Back to Empirical Processes

Lab 5, BIO251

03/03/2014

1 Spaces of Bounded Functions

Theorem. Let $X_n : \Omega_n \mapsto \ell^\infty(T)$ be arbitrary. Then X_n converges weakly to a tight limit iff X_n is asymptotically tight, and the marginals $(X_n(t_1), \ldots, X_n(t_k))$ converge weakly to a limit for every finite subset $t_1, \ldots, t_k$ of T. If X_n is asymptotically tight and it's marginals converge weakly to the marginals $(X(t_1), \ldots, X(t_k))$ of a stochastic process X then there is a version of X with uniformly bounded sample paths and $X_n \rightsquigarrow X$.

Theorem Proof. If X_n is asymptotically tight and the marginals converge weakly, then by Lemma 1 X_n is asymptotically measurable. Now we can apply Prohorov's theorem to get that we can divide X_n into subsequences each of which is converging weakly to a tight limit. If we show that each limit is the same we are done. Because we have the marginal convergence (on all subsequences), and Lemma 2 this finishes the proof in this direction. Now if X_n converges weakly to a tight limit we know that this is equivalent to X_n being asymptotically tight. Furthermore the marginals weak convergence comes as an implication of the continuous mapping theorem (note that the map $f(z) = (z(t_1), \ldots, z(t_k))$ is continuous for $z \in \ell^\infty(T)$).

Finally note that if X_n is asymptotically tight and the marginals converge, by the first part we have that $X_n \rightsquigarrow X$, where X is tight limit. Furthermore, since X is tight, then it concentrates on a σ-compact set $K \subset \ell^\infty(T)$ with $P(X \in K) = 1$. The last implication shows the uniform boundedness of the sample paths.

2 Characterizing Asymptotic Tightness in $\ell^\infty(T)$

While we know how to deal with the marginal weak convergence, characterizations of the asymptotic tightness are needed, in order to be able to prove weak convergence of the empirical process.
We formulate a theorem which is essentially relating the (asymptotic uniform, equi-) continuity of the sample paths $t \mapsto X_n(t)$ to asymptotic tightness.

We have the following theorem:

Theorem. A sequence $X_n : \Omega_n \mapsto \ell^\infty(T)$ is asymptotically tight, if and only if $X_n(t)$ is asymptotically tight in $\mathbb{R}$ for every t, and there exists a semimetric ρ on T such that (T, ρ) is totally bounded and X_n is *asymptotically uniformly ρ-equicontinuous in probability*, i.e. for every $\varepsilon, \eta > 0$, there exist a $\delta > 0$ such that:

$$\limsup_n P^* \left(\sup_{\rho(s,t)<\delta} |X_n(s) - X_n(t)| > \varepsilon \right) < \eta$$

Futhermore we have:

Addendum. If, moreover, $X_n \rightsquigarrow X$, then almost all paths $t \mapsto X(t,\omega)$ are uniformly ρ-continuous; and the semimetric ρ can WLOG be taken equal to any semimetric ρ for which this is true and (T,ρ) is totally bounded.

Let's spend some time on understanding the theorem and the addendum.

It is not surprising that we require asymptotic tightness for each of the coordinate projections $X_n(t)$, since any continuous function $g(X_n)$ should be asymptotically tight if X_n is asymptotically tight, and the projection is a particular example of that. This is true because a continuous image of a compact set is compact.

What does a totally bounded set mean? Totally bounded set, is a set such that for each $\varepsilon > 0$ we can find a finite cover of the set consisting of open balls of radius ε of the set. So the theorem is really saying that we can "chop up" the space T into finite number of balls of radius δ, and on each of these balls, for all n it is very likely that our X_n's restricted to a ball should be very close to each other with high probability.

In other words if we denote the balls with B_i, $i = 1, \dots, k$, we have that the behavior of the processes $X_n(t)$ could be explained by the marginal distributions $(X_n(t_1), \dots, X_n(t_k))$, where $t_i \in B_i$ up to errors of ε, η.

