Note Title 3/28/2006

- Empirical Process Preliminaries

(d, T) pseudometric space. We start by defining $\boxed{\text{covering numbers}}$.

Covering Number -- For each $\varepsilon > 0$, define the covering number $N(\varepsilon, d, T)$ to be the smallest number N for which there exist points $c_1, \dots, c_N$ in T such that

$$\min_i d(t, c_i) \leq \varepsilon \text{ for each } t \in T$$

Note that $N(\varepsilon, d, T)$ is the smallest number of closed balls of radius ε and centers in T whose union covers T.

Example: Let $T = [0,1]$; $d(x,y) = |x-y|$
for $i = 1, 2, \ldots$, it is easy to see that

$$N(1/2^i, d, T) = 2^{i-1}$$

To see this, consider the set

$A = \left\{ \frac{2k-1}{2^i} : k = 1, \ldots, 2^{i-1} \right\}$. Then $\# A = 2^{i-1}$
and for each $t \in T$, $\min_{c_i \in A} |c_i - t| \leq \frac{1}{2^i}$.

No set smaller than A satisfies this.

- Next, we define the closely related notion of packing number.

Packing Number

- For each $\varepsilon > 0$, define the packing number $D(\varepsilon, d, T)$ to be the largest number D for which there exist points $m_1, \ldots, m_D$ in T such that $d(m_i, m_j) > \varepsilon$ $\forall i \neq j$.

Example (cont) -- Take once again $T = [0,1]$ and $d(x,y) = |x-y|$. For $i = 1, 2, \ldots$, it is easy to see that

$$D(2^{-i}, d, T) = 2^{i-1} + 1 = N(2^{-i}, d, T) + 1$$

to see this, consider the set

$$\mathcal{A}' = \left\{ \frac{2k}{2^i} : k = 0, \ldots, 2^{i-1} \right\}$$

then $\mathcal{A}'$ is the largest set that satisfies the packing number requirement.

- As this simple example suggests, there is a general relationship between packing and covering numbers:

Lemma [Packing and Covering numbers]
For each $\varepsilon > 0$ and each pseudometric space (d, T),

$$D(2\varepsilon, d, T) \leq N(\varepsilon, d, T) \leq D(\varepsilon, d, T)$$

- Packing and covering numbers play a crucial role in providing sufficient conditions that lead ultimately to stochastic equicontinuity. Before stating the result that states the link, we define the property of polynomial classes of sets.

Polynomial Class of Sets

- This is a useful property of a class of sets. This property will be eventually linked with a covering number property of a class of functions.

Definition.— Let $\mathcal{D}$ be a class of subsets of a set $\mathcal{X}$, and let $\mathcal{X}_0$ be a set of n points in $\mathcal{X}$. If there exists a polynomial $p(\cdot)$ not depending on $\mathcal{X}_0$ for which

$$\#\{\mathcal{X}_0 \cap D : D \in \mathcal{D}\} \leq p(n),$$

then $\mathcal{D}$ is called a polynomial class of sets.

<u>VC-Class</u> -- Polynomial classes of sets are also called V-C or Vapnik-Chervonenkis classes. These authors showed that:

An empirical process $Z(t)$ indexed by a class of sets of polynomial class obeys a uniform strong law of large numbers. The precise statement is the following: Let $P_n D = n^{-1} \sum_{i=1}^{n} \{X_i \in D\}$ and $PD = Pr\{X_i \in D\}$. Suppose we have an independent sample from P. The statement of the theorem is as follows:

<u>Theorem (Uniform strong convergence of empirical measures indexed on a polynomial class of sets)</u>

- Let P be a probability measure on a space S. For every (permissible) class $\mathcal{D}$ on S with polynomial discrimination,

$$\sup_{\mathcal{D}} |P_n D - PD| \longrightarrow 0 \text{ almost surely}$$

This is Theorem 14 in Pollard (1984), Chapter II. It is a generalization of the Glivenko-Cantelli Theorem.

