{\infty} \frac{\kappa_r}{r!} \left(\frac{t}{\sqrt{n}}\right)^r = \left(\frac{1}{n}\right) \frac{t^2}{2!} + \left(\frac{\kappa_3}{n^{3/2}}\right) \frac{t^3}{3!} + \left(\frac{\kappa_4}{n^2}\right) \frac{t^4}{4!} + o(n^{-2}).$$

Then, it is easy to check

$$\begin{aligned} M_{Z_n}(t) &= \exp\left(n \log(E[\exp(\frac{tX_i}{\sqrt{n}})]\right) = \exp\left(nK_X\left(\frac{t}{\sqrt{n}}\right)\right) \\ &= \exp\left(\frac{t^2}{2!} + \frac{\kappa_3}{\sqrt{n}}\frac{t^3}{3!} + \frac{\kappa_4}{n}\frac{t^4}{4!} + o(n^{-1})\right) \\ &= \exp\left(\frac{t^2}{2}\right) \exp\left(\frac{\kappa_3}{\sqrt{n}}\frac{t^3}{3!} + \frac{\kappa_4}{n}\frac{t^4}{4!} + o(n^{-1})\right). \end{aligned}$$

Using a second-order expansion of exponential function, i.e.

$$\exp\left(x + o(n^{-1})\right) = 1 + x + \frac{x^2}{2} + o(n^{-1})$$
$$\to \exp\left(\frac{\kappa_3}{\sqrt{n}}\frac{t^3}{3!} + \frac{\kappa_4}{n}\frac{t^4}{4!} + o(n^{-1})\right) = 1 + \frac{\kappa_3}{\sqrt{n}}\frac{t^3}{6} + \frac{\kappa_4}{n}\frac{t^4}{24} + \frac{\kappa_3^2}{n}\frac{t^6}{72} + o(n^{-1}),$$

we get

$$M_{Z_n}(t) = \exp\left(\frac{t^2}{2}\right) + \underbrace{\exp\left(\frac{t^2}{2}\right) t^3 \frac{\kappa_3}{6}}_{\int_{-\infty}^{\infty} e^{tx} H_3(x)\phi(x)dx} \times \frac{1}{\sqrt{n}}$$
$$+ \underbrace{\exp\left(\frac{t^2}{2}\right)\left(\frac{\kappa_4}{24}t^4 + \frac{\kappa_3^2}{72}t^6\right)}_{\int_{-\infty}^{\infty} e^{tx} He_4(x)\phi(x)dx\frac{\kappa_4}{24} + \int_{-\infty}^{\infty} e^{tx} H_6(x)\phi(x)dx\frac{\kappa_3^2}{72}} \times \frac{1}{n} + o(n^{-1}).$$

Substituting (7) into the above equation ,

$$
\begin{aligned}
M_{Z_n}(t) &= \int_{-\infty}^{\infty} e^{tx}\phi(x)dx + \left(\int_{-\infty}^{\infty} e^{tx}H_3(x)\phi(x)dx\right)\frac{\kappa_3}{6\sqrt{n}} + \left(\int_{-\infty}^{\infty} e^{tx}H_4(x)\phi(x)dx\frac{\kappa_4}{24}\right. \\
&\quad \left.\int_{-\infty}^{\infty} e^{tx}H_6(x)\phi(x)dx\frac{\kappa_3^2}{72}\right)\frac{1}{n} + o(n^{-1}) \\
&= \int_{-\infty}^{\infty} e^{tx}\left(\phi(x) + \frac{\kappa_3}{6\sqrt{n}}H_3(x)\phi(x) + \frac{1}{n}\left(\frac{\kappa_4}{24}H_4(x) + \frac{\kappa_3^2}{72}H_6(x)\right)\phi(x)\right)dx + o(n^{-1}).
\end{aligned}
$$

From (6),

$$H_{j+1}(x)\phi(x)dx = -dH_j(x)\phi(x).$$

Using this we find

$$
\begin{aligned}
M_{Z_n}(t) &= \int_{-\infty}^{\infty} e^{tx}dG_n(x) \\
&= \int_{-\infty}^{\infty} e^{tx}d\left(\Phi(x) - \frac{\kappa_3}{6\sqrt{n}}H_2(x)\phi(x) - \frac{1}{24n}\left(\kappa_4 H_3(x) + \frac{\kappa_3^2}{3}H_5(x)\right)\phi(x)\right) + o(n^{-1}),
\end{aligned}
$$

and this shows that $G_n(x)$ is approximated by the distribution in the bracket which is the desired results. ■

Remark

- The expansion (5) may not be convergent. It is interpreted as an asymptotic series, meaning that the remainder is of a smaller order than the last included term.
- The expansion (5) can be interpreted as the sum of the normal distribution $\Phi(x)$, a $n^{-1/2}$ correction for the main effect of skewness, and n^{-1} correction for the main effect of kurtosis and the secondary effect of skewness.
- The $n^{-1/2}$ skewness correction is an even function of x which means it changes the distribution function symmetrically about zero.
- The $n^{-1/2}$ skewness correction is an odd function of x.

1.2 Edgeworth Expansion for Smooth Function Model

Theorem 4 (Smooth function model) If $\{X_i\}_{i=1}^n$ *are independent and identically distributed,* $\mu = E[h(X_i)]$, $E||h(X_i)||^4 < \infty$, $g(\cdot)$ *has four continuous derivatives in a neighborhood of* μ, *and* $E(\exp(t||h(X_i)||)) \leq B < 1$, *for* $\hat{\mu} = n^{-1}\sum_{i=1}^n h(X_i)$, $V = E[(h(X_i)-\mu)(h(X_i)-\mu)')$ *and* $G = \frac{\partial}{\partial u} g(u)'\Big|_{u=\mu} = G(\mu)$, *as* $n \to \infty$

$$T_n = \frac{\sqrt{n}(g(\hat{\mu}) - g(\mu))}{\sqrt{G'VG}}.$$

Then,

$$P\left(T_n \leq x\right) = \Phi(x) + \frac{1}{\sqrt{n}} p_1(x)\phi(x) + \frac{1}{n} p_2(x)\phi(x) + o(n^{-1}), \tag{8}$$

uniformly in x, $p_1(x)$ *is an even polynomial of order* 2, *and* $p_2(x)$ *is an odd polynomial of degree* 5, *with coefficients depending on the moments of* $h(X_i)$ *up to order* 4.

- The expansion is identical to that in (2). The only difference is in the coefficients of the polynomials.

- One implication is that when the normal distribution $\Phi(x)$ is used as an approximation to the actual distribution of $P(T_n \leq x)$, the error is $O(n^{-1/2})$.

Corollary 1 Let assumptions in Theorem 4 be true, then the result of Theorem 4 applies to feasible version studentized statistics such as the t-ratio T_n, replacing $G'VG$ with $\hat{G}'\hat{V}\hat{G}$, so long as the variance estimator $\hat{G}'\hat{V}\hat{G}$ can be written as function of sample means as

$$\hat{G} = G(\hat{\mu}) \text{ and } \hat{V} = \frac{1}{n}\sum_{i=1}^{n}\left(h(X_i) - \hat{\mu}\right)\left(h(X_i) - \hat{\mu}\right)'.$$

The polynomials $p_1(x)$ and $p_2(x)$ have same orders as Theorem 4 but their coefficients depending on the moments of $h(X_i)$ up to order 8.

- Sometimes we are interested in the distribution of the absolute value of the t-ratio $|T|$ whose distribution is

$$P(|T_n| \leq x) = P(-x \leq T_n \leq x) = P(T_n \leq x) - P(T_n < -x).$$

From Theorem 4, this equals

$$\begin{aligned} & \Phi(x) + \frac{1}{\sqrt{n}} p_1(x)\phi(x) + \frac{1}{n} p_2(x)\phi(x) \\ & - \left(\Phi(-x) + \frac{1}{\sqrt{n}} p_1(-x)\phi(-x) + \frac{1}{n} p_2(-x)\phi(-x) \right) + o(n^{-1}) \\ = \; & (2\Phi(x) - 1) + \frac{2}{n} p_2(x)\phi(x) + o(n^{-1}). \end{aligned} \tag{9}$$

- Thus, when the normal distribution $(2\Phi(x) - 1)$ is used as an approximation to the actual distribution $P(|T_n| \leq x)$, the error is $O(n^{-1})$.
- There is also a version of the Delta Method for Edgeworth expansions. Especially, if two random variables differ by $O_p(a_n)$ then they have the same Edgeworth expansions up to $O(a_n)$.

