

Weak Convergence in General Metric Spaces

Lecturer: Michael I. Jordan

Scribe: Yueqing Wang

1 General Metric (Norm) Space

The objects of interest are functions from a sample space to a general metric space, where each point is a function. Then we can try to use statistical properties, e.g. goodness of fit, to test certain assumptions.

Example 1 (*Cramér-von Mises*). Let P_n be the empirical probability measures of a random sample X_1, \dots, X_n of real-valued random variables. The *Cramér-von Mises statistic* for testing the (null) hypothesis that the underlying probability measure is a given P is given by

$$\int (P_n f - P f)^2 dP,$$

which can be considered as a measure for the distance between P_n and P . If the distribution of this statistic is known, we can test the hypothesis. P can be very complex. But if the class \mathcal{F} of measurable functions is *P-Donsker*, the *Cramér-von Mises statistic* converges to a Brownian Bridge.

Definition 2 (Uniform Norm). The uniform norm on function spaces is defined as

$$\|Z\| = \sup_{t \in T} |Z(t)|. \quad (1)$$

Example 3. Some commonly used general metric spaces:

- $C[a, b]$. All the continuous functions on $[a, b] \in \mathbf{R}$.
- $D[a, b]$. (*Cadlag* functions). All the functions that have limit from the left and are continuous from the right.
- $\ell^\infty[a, b]$. All bounded functions.

And we have,

$$C[a, b] \subseteq D[a, b] \subseteq \ell^\infty[a, b]$$

Note. $C[a, b]$ is separable, i.e. it has a countable dense subset. $D[a, b]$ isn't separable. Hence, $\ell^\infty[a, b]$ is not separable, neither. Most of the empirical processes are in $D[a, b]$ because of the jumps; most limiting processes are in $C[a, b]$.

2 Weak Convergence

Definition 4 (Random Element). The *Borel σ -field* on a metric space \mathbb{D} is the smallest σ -field that contains the open sets (and then also the closed sets). A function defined relative to (one or two) metric

spaces is called *Borel-measurable* if it is measurable relative to the Borel σ -field(s). A Borel-measurable map $X : \Omega \rightarrow \mathbb{D}$ defined on a probability space (Ω, \mathcal{U}, P) is referred to as a *random element* with values in \mathbb{D} .

Definition 5. *Random Elements* (R.E.) X_n converging weakly to the random element X means $\mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X)$, for all bounded and continuous function f .

Note. For random elements, Continuous Mapping Theorem still holds. If random elements $X_n \xrightarrow{d} X$ and functions $g_n \rightarrow g$ are continuous, it follows that

$$g_n(X_n) \xrightarrow{d} g(X)$$

Definition 6. A random element is *tight* if $\forall \epsilon > 0, \exists$ a compact set K such that

$$\mathbb{P}(X \notin K) \leq \epsilon.$$

Definition 7. $X = \{X_t : t \in T\}$ is a collection of random variables, where $X_t : \Omega \rightarrow \mathbb{R}$ is defined on (Ω, \mathcal{U}, P) . A *sample path* is defined as $t \rightarrow X_t(\omega)$.

Theorem 8 (Converge Weakly to a Tight Random Element). A sequence of maps $X_n : \Omega_n \rightarrow l^\infty(T)$ converge weakly to a tight R.E. iff

- (i) (*Fidi Convergence*) $(X_{n,t_1}, \dots, X_{n,t_k})$ converges weakly in \mathbf{R}^k for each finite set (t_1, \dots, t_k) .
- (ii) (*Asymptotic Partition*) $\forall \epsilon, \eta > 0$, exists a partition of T into finitely many sets T_1, \dots, T_k such that

$$\limsup_{n \rightarrow \infty} \mathbb{P}(\sup_i \sup_{s, t \in T_i} |X_{n,s} - X_{n,t}| \geq \epsilon) \leq \eta.$$

