Donsker's Theorem: A Brief Review

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

Department of Statistics, Indian Institute of Technology (IIT) Kanpur

Februray 12, 2021

1 Introduction

The earliest attempts to study the convergence of empirical processes as individual stochastic processes can be attributed to Doob (1949). Although, Doob heuristically suggested a limiting process based on finite dimensional asymptotic distributions of the process. Donsker (1951, 1952) provided a justification for this heuristic argument by mathematically establishing a weak convergence result. It was noted later that Donsker's approach suffered from several measurability issues that arose from working with the uniform metric. Hence, a new metric, called the *Skorokhod metric*, was defined on the space of *cádlaǵ* functions (right continuous with existing left limit at each point) that yielded a separable metric space and Donsker's result was re-established on this space. The following decades saw a significant amount of literature towards generalizing these results to more abstract settings. This includes the works of R. M. Dudley, David Pollard, Vladimir Vapnik, and Alexey Chervonenkis, among others. The result of their efforts is the extremely rich theory of empirical processes that we have today.

This report gives a brief discussion on the classical Donsker's Theorem and its generalization.

2 Setup

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* is defined as follows,

$$F_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{Z_i}.$$

It can be verified that F_n is in fact a probability measure which assigns a mass $1/n$ to each of the n sample points. 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. The *empirical process* X_n is defined by,

$$X_n = \sqrt{n}(F_n - F).$$

If $\mathcal{G}$ be some collection of 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* $\mathcal{G}$. Now, consider a special case where $\mathcal{X} = \mathbb{R}$ and $\mathcal{G} = \{I_{(-\infty,t]}(\cdot) : t \in \mathbb{R}\}$. Plugging them in the setup above yields the following familiar quantities,

$$\hat{F}_n(t) := F_n I_{(-\infty,t]} = \frac{1}{n}\sum_{i=1}^{n} I_{(-\infty,t]}(Z_i); \quad X_n(t) := X_n F_n I_{(-\infty,t]} = \sqrt{n}(\hat{F}_n(t) - F(t)); \quad t \in \mathbb{R} \tag{1}$$

i.e. empirical distribution function and the associated empirical process. We then want to talk about the weak convergence of X_n as n goes to infinity, considering each X_n as an individual stochastic process.

3 Doob's heuristic approach and Donsker's result

Consider the setup leading to (1) and assume that $F = U[0,1]$. Then, $X_n(t) = \sqrt{n}(\hat{F}_n(t) - t)$ is a stochastic process indexed by $t \in [0,1]$. For any fixed finite collection, $\{t_1, \dots, t_k\}$ in $[0,1]$, the multivariate Central Limit Theorem implies that the law of the random vector $(X_n(t_1), \dots, X_n(t_k))$ converges to the $k-$dimensional Gaussian distribution with 0 mean and co-variance matrix Σ such that $\Sigma(i,j) = \min\{t_i, t_j\} - t_i t_j$. Based on this observation, Doob heuristically suggested that the required limiting process, X, should be such that its finite dimensional distribution is Gaussian with 0 mean and co-variance matrix Σ. In fact, we have such a process with continuous sample paths. It is called the standard *Brownian Bridge* process. Note that although, this argument is for $U[0,1]$, the same holds for all F. For simplicity, let F be continuous. Then, we can always make the transformation $F(Z) = U$ and work with U. This also leads to the observation that irrespective of the parent distribution, F, the heuristic limiting process is same. Let $\Rightarrow$ denote weak convergence and X be the standard Brownian Bridge process, then we have the following result due to Donsker.

Theorem 1. *(Donsker (1952)).* $X_n \Rightarrow X$ *in* $D([0,1], \|\cdot\|_\infty)$

The space $D[0,1]$ is the space of all cádlaǵ functions on $[0,1]$. Clearly, the paths of X_n live in $D[0,1]$ for all n. Hence, to make sense of the Theorem 1 above, we need to study weak convergence on $D[0,1]$. Note that, X_n is not measurable with respect to the Borel $\sigma-$field on $D[0,1]$ generated by the uniform topology. Hence, in order to talk about weak convergence, following remedies can be used. One can consider a smaller $\sigma-$field then the Borel $\sigma-$field. In particular, it can be shown that X_n are measurable with respect to the *projection $\sigma-$field* generated by the projection maps, $x \mapsto x(t)$ for all $t \in [0,1]$. Another way is to use a different metric that can make X_n measurable. Skorokhod came up with a metric to accomplish this. Popularly, the space D endowed with the topology generated by the Skorokhod metric is called as the *Skorokhod space*. Yet another way is to define a more general notion of weak convergence on any metric space using outer expectations. This notion surpasses any measurability issues by requiring only the limiting element to be measurable and facilitates in the generalization of Theorem 1.

4 Generalizations

Let's go back to the general setup where Z_i's take values in $\mathcal{X}$ and empirical process X_n is indexed by an arbitrary class of real-valued measurable functions $\mathcal{G}$. Then, we have the following definition,

Definition 1. *(Donsker's Class). A class $\mathcal{G}$ of measurable functions $g : \mathcal{X} \to \mathbb{R}$ is 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 definition assumes that the maps $g \mapsto X_n g$ are bounded, and thus the empirical process can be viewed as a map to $l^\infty(\mathcal{G})$. The general notion of weak convergence using outer measures extends quite nicely to the space $l^\infty(.)$. Moreover, since $D(.) \subset l^\infty(.)$, the class of indicator functions becomes a special case of a more general theory. There are several results that provide sufficient conditions for determining if a class of functions is Donsker. These conditions depend on the complexity of the class, as measured by *covering numbers* and *bracketing numbers*.

5 Simulation & Conclusion

We have given a brief overview of one of the most important results in probability theory. In addition, a simulation visualizing Theorem 1 have also been posted online along with working codes. Several topics, including the weak convergence in non-separable metric spaces, covering and bracketing numbers, and applications of Theorem 1 have not been covered in detail and calls for future work. Meanwhile, we believe that this report can serve as a nice ice-breaker for someone starting with the subject.

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

6 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.

- Doob, J. L. (1949). Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics*, 20:393–403.
- Donsker, M. D. (1952). Justification and extension of Doob's heuristic approach to the Komogorov-Smirnov theorems. *Ann. Math. Statistics*, 23:277–281.
- Pollard, D. (1984). *Convergence of stochastic processes* (Springer Series in Statistics). Springer-Verlag, New York.
- Vaart, A. (1998). *Asymptotic statistics* (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge University Press, Cambridge.