Donsker's Theorem: A brief review

Abhilash Singh Avik Biswas Sanket Agrawal

Department of Mathematics & Statistics
Indian Institute of Technology Kanpur

MTH516A: Mini Project 2, April 17, 2021

Empirical processes

- Let $Z_1, \ldots Z_n \sim F$ be a random sample taking values in $\mathcal{X}$. Define,

$$\underbrace{F_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{Z_i};}_{\textit{Emprical measure}} \qquad \underbrace{X_n = \sqrt{n}(F_n - F)}_{\textit{Empirical process}}$$

- For any measure μ and measurable f, let μf denote,

$$\mu f := \int f d\mu,$$

- Consider $\mathcal{G} = \{g \text{ defined on } \mathcal{X}, g \text{ measurable}\}$. Then, the collection of random elements $\{X_n g : g \in \mathcal{G}\}$ is called an *empirical process indexed by $\mathcal{G}$*.

Empirical processes

- Let $Z_1, \dots Z_n \sim F$ be a random sample taking values in $\mathcal{X}$. Define,

$$\underbrace{F_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Z_i};}_{\textit{Emprical measure}} \qquad \underbrace{X_n = \sqrt{n}(F_n - F)}_{\textit{Empirical process}}$$

- For any measure μ and measurable f, let μf denote,

$$\mu f := \int f d\mu,$$

- Consider $\mathcal{G} = \{g \text{ defined on } \mathcal{X}, g \text{ measurable}\}$. Then, the collection of random elements $\{X_n g : g \in \mathcal{G}\}$ is called an *empirical process indexed by $\mathcal{G}$*.

Empirical processes

- Let $Z_1, \dots Z_n \sim F$ be a random sample taking values in $\mathcal{X}$. Define,

$$\underbrace{F_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{Z_i};}_{\textit{Emprical measure}} \qquad \underbrace{X_n = \sqrt{n}(F_n - F)}_{\textit{Empirical process}}$$

- For any measure μ and measurable f, let μf denote,

$$\mu f := \int f d\mu,$$

- Consider $\mathcal{G} = \{g \text{ defined on } \mathcal{X}, g \text{ measurable}\}$. Then, the collection of random elements $\{X_n g : g \in \mathcal{G}\}$ is called an *empirical process indexed by $\mathcal{G}$*.

Donsker's class

- Donsker's Class

  A class $\mathcal{G}$ of measurable functions $g : \mathcal{X} \to \mathbb{R}$ is F-Donsker if the process $\{X_n g : g \in \mathcal{G}\}$ indexed by $\mathcal{G}$ converges in distribution in the space $l^\infty(\mathcal{G})$ to a tight random element.

- When is $\mathcal{G}$ $F-$Donsker?

  - $\mathcal{G}$ is finite and consists of square integrable functions: the multi-variate CLT.

  - Else, depends on the size and complexity of the class.

Donsker's class

- Donsker's Class

  A class $\mathcal{G}$ of measurable functions $g : \mathcal{X} \to \mathbb{R}$ is F-Donsker if the process $\{X_n g : g \in \mathcal{G}\}$ indexed by $\mathcal{G}$ converges in distribution in the space $l^\infty(\mathcal{G})$ to a tight random element.

- When is $\mathcal{G}$ $F-$Donsker?
  - $\mathcal{G}$ is finite and consists of square integrable functions: the multi-variate CLT.
  - Else, depends on the size and complexity of the class.

Donsker's class

- Donsker's Class

  A class $\mathcal{G}$ of measurable functions $g : \mathcal{X} \to \mathbb{R}$ is F-Donsker if the process $\{X_n g : g \in \mathcal{G}\}$ indexed by $\mathcal{G}$ converges in distribution in the space $l^\infty(\mathcal{G})$ to a tight random element.

- When is $\mathcal{G}$ $F-$Donsker?

  - $\mathcal{G}$ is finite and consists of square integrable functions: the multi-variate CLT.
  - Else, depends on the size and complexity of the class.

Donsker's class

- **Donsker's Class**

  A class $\mathcal{G}$ of measurable functions $g : \mathcal{X} \to \mathbb{R}$ is F-Donsker if the process $\{X_n g : g \in \mathcal{G}\}$ indexed by $\mathcal{G}$ converges in distribution in the space $l^\infty(\mathcal{G})$ to a tight random element.

- **When is $\mathcal{G}$ $F-$Donsker?**
  - $\mathcal{G}$ is finite and consists of square integrable functions: the multi-variate CLT.
  - Else, depends on the size and complexity of the class.

