Some Asymptotic Theory for the Bootstrap

Study Notes June 28, 2015 Kenneth Chu

This set of notes aims to provide complete proofs of a number of asymptotic results regarding the Bootstrap [3] contained in Bickel and Freedman [1].

1 Bootstrap asymptotics for sample mean

Theorem 1.1 (Bootstrap Central Limit Theorem for I.I.D. sample mean, Theorem 2.1 [1])
Let $(\Omega, \mathcal{A}, \nu)$ be a probability space. Let $X, X_1, X_2, \ldots : \Omega \longrightarrow \mathbb{R}$ be a sequence of independent and identically distributed $\mathbb{R}$-valued random variables defined on Ω with finite expectation value $\mu_X \in \mathbb{R}$ and variance $\sigma_X^2 < \infty$. For each $n \in \mathbb{N}$ be fixed, define:

$$\overline{X}_n \;:\; \Omega \;\longrightarrow\; \mathbb{R} \;:\; \omega \;\longmapsto\; \frac{1}{n}\sum_{i=1}^{n} X_i(\omega).$$

For $n, m \in \mathbb{N}$, define $\mathcal{S}_m^{(n)}$ to be the set of all functions from $\{1, 2, \ldots, m\} \longrightarrow \{1, 2, \ldots, n\}$. Thus, each

$$s \;=\; (s(1), s(2), \ldots, s(m)) \;\in\; \mathcal{S}_m^{(n)}$$

can be regarded as a length-m finite (ordered) sequence of positive integers between 1 and n, inclusive. Note that $\mathcal{S}_m^{(n)}$ is a finite set with $\mid \mathcal{S}_m^{(n)} \mid = n^m$. Endow $\mathcal{S}_m^{(n)}$ with the probability space structure induced by the uniform probability function:

$$P_{\mathcal{S}_m^{(n)}}(s) \;:=\; \frac{1}{n^m}, \quad \text{for each } s \in \mathcal{S}_m^{(n)}.$$

Let $\Omega \times \mathcal{S}_m^{(n)}$ be the product probability space of Ω and $\mathcal{S}_m^{(n)}$. Define:

$$\overline{X}_m^{(n)} \;:\; \Omega \times \mathcal{S}_m^{(n)} \;\longrightarrow\; \mathbb{R} \;:\; (\omega, s) \;\longmapsto\; \frac{1}{m}\sum_{j=1}^{n} X_{s(j)}(\omega).$$

For each $\omega \in \Omega$, define:

$$\Phi_{m,\omega}^{(n)} \;:\; \mathcal{S}_m^{(n)} \;\longrightarrow\; \mathbb{R} \;:\; s \;\longmapsto\; \sqrt{m}\left(\overline{X}_m^{(n)}(\omega, s) - \overline{X}_n(\omega)\right)$$

Then,

$$P\Big(\; \Phi_{m,\omega}^{(n)} \xrightarrow{d} N(0, \sigma_X^2), \text{ as } n, m \to \infty \;\Big) \;=\; \nu\Big(\Big\{\; \omega \in \Omega \;\Big|\; \Phi_{m,\omega}^{(n)} \xrightarrow{d} N(0, \sigma_X^2), \text{ as } n, m \to \infty \;\Big\}\Big) \;=\; 1.$$

Remark 1.2
*For each fixed $\omega \in \Omega$, $\left\{ \Phi_{m,\omega}^{(n)} : \mathcal{S}_m^{(n)} \longrightarrow \mathbb{R} \right\}_{n,m\in\mathbb{N}}$ is a doubly indexed sequence of $\mathbb{R}$-valued random variables. Note that their respective domains $\mathcal{S}_m^{(n)}$ are pairwise distinct probability spaces. The **Bootstrap Central Limit Theorem for I.I.D. sample mean** asserts that for almost every $\omega \in \Omega$, the doubly indexed sequence $\left\{ \Phi_{m,\omega}^{(n)} \right\}$ of $\mathbb{R}$-valued random variables converges in distribution to $N(0, \sigma_X^2)$ as $n, m \longrightarrow \infty$.*

Remark 1.3 *The following results are well known from classical asymptotic theory:*
*By the **Weak Law of Large Numbers**, $\overline{X}_n$ converges in probability to μ_X, as $n \longrightarrow \infty$; in other words,*

$$\lim_{n\to\infty} P\big(\,\big|\overline{X}_n - \mu_X\big| > \varepsilon\,\big) \;=\; \lim_{n\to\infty} \nu\left(\left\{\,\omega \in \Omega \;:\; \big|\overline{X}_n(\omega) - \mu_X\big| > \varepsilon\,\right\}\right) \;=\; 0, \quad \text{for each } \varepsilon > 0.$$

*By the **Strong Law of Large Numbers**, $\overline{X}_n$ converges almost surely to μ_X, as $n \longrightarrow \infty$; in other words,*

$$P\Big(\lim_{n\to\infty} \overline{X}_n = \mu_X\Big) \;=\; \nu\left(\Big\{\,\omega \in \Omega \;\Big|\; \lim_{n\to\infty} \overline{X}_n(\omega) = \mu_X\,\Big\}\right) \;=\; 1.$$

