Some Asymptotic Theory for the Bootstrap

Study Notes July 10, 2015 Kenneth Chu

This set of notes aims to provide complete proofs of a number of asymptotic results regarding the Bootstrap [4] contained in Bickel and Freedman [1].

1 Bootstrap asymptotics for the I.I.D. 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 $m \in \mathbb{N}$, define:

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

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

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

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

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

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

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

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

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

Then,

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

Remark 1.2
*For each fixed $\omega \in \Omega$, $\left\{ Y_{n,\omega}^{(m)} : \mathcal{S}_n^{(m)} \longrightarrow \mathbb{R} \right\}_{n,m\in\mathbb{N}}$ is a doubly indexed sequence of $\mathbb{R}$-valued random variables. Note that their respective domains $\mathcal{S}_n^{(m)}$ 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 $\left\{ Y_{n,\omega}^{(m)} \right\}$ 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,*

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

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

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

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

Proof Let $\mathcal{M}_1(\mathbb{R},\mathcal{B}(\mathbb{R}))$ denotes the set of probability measures on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. Let

$$\mathcal{W}_1^2(\mathbb{R},\mathcal{B}(\mathbb{R})) \;:=\; \left\{\; G\in\mathcal{M}_1(\mathbb{R},\mathcal{B}(\mathbb{R})) \;\;\middle|\;\; \int_{\mathbb{R}} x^2\,\mathrm{d}G(x)<\infty \;\right\}.$$

denote the Wasserstein space (Definition A.2) of order 2 of the measurable space $(\mathbb{R},\mathcal{B}(\mathbb{R}))$, whose underlying topological space (1-dimensional Euclidean space) is a Polish space (i.e. separable complete metric space). Let

$$W_2 \,:\, \mathcal{W}_1^2(\mathbb{R},\mathcal{B}(\mathbb{R}))\times\mathcal{W}_1^2(\mathbb{R},\mathcal{B}(\mathbb{R})) \longrightarrow \mathbb{R} \,:\, (G,G') \longmapsto \inf\left\{\;\sqrt{E(\,|X-Y|^2\,)}\;\;\middle|\;\;(X,Y)\in\Pi(G,G')\;\right\}$$

denote the Wasserstein metric (Theorem A.3) on $\mathcal{W}_1^2(\mathbb{R},\mathcal{B}(\mathbb{R}))$.

For each $m\in\mathbb{N}$, let $F^{(m)}\in\mathcal{M}_1(\mathbb{R},\mathcal{B}(\mathbb{R}))$ denote the distribution of

$$Y^{(m)} \,:\, \Omega \longrightarrow \mathbb{R} \,:\, \omega \longmapsto \sqrt{m}\left(\,\overline{X}^{(m)}(\omega)-\mu_X\,\right).$$

And, for each $\omega\in\Omega$, and each $m,n\in\mathbb{N}$, let $F_n^{(m)}(\omega)\in\mathcal{M}_1(\mathbb{R},\mathcal{B}(\mathbb{R}))$ denote the distribution of

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

Note that $N(0,\sigma_X^2)\in\mathcal{W}_1^2(\mathbb{R},\mathcal{B}(\mathbb{R}))$. By hypothesis, $F^{(m)}\in\mathcal{W}_1^2(\mathbb{R},\mathcal{B}(\mathbb{R}))$ for each $m\in\mathbb{N}$. And, for each $\omega\in\Omega$, $m,n\in\mathbb{N}$, we have $F_n^{(m)}(\omega)\in\mathcal{W}_1^2(\mathbb{R},\mathcal{B}(\mathbb{R}))$, since $\mathcal{S}_m^{(n)}$ is a finite probability space. Therefore, by Theorem A.3 and Claim 3 below, the following inequalities hold: For each $\omega\in\Omega$ and $m,n\in\mathbb{N}$,

$$\begin{aligned} W_2\Big(F_n^{(m)}(\omega),N(0,\sigma_X^2)\Big) &\leq W_2\Big(F_n^{(m)}(\omega),F^{(m)}\Big) + W_2\Big(F^{(m)},N(0,\sigma_X^2)\Big) \\ &\leq W_2(\,F_n(\omega),F\,) + W_2\Big(F^{(m)},N(0,\sigma_X^2)\Big). \end{aligned}$$

Thus, the present Theorem follows by Theorem A.6 and the following two claims:

Claim 1: $W_2\big(F^{(m)},N(0,\sigma_X^2))\big)\longrightarrow 0$, as $m\longrightarrow\infty$.

Claim 2: $\nu\Big(\Big\{\;\omega\in\Omega\;\Big|\;W_2(\,F_n(\omega),F\,)\longrightarrow 0\,,\text{ as } n\longrightarrow\infty\;\Big\}\Big) \;=\; 1.$

Proof of Claim 1: By the Classical Central Limit Theorem, $F^{(m)}\overset{w}{\longrightarrow}N(0,\sigma_X^2)$. Since $E\big[\,Y^{(m)}\,\big]=0$, we have

$$\begin{aligned} \int_{\mathbb{R}} y^2\,\mathrm{d}F^{(m)}(y) &= E\left[\left(Y^{(m)}\right)^2\right] = \mathrm{Var}\Big[\,Y^{(m)}\,\Big] = m\cdot\mathrm{Var}\left[\left(\frac{1}{m}\sum_{i=1}^m X_i\right)-\mu_X\right] \\ &= \frac{m}{m^2}\cdot\sum_{i=1}^m \mathrm{Var}[\,X_i\,] = \frac{1}{m}\cdot m\cdot\sigma_X^2 = \sigma_X^2, \end{aligned}$$

which is the second moment of $N(0,\sigma_X^2)$. Hence, by Definition A.5, we have $F^{(m)}\overset{\mathcal{W}_1^2}{\longrightarrow}N(0,\sigma_X^2)$, and by Theorem A.6, we have $W_2\big(F^{(m)},N(0,\sigma_X^2))\big)\longrightarrow 0$, as $m\longrightarrow\infty$. This completes the proof of Claim 1.

