Introduction to Empirical Process Theory

Lab 1, BIO251

02/03/2014

1 Introduction

1.1 Historical Motivation

Historically, empirical processes theory formally started with the interest of people in the empirical distribution function. The empirical distribution is given by:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$$

Where X_i are iid random elements defined on a measurable space $(\mathcal{X}, \mathcal{A})$. A natural candidate to what the empirical distribution converges as a function, is of course the pointwise expected value of the empirical distribution function, to which it's consistent at each fixed x by the LLN.

Glivenko-Cantelli's theorem (1933) formalizes this intuition, and shows the uniform convergence of $F_n(x)$ to $F(x) = P(X \leq x)$, i.e. $\sup_x |F_n(x) - F| \to 0$.

Note. we will use the notation $\rightsquigarrow$ instead of $\xrightarrow{\mathcal{D}}$ for weak convergence.

It was then clear that if GC-theorem gives rise to some *uniform* law of large numbers, there must be some equivalent of the CLT in that case.

Kolmogorov-Smirnov test nonparametric test statistic for testing whether a distribution comes from a given distribution is one of the seminal works in that direction. The test statistic is $|F_n - F|_\infty$. Kolmogorov and Smirnov derived the asymptotic distribution of the statistic using direct arguments. In fact it turns out that (under the null) $\sqrt{n}\sup_x |F_n(x) - F(x)| = \sqrt{n}|F_n - F|_\infty \rightsquigarrow \sup_x |B(F(x))|$, where $B(x)$ here is the *Brownian Bridge*, which is an example of a *Gaussian Process*.

From the standard CLT we know that for each fixed t: $\sqrt{n}(F_n(t) - F(t)) \rightsquigarrow G(t)$, where we have $\mathrm{Cov}(G(t), G(s)) = F(s \wedge t) - F(s)F(t)$. In fact *Donsker* (1952) showed that something much more powerful is true $G_n = \sqrt{n}(F_n - F) \rightsquigarrow G$, where G here is a 0 mean Gaussian process, with $E(G(s)G(t)) = F(s \wedge t) - F(s)F(t)$. Another way of stating the same result, which underlines the difference between Donsker's result and the CLT is that $G_n \rightsquigarrow G$ in $\ell^\infty(\mathbb{R})$, where for any index set T, $\ell^\infty(T)$ is the collection of all bounded functions $f : T \to \mathbb{R}$.

As it is well known (ha-ha!) the Brownian bridge process (restricted on the unit interval) is a Gaussian process that has exactly the same covariance structure: $s \wedge t - st$. It can be seen that the process G can be re-expressed as $B(F(t))$. Now with another (giant) leap of faith, that we can use something like a continuous mapping theorem, it should be believable that the KS test, asymptotic distribution is correct.

1.2 Abstract Formulation

Now let's take time to view the modern thinking and notation in the empirical process theory. Gradually people realized that the theory for the empirical distribution function, that we just mentioned, i.e. Glivenko-Cantelli's and Dosnker's theorems, can be verified over a much broader classes of processes rather than just the empirical distribution.

Define the *empirical measure* $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$. For a given set C it can be seen that, $\mathbb{P}_n(C) = n^{-1}\#(1 \leq i \leq n : X_i \in C)$. Let $\mathcal{F}$ be a collection of measurable functions. Define a map from $\mathcal{F}$ to $\mathbb{R}$ by:

$$f \mapsto \mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^n f(X_i)$$

The abbreviation $Qf = \int f dQ$ for a measurable function f and a measure Q is common in empirical process theory, and means nothing but the expectation of f under the measure Q. The centered and scaled version of this map, is called the *empirical process* $\mathbb{G}_n$ and is given by:

$$f \mapsto \mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n f - Pf)$$

Where here by P we mean the probability law of the random elements X_i. What can we say is true by LLN and CLT in this case, provided that Pf exists, and $Pf^2 < \infty$?