Theorem 5 (Delta method Edgeworth expansions) Suppose that the distribution of a random variable $\tilde{T}_n$ has the Edgeworth expansion

$$P(\tilde{T}_n \leq x) = \Phi(x) + a_n^{-1} p_1(x)\phi(x) + o(a_n^{-1})$$

and a random variable T_n satisfies $T_n = \tilde{T}_n + o_p(a_n^{-1})$. Then, T_n has the following Edgeworth expansion

$$P(T_n \leq x) = \Phi(x) + a_n^{-1} p_1(x)\phi(x) + o(a_n^{-1}).$$

Proof. [Proof of Theorem 5] From the assumption $T_n = \tilde{T}_n + o_p(a_n^{-1})$, for any $\epsilon > 0$, there is n sufficiently large such that $P(|T_n - \tilde{T}_n| > a_n^{-1}\epsilon) \leq \epsilon$. Then,

$$\begin{aligned} P(T_n \leq\ & x) = P(T_n \leq x, |T_n - \tilde{T}_n| \leq a_n^{-1}\epsilon) + \epsilon \\ \leq\ & P(\tilde{T}_n \leq x + a_n^{-1}\epsilon) + \epsilon \\ =\ & \Phi(x + a_n^{-1}\epsilon) + a_n^{-1} p_1(x + a_n^{-1}\epsilon)\phi(x + a_n^{-1}\epsilon) + \epsilon + o(a_n^{-1}) \\ \leq\ & \Phi(x) + a_n^{-1} p_1(x)\phi(x) + o(a_n^{-1}). \end{aligned}$$

Similarly, one can show that $P(T_n \leq x) \geq \Phi(x) + a_n^{-1} p_1(x)\phi(x) + o(a_n^{-1})$.

■

1.3 Cornish-Fisher Expansions

- For some purposes it is useful to have similar expansions for the inverse of the distribution function, which are the quantiles of the distribution. Such expansions are known as **Cornish-Fisher expansions of finite sample quantile**.
- This will be a key device to prove a high-order refinement of percentile-t bootstrap.

Theorem 6 (Cornish-Fisher expansions) Suppose that the distribution of a random variable T_n has the Edgeworth expansion

$$G_n(x) = P(T_n \leq x) = \Phi(x) + \frac{1}{\sqrt{n}} p_1(x)\phi(x) + \frac{1}{n} p_2(x)\phi(x) + o\left(\frac{1}{n}\right) \tag{10}$$

uniformly in x. For any $\alpha \in (0,1)$, let q_α and z_α be the α-th quantile of $G_n(\cdot)$ and $\Phi(\cdot)$, that is the solutions to $G_n(q_\alpha) = \alpha$ and $\Phi(z_\alpha) = \alpha$. Then,

$$q_\alpha = z_\alpha + \frac{1}{\sqrt{n}} p_{11}(z_\alpha) + \frac{1}{n} p_{21}(z_\alpha) + o\left(\frac{1}{n}\right),$$

where

$$\begin{aligned} p_{11}(x) &= -p_1(x); \\ p_{21}(x) &= -p_2(x) + p_1(x)p_1'(x) - \frac{1}{2}xp_1(x)^2. \end{aligned}$$

- Let $\tilde{q}_\alpha = z_\alpha + \frac{1}{\sqrt{n}}p_{11}(z_\alpha) + \frac{1}{n}p_{21}(z_\alpha)$ be an asymptotically refined approximation of the true quantile q_α. By Edgeworth expansion Delta method, T_n and $T_n + (q_\alpha - \tilde{q}_\alpha) = T_n + o(n^{-1})$ have the same Edgeworth expansion to order $o(n^{-1})$. Thus,

$$\begin{aligned} G_n(\tilde{q}_\alpha) &= P(T_n \le \tilde{q}_\alpha) \\ &= P(T_n + (q_\alpha - \tilde{q}_\alpha) \le q_\alpha) \\ &= P(T_n \le q_\alpha) + o(n^{-1}) \\ &= \alpha + o(n^{-1}). \end{aligned}$$

  In contrast, the standard first-order approximated quantile z_α is with

$$G_n(z_\alpha) = \alpha + O\left(\frac{1}{\sqrt{n}}\right),$$

  which follows by the Edgeworth expansion in (5).

Proof. We derive the result of Theorem 6 using Taylor expansions. Evaluating the Edgeworth expansion (10) at $\tilde{q}_\alpha$, we have

$$G_n(\tilde{q}_\alpha) = \Phi(\tilde{q}_\alpha) + \frac{1}{\sqrt{n}} p_1(\tilde{q}_\alpha)\phi(\tilde{q}_\alpha) + \frac{1}{n} p_2(\tilde{q}_\alpha)\phi(\tilde{q}_\alpha) + o\left(\frac{1}{n}\right).$$

For the first term,

$$\Phi(\tilde{q}_\alpha) = \Phi\left(z_\alpha + \frac{1}{\sqrt{n}} p_{11}(z_\alpha) + \frac{1}{n} p_{21}(z_\alpha)\right),$$

and the second term and third term are

$$\begin{aligned} \frac{1}{\sqrt{n}} p_1(\tilde{q}_\alpha)\phi(\tilde{q}_\alpha) &= \frac{1}{\sqrt{n}} p_1\left(z_\alpha + \frac{1}{\sqrt{n}} p_{11}(z_\alpha)\right) \phi\left(z_\alpha + \frac{1}{\sqrt{n}} p_{11}(z_\alpha)\right) + o\left(\frac{1}{n}\right); \\ \frac{1}{n} p_2(\tilde{q}_\alpha)\phi(\tilde{q}_\alpha) &= \frac{1}{n} p_2(z_\alpha)\phi(z_\alpha) + o\left(\frac{1}{n}\right). \end{aligned}$$

Summing up,

$$\alpha + o\left(\frac{1}{n}\right) = G_n(\tilde{q}_\alpha) = \Phi\left(z_\alpha + \frac{1}{\sqrt{n}}p_{11}(z_\alpha) + \frac{1}{n}p_{21}(z_\alpha)\right) + \frac{1}{\sqrt{n}}p_1\left(z_\alpha + \frac{1}{\sqrt{n}}p_{11}(z_\alpha)\right)\phi\left(z_\alpha + \frac{1}{\sqrt{n}}p_{11}(z_\alpha)\right) + \frac{1}{n}p_2(z_\alpha)\phi(z_\alpha) + o\left(\frac{1}{n}\right).$$

Now, expand $\Phi(\cdot)$ in a second-order Taylor expansion and $p_1(\cdot)$ and $\phi(\cdot)$ in first-order expansions, both about z_α, using $\phi(x)' = -x\phi(x)$, and get

$$\frac{1}{\sqrt{n}}\underbrace{p_1\left(z_\alpha + \frac{1}{\sqrt{n}}p_{11}(z_\alpha)\right)}_{=p_1(z_\alpha)+\frac{1}{\sqrt{n}}p_{11}(z_\alpha)p_1'(z_\alpha)+O(\frac{1}{n})} \times \underbrace{\phi\left(z_\alpha + \frac{1}{\sqrt{n}}p_{11}(z_\alpha)\right)}_{=\phi(z_\alpha)-\frac{1}{\sqrt{n}}z_\alpha p_{11}(z_\alpha)\phi(z_\alpha)+O(\frac{1}{n})}$$
$$= \frac{1}{\sqrt{n}}p_1(z_\alpha)\phi(z_\alpha) + \frac{1}{n}p_1'(z_\alpha)p_{11}(z_\alpha)\phi(z_\alpha) - \frac{1}{n}z_\alpha p_1(z_\alpha)p_{11}(z_\alpha)\phi(z_\alpha) + o\left(\frac{1}{n}\right),$$

and

$$\begin{aligned} & \Phi\left(z_\alpha + \frac{1}{\sqrt{n}} p_{11}(z_\alpha) + \frac{1}{n} p_{21}(z_\alpha)\right) \\ = \; & \Phi\left(z_\alpha\right) + \frac{1}{\sqrt{n}} p_{11}(z_\alpha)\phi(z_\alpha) + \frac{1}{n}\left(p_{21}(z_\alpha) - \frac{z_\alpha p_{11}^2(z_\alpha)}{2}\right)\phi(z_\alpha) + o\left(\frac{1}{n}\right). \end{aligned} \tag{11}$$