- The polynomial property is crucial here because it allows us to "discretize" the search for $\sup_{\mathcal{D}} |P_n D - PD|$ over a finite collection of sets $D_1, D_2, \cdots, D_{k(n)}$. If $\mathcal{D}$ is of polynomial class, then $k(n)$ is a polynomial function of "n", and combined with a Hoeffding-exponential inequality, we get the desired result. See the steps in pages 14-16 of Pollard (1984), Ch. II.

- This "discretization" trick was also present in Huber (1967) as we studied previously.

Intuitively: A class of sets is of polynomial class if it can "pick out" no more than a fixed polynomial number of subsets out of the 2^n possible subsets of any set of n points from an underlying set $\mathcal{X}$.

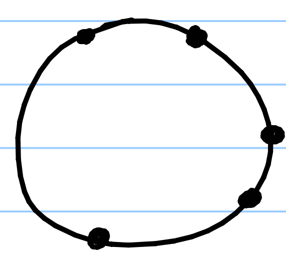
Example: Let $\mathcal{X} = \mathbb{R}$ and take the class of sets $\mathcal{D} = \{ D = [a,b] : a < b \}$ [the class of all closed intervals in $\mathbb{R}$]. Now take $\mathcal{X}_0 = \{ x_i \in \mathbb{R} : i = 1, \dots, n \}$ where $x_i \leq x_{i+1}$ for all i. Then:

$$\#\{ \mathcal{X}_0 \cap D : D \in \mathcal{D} \} \leq \frac{n(n-1)}{2} + 1$$

to see this, suppose $x_i < a \leq x_{i+1}$ for some i. Then there are at most $n-i$ intervals in which to place b so that $[a,b]$ contains a different nonempty subset of $\mathcal{X}_0$. Therefore $\mathcal{D}$ is of polynomial class.

Example: Let $\mathcal{X} = \mathbb{R}^2$, and let $\mathcal{D}$ be the class of quadrants of the form $(-\infty, t_1] \times (-\infty, t_2]$. This is a polynomial class since it can only pick out fewer than $(n+1)^2$ different subsets from any set of n points in $\mathbb{R}^2$: There are at most $(n+1)$ relevant places to place t_1 and $(n+1)$ relevant places to place t_2.

Example: Some classes of sets are too rich to have polynomial property. Suppose $\mathcal{X}$ is a circle in $\mathbb{R}^2$, and let $\mathcal{D}$ be the class of all closed, convex subsets of $\mathcal{X}$. Take any collection of n points in the boundary of $\mathcal{X}$. Then $\mathcal{D}$ can pick out all 2^n subsets of this collection:



$\Big\}$ $\mathcal{D}$ is too rich for $\mathcal{X}$

For such a collection of points $\mathcal{X}_0$, we say that $\mathcal{D}$ "shatters" $\mathcal{X}_0$.

How to verify if a class of sets is of polynomial class?

- Direct verification of the polynomial property is difficult for many interesting classes of sets. Luckily, there are means for easily verifying it in a number of cases. The next result illustrates an important case:

Lemma [Sufficient Condition for Polynomial Property]

- If $\mathcal{G}$ is a finite-dimensional vector space of real-valued functions on a set $\mathcal{X}$, then the class of all sets of the form $\{g \geq r\}$ or $\{g > r\}$ with $g \in \mathcal{G}$ and $r \in \mathbb{R}$ is a polynomial class.

Note: $\{A\}$ denotes the indicator function, which equals one if A is true and zero otherwise.

Example: Let

$$\mathcal{G} = \{ g(x,y;\alpha,\beta,\gamma) = \alpha x + \beta y + \gamma : \alpha,\beta,\gamma \in \mathbb{R} \}$$

where for each $g \in \mathcal{G}$, the pair (x,y) ranges over $\mathbb{R}^2$. Then $\mathcal{G}$ is a 3-dimensional vector space of real-valued functions on $\mathbb{R}^2$. By the previous lemma, the class of all sets of the form

$$\{(x,y) \in \mathbb{R}^2 : g(x,y) \geq r\} \quad \text{or}$$

$$\{(x,y) \in \mathbb{R}^2 : g(x,y) \geq r\}$$

with $g \in \mathcal{G}$ and $r \in \mathbb{R}$ is a polynomial class (note that it is the class of all half-spaces in $\mathbb{R}^2$).