Covering and Packing numbers

- Let (M, d) be any arbitrary semi-metric space.
- An ϵ-cover of the set M with respect to the semi-metric d is a set $\{x_1, \ldots, x_N\}$ such that for any $x \in M$, there exists a $r \in \{1, \ldots, N\}$ with $d(x, x_r) \leq \epsilon$.
- The ϵ-covering number of M is

$$N(\epsilon, M, d) := \inf\{N \in \mathbb{N} : \exists \text{ a cover } x_1, \ldots, x_N \text{ of } M\}.$$

- An ϵ-packing of the set M with respect to the semi-metric d is a set $\{x_1, \ldots, x_D\} \subseteq M$ such that for all $r, r' \in \{1, \ldots, D\}$, we have $d(x_r, x_r') > \epsilon$.
- The ϵ-packing number of M is

$$D(\epsilon, M, d) := \sup\{D \in \mathbb{N} : \exists \text{ a packing } x_1, \ldots, x_D \text{ of } M\}.$$

- Both covering numbers and packing numbers increase as ϵ decreases.

Covering and Packing numbers

- Let (M, d) be any arbitrary semi-metric space.
- An ϵ-cover of the set M with respect to the semi-metric d is a set $\{x_1, \ldots, x_N\}$ such that for any $x \in M$, there exists a $r \in \{1, \ldots, N\}$ with $d(x, x_r) \leq \epsilon$.
- The ϵ-covering number of M is

$$N(\epsilon, M, d) := \inf\{N \in \mathbb{N} : \exists \text{ a cover } x_1, \ldots, x_N \text{ of } M\}.$$

- An ϵ-packing of the set M with respect to the semi-metric d is a set $\{x_1, \ldots, x_D\} \subseteq M$ such that for all $r, r' \in \{1, \ldots, D\}$, we have $d(x_r, x_r') > \epsilon$.
- The ϵ-packing number of M is

$$D(\epsilon, M, d) := \sup\{D \in \mathbb{N} : \exists \text{ a packing } x_1, \ldots, x_D \text{ of } M\}.$$

- Both covering numbers and packing numbers increase as ϵ decreases.

Covering and Packing numbers

- Let (M, d) be any arbitrary semi-metric space.
- An ϵ-cover of the set M with respect to the semi-metric d is a set $\{x_1, \ldots, x_N\}$ such that for any $x \in M$, there exists a $r \in \{1, \ldots, N\}$ with $d(x, x_r) \leq \epsilon$.
- The ϵ-covering number of M is

$$N(\epsilon, M, d) := \inf\{N \in \mathbb{N} : \exists \text{ a cover } x_1, \ldots, x_N \text{ of } M\}.$$

- An ϵ-packing of the set M with respect to the semi-metric d is a set $\{x_1, \ldots, x_D\} \subseteq M$ such that for all $r, r' \in \{1, \ldots, D\}$, we have $d(x_r, x_r') > \epsilon$.
- The ϵ-packing number of M is

$$D(\epsilon, M, d) := \sup\{D \in \mathbb{N} : \exists \text{ a packing } x_1, \ldots, x_D \text{ of } M\}.$$

- Both covering numbers and packing numbers increase as ϵ decreases.

Covering and Packing numbers

- Let (M, d) be any arbitrary semi-metric space.
- An ϵ-cover of the set M with respect to the semi-metric d is a set $\{x_1, \ldots, x_N\}$ such that for any $x \in M$, there exists a $r \in \{1, \ldots, N\}$ with $d(x, x_r) \leq \epsilon$.
- The ϵ-covering number of M is

$$N(\epsilon, M, d) := \inf\{N \in \mathbb{N} : \exists \text{ a cover } x_1, \ldots, x_N \text{ of } M\}.$$

- An ϵ-packing of the set M with respect to the semi-metric d is a set $\{x_1, \ldots, x_D\} \subseteq M$ such that for all $r, r' \in \{1, \ldots, D\}$, we have $d(x_r, x_r') > \epsilon$.
- The ϵ-packing number of M is

$$D(\epsilon, M, d) := \sup\{D \in \mathbb{N} : \exists \text{ a packing } x_1, \ldots, x_D \text{ of } M\}.$$

- Both covering numbers and packing numbers increase as ϵ decreases.