*By the **Central Limit Theorem**, $\sqrt{n}\left(\overline{X}_n - \mu_X\right)$ converges in distribution to $N(0, \sigma_X^2)$.*

A Gaussian Processes

Definition A.1 (Stochastic processes)
Suppose $(\Omega, \mathcal{A}, \mu)$ is a probability space, $(V, \mathcal{F})$ is a measurable space, and T is an arbitrary non-empty set. A ***stochastic process*** *indexed by T defined on Ω with codomain V is a family $\{ X_t : \Omega \longrightarrow V \}_{t\in T}$ indexed by T of V-valued random variables defined on Ω.*

Definition A.2 (Finite-dimensional distributions of a stochastic processes)
Let $\{ X_t : \Omega \longrightarrow (V, \mathcal{F}) \}_{t\in T}$ be a stochastic process. Let $n \in \mathbb{N}$ and $t_1, t_2, \ldots, t_n \in T$ be distinct elements of T. The probability distribution induced on the product measurable space $(V^n, \mathcal{F}^{\otimes n})$ by $(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) : \Omega \longrightarrow V^n$ is called a ***finite-dimensional distribution*** *of the stochastic process.*

Definition A.3 (Komolgorov systems of finite-dimensional distributions & Komolgorov consistency)
Let T be an arbitrary non-empty set, and $\mathcal{D}(T)$ the set of all finite ordered sequences of elements of T whose elements are pairwise distinct; in other words,

$$\mathcal{D}(T) := \left\{ (t_1, t_2, \ldots, t_n) \in \bigcup_{k=1}^{\infty} T^k \;\middle|\; n \in \mathbb{N},\ t_i \neq t_j,\ \text{whenever } i \neq j \right\}.$$

For each $n \in \mathbb{N}$, let $\mathcal{M}_1(R^n, \mathcal{B}(\mathbb{R}^n))$ be the set of all probability measures defined on the product measurable space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. A ***Komolgorov system of finite-dimensional distributions*** *is a $\mathcal{D}(T)$-indexed family $\mathcal{P}$ of probability measures of the following form:*

$$\mathcal{P} = \left\{ P_{(t_1,\ldots,t_n)} \in \mathcal{M}_1(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \;\middle|\; (t_1, \ldots, t_n) \in \mathcal{D}(T) \right\}.$$

Furthermore, $\mathcal{P}$ is said to be ***Komolgorov consistent*** *if it satisfies both of the following conditions:*

- ***permutation invariance:*** *For any $n \in \mathbb{N}$, any $(t_1, \ldots, t_n) \in \mathcal{D}(T)$, any $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$, and any permutation $\pi : \{1, \ldots, n\} \longrightarrow \{1, \ldots, n\}$, the following equality holds:*

$$P_{(t_1,\ldots,t_n)}(B_1 \times \cdots \times B_n) = P_{(t_{\pi(1)},\ldots,t_{\pi(n)})}\big(B_{\pi(1)} \times \cdots \times B_{\pi(n)}\big).$$

- ***projection invariance:*** *For any $n \in \mathbb{N}$, any $(t_1, \ldots, t_{n+1}) \in \mathcal{D}(T)$, and any $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$, the following equality holds:*

$$P_{(t_1,\ldots,t_n,t_{n+1})}(B_1 \times \cdots \times B_n \times \mathbb{R}) = P_{(t_1,\ldots,t_n)}(B_1 \times \cdots \times B_n).$$

Remark A.4
It is obvious that the collection of finite-dimensional distributions of any $\mathbb{R}$-valued stochastic process is a Komolgorov consistent Komolgorov system of finite-dimensional distributions.

Definition A.5
Let $\{ X_t : \Omega \longrightarrow \mathbb{R} \}_{t\in T}$ be an $\mathbb{R}$-valued stochastic process, and

$$\mathcal{P} = \left\{ P_{(t_1,\ldots,t_n)} \in \mathcal{M}_1(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \;\middle|\; (t_1, \ldots, t_n) \in \mathcal{D}(T) \right\}$$

be a Komolgorov system of finite-dimensional distributions. We say that ***the stochastic process $\{ X_t \}$ admits $\mathcal{P}$ as its collection of finite-dimensional distributions*** *if, for each $n \in \mathbb{N}$ and any $(t_1, t_2, \ldots, t_n) \in \mathcal{D}(T)$, the probability distribution induced on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ by the map*

$$(X_{t_1}, \ldots, X_{t_n}) : \Omega \longrightarrow \mathbb{R}^n$$

equals $P_{(t_1,\ldots,t_n)} \in \mathcal{P}$.