Lemma [Constructing Polynomial Classes of sets]

- If $\mathcal{C}$ and $\mathcal{D}$ are polynomial classes of sets, then so are

(i) $\mathcal{C} \cup \mathcal{D}$
(ii) $\{C \cup D : C \in \mathcal{C} ; D \in \mathcal{D}\}$
(iii) $\{C \cap D : C \in \mathcal{C} ; D \in \mathcal{D}\}$
(iv) $\{C^c : C \in \mathcal{C}\}$

- Next, we go from classes of sets to classes of functions. We focus on "Euclidean" classes of functions.

Euclidean Classes of Functions

- Let $(\mathcal{X}, S, \mu)$ be a probability space, and let $\mathcal{F}$ be a class of real-valued functions defined on $\mathcal{X}$ and μ-measurable w.r.t S.

- We now define an envelope F for the class $\mathcal{F}$.

Definition [envelope]
Let F be a real-valued function defined on $\mathcal{X}$ and μ-measurable w.r.t S. Call F an envelope for $\mathcal{F}$ if $|f| \le F \ \forall f \in \mathcal{F}$.

Definition (pseudometric to be used]
Let F be an envelope for $\mathcal{F}$ satisfying $\mu F^2 < \infty$. For $f, g \in \mathcal{F}$ define the pseudo-metric

$$d_\mu(f,g) = \frac{(\mu|f-g|^2)^{1/2}}{\tau} \quad ; \quad \text{where } \tau^2 = \mu F^2$$

- We now define the $\boxed{\text{Euclidean Property}}$

Definition [Euclidean property]
- Call $\mathcal{F}$ Euclidean for the envelope F if there exist positive constants A, V <u>not depending on μ</u> such that

$$N(\varepsilon, d_\mu, \mathcal{F}) \le A\varepsilon^{-V} \quad \forall \ 0 < \varepsilon \le 1$$

Note: what's important in the definition of Euclidean property is that the constants (A, V) not depend on η (they may depend on the envelope F, but must work for all possible measures η).

Example: Consider all sets of the form $[0,c]$ with $0 \le c \le 1$, and let $I_c = \{[0,c]\}$ (the indicator function for $[0,c]$. Then $I_1 = \{[0,1]\}$ is clearly an envelope for this class $\mathcal{F}$.
Take any prob. measure η in $[0,1]$.
Fix $\varepsilon > 0$ and let R be the smallest integer such that

$$\frac{1}{\varepsilon^2} \le R < \frac{1}{\varepsilon^2} + 1$$

For $i = 0, \dots, R$ define the endpoint

$$c_i = \inf\{c : \eta[0,c] \ge i/R\}$$

and the subclass $\mathcal{C} = \{I_{c_i} : i = 0, \dots, R\}$. Note that $\eta F^2 = 1$ for any well-defined probability measure in $[0,1]$

and therefore $d_\mu(f,g) = (\mu|f-g|^2)^{1/2}$ for any $f,g \in \mathcal{F}$. For any $c_j \le c < c_{j+1}$, we have:

$$\min_i d_\mu(I_c, I_{c_i}) = \min_i (\mu|I_c - I_{c_i}|^2)^{1/2}$$

$$= (\min\{\mu(c_j, c], \mu(c, c_{j+1}]\})^{1/2} \le \left(\frac{1}{R}\right)^{1/2} \le \varepsilon$$

- Note that $\#C \le R+1$, therefore:

$$N(\varepsilon, d_\mu, \mathcal{F}) \le R+1 < \frac{1}{\varepsilon^2} + 2 \le 3\varepsilon^{2} \text{ if } 0<\varepsilon\le 1$$

therefore $\mathcal{F}$ is Euclidean for the envelope F.

How to verify the Euclidean Property?

- The previous example verified the Euclidean property directly for a class of functions. Direct verification is often difficult for more interesting classes. Luckily, there exist results in the literature that facilitate verification of Euclidean property. The most important ones establish a link between polynomial classes of sets and Euclidean classes of functions by exploiting the properties of subgraphs of functions.