3 The Donsker Theorems

Theorem 9 (Classical Donsker Theorem). If X_1, \dots are i.i.d. random variables with distribution function F , where F is uniform distribution function on the real line and $\{\mathbb{F}_n\}$ are the empirical processes: $\mathbb{F}_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq t\}}$. Then for fixed (t_1, \dots, t_k) , it follows that,

$$\sqrt{n}(\mathbb{F}_n(t_1) - F(t_1), \dots, \mathbb{F}_n(t_k) - F(t_k)) \xrightarrow{d} (\mathbb{G}_F(t_1), \dots, \mathbb{G}_F(t_k)),$$

where $\{\mathbb{G}_F(t_i)\}$ are zero-mean Gaussian with covariance $t_i \wedge t_j - t_i t_j$.

Theorem 10 (Donsker). If X_1, \dots are i.i.d. random variables with distribution function F , then the sequence of empirical processes $\sqrt{n}(\mathbb{F}_n - F)$ converges in distribution in the space $D[-\infty, \infty]$ to a tight random element \mathbb{G}_F (i.e. a Brownian Bridge), whose marginal distributions are zero-mean normal with covariance function: $\mathbb{E}\mathbb{G}_F(t_i)\mathbb{G}_F(t_j) = F(t_i \wedge t_j) - F(t_i)F(t_j)$.

Denote empirical processes as follows: $\mathbb{G}_n = \sqrt{n}(P_n - P)$ and thus $\mathbb{G}_n f = \sqrt{n}(P f_n - P f)$.

Definition 11 (P-Donsker). \mathcal{F} is *P-Donsker* if \mathbb{G}_n converges weakly to a tight limit process in $l^\infty(\mathcal{F})$ which is a P-Brownian Bridge \mathbb{G}_P with zero mean and covariance function $\mathbb{E}\mathbb{G}_P f \mathbb{G}_P g = P f g - P f P g$.

Definition 12. Define the *Bracketing Integral* as,

$$J_{[]}(\delta, \mathcal{F}, L_2(P)) = \int_0^\delta \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P))} d\epsilon$$

Theorem 13. If $J_{[]} (1, \mathcal{F}, L_2(P)) < \infty$, \mathcal{F} is *P-Donsker*.

Example 14. $\mathcal{F} = \{1_{(-\infty, t]} : t \in \mathbf{R}\}$. By calculating the bracketing number, it follows that $\log N_{[]} \rightarrow \frac{1}{\epsilon^2}$. Hence there exists limits for $J_{[]} (1, \mathcal{F}, L_2(P))$. By the above theorem we know that this function space is *P-Donsker* and the empirical processes will converge to a Brownian Bridge.

Example 15 (Lipschitz Classes are P-Donsker). Let $\mathcal{F} = \{f_\theta : \theta \in \Theta \subset \mathbf{R}^d\}$ be a Lipschitz function class. i.e. given x (fixed), if

$$|f_{\theta_1}(x) - f_{\theta_2}(x)| \leq m(x) \|\theta_1 - \theta_2\|, \forall \theta_1, \theta_2,$$

then,

$$N_{[]}(\epsilon \|m\|_{p,r}, \mathcal{F}, L_r(P)) \leq k \left(\frac{\text{diameter } \Theta}{\epsilon} \right)^d,$$

where k is a certain constant.

Proof. The brackets $(f_\theta - \epsilon m, f_\theta + \epsilon m)$ for θ have size smaller than $2\epsilon \|m\|_{p,r}$. And they cover \mathcal{F} because,

$$f_{\theta_1} - \epsilon m \leq f_{\theta_2} \leq f_{\theta_1} + \epsilon m, \text{ if } \|\theta_1 - \theta_2\| \leq t.$$

Hence, we need at most $\left(\frac{\text{diam } \Theta}{\epsilon}\right)^d$ cubes of size ϵ to cover Θ , and then use balls to cover the cubes. \square

References

References

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press, Cambridge.