Covering and Packing numbers

- Let (M, d) be any arbitrary semi-metric space.
- An ϵ-cover of the set M with respect to the semi-metric d is a set $\{x_1, \dots, x_N\}$ such that for any $x \in M$, there exists a $r \in \{1, \dots, N\}$ with $d(x, x_r) \leq \epsilon$.
- The ϵ-covering number of M is

$$N(\epsilon, M, d) := \inf\{N \in \mathbb{N} : \exists \text{ a cover } x_1, \dots, x_N \text{ of } M\}.$$

- An ϵ-packing of the set M with respect to the semi-metric d is a set $\{x_1, \dots, x_D\} \subseteq M$ such that for all $r, r' \in \{1, \dots, D\}$, we have $d(x_r, x_r') > \epsilon$.
- The ϵ-packing number of M is

$$D(\epsilon, M, d) := \sup\{D \in \mathbb{N} : \exists \text{ a packing } x_1, \dots, x_D \text{ of } M\}.$$

- Both covering numbers and packing numbers increase as ϵ decreases.

Covering and Packing numbers

- Let (M, d) be any arbitrary semi-metric space.
- An ϵ-cover of the set M with respect to the semi-metric d is a set $\{x_1, \ldots, x_N\}$ such that for any $x \in M$, there exists a $r \in \{1, \ldots, N\}$ with $d(x, x_r) \leq \epsilon$.
- The ϵ-covering number of M is

$$N(\epsilon, M, d) := \inf\{N \in \mathbb{N} : \exists \text{ a cover } x_1, \ldots, x_N \text{ of } M\}.$$

- An ϵ-packing of the set M with respect to the semi-metric d is a set $\{x_1, \ldots, x_D\} \subseteq M$ such that for all $r, r' \in \{1, \ldots, D\}$, we have $d(x_r, x_r') > \epsilon$.
- The ϵ-packing number of M is

$$D(\epsilon, M, d) := \sup\{D \in \mathbb{N} : \exists \text{ a packing } x_1, \ldots, x_D \text{ of } M\}.$$

- Both covering numbers and packing numbers increase as ϵ decreases.

Bracketing numbers

- $(\mathcal{G}, \|\cdot\|)$ be a subset of a normed linear space of functions $g : \mathcal{X} \to \mathbb{R}$.
- For probability spaces $(\Omega, \mathcal{F}, F)$, norm is usually $L_r(F)$ norm.
- Given two functions $l(\cdot)$ and $u(\cdot)$, the bracket $[l, u]$ is the set $\{g \in \mathcal{G}: l(x) \leq g(x) \leq u(x), \forall x \in \mathcal{X}\}$. An ϵ-bracket is a bracket $[l, u]$ such that $\|l - u\| < \epsilon$.
- The $\epsilon-$bracketing number $N_{[]}(\epsilon, \mathcal{G}, \|\cdot\|)$ is the min. number of ϵ-brackets needed to cover $\mathcal{G}$. They also increase as ϵ decreases.
- The bracketing integral is defined as

$$J_{[]}(\delta, \mathcal{G}, L_2(F)) := \int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{G} \cup \{0\}, L_2(F))} d\epsilon, \delta > 0.$$

- An envelope function of a class $\mathcal{G}$ is any function $x \mapsto G(x)$ such that $g(x) \leq G(x)$, for every $x \in \mathcal{X}$ and $g \in \mathcal{G}$.

Bracketing numbers

- $(\mathcal{G}, \|\cdot\|)$ be a subset of a normed linear space of functions $g : \mathcal{X} \to \mathbb{R}$.
- For probability spaces $(\Omega, \mathcal{F}, F)$, norm is usually $L_r(F)$ norm.
- Given two functions $l(\cdot)$ and $u(\cdot)$, the bracket $[l, u]$ is the set $\{g \in \mathcal{G}: l(x) \leq g(x) \leq u(x), \forall x \in \mathcal{X}\}$. An ϵ-bracket is a bracket $[l, u]$ such that $\|l - u\| < \epsilon$.
- The $\epsilon-$bracketing number $N_{[]}(\epsilon, \mathcal{G}, \|\cdot\|)$ is the min. number of ϵ-brackets needed to cover $\mathcal{G}$. They also increase as ϵ decreases.
- The bracketing integral is defined as

$$J_{[]}(\delta, \mathcal{G}, L_2(F)) := \int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{G} \cup \{0\}, L_2(F))} d\epsilon, \delta > 0.$$

- An envelope function of a class $\mathcal{G}$ is any function $x \mapsto G(x)$ such that $g(x) \leq G(x)$, for every $x \in \mathcal{X}$ and $g \in \mathcal{G}$.