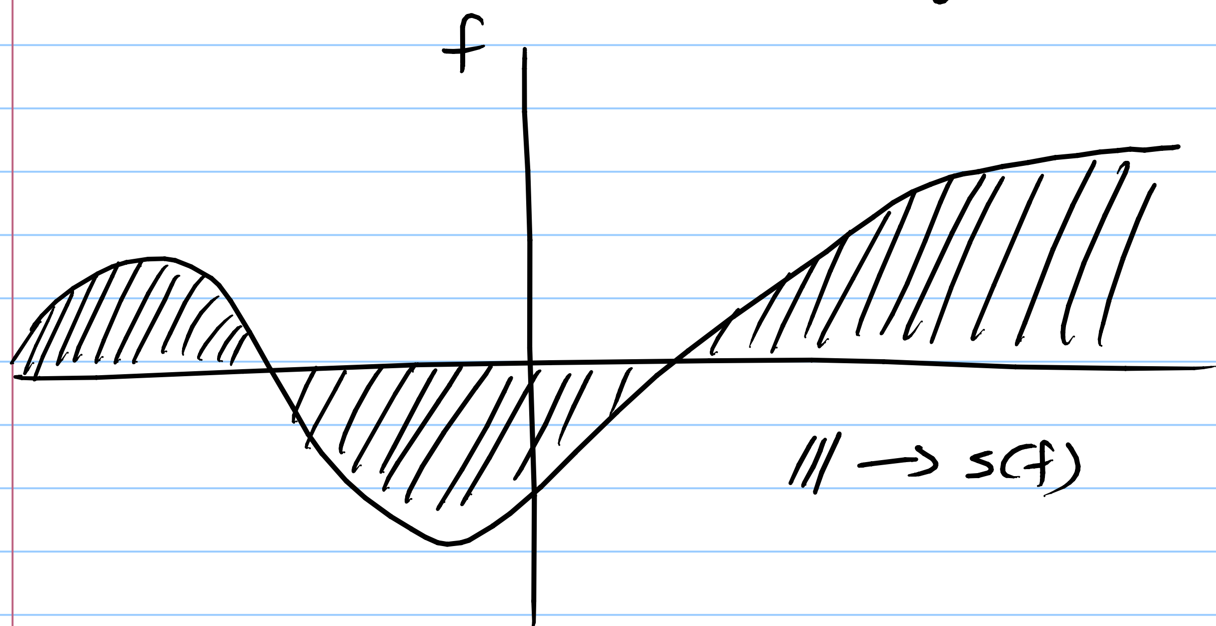
- We define a subgraph next:

Definition (Subgraph)

- For each $f \in \mathcal{F}$ define

$$s(f) = \{(x, t) \in \mathcal{X} \otimes \mathbb{R} : 0 < t < f(x) \text{ or } 0 > t > f(x)\}$$

then $s(f)$ is called the subgraph of f.



Lemma (subgraphs and Euclidean property)

If $\{S(f) : f \in \mathcal{F}\}$ is a polynomial class of sets, then $\mathcal{F}$ is a Euclidean class of functions for every envelope.

Sketch of Proof:

- Take $0 < \varepsilon \leq 1$ and consider the maximal set $\{f \in \mathcal{F} : d_\mu(f_i, f_j) > \varepsilon \ \forall\, i \neq j\}$. Let "$m$" denote the cardinality of this set. Note that $m \geq N(\varepsilon, d_\mu, \mathcal{F})$ because for every $f \in \mathcal{F}$ not in this maximal set, we have $d_\mu(f, f_i) \leq \varepsilon$ for some i in the set.

- The next step is to show that we can find a collection of k points in $\mathcal{X} \otimes \mathbb{R}$ such that the class of sets generated by the subgraphs of $\mathcal{F}$ is capable of picking up "m" different subsets of these k points. We show this by proving that with strictly positive probability, the class of graphs of the maximal set described above picks out m different subsets of a set of K points. Since the subgraphs of $\mathcal{F}$ have polynomial discrimination, this means that

$$m \le B \cdot K^V \text{ for some fixed } V$$

and therefore

$$N(\varepsilon, d_n, \mathcal{F}) \le B \cdot K^V$$

which is the Euclidean property.