Combining these into (11), we get

$$\begin{aligned} G_n(\tilde{q}_\alpha) \; = \; & \Phi\left(z_\alpha + \frac{1}{\sqrt{n}} p_{11}(z_\alpha) + \frac{1}{n} p_{21}(z_\alpha)\right) \\ & + \frac{1}{\sqrt{n}} p_1\left(z_\alpha + \frac{1}{\sqrt{n}} p_{11}(z_\alpha)\right)\phi\left(z_\alpha + \frac{1}{\sqrt{n}} p_{11}(z_\alpha)\right) + \frac{1}{n} p_2(z_\alpha)\phi(z_\alpha) + o\left(\frac{1}{n}\right) \\ = \; & \underbrace{\Phi\left(z_\alpha\right)}_{=\alpha} + \frac{1}{\sqrt{n}}\phi(z_\alpha)\left(p_{11}(z_\alpha) + p_1\left(z_\alpha\right)\right) \\ & + \frac{1}{n}\left(p_{21}(z_\alpha) - \frac{z_\alpha p_{11}^2(z_\alpha)}{2} + p_1'\left(z_\alpha\right) p_{11}(z_\alpha) - z_\alpha p_1(z_\alpha) p_{11}(z_\alpha) + p_2(z_\alpha)\right)\phi(z_\alpha) \\ & + o\left(\frac{1}{n}\right). \end{aligned}$$

For the first three terms to be equal to α, we deduce that

$$\begin{aligned}
p_{11}(x) &= -p_1(x); \\
p_{21}(x) &= -p_2(x) + xp_1(x)p_{11}(x) - p_1'(x)\,p_{11}(x) + \frac{xp_{11}^2(x)}{2} \\
&= -p_2(x) - xp_1^2(x) + \frac{xp_1^2(x)}{2} + p_1'(x)\,p_1(x) \\
&= -p_2(x) + p_1'(x)\,p_{11}(x) - \frac{1}{2}xp_1(x)^2,
\end{aligned}$$

which is the desired result.

■

2 Bootstrap and Asymptotic Refinements

- The distribution of any estimator or statistic is determined by the distribution of the data. While the latter is unknown, it can be estimated by the empirical distribution of the data which is the idea of the bootstrap.
- Let $F(u) = P(w_i \leq u)$ denote the (unknown) distribution of an individual observation w_i, and $W_n = (w_1, \ldots, w_n)'$. Let $G_n(u, F)$ be the distribution of a asymptotically pivotal test-statistic T_n for an estimator $\hat{\theta}_n = \hat{\theta}(W_n)$. That is,

$$G_n = G_n(u, F) = P(\hat{\theta}_n \leq u|F).$$

- In practice, there are two barriers to implement $G_n(u, F)$. The first barrier is that the analytical calculation of the exact distribution $G_n(u, F)$ is infeasible in most cases even if we know what is F, except in certain special cases such as the normal regression model. The second barrier is in general we do not know about F.
- The bootstrap method simultaneously circumvents these two barriers by two clever

ideas. First, we replace F with $\hat{F}_n$, an empirical distribution of w_i, i.e.

$$\hat{F}_n := \hat{F}_n(u) = \frac{1}{n}\sum_{i=1}^{n} 1(w_i \leq u),$$

and obtain the ideal bootstrap estimator of the distribution of T_n(for $\hat{\theta}_n$) as

$$\hat{G}_n^* = G_n(u, \hat{F}_n).$$

- Note that the moments and cumulants generated by $\hat{F}_n$ are exact sample counter parts, i.e. sample moments and sample cumulants, respectively, of population ones generated by $F(u)$.

- Still, G_n^* is unknown in practice. But the bootstrap proposes estimation of G_n^* by simulation $\{W_{n,b}^*\}_{b=1}^B$ from $\hat{F}_n$ which is simply sampling each random draws $W_{n,b}^* = (w_{1,b}^*, \ldots, w_{n,b}^*)'$with replacement from the original data many many times. After the bootstrap sampling, we apply the estimation formula $\hat{\theta}_n = \hat{\theta}(W_n)$ and T_n, and thus obtain i.i.d draws from $\{\hat{\theta}_b^*\}_{b=1}^B$ and $\{\hat{\theta}_b^*\}_{b=1}^B$with $\{T_{n,b}^*\}_{b=1}^B$. By making a large number of B which is close to ∞ of bootstrap draws $\{\hat{\theta}_b\}_{b=1}^B$, we can equivalently calculate any feature of $\hat{G}_n^*$ of interest.

2.1 Percentile-t asymptotic refinements

- Recall that the one-sided asymptotic confidence interval have accuracy to order $O(n^{-1/2})$.
- From Theorem 6, Cornish-Fisher expansion of the true quantile $q_{\alpha,n} := q_\alpha$ for a t-ratio T_n for $\hat{\theta}_n$ in the smooth function model,

$$q_\alpha = z_\alpha + \frac{1}{\sqrt{n}} p_{11}(z_\alpha) + O\left(\frac{1}{n}\right),$$

where $p_{11}(x)$ is an even polynomial of order 2 with coefficients depending on the moments of $W_i = h(X_i)$ generated by $F(u)$ up to order 8.

- The bootstrap quantile $q^*_{\alpha,n} = q^*_\alpha$ of T^*_n from $\{\hat{\theta}^*_b\}_{b=1}^B$ has a similar Cornish-Fisher expansion

$$q^*_\alpha = z_\alpha + \frac{1}{\sqrt{n}} p^*_{11}(z_\alpha) + O_p\left(\frac{1}{n}\right),$$

where $p^*_{11}(x)$ is the same as $p_{11}(x)$ except the moments of W_i are replaced by the corresponding sample moments generated by bootstraped sample $W^*_{n,b}$ (bootstraped distribution $\hat{F}_n$).

- The sample moments in the definitions of $p^*_{11}(z_\alpha)$ are (automatically) estimated at the rate $n^{-1/2}$, i.e. $p^*_{11}(z_\alpha) = p_{11}(z_\alpha) + O_p\left(\frac{1}{n}\right)$. Thus we can replace p^*_{11} with p_{11} without affecting the order of the expansion, i.e.

$$\begin{aligned} q^*_\alpha &= z_\alpha + \frac{1}{\sqrt{n}} p_{11}(z_\alpha) + O_p\left(\frac{1}{n}\right) \\ &= q_\alpha + O_p\left(\frac{1}{n}\right), \end{aligned}$$

  which shows that the boostrap quantiles of q^*_α of the *studentized* t-ratio are within $O_p\left(\frac{1}{n}\right)$ of the exact quantiles of q_α.

- By Edgeworth expansion Delta method, Corollary 1, T_n and $T_n + (q_\alpha - q^*_\alpha) = T_n + O_p(n^{-1})$ have the same Edgeworth expansion to order $O(n^{-1})$. Thus,

$$\begin{aligned} P(T_n &\leq q^*_\alpha) = P(T_n + (q_\alpha - q^*_\alpha) \leq q_\alpha) \\ &= P(T \leq q_\alpha) + O(n^{-1}) \\ &= \alpha + O(n^{-1}), \end{aligned}$$

which is the improved rate of the convergence relative to the one-sided test using asymptotic critical value z_α. Similarly, the coverage of the percentile-t interval is

$$\begin{aligned} P(\theta \in C^{\text{pt}}) &= P(q^*_{\alpha/2} \leq T \leq q^*_{1-\alpha/2}) \\ &= P(T \leq q^*_{1-\alpha/2}) - P(T < q^*_{\alpha/2}) \\ &= (1-\alpha/2) - \alpha/2 + O(n^{-1}) \\ &= 1 - \alpha + O(n^{-1}). \end{aligned}$$

- In the smooth function model the t-test (with correct standard error) has the following performance.