Proof of Claim 2: By hypothesis $\mu_X:=E[\,X\,]\in\mathbb{R}$ and $\sigma_X^2:=\mathrm{Var}[\,X\,]<\infty$. Hence $E\big[\,X^2\,\big]=\mathrm{Var}[\,X\,]+E[\,X\,]^2<\infty$. Thus, by the Strong Law of Large Numbers, we have:

$$\int_{\mathbb{R}} x^2\,\mathrm{d}F_n(\omega)(x) \;=\; \frac{1}{n}\sum_{i=1}^n X_i(\omega)^2 \;\overset{\text{a.e}}{\longrightarrow}\; E\big[\,X^2\,\big] \;=\; \int_{\mathbb{R}} x^2\,\mathrm{d}F(x),$$

equivalently,

$$\nu\left(\left\{\ \omega\in\Omega\ \middle|\ \int_{\mathbb{R}} x^2\,\mathrm{d}F_n(\omega)(x)\longrightarrow\int_{\mathbb{R}} x^2\,\mathrm{d}F(x)\ \right\}\right)\ =\ 1.$$

On the other hand, by the Glivenko-Cantelli Theorem, we have:

$$\nu\left(\left\{\ \omega\in\Omega\ \middle|\ \lim_{n\to\infty}\sup_{t\in\mathbb{R}}|F_n(\omega)(t)-F(t)|\ =\ 0\ \right\}\right)\ =\ 1,$$

which implies trivially

$$\nu\left(\left\{\ \omega\in\Omega\ \middle|\ \lim_{n\to\infty}F_n(\omega)(t)=F(t),\ \text{for each } t\in\mathbb{R}\ \right\}\right)\ =\ 1,$$

which in turn implies

$$\nu\left(\left\{\ \omega\in\Omega\ \middle|\ F_n(\omega)\overset{w}{\longrightarrow}F\ \right\}\right)\ =\ 1.$$

Note that in the above assertion, we used the slight abuse of notation that $F_n(\omega)$ represents both the distribution (measure) $F_n(\omega)\in\mathcal{M}_1(\mathbb{R},\mathcal{B}(\mathbb{R}))$ as well as its cumulative distribution function defined on $\mathbb{R}$. Thus, we see that Theorem A.6, the Glivenko-Cantelli Theorem, and the Strong Law of Large Numbers together imply:

$$\begin{aligned}
&\nu\left(\left\{\ \omega\in\Omega\ \middle|\ W_2(\,F_n(\omega),F\,)\longrightarrow 0\ \right\}\right)\\
=\ &\nu\left(\left\{\ \omega\in\Omega\ \middle|\ F_n(\omega)\overset{\mathcal{W}_1^2}{\longrightarrow}F\ \right\}\right)\\
=\ &\nu\left(\left\{\ \omega\in\Omega\ \middle|\ F_n(\omega)\overset{w}{\longrightarrow}F\ \text{and}\ \int_{\mathbb{R}}x^2\,\mathrm{d}F_n(\omega)(x)\longrightarrow\int_{\mathbb{R}}x^2\,\mathrm{d}F(x)\ \right\}\right)\\
=\ &\nu\left(\left\{\ \omega\in\Omega\ \middle|\ F_n(\omega)\overset{w}{\longrightarrow}F\ \right\}\bigcap\left\{\ \omega\in\Omega\ \middle|\ \int_{\mathbb{R}}x^2\,\mathrm{d}F_n(\omega)(x)\longrightarrow\int_{\mathbb{R}}x^2\,\mathrm{d}F(x)\ \right\}\right)\\
=\ &1.
\end{aligned}$$

This completes the proof of Claim 2.

Claim 3: Let $G\in\mathcal{W}_1^2(\mathbb{R},\mathcal{B}(\mathbb{R}))$ and $m\in\mathbb{N}$. Suppose $Z_1^{(G)},Z_2^{(G)},\ldots,Z_m^{(G)}$ are independent and identically distributed random variables, each having distribution G. Let $G^{(m)}\in\mathcal{M}_1(\mathbb{R},\mathcal{B}(\mathbb{R}))$ be the (empirical) measure of the random variable

$$Y_m^{(G)}\ :=\ \frac{1}{m^{1/2}}\sum_{i=1}^{m}\left(Z_i^{(G)}-\mu_G\right)\ :\ \Omega\ \longrightarrow\ \mathbb{R}\,,$$

where $\mu_G:=\int_{\mathbb{R}}x\,\mathrm{d}G(x)$ is the expectation value of the distribution G. Then, for any $G,H\in\mathcal{W}_1^2(\mathbb{R},\mathcal{B}(\mathbb{R}))$, we have

$$W_2\left(G^{(m)},H^{(m)}\right)\ \leq\ W_2(\,G,H\,)\,.$$

<u>Proof of Claim 3:</u>

□

A Wasserstein Spaces

Proofs of results mentioned in this section can be found in Chapters 1 and 6 of [6].

Suppose $(S,\mathcal{S})$ and $(T,\mathcal{T})$ are two measurable spaces. We will use the following notations:

- $(S \times T, \mathcal{S} \otimes \mathcal{T})$ denotes their product measurable space (see Chapter 10, [5]).
- $\mathcal{M}_1(S, \mathcal{S})$, $\mathcal{M}_1(T, \mathcal{T})$, and $\mathcal{M}_1(S \times T, \mathcal{S} \otimes \mathcal{T})$ denote the sets of probability measures on the respective measurable spaces.
- $\Pi^S : S \times T \longrightarrow S : (s,t) \longmapsto s$ and $\Pi^T : S \times T \longrightarrow T : (s,t) \longmapsto t$ are the canonical projection maps, and

$$\begin{array}{ccccccccc} \Pi^S_* & : & \mathcal{M}_1(S \times T, \mathcal{S} \otimes \mathcal{T}) & \longrightarrow & \mathcal{M}_1(S, \mathcal{S}) & : & \pi & \longmapsto & \left(\; A \in \mathcal{S} \longmapsto \pi\big[(\Pi^S)^{-1}(A)\big] \; \right), \\ \Pi^T_* & : & \mathcal{M}_1(S \times T, \mathcal{S} \otimes \mathcal{T}) & \longrightarrow & \mathcal{M}_1(T, \mathcal{T}) & : & \pi & \longmapsto & \left(\; B \in \mathcal{T} \longmapsto \pi\big[(\Pi^T)^{-1}(B)\big] \; \right) \end{array}$$

  are the corresponding push-forward maps of measures.

Definition A.1 (Coupling measures and couplings (Definition 1.1, [6]))
Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ be two measurable spaces. Let $\mu \in \mathcal{M}_1(S, \mathcal{S})$ and $\nu \in \mathcal{M}_1(T, \mathcal{T})$.