If you want to see the details of the proof, look up Lemma 25 in Pollard (1984), p. 27.

Manageable Classes of Functions

- This class includes Euclidean classes but is larger. Basically, we say that a class of functions $\mathcal{F}$ is manageable for the envelope F if there exists a deterministic function $\lambda(\cdot)$ [called the Capacity Bound] such that:

$$\text{(i)} \quad \int_0^1 \sqrt{\log \lambda(x)}\, dx < \infty$$

$$\text{(ii)} \quad D(x, d_n, \mathcal{F}) \le \lambda(x) \quad \forall \; 0 < x \le 1$$

where $D(x, d_\mu, \mathcal{F})$ denotes (as before) the packing number for $x, d_\mu, \mathcal{F}$. Notice that for a Euclidean class: $N(x, d_\mu, \mathcal{F}) \leq A x^{-V}$ for some (A, V). Therefore by the properties of packing numbers: $D(x, d_\mu, \mathcal{F}) \leq A(2x)^{-V} = \tilde{A} x^{-V}$, so choosing $\lambda(x) = \tilde{A} x^{-V}$ easily satisfies the "manageability" requirements.

- As we will see below, and as Theorem 1 in Andrews (1994) states, manageable empirical processes are stochastically equicontinuous.

<u>Entropy</u>: We call the logarithm of a covering number $N(\varepsilon, d_\mu, \mathcal{F})$ the "entropy" or rather ε-entropy. A class of functions $\mathcal{F}$ satisfies <u>Pollard's entropy condition</u> if:

$$\int_0^1 \sup_\mu \sqrt{\log N(\varepsilon, d_\mu, \mathcal{F})}\, d\varepsilon < \infty$$

for some envelope F, and the search for μ is over the set of all measures that concentrate on a finite set.

Sufficient Conditions for Euclidean and Manageable classes of functions:

- Verifying directly if a class of subgraphs is of polynomial discrimination may be tedious and difficult. Much effort has been devoted to finding easy-to-verify sufficient conditions:

Lemma (Pakes and Pollard (1989))

- Let $\mathcal{F} = \{f(\cdot,t) : t \in T\}$ be a class of real-valued functions defined on a set $\mathcal{X}$ and indexed by a bounded subset $T \subset \mathbb{R}^d$. If there exists a constant C such that

$$|f(x,t) - f(x,t')| \leq C|t-t'| \quad \forall x \in \mathcal{X},\ t,t' \in T$$

then $\mathcal{F}$ is euclidean for a constant envelope.

- Theorem 2 in Andrews (1994) is much more general, it deals with three types of classes of functions that satisfy

Pollard's entropy condition. We will return to this after we present the result that links entropy and stochastic equi-continuity.

- In the meantime, we present an additional result for Euclidean classes of functions:

Lemma

- If $\mathcal{F}$ is Euclidean for an envelope F and $\mathcal{G}$ is Euclidean for an envelope G, then

(i) $\{f \pm g : f \in \mathcal{F}, g \in \mathcal{G}\}$ is Euclidean for the envelope $F + G$

(ii) $\{fg : f \in \mathcal{F}, g \in \mathcal{G}\}$ is Euclidean for the envelope FG

Total Variation, Bounded Variation

- Take a metric space (X, d). A function $\gamma : [a,b] \longrightarrow X$ is of bounded variation if $\exists\, M$ such that for each partition $P = \{a = t_0 < t_1 < \cdots < t_n = b\}$ of $[a,b]$

$$v(\gamma, P) = \sum_{k=1}^{n} d(\gamma(t_k), \gamma(t_{k-1})) \leq M$$

The total variation V_γ of γ is defined by

$$V_\gamma = \sup \{ v(\gamma, P) : P \text{ is a partition of } [a,b] \}$$

Examples: All smooth functions are of bounded variation. A function γ is of bdd. variation if and only if it is the difference between two monotonic functions.