Define $||Q||_{\mathcal{F}} = \sup\{Qf : f \in \mathcal{F}\}$, under this notation a uniform version of the LLN would look like:

$$||\mathbb{P}_n - P||_{\mathcal{F}} \to 0$$

The convergence above is in (outer) probability. A class of functions $\mathcal{F}$ for which this is true is called a *Glivenko-Cantelli class*, or P-Glivenko-Cantelli to stress on the dependence on the measure P.

We can further view the empirical process $\{\mathbb{G}_n f : f \in \mathcal{F}\}$ as a map into $\ell^\infty(\mathcal{F})$ (provided we have assumed $\sup_f |f(x) - Pf| < \infty$ for all x). Therefore it makes sense to search for conditions on $\mathcal{F}$ so that:

$$\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P) \rightsquigarrow \mathbb{G}, \text{ in } \ell^\infty(\mathcal{F})$$

Such a class is called a *Donsker class*, or a P-Donsker class. We can still claim by standard multivariate CLT argument that the finite dimensional distributions of the process, i.e. for any finite set of functions $f_1, \ldots, f_k$ we would have:

$$(\mathbb{G}_n f_1, \ldots, \mathbb{G}_n f_k) \rightsquigarrow N_k(0, \Sigma)$$

Where $\Sigma_{ij} = P(f_i - Pf_i)(f_j - Pf_j)$.

We conclude this section by translating the more abstract second part, to the first part with a simple example.

Example Let $X_1, \ldots, X_n$ be iid random elements in $\mathbb{R}$, and let $\mathcal{F}$ be the collation of all indicators $\{\mathbb{1}((-\infty, t]) : t \in \mathbb{R})\}$. Then we can identiy the empirical measure with the empirical distribution function.

2 Outer Integrals

In order for us to be able to write $\rightsquigarrow$, and talk about processes we need to define these notions (which takes quite a bit of effort).

To elaborate on the difficulty of the weak convergence let $(\mathbb{D}, d)$ is a metric space and P_n and P are Borel probability measures on $(\mathbb{D}, \mathcal{D})$, where $\mathcal{D}$ is the Borel σ-field on $\mathbb{D}$. We say that $P_n \rightsquigarrow P$, if and only if:

$$\int_{\mathbb{D}} f dP_n \to \int_{\mathbb{D}} f dP, \text{ for all } f \in C_b(\mathbb{D})$$

where $C_b(\mathbb{D})$ is the space of bounded and continuous functions $f : \mathbb{D} \mapsto \mathbb{R}$. Equivalently if X_n and X are random variables we have that $X_n \rightsquigarrow X$, if and only if:

$$\mathrm{E}\, f(X_n) \to \mathrm{E}\, f(X), \text{ for all } f \in C_b(\mathbb{D})$$

A key requirement that is hidden here is that the measures P_n is defined for each n of the Borel σ-field $\mathcal{D}$, or equivalently that the random variables X_n, are defined on probability spaces $(\Omega_n, \mathcal{A}_n, P_n)$ such that $X_n^{-1}(D) \in \mathcal{A}_n$ for each $D \in \mathcal{D}$.

Example. An example where this seemingly unimportant restriction fails, is the Skorohod space $\mathbb{D} = D[0,1]$ of all right-continuous functions with left hand limits, equipped with the uniform metric (the sup metric, that it inherits from $\ell^\infty([0,1])$). Consider a one sample empirical process defined as $X : [0,1] \mapsto \mathbb{D}$, defined through $X(\omega) = \mathbb{1}_{[\omega,1]}$. If $[0,1]$ is equipped with the Borel σ-field, then X is not Borel not measurable.

To see this let B_s be the open ball of radius $1/2$ in $\mathbb{D}$ around the function $\mathbb{1}_{[s,1]}$. For any $S \subset [0,1]$ then the set $G = \cup_{s \in S} B_s$ is open. Now $X(\omega) \in B_s$ is equivalent to $\omega = s$ (draw a pic!), implying that $X^{-1}(G) = S$. Therefore if X ought to be measurable every subset of $[0,1]$ should be open which is apparently not true.