- *A* ***coupling (probability) measure*** *of μ and ν is a probability measure $\pi \in \mathcal{M}_1(S \times T, \mathcal{S} \otimes \mathcal{T})$ whose push-forwards under the canonical projection maps are μ and ν respectively; in other words $\pi \in \mathcal{M}_1(S \times T, \mathcal{S} \otimes \mathcal{T})$ is a coupling measure of $(\mu, \nu) \in \mathcal{M}_1(S, \mathcal{S}) \times \mathcal{M}_1(T, \mathcal{T})$ if π satisfies:*

$$\Pi^S_*(\pi) = \mu \quad \text{and} \quad \Pi^T_*(\pi) = \nu.$$

  In this case, μ and ν are called the ***marginal (probability) measures*** *of π. We denote by $\Pi(\mu, \nu)$ the subset of $\mathcal{M}_1(S \times T, \mathcal{S} \otimes \mathcal{T})$ of all coupling probability measures of μ and ν.*

- *A* ***coupling*** *of μ and ν is an $(S \times T)$-valued random variable*

$$Z \,=\, (X, Y) \,:\, (\,\Omega, \mathcal{A}, P_\Omega\,) \longrightarrow (\,S \times T, \mathcal{S} \otimes \mathcal{T}\,)$$

  whose induced measure on $(\,S \times T, \mathcal{S} \otimes \mathcal{T}\,)$ is a coupling probability measure of μ and ν. More precisely,

$$\begin{array}{ccl} \mu(A) & = & P_X(A) \,=\, P_\Omega\big(X^{-1}(A)\big) \,=\, P_\Omega\big((\Pi^S \circ Z)^{-1}(A)\big) \,=\, P_\Omega\big(\,Z^{-1}\big[(\Pi^S)^{-1}(A)\big]\,\big), \quad \text{for each } A \in \mathcal{S} \\ \nu(B) & = & P_Y(B) \,=\, P_\Omega\big(Y^{-1}(B)\big) \,=\, P_\Omega\big((\Pi^T \circ Z)^{-1}(B)\big) \,=\, P_\Omega\big(\,Z^{-1}\big[(\Pi^T)^{-1}(B)\big]\,\big), \quad \text{for each } B \in \mathcal{T} \end{array}$$

Definition A.2 (Wasserstein distances and Wasserstein spaces (Definitions 6.1 and 6.4, [6]))
Let $p \in [1, \infty)$. Let (S, ρ) be a Polish space (i.e. separable complete metric space), and $\mathcal{S}$ its Borel σ-algebra.

- *The* ***Wasserstein distance of order*** p *is, by definition, the map $W_p : \mathcal{M}_1(S, \mathcal{S}) \times \mathcal{M}_1(S, \mathcal{S}) \longrightarrow \mathbb{R} \cup \{\, +\infty \,\}$ given by:*

$$\begin{array}{rcl} W_p(\mu, \nu) & := & \displaystyle\inf_{\pi \in \Pi(\mu,\nu)} \left\{ \left(\int_{S \times S} \rho(x,y)^p \, \mathrm{d}\pi(x,y) \right)^{1/p} \right\} \\ & = & \inf \left\{ \left(E[\,\rho(X,Y)^p\,] \right)^{1/p} \in \mathbb{R} \cup \{\,+\infty\,\} \;\middle|\; \begin{array}{c} X\,,\,Y : (\Omega, \mathcal{A}, \pi) \longrightarrow (S, \mathcal{S}) \text{ are } S\text{-valued} \\ \text{random variables with } X^*(\pi) = \mu, Y^*(\pi) = \nu \end{array} \right\}. \end{array}$$

- *The* ***Wasserstein space of order*** p *is defined to be:*

$$\mathcal{W}^p_1(S, \mathcal{S}) \;:=\; \left\{ \mu \in \mathcal{M}_1(S, \mathcal{S}) \;\middle|\; \int_S \rho(x_0, x)^p \, \mathrm{d}\mu(x) \,<\, \infty \right\},$$

  where $x_0 \in S$ is an arbitrary point in S ($\mathcal{W}^p_1(S, \mathcal{S})$ is independent of the choice of $x_0 \in S$). Thus, $\mathcal{W}^p_1(S, \mathcal{S})$ is the set of probability measures on $(S, \mathcal{S})$ with finite moment of order p.

Theorem A.3 (Wasserstein metrics (Definition 6.4 and Theorem 6.18, [6]))

- *The Wasserstein space* $\mathcal{W}_1^p(S,\mathcal{S})$ *is independent of the choice of the point* $x_0 \in S$ *in its definition.*
- *The Wasserstein distance* W_p *restricts to a metric on* $\mathcal{W}_1^p(S,\mathcal{S}) \times \mathcal{W}_1^p(S,\mathcal{S})$.
- *For a Polish space (i.e. separable complete metric space)* (S,ρ) *with Borel* σ*-algebra* $\mathcal{S}$*, the Wassertein space* $\mathcal{W}_1^p(S,\mathcal{S})$*, when metrized by the Wasserstein metric* W_p*, is itself a Polish space.*

Definition A.4 (Weak convergence in metric spaces (Chapter 1, [3]))
Suppose:

- (S,ρ) *is a metric space and* $\mathcal{S}$ *is its Borel* σ*-algebra.*
- $\mathcal{M}_1(S,\mathcal{S})$ *denotes the set of probability measures defined on* $(S,\mathcal{S})$.
- $\mu \in \mathcal{M}_1(S,\mathcal{S})$ *and* $\{\,\mu_k\,\}_{k\in\mathbb{N}} \subset \mathcal{M}_1(S,\mathcal{S})$.

Then, $\{\,\mu_k\,\}_{k\in\mathbb{N}}$ *is said to* ***converge weakly*** *to* μ *if, for each* $f \in C_b(S,\mathbb{R})$,

$$\int_S f(x)\,\mathrm{d}\mu_k(x) \;\longrightarrow\; \int_S f(x)\,\mathrm{d}\mu(x), \;\; \text{as } k \longrightarrow \infty,$$

where $C_b(S,\mathbb{R})$ *denotes the set of all bounded continuous* $\mathbb{R}$*-valued functions on* S*. We write* $\mu_k \xrightarrow{w} \mu$ *for* μ_k *converging weakly to* μ.

Definition A.5 (Weak convergence in Wassertein spaces (Definition 6.8, [6]))
Suppose:

- (X,ρ) *is a Polish space, and* $\mathcal{S}$ *is its Borel* σ*-algebra.*
- $p \in [1,\infty)$ *and* $\mathcal{W}_1^p(S,\mathcal{S})$ *is the corresponding Wasserstein space of order* p.
- $\mu \in \mathcal{W}_1^p(S,\mathcal{S})$ *and* $\{\,\mu_k\,\}_{k\in\mathbb{N}} \subset \mathcal{W}_1^p(S,\mathcal{S})$.