Type I Class of functions

- Class $\mathcal{F}$ on $\mathcal{W}$ is type I class if it is of the form
$$\mathcal{F} = \{f : f(w) = h(w'\theta) \ \forall w \in \mathcal{W}, \text{ some } \theta \in \Theta \subset \mathbb{R}^{k} \text{ and } h \in V_K\}$$

where V_K is some set of functions from $\mathbb{R} \to \mathbb{R}$ with total variation less than or equal to $K < \infty$.

* Type I Class -- Useful to deal with non-differentiable M-estimation problems.

Type II Class of functions

- Class $\mathcal{F}$ on $\mathcal{W}$ is type II if every $f \in \mathcal{F}$ is of the form: $f(\cdot) = f(\cdot, \tau)$ for some $\tau \in \mathcal{T}$, with $\mathcal{T}$ a bdd. subset of Euclidean space, and $f(\cdot, \tau)$ is Lipschitz in τ:

$$|f(\cdot,\tau_1) - f(\cdot,\tau_2)| \leq B(\cdot)\|\tau_1 - \tau_2\|$$

$$\forall\ \tau_1, \tau_2 \in \mathcal{T}$$

and some function $B(\cdot): \mathcal{W} \longrightarrow \mathbb{R}$.

<u>Type III</u> Class $\longrightarrow$ We'll discuss it in nonparametric estimation problems.

<u>Theorem 2</u>

If $\mathcal{F}$ is class I or II, then Pollard's entropy condition holds with:

A) Envelope $F(\cdot) = \max\{1, \sup_{f \in \mathcal{F}} |f(\cdot)|\}$

for type I class

B) Envelope $F(\cdot) = \max\{1, \sup_{f \in \mathcal{F}} |f(\cdot)|, B(\cdot)\}$

for type II class

- We're ready for Theorem 1 in Andrews (1994)

- Recall, we have:

$\mathcal{M} = \{m(\cdot, \tau) : \tau \in \mathcal{T}\}$ class of functions

$$v_T(\cdot) = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} [m(W_t, \tau) - E\, m(W_t, \tau)]$$

for $\tau \in \mathcal{T}$

Then:

$v_T(\cdot)$ is Euclidean if $\mathcal{M}$ is a Euclidean class of functions

$v_T(\cdot)$ "satisfies Pollard's entropy condition" if $\mathcal{M}$ does.

<u>Theorem 1</u>

- If: (A) $\mathcal{M}$ satisfies Pollard's entropy condition with some envelope $\overline{M}(\cdot)$

(B) $\lim_{T \to \infty} (1/T) \sum_{t=1}^{T} E\, \overline{M}(W_t)^{2+\delta} < \infty$ for some $\delta > 0$

(C) $\{W_t : t \le T, T \ge 1\}$ is m-dependent

Then $\{v_T(\cdot) : T \ge 1\}$ is stoch. equicontinuous.

quitar esta parte

Euclidean and Manageable Processes

- We go from Euclidean and Manageble classes of Functions to Euclidean and Manageble empirical processes immediately, by extending the concepts to sequences of classes of functions:

The sequence of classes $\{\mathcal{F}_n\}$ is called Euclidean for envelope F (not depending on "n") if there exist positive constants A and V not depending on n such that:

$$\sup_n N(\varepsilon, d_n, \mathcal{F}_n) \leq A\varepsilon^{-V} \quad \forall\ 0 < \varepsilon \leq 1$$

- We extend the definition of managable-class analogously. An empirical process $Z_n(\cdot)$ is Euclidean if $\{Z_n(\cdot)\}$ is a Euclidean sequence of classes for some envelope F. $Z_n(\cdot)$ is Manageable if the sequence of classes $\{Z_n(\cdot)\}$ is manageable for some envelope F.

Packing Numbers and Stochastic Equicontinuity

- The relationship between stochastic equicontinuity of an empirical process $Z(t)$ and the packing numbers of T is illustrated in Lemma 3.4 in Pollard (1990). Before presenting this lemma, we introduce the concept of Orlicz Norm and a preliminary Lemma [Lemma 3.2 in Pollard (1990)]

Orlicz-Norm of a Random Variable

- Suppose Φ is a convex, increasing function in $\mathbb{R}^+$ with $0 \le \Phi(0) \le 1$. The Orlicz norm of a R.V Z, denoted $\|Z\|_\Phi$ is:

$$\|Z\|_\Phi = \inf\{c>0 : E\,\Phi(|Z|/c) \le 1\}$$

For example, if $\Phi(z) = |z|^p$, then $\|Z\|_\Phi = \|Z\|_p$

* We can think of an Orlicz norm as $E\,\Phi(|Z|)$ if this expectation exists.