Bracketing numbers

- $(\mathcal{G}, \|\cdot\|)$ be a subset of a normed linear space of functions $g : \mathcal{X} \to \mathbb{R}$.
- For probability spaces $(\Omega, \mathcal{F}, F)$, norm is usually $L_r(F)$ norm.
- Given two functions $l(\cdot)$ and $u(\cdot)$, the bracket $[l, u]$ is the set $\{g \in \mathcal{G}: l(x) \le g(x) \le u(x), \forall x \in \mathcal{X}\}$. An ϵ-bracket is a bracket $[l, u]$ such that $\|l - u\| < \epsilon$.
- The $\epsilon-$bracketing number $N_{[]}(\epsilon, \mathcal{G}, \|\cdot\|)$ is the min. number of ϵ-brackets needed to cover $\mathcal{G}$. They also increase as ϵ decreases.
- The bracketing integral is defined as

$$J_{[]}(\delta, \mathcal{G}, L_2(F)) := \int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{G} \cup \{0\}, L_2(F))} d\epsilon, \delta > 0.$$

- An envelope function of a class $\mathcal{G}$ is any function $x \mapsto G(x)$ such that $g(x) \le G(x)$, for every $x \in \mathcal{X}$ and $g \in \mathcal{G}$.

Bracketing numbers

- $(\mathcal{G}, \|\cdot\|)$ be a subset of a normed linear space of functions $g : \mathcal{X} \to \mathbb{R}$.
- For probability spaces $(\Omega, \mathcal{F}, F)$, norm is usually $L_r(F)$ norm.
- Given two functions $l(\cdot)$ and $u(\cdot)$, the bracket $[l, u]$ is the set $\{g \in \mathcal{G}: l(x) \leq g(x) \leq u(x), \forall x \in \mathcal{X}\}$. An ϵ-bracket is a bracket $[l, u]$ such that $\|l - u\| < \epsilon$.
- The $\epsilon-$bracketing number $N_{[]}(\epsilon, \mathcal{G}, \|\cdot\|)$ is the min. number of ϵ-brackets needed to cover $\mathcal{G}$. They also increase as ϵ decreases.
- The bracketing integral is defined as

$$J_{[]}(\delta, \mathcal{G}, L_2(F)) := \int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{G} \cup \{0\}, L_2(F))} d\epsilon, \delta > 0.$$

- An envelope function of a class $\mathcal{G}$ is any function $x \mapsto G(x)$ such that $g(x) \leq G(x)$, for every $x \in \mathcal{X}$ and $g \in \mathcal{G}$.

Bracketing numbers

- $(\mathcal{G}, \|\cdot\|)$ be a subset of a normed linear space of functions $g : \mathcal{X} \to \mathbb{R}$.
- For probability spaces $(\Omega, \mathcal{F}, F)$, norm is usually $L_r(F)$ norm.
- Given two functions $l(\cdot)$ and $u(\cdot)$, the bracket $[l, u]$ is the set $\{g \in \mathcal{G}: l(x) \leq g(x) \leq u(x), \forall x \in \mathcal{X}\}$. An ϵ-bracket is a bracket $[l, u]$ such that $\|l - u\| < \epsilon$.
- The $\epsilon-$bracketing number $N_{[]}(\epsilon, \mathcal{G}, \|\cdot\|)$ is the min. number of ϵ-brackets needed to cover $\mathcal{G}$. They also increase as ϵ decreases.
- The bracketing integral is defined as

$$J_{[]}(\delta, \mathcal{G}, L_2(F)) := \int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{G} \cup \{0\}, L_2(F))} d\epsilon, \delta > 0.$$

- An envelope function of a class $\mathcal{G}$ is any function $x \mapsto G(x)$ such that $g(x) \leq G(x)$, for every $x \in \mathcal{X}$ and $g \in \mathcal{G}$.

Bracketing numbers

- $(\mathcal{G}, \|\cdot\|)$ be a subset of a normed linear space of functions $g : \mathcal{X} \to \mathbb{R}$.
- For probability spaces $(\Omega, \mathcal{F}, F)$, norm is usually $L_r(F)$ norm.
- Given two functions $l(\cdot)$ and $u(\cdot)$, the bracket $[l, u]$ is the set $\{g \in \mathcal{G}$: $l(x) \leq g(x) \leq u(x), \forall x \in \mathcal{X}\}$. An ϵ-bracket is a bracket $[l, u]$ such that $\|l - u\| < \epsilon$.
- The $\epsilon-$bracketing number $N_{[]}(\epsilon, \mathcal{G}, \|\cdot\|)$ is the min. number of ϵ-brackets needed to cover $\mathcal{G}$. They also increase as ϵ decreases.
- The bracketing integral is defined as

$$J_{[]}(\delta, \mathcal{G}, L_2(F)) := \int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{G} \cup \{0\}, L_2(F))} d\epsilon, \delta > 0.$$

- An envelope function of a class $\mathcal{G}$ is any function $x \mapsto G(x)$ such that $g(x) \leq G(x)$, for every $x \in \mathcal{X}$ and $g \in \mathcal{G}$.