Then, $\{\,\mu_k\,\}_{k\in\mathbb{N}}$ *is said to* ***converge weakly in*** $\mathcal{W}_1^p(S,\mathcal{S})$ *to* μ *if, for some (hence any)* $x_0 \in S$*, we have:*

$$\mu_k \xrightarrow{w} \mu \quad \text{and} \quad \int_S \rho(x_0,x)^p\,\mathrm{d}\mu_k(x) \;\longrightarrow\; \int_S \rho(x_0,x)^p\,\mathrm{d}\mu(x), \;\; \text{as } k \longrightarrow \infty.$$

We write $\mu_k \xrightarrow{\mathcal{W}_1^p} \mu$ *for* μ_k *converging weakly to* μ *in* $\mathcal{W}_1^p(S,\mathcal{S})$.

Theorem A.6 (Wasserstein metrics metrize weak convergence in Wassertein spaces (Theorem 6.9, [6]))
Suppose:

- (X,ρ) *is a Polish space, and* $\mathcal{S}$ *is its Borel* σ*-algebra.*
- $p \in [1,\infty)$, $(\,\mathcal{W}_1^p(S,\mathcal{S})\,,W_p\,)$ *is the corresponding Wasserstein space of order* p*, metrized by the Wasserstein metric* W_p *defined on it.*
- $\mu \in \mathcal{W}_1^p(S,\mathcal{S})$ *and* $\{\,\mu_k\,\}_{k\in\mathbb{N}} \subset \mathcal{W}_1^p(S,\mathcal{S})$.

Then,

$$\mu_k \xrightarrow{\mathcal{W}_1^p} \mu \quad \text{if and only if} \quad W_p(\mu_k,\mu) \longrightarrow 0.$$

To conclude this Appendix, we present several technical results regarding $\mathcal{W}_1^2(\mathbb{R},\mathcal{B}(\mathbb{R}))$ that are used in the main text.

Lemma A.7

Suppose:

- $m \in \mathbb{N}$ *is a positive integer.*
- $X_1, X_2, \ldots, X_m : \Omega_X \longrightarrow \mathbb{R}$ *are independent and identically distributed* $\mathbb{R}$*-valued random variables defined on the same probability space* Ω_X *such that* $E[\,X_i\,] \;=\; 0$*, for each* $i = 1, 2, \ldots, m$.
- $Y_1, Y_2, \ldots, Y_m : \Omega_Y \longrightarrow \mathbb{R}$ *are independent and identically distributed* $\mathbb{R}$*-valued random variables defined on the same probability space* Ω_Y *such that* $E[\,Y_i\,] \;=\; 0$*, for each* $i = 1, 2, \ldots, m$.
- $G^{(1)} \in \mathcal{W}_1^2(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *denotes the common probability distribution of* X_i*, for* $i = 1, 2, \ldots, m$*, and* $G^{(m)} \in \mathcal{W}_1^2(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *denotes the probability distribution on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *induced by:*
$$\frac{1}{\sqrt{m}} \sum_{i=1}^{m} X_i \,:\, \Omega_X \longrightarrow \mathbb{R}\,.$$
- $H^{(1)} \in \mathcal{W}_1^2(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *denotes the common probability distribution of* Y_i*, for* $i = 1, 2, \ldots, m$*, and* $H^{(m)} \in \mathcal{W}_1^2(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *denotes the probability distribution on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *induced by:*
$$\frac{1}{\sqrt{m}} \sum_{i=1}^{m} Y_i \,:\, \Omega_Y \longrightarrow \mathbb{R}\,.$$

Then,
$$W_2\Big(G^{(m)}, H^{(m)}\Big) \;\;\leq\;\; W_2\Big(G^{(1)}, H^{(1)}\Big)\,.$$

Proof First, we make two observations:

Claim 1:
$$\begin{aligned}
&\inf\left\{\; \int_{\mathbb{R}^2} (x-y)^2 \,\mathrm{d}\mu(x,y) \in [0,\infty) \;\middle|\; \mu \in \Pi\Big(G^{(m)}, H^{(m)}\Big) \right\} \\
\leq\;\; &\inf\left\{\; \int_{\mathbb{R}^{2m}} \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} x_i \;-\; \frac{1}{\sqrt{m}} \sum_{i=1}^{m} y_i \right)^2 \mathrm{d}\mu_1(x_1,y_1) \,\cdots\, \mathrm{d}\mu_m(x_m,y_m) \;\middle|\; (\mu_1, \ldots, \mu_m) \in \Pi\Big(G^{(1)}, H^{(1)}\Big)^m \right\},
\end{aligned}$$

where $(\mu_1, \ldots, \mu_m) \in \Pi\big(G^{(1)}, H^{(1)}\big)^m$ means that each of $\mu_1, \ldots, \mu_m \in \Pi\big(G^{(1)}, H^{(1)}\big) \subset \mathcal{M}_1(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$, and that the m $\mathbb{R}^2$-valued random variables respectively corresponding to $\mu_1, \ldots, \mu_m$ are independent.

Claim 2: If $\mu_1, \mu_2, \ldots, \mu_m \in \mathcal{M}_1(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ such that
$$\int_{\mathbb{R}^2} x \,\mathrm{d}\mu_i(x,y) \;=\; \int_{\mathbb{R}^2} y \,\mathrm{d}\mu_i(x,y) \;=\; 0\,, \quad \text{for each } i = 1, 2, \ldots, m,$$
then
$$\int_{\mathbb{R}^{2m}} \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} x_i \;-\; \frac{1}{\sqrt{m}} \sum_{i=1}^{m} y_i \right)^2 \mathrm{d}\mu_1(x_1,y_1) \,\cdots\, \mathrm{d}\mu_m(x_m,y_m) \;\;=\;\; \frac{1}{m} \sum_{i=1}^{m} \int_{\mathbb{R}^2} (x_i - y_i)^2 \,\mathrm{d}\mu_i(x_i,y_i).$$

Proof of Claim 1:

