Donsker's Theorem: A Brief Review

Abhilash Singh (191003), Avik Biswas (191029), Sanket Agrawal (191124)

Department of Mathematics & Statistics, Indian Institute of Technology Kanpur

April 17, 2021

1 Introduction

In our previous report, we presented a brief overview of one of the classical results in Probability theory and Statistics, the Donsker's Theorem, which establishes the weak convergence of empirical processes indexed by the indicator functions to a Brownian Bridge process in the Skorokhod space. We ended our discussion with the generalization of Donsker's theorem to arbitrary real-valued measurable functions and defined the Donsker's class of functions. In the passing, we also mentioned about sufficient conditions based on the size and complexity of the class to determine if a class of functions is Donsker. The present report continues the above discussion and talks about these conditions in a little more detail. We introduce the notions of *covering numbers* and *bracketing numbers* to measure the size and complexity of a class and present some results from the literature.

As previously, we mainly follow the lecture notes of Prof. Bodhisattava Sen, Columbia University.

2 Recall

Consider a random sample $Z_1, \dots Z_n$ taking values in an arbitrary space, $\mathcal{X}$ and following the law F. The corresponding *empirical measure* and *empirical process* are respectively defined as,

$$F_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{Z_i}, \qquad\qquad X_n = \sqrt{n}(F_n - F).$$

For any measurable function g defined on $\mathcal{X}$, denote by Fg (and similarly $F_n g$), the integral of g with respect to F (or F_n). All $F_n g$ and Fg are random elements defined on the original sample space, say Ω. If $\mathcal{G}$ be some collection of arbitrary real-valued measurable functions defined on $\mathcal{X}$, then the collection of random elements $\{X_n g : g \in \mathcal{G}\}$ is called an *empirical process indexed by* $\mathcal{G}$.

Definition 1. *(Donsker's Class). A class $\mathcal{G}$ of measurable functions $g : \mathcal{X} \to \mathbb{R}$ is F-Donsker if the empirical 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.*

The aim is to determine whether a given class of functions $\mathcal{G}$ is Donsker. When $\mathcal{G}$ is finite and consists of square integrable functions, the multi-variate CLT implies that $\mathcal{G}$ is always Donsker.

3 Covering and Bracketing numbers

In this section, we give the definitions of different quantities that measure the size and complexity of a function class. Let (M, d) be any arbitrary semi-metric space.

Definition 2. *(ϵ-cover and Covering number) A 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 $d(x, x_r) \le \epsilon$.*
The ϵ-covering number of M is

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

Hence, the ϵ-covering number of M is the minimum number of open balls of radius ϵ needed to cover M.

Definition 3. *(ϵ-packing and Packing number) 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 of the above notions are closely related and useful for measuring the size of a set. The ϵ-packing number of M is the maximum number of disjoint balls of radius ϵ that can be accommodated in M. Hence, a maximal ϵ-packing of M is also an ϵ-cover of M and thus we have, $N(\epsilon, M, d) \le D(\epsilon, M, d)$.

Now, let $(\mathcal{G}, \|\cdot\|)$ be a subset of a normed linear space of functions $g : \mathcal{X} \to \mathbb{R}$ on some $\mathcal{X}$. When dealing with probability spaces $(\mathcal{X}, \mathcal{F}, F)$, we will define covering numbers for the set $\mathcal{G}$ with respect to the metric associated with $L_r(F)$ norm. For the rest of the report, reader may assume that we are in such probability spaces, although the following definitions hold more generally.

Definition 4. *(ϵ-bracket and Bracketing number) Given two functions $l(\cdot)$ and $u(\cdot)$, the bracket $[l, u]$ is the set $\{g \in \mathcal{G}$: $l(x) \leq g(x) \leq u(x)$, for all $x \in \mathcal{X}$ $\}$. An ϵ-bracket is a bracket $[l, u]$ such that $\|l - u\| < \epsilon$. The bracketing number $N_{[]}(\epsilon, \mathcal{G}, \|\cdot\|)$ is the minimum number of ϵ-brackets needed to cover $\mathcal{G}$.*

Definition 5. *(Bracketing integral) The bracketing integral is defined as*

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

Definition 6. *(Envelope function) 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 $g \in \mathcal{G}$.*

4 Donsker's class of functions

Assume the setup described in Section 2. Further assume that a square-integrable (measurable) envelope G exists for $\mathcal{G}$. Then,

Result 1. *$\mathcal{G}$ 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*

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

We have the following sufficient condition for $\mathcal{G}$ to be $F-$Donsker in terms of bracketing integrals.