The addendum, seems also to make intuitive sense. If the sample paths of the sequence of processes become to act similarly on the balls B_i then so must be true for the sample paths of X, hence continuity should be expected. Furthermore, if many metrics ρ are making (T,ρ) totally bounded and the sample paths of X uniformly ρ-continuous it would be "unnatural" if we could only use some of these, to show the asymptotic uniform equicontinuity. Fortunately the addendum guarantees that that's not the case.

There is a question of what semimetrics one should try in practice. Examples could be $\rho_0(s,t) = \mathrm{E}\arctan|X(s) - X(t)|$, $\rho_p(s,t) = (\mathrm{E}\,|X(s) - X(t)|^p)^{1/(p\vee 1)}$, $0 < p < \infty$. Checking that these are semimetrics is left as an exercise (Hints: consider the function $\arctan(x){+}\arctan(c{-}x)$, for $0 \le x \le c$, $|a+b|^p \le |a|^p + |b|^p$ for $p < 1$, and Minkowski's inequality).

It can be shown that ρ_0, would do the job, however it might not be convenient to use. For the ρ_p metrics it is not clear wether they would work, as the expectations need not even be finite.

Turns out however, that for the Gaussian process, we shouldn't be worried about using ρ_p.

2.1 Gaussian Processes

Def. A stochastic process X is called, *Gaussian* if each of its finite dimensional marginals $(X(t_1), \dots, X(t_k))$ has a multivariate normal distribution on Euclidean space.

Theorem. Let X be a Gaussian process with "intrinsic" semimetrics ρ_p, and let X_n be a sequence of random elements with values in $\ell^\infty(T)$. Then there exists a version of X which is a tight Borel measurable map into $\ell^\infty(T)$ and X_n converges weakly to X if and only if for some p (and then for all p):

(i) The marginals of X_n converge weakly to the corresponding marginals of X

(ii) X_n is asymptotically equicontinuous in probability with respect to ρ_p

(iii) T is totally bounded for ρ_p

Typically one uses ρ_2 as a semimetric as it is the easiest to work with.

The sufficiency of these conditions should be evident from the theorem above. However in the Gaussian process case, we have also that these conditions are necessary, i.e. we can always use any of the semimetrics ρ_p. This is not obvious form the theorem but we don't discuss it further here.

3 Back to Empirical Processes

Recal the notations:

$$\mathbb{P}_n = n^{-1}\sum_{i=1}^{n}\delta_{X_i}$$
$$\mathcal{F} = f : \mathcal{X} \mapsto \mathbb{R} : f \text{ measurable}$$
$$f \mapsto \mathbb{P}_n f = n^{-1}\sum_{i=1}^{n} f(X_i), \text{ with } Qf = \int f dQ$$

The centered and scaled version of $\mathbb{P}_n f$ is the empirical process indexed by $\mathcal{F}$:

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

We are interested in giving conditions on the classes $\mathcal{F}$ such that we have:

- $||\mathbb{P}_n - P||_{\mathcal{F}} = \sup_{f\in\mathcal{F}} ||(\mathbb{P}_n - P)f|| \to 0$, outer almost surely
- $\mathbb{G}_n \rightsquigarrow \mathbb{G}$, in $\ell^\infty(\mathcal{F})$

Clearly we have:

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

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

Therefore if there was a process $\mathbb{G}$ to which the empirical process should converge it ought to be a Gaussian process, $\{\mathbb{G}f : f \in \mathcal{F}\}$ with zero mean and covariance function:

$$\mathrm{E}\,\mathbb{G}f_1\mathbb{G}f_2 = P(f_1 - Pf_1)(f_2 - Pf_2) = Pf_1f_2 - Pf_1Pf_2$$

According to one of the lemmas from last time, this and tightness completely characterize the limiting process.

4 Maximal Inequalities

In order for us to deal with properties like the, stochastic equicontinuity we need to establish several very useful inequalities.