Results

Suppose that a square-integrable (measurable) envelope G exists for $\mathcal{G}$.

Result

$\mathcal{G}$ is $F-$Donsker if and only if there exists a semi-metric $d(\cdot,\cdot)$ on $\mathcal{G}$ such that $(\mathcal{G}, d)$ is totally bounded and

$$\lim_{\delta \to 0} \limsup_{n \to \infty} Pr^* \left(\sup_{d(f,g) \leq \delta; f,g \in \mathcal{G}} |X_n f - X_n g| > \epsilon \right) = 0,$$

for every $\epsilon > 0$. (Asymptotically equicontinuous in probability)

A possible choice for the semi-metric d is $d^2(f,g) = Var_F(f(Z) - g(Z))$

Results

Suppose that a square-integrable (measurable) envelope G exists for $\mathcal{G}$.

Result

$\mathcal{G}$ is $F-$Donsker if and only if there exists a semi-metric $d(\cdot,\cdot)$ on $\mathcal{G}$ such that $(\mathcal{G}, d)$ is totally bounded and

$$\lim_{\delta\to 0} \limsup_{n\to\infty} Pr^* \left(\sup_{d(f,g)\le\delta; f,g\in\mathcal{G}} |X_n f - X_n g| > \epsilon \right) = 0,$$

for every $\epsilon > 0$. (Asymptotically equicontinuous in probability)

A possible choice for the semi-metric d is $d^2(f,g) = Var_F(f(Z) - g(Z))$

Results

Suppose that a square-integrable (measurable) envelope G exists for $\mathcal{G}$.

Result

$\mathcal{G}$ is $F-$Donsker if and only if there exists a semi-metric $d(\cdot,\cdot)$ on $\mathcal{G}$ such that $(\mathcal{G}, d)$ is totally bounded and

$$\lim_{\delta\to 0} \limsup_{n\to\infty} Pr^* \left(\sup_{d(f,g)\le\delta; f,g\in\mathcal{G}} |X_n f - X_n g| > \epsilon \right) = 0,$$

for every $\epsilon > 0$. (Asymptotically equicontinuous in probability)

A possible choice for the semi-metric d is $d^2(f,g) = Var_F(f(Z) - g(Z))$

Results

Result (Bracketing integrals)

Suppose that $J_{[]}(1, \mathcal{G}, L_2(F)) < \infty$. *Then* $\mathcal{G}$ *is* $F-$*Donsker.*

$$J_{[]}(1, \mathcal{G}, L_2(F)) = \int_0^1 \sqrt{\log N_{[]}(\epsilon, \mathcal{G} \cup \{0\}, L_2(F))} d\epsilon$$

- The integrand is a decreasing function of ϵ. So the convergence depends on the size of bracketing numbers as $\epsilon \downarrow 0$.
- We know that $\int_0^1 \epsilon^{-r} d\epsilon$ converges for $r < 1$ and diverges for $r \geq 1$.
- The result is in force if the bracketing entropy, i.e. the logarithm of the bracketing number is bounded by,

$$K\epsilon^{-r} \text{ for some } K \text{ and } r < 2.$$

Results

Result (Bracketing integrals)

Suppose that $J_{[\,]}(1, \mathcal{G}, L_2(F)) < \infty$. *Then* $\mathcal{G}$ *is* $F-$*Donsker.*

$$J_{[\,]}(1, \mathcal{G}, L_2(F)) = \int_0^1 \sqrt{\log N_{[\,]}(\epsilon, \mathcal{G} \cup \{0\}, L_2(F))} d\epsilon$$

- The integrand is a decreasing function of ϵ. So the convergence depends on the size of bracketing numbers as $\epsilon \downarrow 0$.
- We know that $\int_0^1 \epsilon^{-r} d\epsilon$ converges for $r < 1$ and diverges for $r \geq 1$.
- The result is in force if the bracketing entropy, i.e. the logarithm of the bracketing number is bounded by,

$$K\epsilon^{-r} \text{ for some } K \text{ and } r < 2.$$

Results

Result (Bracketing integrals)

Suppose that $J_{[]}(1, \mathcal{G}, L_2(F)) < \infty$. Then $\mathcal{G}$ is $F-$Donsker.