- The results that follow focus on a particular Orlicz norm: the one that corresponds to

$$\psi(x) = \frac{1}{5} \exp\{x^2\}$$

which we denote by $\|z\|_\psi$. The space of r.v's with finite $\|\cdot\|_\psi$-norm is denoted $\mathcal{L}^\psi$.

Lemma 3.2 in Pollard (1990)

- For any finite collection of m random variables, $z_1, z_2, \ldots, z_m$,

$$\left\| \max_{i \le m} |z_i| \right\|_\psi \le \sqrt{2 + \log m}\, \max_{i \le m} \|z_i\|_\psi$$

Proof: See Pollard (1990), p. 10.

- We will focus on the following result, which uses the previous lemma.

(Hint of)

<u>Lemma 3.4 in Pollard (1990)</u> [Link between packing numbers and stochastic equicontinuity]

- If a stochastic process $Z(\cdot)$ has continuous sample paths and satisfies:

$$\|Z(s) - Z(t)\|_\psi \le d(s,t) \quad \forall\, s,t \in T,$$

and if there exists $t_0 \in T$ such that $\sup_{t \in T} d(t, t_0) = \delta$, then

$$\left\| \sup_T |Z(t) - Z(t_0)| \right\|_\psi$$

$$\le \sum_{i=0}^{\infty} \frac{\delta}{2^i} \sqrt{2 + \log D(\delta/2^{i+1}, d, T)}$$

(where $D(\cdot)$ stands for packing # as before)

Proof: The key to the proof is a so-called "chaining" argument, which is a way of "discretizing" the problem.

- If $\delta = \infty$, then the result is trivially true, so we focus on $\delta < \infty$, which is compatible with a totally bounded pseudometric space (remember the statement of the Functional Central Limit Theorem).

- Define $\delta_i \equiv \delta / 2^i$ for $i = 0, 1, 2, \ldots$. Now construct a sequence of maximal subsets $\{t_0\} \equiv T_0, T_1, T_2, \ldots,$ with the property that:

$$d(s,t) > \delta_i \text{ if } s,t \in T_i \text{ and } s \neq t.$$

By definition of maximality, $s \notin T_i$ implies that $\min_{t \in T_i} d(s,t) \leq \delta_i$. Note that $\# T_i \leq D(\delta_i, d, T)$.

- Next, approximate $\sup_T |Z(t) - Z(t_0)|$ by $\max_{T_m} |Z(t) - Z(t_0)|$ for some "m".

- By the maximality property of the sets T_i, for each $t \in T_m$ there exists a sequence of points leading from t to t_0:

$$t = t_m, t_{m-1}, t_{m-2}, \ldots, t_1, t_0$$

such that:

(i) $t_i \in T_i$ for each element of the sequence
(ii) $d(t_i, t_{i-1}) \leq \delta_i$ " " " " "

- A simple triangle inequality yields:

$$\max_{T_m} |Z(t) - Z(t_0)| \leq \max_{T_m} \sum_{i=1}^{m} |Z(t_i) - Z(t_{i-1})|$$

$$\leq \sum_{i=1}^{m} \max_{T_i} |Z(t_i) - Z(t_{i-1})|$$

- Any well-defined norm (such as an Orlicz norm) must preserve this inequality. Combining this with a further triangle inequality we get:

$$\left\| \max_{T_m} |Z(t) - Z(t_0)| \right\|_\psi \leq \sum_{i=1}^{m} \left\| \max_{T_i} |Z(t_i) - Z(t_{i-1})| \right\|_\psi$$