Theorem 7 Under the assumption of Corollary 1,

$$q^*_{1-\alpha} = q_{1-\alpha} + o_p(n^{-1}),$$

where $q^*_{1-\alpha}$ *and* $q_{1-\alpha}$ *are* $(1-\alpha)$*th quantile of the distribution of* $|T^*_n|$ *and* $|T_n|$*, respectively. Also, the asymptotic test "Reject* H_0*" in favor of* H_1 *if* $|T_n| \geq z_{1-\alpha}$*" has accuracy*

$$P\left(|T_n| > z_{1-\alpha/2} | H_0\right) = 1 - \alpha + O(n^{-1}),$$

and the symmetric two-sided (percentile) bootstrap t-test "Reject H_0 *in favor of* H_1 *if*

$|T| \geq q^*_{1-\alpha}$ *has accuracy*

$$P\left(|T_n| > q^*_{1-\alpha}|H_0\right) = 1 - \alpha + o(n^{-1}).$$

Proof. We haven shown in (9)

$$P\left(|T_n| \leq x|H_0\right) = (2\Phi(x) - 1) + \frac{2}{n}p_2(x)\phi(x) + o(n^{-1}),$$

which means the asymptotic test has accuracy of order $O(n^{-1})$. Given the Edgeworth expansion, we apply the Cornish-Fisher expansion in Theorem 6 for the α-th quantile q_α of the distribution of $|T_n|$ takes the form

$$q_\alpha = z_\alpha + \frac{1}{n}p_{21}(z_\alpha) + o(n^{-1}).$$

Also, the bootstrap quantile q^*_α has the Cornish-Fisher expansion,

$$\begin{aligned} q^*_\alpha &= z_\alpha + n^{-1}p^*_{21}(z_\alpha) + o(n^{-1}); \\ &= z_\alpha + n^{-1}p_{21}(z_\alpha) + o_p(n^{-1}); \\ &= q_\alpha + o_p(n^{-1}), \end{aligned}$$

where $p_{21}^{*}(x)$ is the same as $p_{21}(x)$ except the moment of $h(X_i)$ are replaced by the corresponding sample moments. Then, the bootstrap test has rejection probability, using the Edgeworth expansion Delta method in Theorem **??** as

$$\begin{aligned} P\left(|T_n| \geq q_{1-\alpha}^{*}| \, H_0\right) &= P\left(|T_n| + (q_{1-\alpha} - q_{1-\alpha}^{*}) \geq q_{1-\alpha}| \, H_0\right) \\ &= P\left(|T| \geq q_{1-\alpha}| \, H_0\right) + o(n^{-1}) \\ &= 1 - \alpha + o(n^{-1}) \end{aligned}$$

as claimed! ■

3 Review: Edgeworth Expansions for Sample Mean

3.1 When σ^2 is known

Let $\{X_i\}_{i=1}^n$ be an i.i.d sequence of random variables having mean μ and (known) variance σ^2, and define

$$\begin{aligned} Z_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(X_i - E[X_i])}{Var[X_i]^{1/2}} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} \end{aligned}$$

which is a studentized statistics. Assuming $E(X_i^4) < \infty$, and the characteristic function of X_i, $\varphi_X(\cdot)$, satisfies $\limsup_{|t|\to\infty} |\varphi_X(t)| < 1$, we have

$$\begin{aligned} G_n(x; F) &= P(Z_n \le x) \\ &= \Phi(x) - \frac{1}{\sqrt{n}} p_1(x)\phi(x) - \frac{1}{n} p_2(x)\phi(x) + o\left(\frac{1}{n}\right), \end{aligned}$$

where

$$p_1(x) = \frac{\kappa_3}{6}(x^2 - 1);$$
$$p_2(x) = \frac{\kappa_4}{24}(x^3 - 3x) + \frac{\kappa_3^2}{72}(x^5 - 10x^3 + 15x).$$

Also, if we apply Cornish-Fisher expansion to the finite-sample $(1-\alpha)$-th quantile $q_{Z_n,\alpha}$, we get

$$q_\alpha = z_\alpha + \frac{1}{\sqrt{n}} p_{11}(z_\alpha) + \frac{1}{n} p_{22}(z_\alpha) + o\left(\frac{1}{n}\right),$$

where z_α is $(1-\alpha)$th quantile of standard normal, and

$$p_{11}(x) = -p_1(x);$$
$$p_{22}(x) = -p_2(x) + p_1(x)p_1'(x) - \frac{1}{2}xp_1(x)^2.$$

- Now, let $\hat{F}(x) = \frac{1}{n}\sum_{i=1}^n 1(X_i \leq x)$ be an empirical cdf of $\{X_i\}_{i=1}^n$. Then, one can always simulate a b-th bootstraped sample $\{X_{i,b}^*\}_{i=1}^n$ from $\hat{F}(\cdot)$. Note that conditioning on $\{X_i\}_{i=1}^n$, the expected value of the bootstraped sample $\{X_{i,b}^*\}_{b=1}^B$ is

$$E^*[X_{i,b}^*] := \int_{\mathbb{R}} X_{i,b}^* d\hat{F} = \frac{1}{n}\sum_{i=1}^n X_i = \hat{\mu}_1.$$

Similarly,

$$\begin{aligned} Var^*[X_{i,b}^*] &= E^*[X_{i,b}^{*2}] - E^*[X_{i,b}^*]^2 \\ &= \frac{1}{n}\sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 \\ &= \hat{\mu}_2, \end{aligned}$$

and corresponding 3rd to 6th cumulants of $\{X_{i,b}^*\}_{b=1}^B$ are

$$\begin{aligned} \hat{\kappa}_3 &= \hat{\mu}_3,\ \hat{\kappa}_4 = \hat{\mu}_4 - 3\hat{\mu}_2^2; \\ \hat{\kappa}_5 &= \hat{\mu}_5 - 10\hat{\mu}_3\hat{\mu}_2; \\ \hat{\kappa}_6 &= \hat{\mu}_6 - 15\hat{\mu}_4\hat{\mu}_2 - 10\hat{\mu}_3^2 + 30\hat{\mu}_2^3. \end{aligned}$$

- Let

$$Z^*_{n,b} = \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n (X^*_{i,b} - E^*[X^*_{i,b}])}{\sigma}$$
$$= \frac{1}{\sqrt{n}} \frac{\sum_{i=1}^n (X^*_{i,b} - \bar{X})}{\sigma}$$

be a bootstrap analogue of the sample statistic.

- Note that the bootstrap simulation of $\{Z^*_{n,b}\}_{b=1}^B$ enables us to numerically calculate its cdf

$$G_n(x, \hat{F}) = P(Z^*_{n,b} \leq x)$$

with its $(1-\alpha)$th quantile $q_{Z^*_n,\alpha}$. Also, one may show that there is also an Edgeworth Expansion of probability distribution the random variable $Z^*_{n,b}$ as below

$$G_n(x, \hat{F}) = P(Z^*_{n,b} \leq x) = \Phi(x) - \frac{1}{\sqrt{n}}\hat{p}_1(x)\phi(x) - \frac{1}{n}\hat{p}_2(x)\phi(x) + o_p\left(\frac{1}{n}\right),$$

with

$$\hat{p}_1(x) = \frac{\hat{\kappa}_3}{6}(x^2-1);$$
$$\hat{p}_2(x) = \frac{\hat{\kappa}_4}{24}(x^3-3x) + \frac{\hat{\kappa}_3^2}{72}(x^5-10x^3+15x),$$

and corresponding Cornish-Fisher expansion to $(1-\alpha)$th quantile $q_{Z_n^*,\alpha}$,

$$q_{Z_n^*,\alpha} = z_\alpha + \frac{1}{\sqrt{n}}\hat{p}_{11}(x) + \frac{1}{n}\hat{p}_{22}(x) + o_p\left(\frac{1}{n}\right),$$

with

$$\hat{p}_{11}(x) = -\hat{p}_1(x);$$
$$\hat{p}_{22}(x) = -\hat{p}_2(x) + \hat{p}_1(x)\hat{p}_1'(x) - \frac{1}{2}x\hat{p}_1(x)^2.$$

- Since $\hat{\kappa}_r = \kappa_r + O_p(\frac{1}{\sqrt{n}})$, this implies

$$q_{Z_n^*,\alpha} = q_{Z_n,\alpha} + O_p\left(\frac{1}{n}\right),$$

we conclude that a simulation of bootstraped $q_{Z_n^*,\alpha}$ given data set $\{X_i\}_{i=1}^n$ automatically approximate the true quantile $q_{Z_n,\alpha}$ upto the order $O_p(\frac{1}{n})$.