Definition. *Orlicz norm* of a non-decreasing, convex function ψ with $\psi(0) = 0$ for a random variable X: $||X||_\psi$ is defined as:

$$||X||_\psi = \inf\{C > 0 : \mathrm{E}\,\psi\left(\frac{|X|}{C}\right) \leq 1\}$$

Verifying that this is a norm is left as an exercise. Hint: consider multiplying by $\frac{||X||_\psi}{||X||_\psi+||Y||_\psi}$ and $\frac{||Y||_\psi}{||X||_\psi+||Y||_\psi}$.

Examples of these norms include $\psi(x) = x^p$, $p \geq 1$ which gives the usual L_p norm – $(\mathrm{E}\,||X||_p)^{1/p}$. Another example which we are going to be using more is the function $\psi_p(x) = \exp(x^p) - 1$. The latter function gives much more weight to the tails of the distribution. We have that $x^p \leq \psi_p(x)$ which implies that $||X||_p \leq ||X||_{\psi_p}$.

We have the following two inequalities for the ψ_p norms:

$$
\begin{aligned}
&||X||_{\psi_p} \leq ||X||_{\psi_q}(\log 2)^{p/q}, \text{ for } p \leq q \\
&||X||_p \leq p!||X||_{\psi_1}
\end{aligned}
$$

They are left as an exercise. Hints: for the first one note that the function ϕ for which $\psi_p(x(\log 2)^{1/p}) = \phi(\psi_q(x(\log 2)^{1/q}))$ is a concave function with $\phi(1) = 1$, so you can use Jensen's inequality. For the second one just use a Taylor expansion. These inequalities imply that up to constants (which are irrelevant for our purposes) the ψ_p give better bounds than the L_p norms.

We can use Orlicz norms in conjunction with Markov's inequality to obtain tail bounds of random variables:

$$
P(|X| > x) = P(\psi(|X|/||X||_\psi) > \psi(x/||X||_\psi)) \leq \frac{\mathrm{E}\,\psi(|X|/||X||_\psi)}{\psi(x/||X||_\psi)} \leq \frac{1}{\psi(x/||X||_\psi)}
$$

When ψ is ψ_p this leads to abound of the sort $\exp(-Cx^p)$. In the case when ψ is ψ_2 we get a a Gaussian tail bound, which explains our interest in this type of norms. Conversely, having a random variable with such a tail bound, shows that $||X||_{\psi_p}$ is finite, as we can reassure from the following lemma:

Lemma. Let X be a random variable with $P(|X| > x) \leq K\exp(-Cx^p)$, for every x for some constants K and C, and for $p \geq 1$. Then its Orlicz norm satisfies $||X||_{\psi_p} \leq ((1+K)/C)^{1/p}$.

Proof. We have the following series of inequalities:

$$
\begin{aligned}
\mathrm{E}\exp(|X|^p/R) - 1 &= \mathrm{E}\int_0^{|X|^p} \frac{1}{R}\exp(x/R)dx = \mathrm{E}\int_0^\infty \mathbb{1}_{x<|X|^p}\frac{1}{R}\exp(x/R)dx \\
&= \int_0^\infty P(|X| > x^{1/p})\frac{1}{R}\exp(x/R)dx \leq \int_0^\infty K\exp(-Cx)\frac{1}{R}\exp(x/R)dx \\
&\leq K/R\frac{1}{C-1/R} = \frac{K}{CR-1}
\end{aligned}
$$

When $R = (1+K)/C$ we have that the expectation above is bounded by 1, which completes the proof. (Technically we had to consider the case when $1/||X||_\psi > C$ but if this holds the bound we claim will still hold)

Note that for L_p norms we have the following:

$$
\left|\left|\max_{1\leq i\leq m} X_i\right|\right|_p = \left(\mathrm{E}\max_{1\leq i\leq m}|X_i|^p\right)^{1/p} \leq m^{1/p}\max_{1\leq i\leq m}||X_i||_p
$$

Next time we continue with a similar inequality for ψ norms.