Theorem A.6 (Komolgorov's Existence Theorem, Theorem 36.2, [2])
Let

$$\mathcal{P} \;=\; \left\{\; P_{(t_1,\ldots,t_n)} \in \mathcal{M}_1(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n)) \;\middle\vert\; (t_1,\ldots,t_n) \in \mathcal{D}(T) \;\right\}.$$

be a Komolgorov system of finite-dimensional distributions. Then, there exists a stochastic process

$$\{\, X_t : (\Omega, \mathcal{A}, \mu) \;\longrightarrow\; (\mathbb{R}, \mathcal{B}(\mathbb{R})) \,\}_{t\in T}$$

which admits $\mathcal{P}$ as its collection of finite-dimensional distributions if and only if $\mathcal{P}$ is Komolgorov consistent.

Definition A.7 (Mean and covariance functions of $\mathbb{R}$-valued stochastic processes)
Let $\{\, X_t : \Omega \longrightarrow \mathbb{R} \,\}_{t\in T}$ be an $\mathbb{R}$-valued stochastic process.

- *If, for each $t \in T$, we have $E(\, X_t \,) \in \mathbb{R}$, then the function*

$$a_X \,:\, T \;\longrightarrow\; \mathbb{R} \,:\, t \;\longmapsto\; E(\, X_t \,)$$

  *is called the **mean** function of the $\mathbb{R}$-valued stochastic process $\{\, X_t \,\}$.*

- *In addition, if, for each $t_1, t_2 \in T$, we have $0 \leq \mathrm{Cov}(X_{t_1}, X_{t_2}) < \infty$, then the function*

$$\Sigma_X \,:\, T \times T \;\longrightarrow\; \mathbb{R} \,:\, (t_1, t_2) \;\longmapsto\; \mathrm{Cov}(X_{t_1}, X_{t_2})$$

  *is called the **covariance** function of the $\mathbb{R}$-valued stochastic process $\{\, X_t \,\}$.*

Definition A.8 (Gaussian processes)
*An $\mathbb{R}$-valued stochastic process $\{\, X_t : \Omega \longrightarrow \mathbb{R} \,\}_{t\in T}$ is said to be **Gaussian** if each of its finite-dimensional distribution is a Gaussian distribution defined on some finite-dimensional Euclidean space.*

Theorem A.9
Let T be an arbitrary non-empty set, $a : T \longrightarrow \mathbb{R}$ an arbitrary $\mathbb{R}$-valued function defined on T, and $\Sigma : T \times T \longrightarrow [0, \infty)$ a non-negative $\mathbb{R}$-valued function defined on $T \times T$. Then, there exists a Gaussian process whose mean and covariance functions are a and Σ, respectively.

Theorem A.10
The mean and covariance functions of a Gaussian process together completely determine its collection of finite-dimensional distributions.

Definition A.11 (Brownian motion, a.k.a. Wiener process)
*A **Brownian motion**, or **Wiener process**, is a stochastic process $\{\, W_t : (\Omega, \mathcal{A}, \mu) \longrightarrow \mathbb{R} \,\}_{t\geq 0}$ indexed by the non-negative real line satisfying the following conditions:*

- *At $t = 0$, the process takes value 0 with probability 1; more precisely:*

$$P(\, W_0 = 0 \,) \;=\; \mu(\{\, \omega \in \Omega \mid W_0(\omega) = 0 \,\}) \;=\; 1.$$

- *The process $\{\, W_t \,\}$ has independent increments; more precisely: for any $0 \leq t_1 \leq t_2 \leq \cdots \leq t_n < \infty$,*

$$W_{t_n} - W_{t_{n-1}}, \quad W_{t_{n-1}} - W_{t_{n-2}}, \quad \ldots \;, \quad W_{t_2} - W_{t_1} \;:\; \Omega \longrightarrow \mathbb{R}$$

  are independent random variables.

- *For $0 \leq t_1 < t_2 < \infty$, the increment $W_{t_2} - W_{t_1}$ follows a Gaussian distribution with mean 0 and variance $t_2 - t_1$.*

Definition A.12 (Brownian bridge)
*A **Brownian bridge** is a Gaussian process $\{\, W_t^\circ : (\Omega, \mathcal{A}, \mu) \longrightarrow \mathbb{R} \,\}_{t\in[0,1]}$ indexed by the closed unit interval in $\mathbb{R}$ satisfying the following conditions:*

- *For each $t \in [0, 1]$, we have $E(\, W_t^\circ \,) = 0$.*
- *For any $t_1, t_2 \in [0, 1]$, we have $\mathrm{Cov}\big(W_{t_1}^0, W_{t_2}^\circ\big) = \min\{t_1, t_2\} - t_1 t_2$.*

References

[1] Bickel, P. J., and Freedman, D. A. Some asymptotic theory for the bootsrap. *The Annals of Statistics 9*, 6 (1981), 1196–1217.

[2] Billingsley, P. *Probability and Measure*, third ed. John Wiley & Sons, 1995.

[3] Efron, B. Bootsrap methods: another look at the jackknife. *The Annals of Statistics 7*, 1 (1979), 1–26.