3.2 When σ^2 is unknown.

- Define

$$T_n(X_1, \ldots, X_n) = T_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(X_i - \mu)}{\hat{\sigma}};$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} X_i \right)^2.$$

- Deriving the Edgeworth expansion for t_n is much more involved than Z_n. This is because the form of $T_n(X_1, \ldots, X_n)$ is a non-linear function of the original sample.

Theorem 8 For a sequence d-dimensional i.i.d. random vector $W_i = (W_{i,1}, W_{i,2}, \ldots, W_{i,d})'$ define $Z = (Z_1, Z_2, \ldots, Z_d)' = \sqrt{n}(\bar{W} - \beta)$. Let $S_n := S_n(Z)$ be a statistic such that $S_n \xrightarrow{d} N(0,1)$ such as T_n. Define a sequence of approximating statistic $U_n := U_{n,r}(Z)$,

$U_{n,r}$ is r-th order polynomials in Z:

$$U_n = \sum_{i=1}^{d} b_i Z_i + n^{-1/2} \sum_{i_1,i_2=1}^{d} b_{i_1 i_2} Z_{i_1} Z_{i_2} + \dots$$
$$+ n^{-(r-1)} \sum_{i_1,i_2,\dots,i_r=1}^{d} b_{i_1 i_2 \dots i_r} Z_{i_1} \dots Z_{i_r},$$

for some positive integer r and non-random coefficients b_j's. Assume that each component of W_i has finite $(j \times r)$-th moments. Then,

i) The j-th cumulant of the approximating statistic U_n has the form

$$\kappa_{j,n} = n^{-(j-2)/2}(k_{j,1} + \frac{1}{n} k_{j,2} + \frac{1}{n^2} k_{j,3} + \dots),$$

where the coefficients $k_{j,l}$ depends only on the coefficients b'_ms in U_n, on the moments of each component of W_i up to the $(j \times r)$-th, and the value of $k_{j,l}$ does not depend on r for $l \leq r + 2 - j$.

iii) If $S_n = U_n + o_p(\frac{1}{n})$, S_n admits the following form of Edgeworth expansion:

$$P(S_n \leqq x) = \Phi(x) + \frac{1}{\sqrt{n}} q_1(x)\phi(x) + \frac{1}{n} q_2(x)\phi(x) + o(\frac{1}{n}),$$

uniformly in x, where

$$\begin{aligned} q_1(x) &= -(k_{1,2} + \frac{1}{6} k_{3,1}(x^2 - 1)); \\ q_2(x) &= -x \left\{ \frac{1}{2}(k_{2,2} + k_{1,2}^2) + \frac{1}{24}(k_{4,1} + 4k_{1,2}k_{3,1})(x^2 - 3) \right. \\ &\quad \left. + \frac{1}{72} k_{3,1}^2 (x_4 - 10x^2 + 15) \right\}. \end{aligned}$$

- Define

$$W_i = (X_i, X_i^2)' \text{ and } \bar{W} = \frac{1}{n} \sum_{i=1}^{n} W_i.$$

Without loss of generality, assume $E[W_i] = \beta = (0, 1)'$, i.e. $E[X_i] = 0$ and $E[X_i^2] = 1$.

- Using

$$\frac{x}{\sqrt{1+x}} = x(1+x)^{-1/2} = x\left(1 - \frac{1}{2}x + \frac{1}{4}x^2\right) + O(x^4),$$

as $x \to 0$, it is easy to check that

$$\begin{aligned} T_n &= \sqrt{n}A(\bar{W}) = \frac{\sqrt{n}\bar{X}}{\sqrt{1 + n^{-1}\sum_{i=1}^{n}(X_i^2 - 1) - \bar{X}^2}} \\ &= \underbrace{\sqrt{n}\bar{X}\left[1 - \frac{1}{2n}\sum_{i=1}^{n}(X_i^2 - 1) + \frac{1}{2}\bar{X}^2 + \frac{3}{8}\left\{\frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 1)\right\}^2\right]}_{=U_n} + O_p(n^{-3/2}), \end{aligned}$$

and directly compute the cumulants of U_n as

$$\begin{aligned}
\kappa_{1,n} &= E[U_n] = -\frac{1}{2\sqrt{n}}\gamma + O(n^{-3/2}) \\
\kappa_{2,n} &= E[U_n^2] - E[U_n]^2 = 1 + \frac{1}{4n}(7\gamma^2 + 12) + O(n^{-2}) \\
\kappa_{3,n} &= E[U_n^3] - 3E[U_n^2]E[U_n] + 2(E[U_n])^3 \\
&= -\frac{2}{\sqrt{n}}\gamma + O(n^{-3/2}) \\
\kappa_{4,n} &= \frac{1}{n}(12\gamma^2 - 2\kappa + 6) + O(n^{-2}),
\end{aligned}$$

where $\gamma = E(X-\mu)^3/\sigma^3$ and $\kappa = E(X-\mu)^4/\sigma^4 - 3$.

- Compare the expansions of cumulants in Theorem-i), and we get

$$k_{1,2} = -\frac{1}{2}\gamma,\ k_{2,2} = \frac{1}{4}(7\gamma^2 + 12),\ k_{3,1} = -2\gamma,\ k_{4,1} = 12\gamma^2 - 2\kappa + 6,$$

- By Theorem-ii), the exact forms in the Edgeworth Expansion of t_n are given as as

$$
\begin{aligned}
q_1(x) &= \frac{1}{6}\gamma(2x^2+1); \\
q_2(x) &= x\left\{\frac{1}{12}\kappa(x^2-3) - \frac{1}{18}\gamma^2(x^4+2x^2-3) - \frac{1}{4}(x^2+3)\right\}.
\end{aligned}
$$

- One may also show that the Edgeworth Expansion of

$$
P(T^*_{n,b} \le x) = \Phi(x) + \frac{1}{\sqrt{n}}\hat{q}_1(x)\phi(x) + \frac{1}{n}\hat{q}_2(x)\phi(x) + o_p(\frac{1}{n}),
$$

with

$$
\begin{aligned}
T^*_{n,b} &= \frac{1}{\sqrt{n}}\frac{\sum_{i=1}^n (X^*_{i,b} - \bar{X})}{\hat{\sigma}^*},\ (\sigma^*)^2 = \frac{\sum_{i=1}^n (X^*_{i,b} - \bar{X}^*)}{n}; \\
\hat{q}_1(x) &= \frac{1}{6}\hat{\gamma}(2x^2+1); \\
\hat{q}_2(x) &= x\left\{\frac{1}{12}\hat{\kappa}(x^2-3) - \frac{1}{18}\hat{\gamma}^2(x^4+2x^2-3) - \frac{1}{4}(x^2+3)\right\}.
\end{aligned}
$$

4 Higher order properties of the Wild bootstrap under misspecification (Kline and Santos, 2012)

- Consider a linear regression model

$$y_i = x_i'\beta_0 + u_i \text{ for } i = 1, \ldots, n.$$

with $x_i = (1, \tilde{x}_i')'$, $(y_i, \tilde{x}_i') = (y_i, \tilde{x}_{i1}, ..., \tilde{x}_{ik})' \in \mathbb{R}^{k+1}$ are i.i.d data and estimate $\hat{\beta} = \left[\sum_{i=1}^n x_i x_i'\right]^{-1} \left[\sum_{i=1}^n x_i y_i\right]$.

- **Basic Assumptions**

i) $E[x_i x_i']$ and $E[u_i^2 x_i x_i']$ are finite full rank (positive definite) matrices with $\beta_0 = [E[x_i x_i']]^{-1} E[x_i$

ii) $Eu_i^4 < \infty$ and $E\|x_i\|^4 < \infty$.

- If Assumptions i)–ii) hold, for any non-zero $c \in \mathbb{R}^k$, we have

$$Z_n, T_n \xrightarrow{d} N(0,1);$$
$$Z_n := \frac{\sqrt{n}(c'\hat{\beta} - c'\beta_0)}{\sigma};$$
$$T_n := \frac{\sqrt{n}(c'\hat{\beta} - c'\beta_0)}{\hat{\sigma}},$$

where

$$\sigma^2 = c' H_0^{-1} \Sigma_0 H_0^{-1} c,\ H_0 = E(x_i x_i'),\ \Sigma_0 = E[u_i^2 x_i x_i'];$$
$$\hat{\sigma}^2 = c' H_n^{-1} \Sigma_n(\hat{\beta}) H_n^{-1} c\ ,\ H_n = \frac{1}{n}\sum_{i=1}^{n} x_i x_i',\ \Sigma_n(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n} x_i x_i'(y_i - x_i'\hat{\beta})^2.$$

- Note that the population regression coefficient β_0 always satisfies

$$E[x_i u_i] = E[x_i(y_i - x_i'\beta_0)] = 0,$$

by construction.