The previous lemma yields:

$$\left\| \max_{T_i} |Z(t_i) - Z(t_{i-1})| \right\|_\psi \le \sqrt{2 + \log(\# T_i)} \max_{T_i} \left\| |Z(t_i) - Z(t_{i-1})| \right\|_\psi$$

$$\le \sqrt{2 + \log D(\delta_i, t, T)} \max_{T_i} d(t_i, t_{i-1})$$

$$\le \sqrt{2 + \log D(\delta_i, t, T)} \, \delta_i$$

The second inequality uses $\# T_i \le \log D(\delta_i, t, T)$ and the assumption $\| Z(t) - Z(s) \|_\psi \le d(s,t)$. The last inequality uses the construction of $T_1, T_2, \ldots$

Therefore:

$$\left\| \max_{T_m} |Z(t) - Z(t_0)| \right\|_\psi \le \sum_{i=1}^{m} \delta_i \sqrt{2 + \log D(\delta_i, t, T)}$$

— Now, recall that we approximated $\sup_T |Z(t) - Z(t_0)|$ with $\left\| \max_{T_m} |Z(t) - Z(t_0)| \right\|$.

Due to the continuity of the sample paths of $Z(\cdot)$, we can appeal to a monotone convergence theorem and let $m \to \infty$ in order to converge to $\sup_T |Z(t) - Z(t_0)|$

$$\lim_{m\to\infty} \left\| \max_{T_m} |Z(t) - Z(t_0)| \right\|_\psi \le \sum_{i=1}^{\infty} \delta_i \sqrt{2 + \log D(\delta_i, t, T)}$$

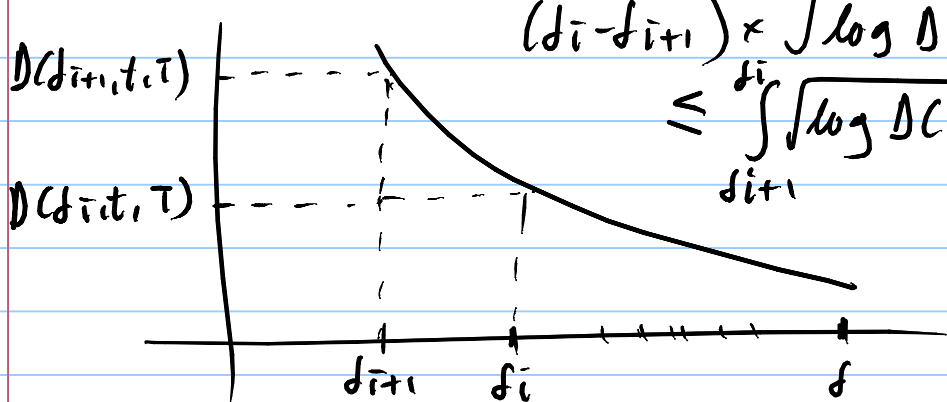
- Note that if $x < \delta$, then $D(x, t, T) \ge 2$. Now, it's easy to show that:

$$\sqrt{2 + \log(1+D)} < 2.2 \cdot \log(D) \quad \text{for } D \ge 2$$

Now, use the fact that $\delta_i = 2(\delta_i - \delta_{i+1})$ to get

$$\sum_{i=1}^{\infty} \delta_i \sqrt{2 + \log D(\delta_i, t, T)} \le 4.4 \times \sum_{i=1}^{\infty} (\delta_i - \delta_{i+1}) \sqrt{\log D(\delta_i, t, T)}$$

- Now, notice that $D(\delta_i, t, T)$ is a decreasing function of δ_i. Therefore:



$$(\delta_i - \delta_{i+1}) \times \sqrt{\log D(\delta_i, t, T)} \le \int_{\delta_{i+1}}^{\delta_i} \sqrt{\log D(x, t, T)}\, dx$$

Therefore,

$$\sum_{i=1}^{\infty} \delta_i \sqrt{2+\log D(\delta_i, t, T)} \le 4.4 \sum_{i=1}^{\infty} \int_{\delta_{i+1}}^{\delta_i} \sqrt{\log D(x, t, T)}\, dx$$

$$= 4.4 \int_0^{\delta/2} \sqrt{\log D(x, t, T)}\, dx$$

$$\le 4.4 \int_0^{\delta} \sqrt{\log D(x, t, T)}\, dx$$

$\square$

Manageable Process