Result 2. *Suppose that $J_{[]}(1, \mathcal{G}, L_2(F)) < \infty$. Then $\mathcal{G}$ is $F-$Donsker.*

Observe that the bracketing numbers (and covering and packing numbers) decrease as ϵ increases. Then the integrand can be scaled as ϵ^{-r} for some $r > 0$. Also, $\int_0^1 \epsilon^{-r} d\epsilon$ converges for $r < 1$ and diverges for $r \geq 1$. Hence, the above result is in force if the bracketing entropy, i.e. the log of bracketing numbers is bounded by $K\epsilon^{-r}$ for some K and $r < 2$.

Result 3. *(Classical Donsker's Theorem) In the setup described in Section 2, let $\mathcal{X} = \mathbb{R}$ and $\mathcal{G} = \{\mathbf{1}_{(-\infty,t]}(\cdot) : t \in \mathbb{R}\}$. 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. To use Result 2, we just need to show that the log of the $\epsilon-$bracketing number of the class $\mathcal{G} = \{\mathbf{1}_{(-\infty,t]}(\cdot) : t \in \mathbb{R}\}$ is bounded by $K\epsilon^{-2}$ for some K. Choose $-\infty = t_0 < t_1 \cdots < t_k = \infty$ such that $F(t_i-) - F(t_{i-1}) < \epsilon^2$ for each $i = 1, \ldots, k$. Then, consider brackets of the form $[\mathbf{1}_{(-\infty,t_{i-1}]}, \mathbf{1}_{(-\infty,t_i)}]$. Clearly, these brackets cover $\mathcal{G}$. Further, $\|\mathbf{1}_{(-\infty,t_{i-1}]} - \mathbf{1}_{(-\infty,t_i)}\|_2 = \|\mathbf{1}_{(t_{i-1},t_i)}\|_2 = (F(t_i-) - F(t_{i-1}))^{1/2} < \epsilon$. Now, these brackets can be considered as partitioning the space $[0, 1]$ into disjoint pieces of size ϵ^2. Then, there are $1/\epsilon^2$ such pieces. So, when each piece has size less than ϵ^2, we can choose k to be less than $2/\epsilon^2$. From this, we get that, $N_{[]}(\epsilon, \mathcal{G}, L_2(F)) < 2/\epsilon^2$. And thus, the log of bracketing numbers is also bounded by $2/\epsilon^2$. Consequently, the bracketing integral is finite and so the class $\mathcal{G}$ is $F-$Donsker. $\square$

Another sufficient condition based on uniform covering numbers is the following,

Result 4. *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.

5 Applications

5.1 Goodness of fit

An important application of the empirical distribution is the testing of goodness-of-fit. Some popular goodness of fit statistic are,

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

Both of the above are sequences of continuous functional of empirical process. Hence, it follows from the continuous mapping theorem and the classical Donsker's theorem that the sequences of *Kolmogorov-Smirnov* statistics and *Cramér-von Mises* statistics converge in distribution to appropriate functions of a Brownian Bridge process. Moreover, for any continuous distribution function F, the limiting process remains the same due to inverse transform property.

5.2 $M-$estimators

In estimation theory, $M-$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. The theory of empirical processes is much widely applied to study the asymptotic behaviour of such estimators in general. Important results such as the *argmax continuity mapping theorem* and *asymptotic normality* of $M-$estimators utilize the rich theory of empirical processes. We, however, will not cover the details in the present report owing to our rudimentary understanding of the topic.

6 Conclusion

In this report, we presented a brief discussion on different measures of size and complexity of a set. We also presented some important results from the empirical processes theory concerning Donsker's class of functions. These included the sufficient condition for a class to be Donsker on the basis of bracketing integrals. We also worked out this condition for the special case of indicator functions and proved the classical Donsker's theorem.

Acknowledgement: We express our sincerest gratitude towards Prof. S. S. Dhar, Indian Institute of Technology Kanpur, for this opportunity to read on such an interesting topic and for his invaluable support throughout the project.

7 References

The excellent lecture notes on empirical processes by Prof. Bodhisattava Sen, Columbia University have been the primary source of consultation for all three authors. Additionally, following were also referred.

- 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