- This does not necessarily imply

$$E[u|x_i] = 0$$

unless $E[y|x_i = x] = x'\beta_0$ which implies there is no model misspecification for the linear functional form of $E[y|x_i = x] = g(x)$.

- Note that although $g(x) \neq x_i'\beta_0$, the linear regression function $x'\beta_0$ can be still meaningful target in the sense that it becomes best (MSE minimizing) linear approximation of $g(x)$. For details, see the seminar papers "*Using least square to approximate* *unknown regression functions (1980)* " and "*Maximum likelihood estimation of misspecified. models (1982)*" by H. White.

- Kline and Santos (2012, JOE) focuses on the approximating nature of linear regression function and connects it to the asymptotic refinement of bootstrap inference. To see this, we begin with finding Edgeworth expansions of the sample statistic T_n under both correct specification and misspecification. We refine the previous assumptions as below
  ii)' $E[\|x_i x_i'\|^v] < \infty$ and $E[||u_i^2 x_i x_i'||^v] < \infty$ for some $v \geq 9$.
  iii) For $w_i := (\tilde{x}_i', x_i' u_i,$vech$(x_i x_i')'$, vech$(x_i x_i' u_i^2))'$, the characteristic function of w_i satisfies $\limsup_{|t|\to\infty} |\varphi_w(t)| < 1$.

- Further define

$$\gamma_0 = E[(c' H_0^{-1} x_i)^2 x_i u_i] \text{ and } \gamma_1 = E[(c' H_0^{-1} x_i)(x_i' H_0^{-1} x_i) u_i]$$
$$\lambda = E[(c' H_0^{-1} x_i)^3 u_i^3].$$

Theorem 9 Under Assumptions i), ii)', and iii), the following Edgeworth approximations of T_n hold uniformly in $z \in \mathbb{R}$.

i) If the linear regression is correctly specified with $E[u_i|x_i] = 0,$

$$P(T_n \le z) = \Phi(z) + \frac{1}{\sqrt{n}} p_{11}(z)\phi(z) + o\left(\frac{1}{\sqrt{n}}\right);$$
$$p_{11}(z) = -\frac{\lambda}{6\sigma^3}(2z^2 + 1).$$

ii) If the linear regression is misspecified. with $E[u_i|x_i] \neq 0$*, and only satisfying* $E[x_i u_i] = 0,$

$$P(T_n \le z) = \Phi(z) + \frac{1}{\sqrt{n}}(p_{11}(z) + p_{12}(z)\phi(z)) + o\left(\frac{1}{\sqrt{n}}\right);$$
$$p_{12}(z) = \frac{\lambda}{\sigma^3}(c' H_0^{-1} \Sigma_0 H_0^{-1} \gamma_0 (z^2 + 1) - \gamma_1 \sigma^2).$$

- The result of Theorem shows that there exists an extra $O(n^{-1/2})$ term in the Edgeworth expansion of T_n when the conditional mean function is misspecified.
- Thus, higher-order refinement of bootstraped inference depends on how the bootstrap procedure mimics the misspecified. nature of population.
- Let us first consider the non-parametric i.i.d boostrap procedure. Bootstrap a b-th

sample $\{(y^*_{i,b}, \tilde{x}^{*\prime}_{i,b})\}_{i=1}^n$ and compute b-th boostrap OLS estimator

$$\hat{\beta}^*_b = \left[\sum_{i=1}^n x^*_{i,b}x^{*\prime}_{i,b}\right]^{-1}\left[\sum_{i=1}^n x^*_{i,b}y^*_{i,b}\right].$$

By construction, each Bootstrap sample satisfies

$$\begin{aligned} y^*_{i,b} &= x^{*\prime}_{i,b}\hat{\beta} + u^*_{i,b} \text{ for } i = 1, \ldots, n. \\ E^*[x^*_{i,b}u^*_{i,b}] &= E^*[x^*_{i,b}y^*_{i,b}] - E^*[x^*_{i,b}x^{*\prime}_{i,b}\hat{\beta}] \\ &= n^{-1}\sum_{i=1}^n x_i y_i - n^{-1}\sum_{i=1}^n x_i x_i'\hat{\beta} \\ &= 0. \end{aligned}$$

but not necessarily

$$E^*[u^*_i|x^*_{i,b}] = 0$$

because it can be

$$E^*[y^*_{i,b}|x^*_{i,b}] \neq x^{*\prime}_{i,b}\hat{\beta}^*_b.$$

Hall and Horowitz (1996) proves the asymptotic refinement of non-parametric i.i.d bootstrap in the context of the moment condition model $E[x_i(y_i - x_i'\beta)] = 0$

- Let us consider the second boostrap procedure, so called "wild"-bootstrap proposed by Wu (1986) and Liu (1988). The "*wildness"* feature of this bootstrap procedure is because it keeps the regressor $\{\tilde{x}_i\}_{i=1}^n$, and only generates the b-th Bootstraped errors and dependent variables as below:

$$y_{i,b}^* = x_i'\hat{\beta} + \hat{u}_{i,b}^*, \text{ with } \hat{u}_{i,b}^* = \hat{u}_i V_{i,b},$$

where $V_b := \{V_{i,b}\}_{i=1}^n$ is an i.i.d sample independent (over i and b) of original data $(y_i, \tilde{x}_i')'$ such that $E[V_{i,b}] = 0$ and $E[V_{i,b}^2] = 1$. Then, the corresponding Bootstrap OLS estimator is

$$\hat{\beta}_b^* = \left[\sum_{i=1}^n x_i x_i'\right]^{-1} \left[\sum_{i=1}^n x_i y_{i,b}^*\right].$$

- It is easy to check

$$
\begin{aligned}
E^*[y^*_{i,b}|x_i] &= x_i'\hat{\beta} + E^*[\hat{u}^*_{i,b}|x_i] \\
&= x_i'\hat{\beta} + E^*[V_{i,b}]E[\hat{u}_i|x_i] \\
&= 0.
\end{aligned}
$$

  Thus, the wild Bootstrap approximation of $\hat{F}$ and corresponding conditional expectation treats $x_i'\hat{\beta}$ to be a true specification of $E^*[y^*_{i,b}|x_i] = 0$. This is very nice when the population model of $E[y_i|x_i]$ is correctly specified with $x_i'\beta$, and the asymptotic refinement is proved in Mammen (1993) and Djogbenou, MacKinnon, Nielsen (2019).

- Define the wild bootstrap statistic T_n^* as

$$
\begin{aligned}
T^*_{n,b} &= \frac{\sqrt{n}(c'\hat{\beta}^*_b - c'\hat{\beta})}{\hat{\sigma}^*_b}; \\
(\hat{\sigma}^*_b)^2 &= c' H_n^{-1}\, \Sigma_n(\hat{\beta}^*_b) H_n^{-1}\, c.
\end{aligned}
$$

Theorem 10 Assume Assumptions i), ii)', and iii), and $E[V_{i,b}] = 0$, $E[V_{i,b}^2] = 1$, *and* $E[|V_{i,b}|^v] < \infty$ *for* $v \geq 9$. *For* $U_{i,b} = (W_{i,b}, W_{i,b}^2)'$, *its characteristic function satisfies*

$\limsup_{|t|\to\infty} |\varphi_u(t)| < 1$. ***Then,***

$$P(T^*_{n,b} \le z) = \Phi(z) + \frac{1}{\sqrt{n}}\hat{p}_{11}(z)\phi(z) + o_p\left(\frac{1}{\sqrt{n}}\right);$$
$$\hat{p}_{11}(z) = -\frac{\hat{\lambda}}{6\hat{\sigma}^3}(2z^2+1).$$

5 Review of Efron (1979)'s i.i.d Bootstrap

We begin by reviewing the formulation of the i.i.d bootstrap method of Efron (1979).