First, note that we have the following set inclusion (of subsets of non-negative real numbers):

$$\begin{aligned}
&\left\{ \int_{\mathbb{R}^2} (x-y)^2 \,\mathrm{d}\mu(x,y) \in [0,\infty) \;\middle|\; \mu \in \Pi\Big(G^{(m)}, H^{(m)}\Big) \right\} \\
\supseteq\; &\left\{ \int_{\mathbb{R}^{2m}} \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} x_i \;-\; \frac{1}{\sqrt{m}} \sum_{i=1}^{m} y_i \right)^2 \mathrm{d}\mu_1(x_1,y_1) \cdots \mathrm{d}\mu_m(x_m,y_m) \;\middle|\; (\mu_1,\ldots,\mu_m) \in \Pi\Big(G^{(1)}, H^{(1)}\Big)^m \right\},
\end{aligned}$$

due to the following implication:

$$\left.\begin{array}{c}
\mathcal{L}(X_1,Y_1),\;\ldots\;,\mathcal{L}(X_m,Y_m) \;\in\; \Pi\big(G^{(1)},H^{(1)}\big) \\
\text{and independence of the } m\ \mathbb{R}^2\text{-valued} \\
\text{random variables } (X_1,Y_1),\;\ldots\;,(X_m,Y_m)
\end{array}\right\} \;\Longrightarrow\; \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} X_i \,,\, \frac{1}{\sqrt{m}} \sum_{i=1}^{m} Y_i \right) \;\sim\; \Pi\Big(G^{(m)}, H^{(m)}\Big).$$

Claim 1 now follows, since $\inf A \geq \inf B$, for $A \subset B \subset \mathbb{R}$.

Proof of Claim 2:

$$\begin{aligned}
&\int_{\mathbb{R}^{2m}} \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} x_i \;-\; \frac{1}{\sqrt{m}} \sum_{i=1}^{m} y_i \right)^2 \mathrm{d}\mu_1(x_1,y_1) \cdots \mathrm{d}\mu_m(x_m,y_m) \\
=\;& \frac{1}{m} \int_{\mathbb{R}^{2m}} \left[\sum_{i=1}^{m} (x_i - y_i) \right]^2 \mathrm{d}\mu_1(x_1,y_1) \cdots \mathrm{d}\mu_m(x_m,y_m) \\
=\;& \frac{1}{m} \int_{\mathbb{R}^{2m}} \left[\sum_{i=1}^{m} (x_i - y_i)^2 + \sum\sum_{i\neq j} (x_i - y_i)(x_j - y_j) \right] \mathrm{d}\mu_1(x_1,y_1) \cdots \mathrm{d}\mu_m(x_m,y_m) \\
=\;& \frac{1}{m} \cdot \sum_{i=1}^{m} \int_{\mathbb{R}^2} (x_i - y_i)^2 \,\mathrm{d}\mu_i(x_i,y_i) + \frac{1}{m} \cdot \sum\sum_{i\neq j} \left(\int_{\mathbb{R}^2} (x_i - y_i) \,\mathrm{d}\mu_i(x_i,y_i) \right) \cdot \left(\int_{\mathbb{R}^2} (x_j - y_j) \,\mathrm{d}\mu_j(x_j,y_j) \right) \\
=\;& \frac{1}{m} \sum_{i=1}^{m} \int_{\mathbb{R}^2} (x_i - y_i)^2 \,\mathrm{d}\mu_i(x_i,y_i) + \frac{1}{m} \cdot \sum\sum_{i\neq j} \left(E[\,X_i\,] - E[\,Y_i\,] \right) \cdot \left(E[\,X_j\,] - E[\,Y_j\,] \right) \\
=\;& \frac{1}{m} \sum_{i=1}^{m} \int_{\mathbb{R}^2} (x_i - y_i)^2 \,\mathrm{d}\mu_i(x_i,y_i).
\end{aligned}$$

This proves Claim 2.

By Claims 1 and 2 above, we have:

$$\begin{aligned}
& W_2\Big(G^{(m)}, H^{(m)}\Big) \\
= \;& \inf\left\{ \int_{\mathbb{R}^2} (x-y)^2 \,\mathrm{d}\mu(x,y) \in [0,\infty) \;\middle|\; \mu \in \Pi\Big(G^{(m)}, H^{(m)}\Big) \right\} \\
\leq \;& \inf\left\{ \int_{\mathbb{R}^{2m}} \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} x_i \;-\; \frac{1}{\sqrt{m}} \sum_{i=1}^{m} y_i \right)^2 \mathrm{d}\mu_1(x_1,y_1) \cdots \mathrm{d}\mu_m(x_m,y_m) \;\middle|\; (\mu_1,\ldots,\mu_m) \in \Pi\Big(G^{(1)}, H^{(1)}\Big)^m \right\} \\
= \;& \inf\left\{ \frac{1}{m} \sum_{i=1}^{m} \int_{\mathbb{R}^2} (x_i - y_i)^2 \,\mathrm{d}\mu_i(x_i,y_i) \;\middle|\; (\mu_1,\ldots,\mu_m) \in \Pi\Big(G^{(1)}, H^{(1)}\Big)^m \right\} \\
= \;& \frac{1}{m} \cdot \sum_{i=1}^{m} \inf\left\{ \int_{\mathbb{R}^2} (x_i - y_i)^2 \,\mathrm{d}\mu_i(x_i,y_i) \;\middle|\; \mu_i \in \Pi\Big(G^{(1)}, H^{(1)}\Big) \right\} \;=\; \frac{1}{m} \cdot \sum_{i=1}^{m} W_2\Big(G^{(1)}, H^{(1)}\Big) \\
= \;& W_2\Big(G^{(1)}, H^{(1)}\Big).
\end{aligned}$$

This proves the present Lemma. □

Lemma A.8

Suppose:

- $G,\ H \in \mathcal{W}_1^2(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, *with* $\mu_G := \int_{\mathbb{R}} x \,\mathrm{d}G(x)$, *and* $\mu_H := \int_{\mathbb{R}} x \,\mathrm{d}H(x)$.
- $(\,X\,, Y)$ *is an* $\mathbb{R}^2$*-valued random variable such that the marginal distributions of* X *and* Y *are* G *and* H*, respectively. And,* $G^{(1)}$ *and* $H^{(1)}$ *are the marginal distributions of the* $\mathbb{R}$*-valued random variables* $X - \mu_G$ *and* $Y - \mu_H$*, respectively.*