$$J_{[]}(1, \mathcal{G}, L_2(F)) = \int_0^1 \sqrt{\log N_{[]}(\epsilon, \mathcal{G} \cup \{0\}, L_2(F))} d\epsilon$$

- The integrand is a decreasing function of ϵ. So the convergence depends on the size of bracketing numbers as $\epsilon \downarrow 0$.
- We know that $\int_0^1 \epsilon^{-r} d\epsilon$ converges for $r < 1$ and diverges for $r \geq 1$.
- The result is in force if the bracketing entropy, i.e. the logarithm of the bracketing number is bounded by,

$$K\epsilon^{-r} \text{ for some } K \text{ and } r < 2.$$

Results

Result (Bracketing integrals)

Suppose that $J_{[]}(1, \mathcal{G}, L_2(F)) < \infty$. *Then* $\mathcal{G}$ *is* $F-$*Donsker.*

Result (Uniform covering numbers)

Suppose that,

$$\int_0^1 \sup_Q \sqrt{\log N(\epsilon \| G \|_{Q,2}, \mathcal{G}, L_2(Q))} d\epsilon < \infty.$$

Then $\mathcal{G}$ *is* $F-$*Donsker.*

Here the supremum is taken over all probability measures Q for which the class $\mathcal{G}$ is not identically zero.

Class of Indicator functions

Result (Classical Donsker's Theorem)

let $\mathcal{X} = \mathbb{R}$ *and* $\mathcal{G} = \{1_{(-\infty,t]}(\cdot) : t \in \mathbb{R}\}$*. Then* $\mathcal{G}$ *is* $F-$*Donsker, i.e.,* $\{X_n g : g \in \mathcal{G}\}$ *converges to a tight element in the space* $l^\infty(\mathcal{G})$*.*

- *Proof:* We will use the Bracketing integral condition to show this.
- It suffices to show that the bracketing entropy grows no faster than K/ϵ^2 for some constant K.
- Choose $-\infty = t_0 < t_1 \cdots < t_k = \infty$ such that $F(t_i-) - F(t_{i-1}) < \epsilon^2$ for each $i = 1, \ldots, k$. k can be chosen to be less than $2/\epsilon^2$.
- Consider brackets of the form $[1_{(-\infty,t_{i-1}]}, 1_{(-\infty,t_i)}]$. Clearly, these brackets cover $\mathcal{G}$.
- $\|1_{(-\infty,t_{i-1}]} - 1_{(-\infty,t_i)}\|_2 = \|1_{(t_{i-1},t_i)}\|_2 = (F(t_i-) - F(t_{i-1}))^{1/2} < \epsilon$.
- So, $N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$, whence, $\log N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$. □

Class of Indicator functions

Result (Classical Donsker's Theorem)

let $\mathcal{X} = \mathbb{R}$ and $\mathcal{G} = \{1_{(-\infty,t]}(\cdot) : t \in \mathbb{R}\}$. Then $\mathcal{G}$ is $F-$Donsker, i.e., $\{X_n g : g \in \mathcal{G}\}$ converges to a tight element in the space $l^\infty(\mathcal{G})$.

- *Proof:* We will use the Bracketing integral condition to show this.
- It suffices to show that the bracketing entropy grows no faster than K/ϵ^2 for some constant K.
- Choose $-\infty = t_0 < t_1 \cdots < t_k = \infty$ such that $F(t_i-) - F(t_{i-1}) < \epsilon^2$ for each $i = 1, \dots, k$. k can be chosen to be less than $2/\epsilon^2$.
- Consider brackets of the form $[1_{(-\infty,t_{i-1}]}, 1_{(-\infty,t_i)}]$. Clearly, these brackets cover $\mathcal{G}$.
- $\|1_{(-\infty,t_{i-1}]} - 1_{(-\infty,t_i)}\|_2 = \|1_{(t_{i-1},t_i)}\|_2 = (F(t_i-) - F(t_{i-1}))^{1/2} < \epsilon$.
- So, $N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$, whence, $\log N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$. □

Class of Indicator functions

Result (Classical Donsker's Theorem)

let $\mathcal{X} = \mathbb{R}$ and $\mathcal{G} = \{1_{(-\infty,t]}(\cdot) : t \in \mathbb{R}\}$. Then $\mathcal{G}$ is $F-$Donsker, i.e., $\{X_n g : g \in \mathcal{G}\}$ converges to a tight element in the space $l^{\infty}(\mathcal{G})$.