- Assume that $X_1, X_2, \ldots,$ is a sequence of i.i.d random variable with common distribution F.
- Suppose $\mathcal{X}_n = \{X_{1,}X_2, \ldots, X_n\}$ is the data at hand.
- Let $T_n = t_n(\mathcal{X}_n, F)$ be a random variable of interest. We again consider the simplest example where $T_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$ with one parameter of interests $\mu = E[X_1]$.
- The bootstrap version T_n^* of T_n based on bootstrap sample $\mathcal{X}_n^* = \{X_1^*, X_2^*, \ldots, X_n^*\}$ with replacement from $\mathcal{X}_n$ is

$$T_n^* = \frac{\sqrt{n}(\bar{X}_n^* - E^*[X_1^*])}{(Var^*[X_1^*])^{1/2}} = \frac{\sqrt{n}(\bar{X}_n^* - \bar{X}_n)}{s_n}$$

  where $\bar{X}_n^* = \frac{1}{n}\sum_{i=1}^n X_i^*$ and $s_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X}_n)^2$.
- When $E[X_1^2] < \infty$, then the bootstrapped-quantile q_α^* is a "consistent" approximation of the true quantile q_α.

Theorem 11 (Berry-Essen) Let $W_1, W_2, \ldots, W_n$ *be a collection of* n *independent (but not necessarily identically distributed) random variables with* $EW_j = 0$ *and* $EW_j < \infty$ *for all* $1 \leq j \leq n$*. If* $\sigma_n^2 = n^{-1} \sum_{j=1}^n EW_j^2 > 0$*, then*

$$\sup_{x \in \mathbb{R}} \left| P\left(\frac{1}{\sqrt{n}\sigma_n} \sum_{j=1}^n W_j \leq x \right) - \Phi(x) \right| \leq (2.75) \frac{1}{n^{2/3}} \sum_{j=1}^n (E|W_j|^3/\sigma_n^3),$$

where $\Phi(x)$ *denotes the distribution function of the standard normal distribution on* $\mathbb{R}$*.*

- The non-asynmptotic bound holds for any random variable W_j with arbitary probability measure.

Theorem 12 If $X_1, X_2, \ldots$ *are i.i.d with* $\sigma^2 = Var(X_1) \in (0, \infty)$*, then* $\sup_{x \in \mathbb{R}} |P^*(T_n^* \leq x) - \Phi($
$o(1)$ *as* $n \to \infty$ *a.s.*

$$\sup_{x \in \mathbb{R}} |P^*(T_n^* \leq x) - \Phi(x)| = o(1) \text{ as } n \to \infty \text{ a.s.}$$

Since $X_1^*, X_2^*, \ldots, X_n^*$ are i.i.d, by Theorem 11,

$$\sup_{x\in\mathbb{R}} |P^*(T_n^* \leq x) - \Phi(x)| \leq (2.75)\hat{\Delta}_n,$$

where $\hat{\Delta}_n = E^*|X_1^* - \bar{X}_n|^3/(s_n^3\sqrt{n})$ and $s_n^2 = E^*(X_1^* - \bar{X}_n)^2$. By the SLLN, it is easy to check that $\hat{\Delta}_n = o(1)$ as $n \to \infty$.

- Arcones and Giné (1989, 1991) proves that Theorem 12 holds for any resample size $m_n \to \infty$ (instead of n) at a rate faster than $\log\log n$.

- Note that by the (CLT), T_n also converges in distribution to $N(0, 1)$. Hence it follows that

$$\sup_{x\in\mathbb{R}} |P^*(T_n^* \leq x) - P(T_n \leq x)| = o(1) \text{ as } n \to \infty \text{ a.s.}$$

6 Inadequacy of i.i.d Bootstrap for Dependent Data

- We point out that the general perception that the bootstrap is an "omnibus" method, giving accurate perception in all problems automatically, is misleading.
- A prime example of this appears in Singh (1981).

Definition 3 $\{X_n\}_{n\geq 1}$ *is called m-dependent for some integer* $m \geq 0$ *if* $\{X_1, \ldots, X_k\}$ *and* $\{X_{k+m+1}, \ldots, \}$ *are independent for all* $k \geq 1$.

Example 1 An i.i.d sequence of random variables $\{\epsilon_n\}_{n\geq 1}$ *is 0-dependent.*

Example 2 $X_n = \epsilon_n + 0.5\epsilon_{n+1}$ *is* 1*-dependent.*

- Let σ_m^2=Var(X$_1) + 2\sum_{i=1}^{m-1} cov(X_1, X_{1+i})$ and $\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$. If $\sigma_m^2 \in (0,\infty)$, then the following CLT for m-dependent r.v. holds

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma_m^2).$$

- Suppose we want to estimate the sampling distribution of $T_n := \sqrt{n}(\bar{X}_n - \mu)$ using the i.i.d bootstrap. Note that the true (asymptotic) distribution of T_n depends on σ_m^2. The bootstrap version of T_n^* of T_n is given by

$$T_n^* := \sqrt{n}(\bar{X}_n^* - \bar{X}_n),$$

where $\bar{X}_n^* = n^{-1}\sum_{i=1}^n X_i^*$.

- Conditioning on the original sample $\mathcal{X}_n^*$, the distribution of T_n^* still converges to a normal distribution, but with a "*wrong*" variance as shown below.

Theorem 13 Suppose $\{X_n\}_{n\geq 1}$ is a sequence of stationary m-dependent random variables with $EX_1 = \mu$ and $\sigma^2 = var(X_1) \in (0, \infty)$. Then,

$$\sup_x |P^*(T_n^* \leq x) - \Phi(\frac{x}{\sigma})| = o(1) \textit{ as } n \to \infty \textit{ a.s.}$$

Proof. Note that conditional on $\mathcal{X}_n^*$, $X_1^*, X_2^*, \ldots, X_n^*$ are i.i.d random variables. As in the proof of Theorem 12, by the Berry-Essen Theorem, we get the result of Theorem because

$$s_n^2 = E^*(X_1^* - \bar{X}_n)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \to \sigma^2 \text{ as } n \to \infty \text{ a.s.}$$

and

$$\frac{1}{n^{3/2}}\sum_{i=1}^{n}|X_i|^3 \to 0 \text{ as } n \to \infty \text{ a.s.}$$

hold by SLLN for m-dependent random variables. ■

Corollary Under the condition of the above theorem, if $\sum_{i=1}^{m-1} cov(X_1, X_{1+i}) \neq 0$ and σ_∞^2, then for any $x \neq 0$,

$$\lim_{n\to\infty} |P^*(T_n^* \leq x) - P(T_n \leq x)| = |\Phi(\frac{x}{\sigma_m}) - \Phi(\frac{x}{\sigma})| \neq 0 \text{ a.s.}$$

- Thus, for all $x \neq 0$, the i.i.d bootstrap estimator $P^*(T_n^* \leq x)$ of $P(T_n \leq x)$ has a non-zero error in the limit, and thus the bootstrap critical value $\hat{q}_\alpha$ is no-longer a consistent estimator for q_α.
- The i.i.d bootstrap method drastically fails for dependent data. This is because it

ignores the dependent structure of the sequence $\{X_n\}$ completely.

7 Bootstrap Methods for Dependent Data

7.1 Bootstrap Based on I.I.D Innovations

- Suppose $\{X_n\}$ is a sequence of random variables satisfying the following AR(p) process:

$$X_n = \beta_1 X_{n-1} + \ldots + \beta_p X_{n-p} + \epsilon_n \text{ for } n > p, \tag{12}$$

  where $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$ is the vector of the autoregressive parameters, and $\{\epsilon_n\}_{n>p}$ is an i.i.d sequence of random variables with common distribution F that are independent of the random variable $X_1, \ldots, X_p$. Assume $E(\epsilon_1) = 0$

- Let $\hat{\beta}_n$ be an OLS estimator of β based on $\mathcal{X}_n$. Suppose we want to approximate the sampling distribution of random variables $T_n = t_n(\mathcal{X}_n, F, \beta)$, e.g. $T_n = \sqrt{n}(\hat{\beta} - \beta)$. Define the residuals $\hat{\epsilon}_i = X_i - \hat{\beta}_1 X_{n-1} - \ldots + \hat{\beta}_p X_{n-p}$, $p < i \leq n$.