Then,

$$W_2\Big(G^{(1)}, H^{(1)}\Big) \;+\; (\,\mu_G - \mu_H\,)^2 \;\;\leq\;\; W_2(\,G, H\,)\,,$$

and, in particular,

$$W_2\Big(G^{(1)}, H^{(1)}\Big) \;\;\leq\;\; W_2(G,H)\,.$$

Proof

$$\begin{aligned}
W_2\Big(G^{(1)}, H^{(1)}\Big) \;&=\; \inf\left\{ \int_{\mathbb{R}^2} (x-y)^2 \,\mathrm{d}\mu(x,y) \;\middle|\; \mu \in \Pi(G^{(1)}, H^{(1)}) \right\} \\
&\leq\; \inf\left\{ \int_{\mathbb{R}^2} (\xi - \mu_G - \eta + \mu_H)^2 \,\mathrm{d}\nu(\xi,\eta) \;\middle|\; \nu \in \Pi(G,H) \right\} \\
&=\; \inf\left\{ \int_{\mathbb{R}^2} \left[\,(\xi-\eta) - (\mu_G - \mu_H)\right]^2 \,\mathrm{d}\nu(\xi,\eta) \;\middle|\; \nu \in \Pi(G,H) \right\} \\
&=\; \inf\left\{ \int_{\mathbb{R}^2} \left[\,(\xi-\eta)^2 - 2\,(\xi-\eta)(\mu_G - \mu_H) + (\mu_G - \mu_H)^2\,\right] \mathrm{d}\nu(\xi,\eta) \;\middle|\; \nu \in \Pi(G,H) \right\} \\
&=\; \inf\left\{ \int_{\mathbb{R}^2} (\xi-\eta)^2 \,\mathrm{d}\nu(\xi,\eta) - 2\,(\mu_G - \mu_H) \int_{\mathbb{R}^2} (\xi-\eta) \,\mathrm{d}\nu(\xi,\eta) + (\mu_G - \mu_H)^2 \;\middle|\; \nu \in \Pi(G,H) \right\} \\
&=\; \inf\left\{ \int_{\mathbb{R}^2} (\xi-\eta)^2 \,\mathrm{d}\nu(\xi,\eta) - (\mu_G - \mu_H)^2 \;\middle|\; \nu \in \Pi(G,H) \right\} \\
&=\; \inf\left\{ \int_{\mathbb{R}^2} (\xi-\eta)^2 \,\mathrm{d}\nu(\xi,\eta) \;\middle|\; \nu \in \Pi(G,H) \right\} \;-\; (\mu_G - \mu_H)^2
\end{aligned}$$

This establishes the following inequality:

$$W_2\Big(G^{(1)}, H^{(1)}\Big) \;+\; (\,\mu_G - \mu_H\,)^2 \;\;\leq\;\; W_2(\,G, H\,)\,,$$

and completes the proof of the present Lemma. $\square$

B A stochastic process $\{\,X_t : \Omega \longrightarrow V\,\}_{t\in T}$ and its equivalent V^T-valued random variable $X : \Omega \longrightarrow V^T$

Let Ω, T, and V be non-empty sets. Let $\{\,X_t : \Omega \longrightarrow V\,\}_{t\in T}$ be a T-index family of maps, each of which maps from Ω into V. Note that this family of maps is set-theoretically equivalent (in the sense that either one completely determines the other) to the following (V^T)-valued map defined on Ω:

$$X \,:\, \Omega \,\longrightarrow\, V^T \,:\, \omega \,\longmapsto\, (t \longmapsto X_t(\omega))\,,$$

where $V^T = \prod_{t\in T} V$ denotes the set of all (arbitrary) V-valued functions defined on T. In this section, we aim to establish the following two results:

- Suppose $(\Omega, \mathcal{A})$ and $(V, \mathcal{F})$ are measurable space structures on Ω and V, respectively. Then, $X : \Omega \longrightarrow V^T$ is $\big(\mathcal{A}, \sigma[(V,\mathcal{F})^T]\big)$-measurable if and only if $X_t : \Omega \longrightarrow V$ is $(\mathcal{A}, \mathcal{F})$-measurable for each $t \in T$. Here, $\sigma[(V,\mathcal{F})^T]$ denotes the product σ-algebra on V^T, which is by definition the smallest σ-algebra on V^T such that, for each $t \in T$, the projection map (or evaluation map)

  $$\pi_t \,:\, V^T \,\longrightarrow\, V \,:\, x \,\longmapsto\, x(t)$$

  is $\big(\sigma[(V,\mathcal{F})^T], \mathcal{F}\big)$-measurable.
- An immediate corollary of the above result is that: Suppose $(\Omega, \mathcal{A}, \mu)$ is a probability space structure on Ω, $(V, \mathcal{F})$ is a measurable space structure on V, and $\sigma[(V,\mathcal{F})^T]$ is the product σ-algebra on V^T. Then, $X : (\Omega, \mathcal{A}, \mu) \longrightarrow (V^T, \sigma[(V,\mathcal{F})^T])$ is V^T-valued random variable if and only if $\{\,X_t : (\Omega, \mathcal{A}, \mu) \longrightarrow (V, \mathcal{F})\,\}_{t\in T}$ is a stochastic process.

Definition B.1 (The product σ-algebra of a Cartesian product of measurable spaces)
Let T be an arbitrary non-empty set. For each $t \in T$, let $(V_t, \mathcal{F}_t)$ be a measurable space (in particular, $V_t \neq \varnothing$). Let $\prod_{t\in T} V_t$ be the Cartesian product of $\{\,V_t\,\}_{t\in T}$. In other words,

$$\prod_{t\in T} V_t \;:=\; \left\{\; v : T \longrightarrow \bigsqcup_{t\in T} V_t \;\middle|\; v(t) \in V_t,\ \text{for each } t \in T \;\right\}.$$

That $\prod_{t\in T} V_t \neq \varnothing$ follows from the Axiom of Choice. For each $t \in T$, let

$$\pi_t \,:\, \prod_{\tau\in T} V_\tau \,\longrightarrow\, V_t \,:\, v \,\longmapsto\, v(t)$$

be the projection map from $\prod_{\tau\in T} V_\tau$ onto V_t. The ***product σ-algebra*** *on $\prod_{t\in T} V_t$ is the following:*

$$\sigma\left(\left\{\; \pi_t^{-1}(F) \,\subset\, \prod_{\tau\in T} V_\tau \;\middle|\; F \in \mathcal{F}_t\,,\, t \in T \;\right\}\right) \;\subset\; \mathrm{PowerSet}\left(\prod_{t\in T} V_t\right).$$