- *Proof:* We will use the Bracketing integral condition to show this.
- It suffices to show that the bracketing entropy grows no faster than K/ϵ^2 for some constant K.
- Choose $-\infty = t_0 < t_1 \cdots < t_k = \infty$ such that $F(t_i-) - F(t_{i-1}) < \epsilon^2$ for each $i = 1, \ldots, k$. k can be chosen to be less than $2/\epsilon^2$.
- Consider brackets of the form $[1_{(-\infty,t_{i-1}]}, 1_{(-\infty,t_i)}]$. Clearly, these brackets cover $\mathcal{G}$.
- $\|1_{(-\infty,t_{i-1}]} - 1_{(-\infty,t_i)}\|_2 = \|1_{(t_{i-1},t_i)}\|_2 = (F(t_i-) - F(t_{i-1}))^{1/2} < \epsilon$.
- So, $N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$, whence, $\log N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$. □

Class of Indicator functions

Result (Classical Donsker's Theorem)

let $\mathcal{X} = \mathbb{R}$ and $\mathcal{G} = \{1_{(-\infty,t]}(\cdot) : t \in \mathbb{R}\}$. Then $\mathcal{G}$ is $F-$Donsker, i.e., $\{X_n g : g \in \mathcal{G}\}$ converges to a tight element in the space $l^\infty(\mathcal{G})$.

- *Proof:* We will use the Bracketing integral condition to show this.
- It suffices to show that the bracketing entropy grows no faster than K/ϵ^2 for some constant K.
- Choose $-\infty = t_0 < t_1 \cdots < t_k = \infty$ such that $F(t_i-) - F(t_{i-1}) < \epsilon^2$ for each $i = 1, \ldots, k$. k can be chosen to be less than $2/\epsilon^2$.
- Consider brackets of the form $[1_{(-\infty,t_{i-1}]}, 1_{(-\infty,t_i)}]$. Clearly, these brackets cover $\mathcal{G}$.
- $\|1_{(-\infty,t_{i-1}]} - 1_{(-\infty,t_i)}\|_2 = \|1_{(t_{i-1},t_i)}\|_2 = (F(t_i-) - F(t_{i-1}))^{1/2} < \epsilon$.
- So, $N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$, whence, $\log N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$. □

Class of Indicator functions

Result (Classical Donsker's Theorem)

let $\mathcal{X} = \mathbb{R}$ and $\mathcal{G} = \{1_{(-\infty,t]}(\cdot) : t \in \mathbb{R}\}$. Then $\mathcal{G}$ is $F-$Donsker, i.e., $\{X_n g : g \in \mathcal{G}\}$ converges to a tight element in the space $l^\infty(\mathcal{G})$.

- *Proof:* We will use the Bracketing integral condition to show this.
- It suffices to show that the bracketing entropy grows no faster than K/ϵ^2 for some constant K.
- Choose $-\infty = t_0 < t_1 \cdots < t_k = \infty$ such that $F(t_i-) - F(t_{i-1}) < \epsilon^2$ for each $i = 1, \ldots, k$. k can be chosen to be less than $2/\epsilon^2$.
- Consider brackets of the form $[1_{(-\infty,t_{i-1}]}, 1_{(-\infty,t_i)}]$. Clearly, these brackets cover $\mathcal{G}$.
- $\|1_{(-\infty,t_{i-1}]} - 1_{(-\infty,t_i)}\|_2 = \|1_{(t_{i-1},t_i)}\|_2 = (F(t_i-) - F(t_{i-1}))^{1/2} < \epsilon$.
- So, $N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$, whence, $\log N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$. □

Class of Indicator functions

Result (Classical Donsker's Theorem)

let $\mathcal{X} = \mathbb{R}$ and $\mathcal{G} = \{1_{(-\infty,t]}(\cdot) : t \in \mathbb{R}\}$. Then $\mathcal{G}$ is $F-$Donsker, i.e., $\{X_n g : g \in \mathcal{G}\}$ converges to a tight element in the space $l^\infty(\mathcal{G})$.

- *Proof:* We will use the Bracketing integral condition to show this.
- It suffices to show that the bracketing entropy grows no faster than K/ϵ^2 for some constant K.
- Choose $-\infty = t_0 < t_1 \cdots < t_k = \infty$ such that $F(t_i-) - F(t_{i-1}) < \epsilon^2$ for each $i = 1, \ldots, k$. k can be chosen to be less than $2/\epsilon^2$.
- Consider brackets of the form $[1_{(-\infty,t_{i-1}]}, 1_{(-\infty,t_i)}]$. Clearly, these brackets cover $\mathcal{G}$.
- $\|1_{(-\infty,t_{i-1}]} - 1_{(-\infty,t_i)}\|_2 = \|1_{(t_{i-1},t_i)}\|_2 = (F(t_i-) - F(t_{i-1}))^{1/2} < \epsilon$.
- So, $N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$, whence, $\log N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$. □

Class of Indicator functions

Result (Classical Donsker's Theorem)

let $\mathcal{X} = \mathbb{R}$ and $\mathcal{G} = \{1_{(-\infty,t]}(\cdot) : t \in \mathbb{R}\}$. Then $\mathcal{G}$ is $F-$Donsker, i.e., $\{X_n g : g \in \mathcal{G}\}$ converges to a tight element in the space $l^\infty(\mathcal{G})$.