- Draw a simple random sample of size n, $\epsilon^*_{p+1}, \epsilon^*_{p+2}, \ldots, \epsilon^*_n$ from $\{\hat{\epsilon}_i\}^n_{p+1}$ with replacement and define the bootstrap peudo-observations using the model structure in (12)

$$X_i^* = X_i \text{ for } i = 1, \ldots, p, \text{ and}$$
$$X_i^* = \hat{\beta}_1 X_{i-1}^* + \ldots + \hat{\beta}_p X_{i-p}^* + \hat{\epsilon}_i \text{ for } p < i \leq n.$$

- By construction, ϵ_i^* are i.i.d and $E^*\epsilon_i^* = 0$. The bootstrap version of the random variables is defined as $T_n^* = t_n(\mathcal{X}_n^*, \hat{F}_n, \hat{\beta})$ where F_n denotes the empirical distribution of $\hat{\epsilon}_i$.
- When $\{X_n\}$ is *stationary,* Bose (1988) shows that under suitable regularity conditions, T_n^* provides a more accurate approximation than the normal distribution.
- However, there is a series of work that the i.i.d-innovation bootstrap is very sensitive to the values of the AR-parameters. If β is such that the roots of the characteristic equations $z^p + \beta_1 z^{p-1} + \ldots + \beta_p = 0$ (closely) lie on the unit circle, then the i.i.d bootstrap fails.

7.2 Moving Block Bootstrap

- Künsch (1989) and Liu and Singh (1992) independently formulated a resampling scheme moving block bootstrap (MBB).

- It is applicable to dependent data without any parametric model assumptions.
- In contrast to a resampling a single observations, the MBB resamples blocks of (consecutive) observations at a time to preserve the dependent structure of the original data.

Definition 4 A sequence of random vectors $\{X_i\}_{i\in\mathbb{Z}}$ is called stationary if for every $i_1 < i_2 < \ldots < i_k$, $k \in \mathbb{N}$, and for every $m \in \mathbb{Z}$, the (joint) distributions of $(X_{i_1}, \ldots, X_{i_k})'$ and $(X_{i_1+m}, \ldots, X_{i_k+m})'$ are the same.

- Let $X_1, X_2,$ be a sequence of stationary random variables, and let $\mathcal{X}_n = \{X_{1,}X_2, \ldots, X_n\}$ be the data at hand.
- Suppose $l := l_n \in [1, n]$ is an integer. We typically require that

$$l \to \infty \text{ and } \frac{l}{n} \to 0 \text{ as } n \to \infty.$$

- Let $\mathcal{B}_i = (X_i, \ldots, X_{i+l-1})$ denote the block of length l starting with X_i, $1 \leq i \leq N$ where $N = n - l + 1.$
- Let b denotes the smallest integer such that $bl \geq n$. To obtain the MBB samples, we randomly select a b number of blocks from the collection $\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_N\}$ with

replacement. Let $\mathcal{B}_1^*, \ldots, \mathcal{B}_b^*$ be the resampled block and $\mathcal{B}_i^* = (X_{(i-1)l+1}^*, \ldots, X_{il}^*)$ for $i = 1, \ldots, b$. Then, $X_1^*, \ldots, X_n^*$ be the first n values of the resampled blocks $\mathcal{B}_1^*, \ldots, \mathcal{B}_b^*$.

- The MBB version θ_n^* of $\hat{\theta}_n = T(F_n)$ is defined as

$$\theta_n^* = T(F_n^*),$$

where F_n^* denotes the (joint) empirical distribution of $X_1^*, \ldots, X_n^*$. For example, $\hat{\theta}_n = n^{-1}\sum_{i=1}^n X_i$ and $\theta_n^* = n^{-1}\sum_{i=1}^n X_i^*$.

7.3 Nonoverlapping Block Bootstrap (NBB)

- Define blocks

$$\mathcal{B}_i = (X_{(i-1)l+1}, \ldots, X_{il})' \text{ for } i = 1, \ldots, b,$$

where b is the largest integer satisfying $lb \leq n$.

- The blocks in the MBB overlap, but the blocks in NBB do not. As a result, the collection of blocks from which the bootstrap blocks are selected $\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_b\}$ is smaller than that of NBB.

- The next step in implementing the NBB is exactly the same as for the MBB. Select a simple random sample of blocks $\mathcal{B}_1^*, \mathcal{B}_2^*, \ldots, \mathcal{B}_b^*$ from $\{\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_b\}$, and obtain the bootstrap version of the estimator $\theta_n^* = T(F_n^*)$, say, $\theta_n^* = m^{-1}\sum_{i=1}^m X_i^*$ with $m = bl$.
- Note that under the NBB, the block variables $\mathcal{B}_1^*, \mathcal{B}_2^*, \ldots, \mathcal{B}_b^*$ are i.i.d with common distribution

$$P(\mathcal{B}_1^* = (X_{(j-1)l+1}, \ldots, X_{jl})') = P((X_1^*, \ldots, X_l^*)' = (X_{(j-1)l+1}, \ldots, X_{jl})') = 1/b$$

for $j = 1, \ldots, b$. Hence,

$$\begin{aligned} E^*(\theta_n^*) &= E^*\left[m^{-1}\sum_{i=1}^m X_i^*\right] = E^*\left[(bl)^{-1}\sum_{i=1}^{bl} X_i^*\right] \\ &= E^*\left[l^{-1}\sum_{i=1}^{l} X_i^*\right] = \frac{1}{bl}\sum_{j=1}^{b}\sum_{i=1}^{l} X_{(j-1)l+i} \\ &= \frac{n}{bl}\bar{X} - \sum_{i=lb+1}^{n} X_i, \end{aligned}$$

which is equal to $\bar{X}$ when $n = bl$.

- We now check the bootstrap principle of NBB in an attempt to recreate the relation between population and the sample. For simplicity let $n = bl$. Because of the stationarity, each block $\mathcal{B}_i$ for $i = 1, \ldots, b$ has the same (identical) l-dimesional joint distribution P_l.

- Also, from the weak dependence of the original sample $\{X_n\}_{n \geq 1}$, these blocks are approximately independent random vectors with common distribution P_l. Thus, we may have the following approximation of $P_n \approx P_l^b := P_l \otimes \cdots \otimes P_l$, where $P_n = F_n$ is the underlying true population of $\mathcal{X}_n$.Let $\tilde{P}_l$ be the empirical distribution of $\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_b$. Then, the joint distribution of the bootstrap observations $\{X_i^*\}_{i=1}^n$ is given $\tilde{P}_l^b := \tilde{P}_l \otimes \cdots \otimes \tilde{P}_l$.

- For the random quantity $\theta_n := T(\mathcal{X}_n, F_n)$ with $F_n = P_n$ is the underlying true population of $\mathcal{X}_n$,

$$\begin{aligned} P_n(T(\mathcal{X}_n, P_n) &\leq x) \approx P_l^b(T(\mathcal{X}_n, P_l^b) \leq x) \\ &\approx \tilde{P}_l^b(T(\mathcal{X}_n^*, \tilde{P}_l^b) \leq x) \\ &= P_n^*(T(\mathcal{X}_n, F_n^*) \leq x). \end{aligned}$$

8 References

1. Ch.2∼3 in Hall, Peter. "The bootstrap and Edgeworth expansion". Springer Science & Business Media, 2013.
2. Ch. 6.22∼6.24 and Ch.10 in Hansen, Bruce. "Econometrics" https://www.ssc.wisc.edu/~bhansen/econometric
3. Djogbenou, A. A., MacKinnon, J. G., & Nielsen, M. Ø. (2019). Asymptotic theory and wild bootstrap inference with clustered errors. Journal of Econometrics.
4. Horowitz, Joel L. "The bootstrap." Handbook of econometrics". Vol. 5. Elsevier, 2001. 3159-3228.
5. Horowitz, Joel L. "Bootstrap methods in econometrics." Annual Review of Economics 11 (2019).
6. Hall, P., & Horowitz, J. L. (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. Econometrica: Journal of the Econometric Society, 891-916.
7. Kline, P., & Santos, A. (2012). Higher order properties of the wild bootstrap under

misspecification. Journal of Econometrics, 171(1), 54-70.

8. Lahiri, S. N. (2013). Resampling methods for dependent data. Springer Science & Business Media.