Clearly, it is the smallest σ-algebra on $\prod_{t\in T} V_t$ with respect to which each projection map $\pi_t : \prod_{t\in T} V_t \longrightarrow (V_t, \mathcal{F}_t)$ is measurable. We denote the product σ-algebra on $\prod_{t\in T} V_t$ by

$$\sigma\left(\prod_{t\in T}(V_t, \mathcal{F}_t)\right).$$

Theorem B.2

Suppose Ω, T, and V are non-empty sets. Let $\{\,X_t : \Omega \longrightarrow V\,\}_{t\in T}$ be a T-indexed family of V-valued maps defined on Ω. Then, the following statements are true:

1. *The family $\{\,X_t : \Omega \longrightarrow V\,\}_{t\in T}$ of maps is set-theoretically equivalent (in the sense that either completely determines the other) to the following (V^T)-valued map defined on Ω:*

$$X \,:\, \Omega \,\longrightarrow\, V^T \,:\, \omega \,\longmapsto\, (t \longmapsto X_t(\omega))\,,$$

where $V^T = \prod_{t\in T} V$ denotes the set of all (arbitrary) V-valued functions defined on T.

2. *Suppose:*
   - *$(\Omega, \mathcal{A})$ and $(V, \mathcal{F})$ are measurable space structures on Ω and V, respectively.*
   - *$W \subset V^T$ is a subset of V^T such that $X(\Omega) = \bigcup_{t\in T} X_t(\Omega) \subset W$.*
   - *$(W, \mathcal{G})$ is a measurable space structure on W such that, for each $t \in T$, the projection map*

   $$\pi_t \,:\, W \,\longrightarrow\, V \,:\, w \,\longmapsto\, w(t)$$

   is $(\mathcal{G}, \mathcal{F})$-measurable.

   Then, $(\mathcal{A}, \mathcal{G})$-measurability of $X : \Omega \longrightarrow W$ implies $(\mathcal{A}, \mathcal{F})$-measurability of $X_t : \Omega \longrightarrow V$ for each $t \in T$.

3. *Suppose:*
   - *$(\Omega, \mathcal{A})$ and $(V, \mathcal{F})$ are measurable space structures on Ω and V, respectively.*
   - *$\sigma[(V,\mathcal{F})^T]$ is the product σ-algebra on $V^T = \prod_{t\in T} V$ generated by the collection of projection maps*

   $$\left\{\, \pi_t \,:\, V^T = \prod_{\tau\in T} V \,\longrightarrow\, V \,:\, w \,\longmapsto\, w(t) \,\right\}_{t\in T}.$$

   Then, $X : \Omega \longrightarrow V^T$ is $\left(\mathcal{A}, \sigma[(V,\mathcal{F})^T]\right)$-measurable if and only if $X_t : \Omega \longrightarrow V$ is $(\mathcal{A}, \mathcal{F})$-measurable for each $t \in T$.

Proof

1. The proof of this result is routine and we omit it.

2. Suppose $X : \Omega \longrightarrow W$ is $(\mathcal{A}, \mathcal{G})$-measurable. Note that $X_t = \pi_t \circ X$, where

$$\pi_t \,:\, V^T = \prod_{t\in T} V \,\longrightarrow\, V \,:\, v \,\longrightarrow\, v(t)$$

   is the projection from $V^T = \prod_{\tau\in T} V$ onto the t-th factor. By hypothesis, $\pi_t : W \longrightarrow V$ is $(\mathcal{G}, \mathcal{F})$-measurable for each $t \in T$. This implies, for each $t \in T$, $X_t = \pi_t \circ X$ is $(\mathcal{A}, \mathcal{F})$-measurable, being a composition of two measurable maps.

3. Since, for each $t \in T$, the projection map $\pi_t : V^T \longrightarrow V$ is $\left(\sigma[(V,\mathcal{F})^T], \mathcal{F}\right)$-measurable (by construction of the σ-algebra $\sigma[(V,\mathcal{F})^T]$ on V^T), the preceding result immediately implies the following implication:

$$(\mathcal{A}, \sigma[(V,\mathcal{F})^T])\text{-measurability of } X : \Omega \longrightarrow V^T \quad\Longrightarrow\quad (\mathcal{A}, \mathcal{F})\text{-measurability of } X_t : \Omega \longrightarrow V, \text{ for each } t \in T.$$

Conversely, suppose X_t is $(\mathcal{A},\mathcal{F})$-measurable for each $t \in T$. Recall that the product σ-algebra on V^T is generated by sets of the form:

$$\pi_t^{-1}(F)\,, \quad \text{for some } t \in T \text{ and } F \in \mathcal{F}.$$

It follows that, for each $t \in T$ and each $F \in \mathcal{F}$, we have

$$X^{-1}\big(\pi_t^{-1}(F)\big) \;=\; (X^{-1} \circ \pi_t^{-1})(F) \;=\; (\pi_t \circ X)^{-1}(F) \;=\; X_t^{-1}(F) \;\subset\; \Omega$$

is $\mathcal{A}$-measurable, since $X_t : (\Omega,\mathcal{A}) \longrightarrow (V,\mathcal{F})$ is $(\mathcal{A},\mathcal{F})$-measurable by hypothesis. This proves that $X : \Omega \longrightarrow V^T$ is $(\mathcal{A}, \sigma[(V,\mathcal{F})^T])$-measurable. □

Definition B.3 (Stochastic processes)
*A **stochastic process** is a family, indexed by some non-empty set T,*

$$\{\; X_t \;:\; (\Omega,\mathcal{A},\mu) \;\longrightarrow\; (V,\mathcal{F}) \;\}_{t\in T}$$

*of $(\mathcal{A},\mathcal{F})$-measurable maps, where the common domain $(\Omega,\mathcal{A},\mu)$ is a probability space and the common codomain $(V,\mathcal{F})$ is a measurable space. The common codomain $(V,\mathcal{F})$ is called the **state space** of the stochastic process.*

Corollary B.4
Suppose:

- *$(\Omega,\mathcal{A},\mu)$ is a probability space and $(V,\mathcal{F})$ is a measurable space.*
- *T is a non-empty set and $W \subset V^T = \prod_{t\in T} V$.*
- *$(W,\mathcal{G})$ is a measurable space structure on W such that, for each $t \in T$, the projection map*

  $$\pi_t \;:\; W \;\longrightarrow\; V \;:\; w \;\longmapsto\; w(t)$$

  is $(\mathcal{G},\mathcal{F})$-measurable.