This surprising fact happens because the σ-field on $\mathcal{D}$ is simply too big. Therefore people started asking themselves alternative ways to define, weak convergence.

So one idea to approach this problem is to define the *outer integral*. Let $(\Omega, \mathcal{A}, P)$ is a probability space, and $T : \Omega \mapsto \bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$. The outer integral of T with respect to P is defined through:

$$\mathrm{E}^*\, T = \inf\{\mathrm{E}\, U : U \geq T, U : \Omega \mapsto \bar{\mathbb{R}} \text{ measurable and } \mathrm{E}\, U \text{ exists}\}$$

Similarly we can define the *outer probability*:

$$P^*(B) = \inf\{P(A), A \supset B, A \in \mathcal{A}\}$$

The *inner probability* and *inner expectation* can be defined in a similar way, or directly by: $\mathrm{E}_*\, T = -\mathrm{E}^* -T$, and $P_*(B) = 1 - P^*(\Omega - B)$.

We next prove a useful lemma, and comment on properties of the outer expectation.

Measurable Cover Function. For any map $T : \Omega \mapsto \bar{\mathbb{R}}$, there exists a measurable function T^* with:

(i) $T^* \geq T$

(ii) $T^* \leq U$ a.s., for every measurable $U : \Omega \mapsto \bar{\mathbb{R}} =$ with $U \geq T$ a.s.

For any T^* with the above requirements we have $\mathrm{E}^* T = ET^*$, provided that $\mathrm{E}\, T^*$ exists. If $\mathrm{E}^* T < \infty$ then the last requirement is necessarily true.

Proof:
Choose a measurable sequence $U_m \geq T$ with $\mathrm{E} \arctan U_m \downarrow \mathrm{E}^* \arctan T$, and set:

$$T^*(\omega) = \lim_{m \to \infty} \inf_{1 \leq k \leq m} U_k(\omega)$$

As defined the function T^* is measurable (why?), and takes values on the extended real line, and $T^* \geq T$. By monotone convergence $\mathrm{E} \arctan T^* = \mathrm{E}^* \arctan T$. For every measurable $U \geq T$ we have that $\arctan U \wedge T^* \geq \arctan T$, and therefore $\mathrm{E} \arctan U \wedge T^* \geq \mathrm{E}^* \arctan T = \mathrm{E} \arctan T^*$. This of course implies that $U \geq T^*$ a.s.

If $\mathrm{E}\, T^*$ exists then we get that $\mathrm{E}\, T^* \overset{\text{(i)}}{\geq} \mathrm{E}^* T \overset{\text{(ii)}}{\geq} \mathrm{E}\, T^*$. If $\mathrm{E}^* T < \infty$, then there exists a measurable function U with $\mathrm{E}\, U^+ < \infty$, which implies $\mathrm{E}(T^*)^+ \leq \mathrm{E}\, U^+ < \infty$, and thus $\mathrm{E}\, T^*$ exists.

After this fact we state (without proof) some properties of outer expectations:

Lemma The following statements are true for arbitrary maps $S, T : \Omega \mapsto \bar{\mathbb{R}}$.

(i) $(S + T)^* \leq S^* + T^*$, with equality when S is measurable

(ii) $(S - T)^* \geq S^* - T^*$

(iii) $|S^* - T^*| \leq |S - T|^*$

(iv) $(S \vee T)^* = S^* \vee T^*$

(v) $(S \wedge T)^* \leq S^* \wedge T^*$

Fubini's Theorem Let T be defined on a product probability space. Then $\mathrm{E}_* T \leq E_{1*} E_{2*} T \leq E_1^* E_2^* T \leq \mathrm{E}^* T$.

Monotone and dominated convergence theorems stay true for outer expectations.