- *Proof:* We will use the Bracketing integral condition to show this.
- It suffices to show that the bracketing entropy grows no faster than K/ϵ^2 for some constant K.
- Choose $-\infty = t_0 < t_1 \cdots < t_k = \infty$ such that $F(t_i-) - F(t_{i-1}) < \epsilon^2$ for each $i = 1, \ldots, k$. k can be chosen to be less than $2/\epsilon^2$.
- Consider brackets of the form $[1_{(-\infty,t_{i-1}]}, 1_{(-\infty,t_i)}]$. Clearly, these brackets cover $\mathcal{G}$.
- $\|1_{(-\infty,t_{i-1}]} - 1_{(-\infty,t_i)}\|_2 = \|1_{(t_{i-1},t_i)}\|_2 = (F(t_i-) - F(t_{i-1}))^{1/2} < \epsilon$.
- So, $N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$, whence, $\log N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$. □

Goodness of Fit

- F_n is a natural estimator of the population distribution function F.
- An analysis of the discrepancy between the two can be used to test the hypothesis about the underlying population.
- Some popular goodness of fit statistic are,

$$\sqrt{n}\|F_n - F\|_\infty, \qquad \textit{(Kolmogorov-Smirnov)},$$
$$n\int (F_n - F)^2 dF, \qquad \textit{(Cramér-von Mises)}.$$

- Donsker's result along with the Continuous mapping theorem gives the asymptotic distribution of both statistics.

Goodness of Fit

- F_n is a natural estimator of the population distribution function F.
- An analysis of the discrepancy between the two can be used to test the hypothesis about the underlying population.
- Some popular goodness of fit statistic are,

$$\sqrt{n}\|F_n - F\|_\infty, \qquad \textit{(Kolmogorov-Smirnov)},$$
$$n\int (F_n - F)^2 dF, \qquad \textit{(Cramér-von Mises)}.$$

- Donsker's result along with the Continuous mapping theorem gives the asymptotic distribution of both statistics.

Goodness of Fit

- F_n is a natural estimator of the population distribution function F.
- An analysis of the discrepancy between the two can be used to test the hypothesis about the underlying population.
- Some popular goodness of fit statistic are,

$$\sqrt{n}\|F_n - F\|_\infty, \qquad \textit{(Kolmogorov-Smirnov)},$$

$$n\int (F_n - F)^2 dF, \qquad \textit{(Cramér-von Mises)}.$$

- Donsker's result along with the Continuous mapping theorem gives the asymptotic distribution of both statistics.

M-estimators

- $M-$estimators are those which are obtained as result of a maximization or a minimization procedure.
- Familiar examples include the *OLS estimates* in regression model, *maximum likelihood estimators*, and also various location estimators like the *sample mean* and *sample quantiles*.
- Important results such as the *argmax continuity mapping theorem* and *asymptotic normality* of $M-$estimators utilize the rich theory of empirical processes.

M-estimators

- $M-$estimators are those which are obtained as result of a maximization or a minimization procedure.
- Familiar examples include the *OLS estimates* in regression model, *maximum likelihood estimators*, and also various location estimators like the *sample mean* and *sample quantiles*.
- Important results such as the *argmax continuity mapping theorem* and *asymptotic normality* of $M-$estimators utilize the rich theory of empirical processes.

M-estimators

- $M-$estimators are those which are obtained as result of a maximization or a minimization procedure.
- Familiar examples include the *OLS estimates* in regression model, *maximum likelihood estimators*, and also various location estimators like the *sample mean* and *sample quantiles*.
- Important results such as the *argmax continuity mapping theorem* and *asymptotic normality* of $M-$estimators utilize the rich theory of empirical processes.

References

- Sen, B. A Gentle Introduction to Empirical Process Theory and Applications. Lecture Notes. Columbia University.
- Vaart, A. (1998). *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.
- Vaart, A. and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. Springer-Verlag, New York

Questions????