If $X : (\Omega,\mathcal{A},\mu) \longrightarrow \big(V^T, \sigma[(V,\mathcal{F})^T]\big)$ is a V^T-valued random variable (i.e. X is $\big(\mathcal{A}, \sigma[(V,\mathcal{F})^T]\big)$-measurable), then its set-theoretically equivalent T-indexed family of V-valued maps defined on Ω

$$\left\{\begin{array}{ccccl} X_t & : & (\,\Omega,\mathcal{A},\mu\,) & \longrightarrow & (\,V,\mathcal{F}\,) \\ & & \omega & \longmapsto & (\pi_t \circ X)(\omega) = \pi_t(X(\omega)) = X(\omega)(t) \end{array}\right\}_{t\in T}$$

is a stochastic process (i.e. X_t is $(\mathcal{A},\mathcal{F})$-measurable for each $t \in T$).

Corollary B.5
Suppose:

- *T, Ω, V are non-empty sets.*
- *$(\Omega,\mathcal{A},\mu)$ is a probability space structure on Ω, $(V,\mathcal{F})$ is a measurable space structure on V.*
- *$\sigma[(V,\mathcal{F})^T]$ denotes the corresponding product σ-algebra on $V^T = \prod_{t\in T} V$.*

Let $\{\, X_t : \Omega \longrightarrow V \,\}_{t\in T}$ be a T-indexed family of V-valued maps defined on Ω, and let

$$X \;:\; \Omega \;\longrightarrow\; V^T \;:\; \omega \;\longmapsto\; (t \longmapsto X_t(\omega))$$

be its set-theoretically equivalent (V^T)-valued map defined on Ω. Then,

$$\{\; X_t \;:\; (\Omega,\mathcal{A},\mu) \;\longrightarrow\; (V,\mathcal{F}) \;\}_{t\in T}$$

is a stochastic process if and only if

$$X \;:\; (\Omega,\mathcal{A},\mu) \;\longrightarrow\; \big(V^T, \sigma[(V,\mathcal{F})^T]\big)$$

is a (V^T)-valued random variable.

C Uniqueness of the "full distribution" of a stochastic process $\{ X_t : \Omega \longrightarrow V \}_{t\in T}$ given its finite-dimensional distributions

Definition C.1 (Finite-dimensional distributions of a stochastic process)
Let $\{ X_t : (\Omega,\mathcal{A},\mu) \longrightarrow (V,\mathcal{F}) \}_{t\in T}$ be a stochastic process. Let $n \in \mathbb{N}$ and $t_1, t_2, \ldots, t_n \in T$ be distinct elements of T. Let $\mathcal{P}_{\left(X_{t_1},\ldots,X_{t_n}\right)} \in \mathcal{M}_1(V^n,\mathcal{F}^{\otimes n})$ denote the probability measure induced on the product measurable space $(V^n,\mathcal{F}^{\otimes n})$ by the random variable

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_n}) : (\Omega,\mathcal{A},\mu) \longrightarrow \left(V^n,\mathcal{F}^{\otimes n}\right)$$

*$\mathcal{P}_{\left(X_{t_1},\ldots,X_{t_n}\right)}$ is called a **finite-dimensional distribution** of the stochastic process.*

Theorem C.2
Let $(V,\mathcal{F})$ be a measurable space, and $\sigma\left[(V,\mathcal{F})^T\right]$ the product σ-algebra on $V^T = \prod_{t\in T} V$. Let

$$\{ X_t : (\Omega_X,\mathcal{A}_X,\mu_X) \longrightarrow (V,\mathcal{F}) \}_{t\in T} \quad \text{and} \quad \{ Y_t : (\Omega_Y,\mathcal{A}_Y,\mu_Y) \longrightarrow (V,\mathcal{F}) \}_{t\in T}$$

be two stochastic processes with the same index set T and the same state space $(V,\mathcal{F})$. Let

$$X : (\Omega_X,\mathcal{A}_X,\mu_X) \longrightarrow \left(V^T, \sigma\Big[(V,\mathcal{F})^T\Big]\right) \quad \text{and} \quad Y : (\Omega_Y,\mathcal{A}_Y,\mu_Y) \longrightarrow \left(V^T, \sigma\Big[(V,\mathcal{F})^T\Big]\right)$$

be their respective $\left(V^T, \sigma\Big[(V,\mathcal{F})^T\Big]\right)$-valued random variables. Let $\mathcal{P}_X,\, \mathcal{P}_Y \in \mathcal{M}_1\Big(V^T, \sigma\Big[(V,\mathcal{F})^T\Big]\Big)$ be the probability measures induced on $\left(V^T, \sigma\Big[(V,\mathcal{F})^T\Big]\right)$ by X and Y, respectively. Then,

$$\mathcal{P}_X = \mathcal{P}_Y \in \mathcal{M}_1\Big(V^T, \sigma\Big[(V,\mathcal{F})^T\Big]\Big)$$

if and only if

$$\mathcal{P}_{\left(X_{t_1},X_{t_2},\ldots,X_{t_n}\right)} = \mathcal{P}_{\left(Y_{t_1},Y_{t_2},\ldots,Y_{t_n}\right)} \in \mathcal{M}_1\big(V^n,\mathcal{F}^{\otimes n}\big), \;\; \text{for each } n\in\mathbb{N} \text{ and pairwise distinct } t_1,t_2,\ldots,t_n \in T.$$

Proof

□

D Existence of a stochastic process given its finite-dimensional distributions: Komolgorov's Existence Theorem

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

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

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

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

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.

E Gaussian Processes

Definition E.1 (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 (univariate or multivariate) Gaussian distribution.*

Definition E.2 (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 $\{\,X_t\,\}$.*

Theorem E.3
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 E.4
The mean and covariance functions of a Gaussian process together completely determine its collection of finite-dimensional distributions.

Definition E.5 (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 E.6 (Brownian bridge)
*A **Brownian bridge** is a Gaussian process $\{\,W^\circ_t : (\Omega, \mathcal{A}, \mu) \longrightarrow \mathbb{R}\,\}_{t\in[0,1]}$ indexed by the closed unit interval in $\mathbb{R}$ satisfying the following conditions:*

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

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] Billingsley, P. *Convergence of Probability Measures*, second ed. John Wiley & Sons, 1999.

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

[5] Jacod, J., and Protter, P. *Probability Essentials*. Springer-Verlag, New York, 2004. Universitext.

[6] Villani, C. *Optimal Transport: Old and New*, first ed., vol. 88 of *A Series of Comprehensive Studies in Mathematics*. Springer, 2009.