Glivenko-Cantelli and Donsker Results

Yangjianchen Xu

Department of Biostatistics
University of North Carolina at Chapel Hill

07/21/2021

Overview

- Glivenko-Cantelli and Donsker results
  - With bracketing
  - Without bracketing
- Results in previous presentations will be frequently used
  - Maximal inequalities
  - Symmetrization
  - Results on weak convergence

Glivenko-Cantelli and Donsker class

With the notation $\|Q\|_{\mathcal{F}} = \sup\{|Qf| : f \in \mathcal{F}\}$, the uniform version of the law of large numbers and central limit theorem becomes

Definition 1 (GC class)

A class $\mathcal{F}$ is called a *Glivenko-Cantelli class*, or also P-Glivenko-Cantelli class to bring out the dependence on the underlying measure P, if

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

where the convergence is in outer probability or is outer almost surely.

Definition 2 (Donsker class)

A class $\mathcal{F}$ is called a Donsker class, or P-Donsker class, if

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

where the limit $\mathbb{G}$ is a tight Borel measurable element in $\ell^\infty(\mathcal{F})$.

Complexity of $(\mathcal{F}, \|\cdot\|)$

- Whether a given class $\mathcal{F}$ is a Glivenko-Cantelli or Donsker class depends on the size of the class.
  - A finite class of square integrable functions is always Donsker.
  - The class of all square integrable, uniformly bounded functions is almost never Donsker.
- Covering number $N(\epsilon, \mathcal{F}, \|\cdot\|)$
  - minimum number of balls $B(f;\epsilon) := \{g : \|g - f\| < \epsilon\}$ to cover $\mathcal{F}$
  - entropy: $\log N(\epsilon, \mathcal{F}, \|\cdot\|)$
- Bracketing number $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$
  - minimum number of brackets $[l, u]$ with $\|l - u\| < \epsilon$ to cover $\mathcal{F}$
  - entropy with bracketing: $\log N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$
- Simple sufficient conditions for a class to be Glivenko-Cantelli or Donsker can be given in terms of the rate of increase as ϵ tends to zero.

Glivenko-Cantelli results with bracketing

Theorem 3

Let $\mathcal{F}$ be a class of measurable functions such that $N_{[]}(\epsilon, \mathcal{F}, L_1(P)) < \infty$ for every $\epsilon > 0$. Then $\mathcal{F}$ is Glivenko-Cantelli.

Proof. Fix $\epsilon > 0$. Choose finitely many ϵ-brackets $[l_i, u_i]$ whose union contains $\mathcal{F}$ and such that $P(u_i - l_i) < \epsilon$ for every i. Then, for every $f \in \mathcal{F}$, there is a bracket such that

$$(\mathbb{P}_n - P) f \leq (\mathbb{P}_n - P) u_i + P(u_i - f) \leq (\mathbb{P}_n - P) u_i + \epsilon$$

Consequently,

$$\sup_{f \in \mathcal{F}} (\mathbb{P}_n - P) f \leq \max_i (\mathbb{P}_n - P) u_i + \epsilon$$

The right side converges almost surely to ϵ by the strong law of large numbers for real variables. Combination with a similar argument for $\inf_{f \in \mathcal{F}} (\mathbb{P}_n - P) f$ yields that limsup $\|\mathbb{P}_n - P\|_{\mathcal{F}}^{*} \leq \epsilon$ almost surely, for every $\epsilon > 0$.

Glivenko-Cantelli results without bracketing

Theorem 4

Let $\mathcal{F}$ be a P-measurable class of measurable functions with envelope F such that $P^*F < \infty$. Let $\mathcal{F}_M$ be the class of functions $f1\{F \leq M\}$ when f ranges over $\mathcal{F}$. If $\log N(\epsilon, \mathcal{F}_M, L_1(\mathbb{P}_n)) = o_P^*(n)$ for every ϵ and $M > 0$, then $\|\mathbb{P}_n - P\|_{\mathcal{F}}^* \to 0$ both almost surely and in mean. In particular, $\mathcal{F}$ is Glivenko-Cantelli.

Lemma 5 (Symmetrization)

For every nondecreasing, convex $\Phi : \mathbb{R} \mapsto \mathbb{R}$ and class of measurable functions $\mathcal{F}$,

$$\mathrm{E}^*\Phi(\|\mathbb{P}_n - P\|_{\mathcal{F}}) \leq \mathrm{E}^*\Phi(2\|\mathbb{P}_n^o\|_{\mathcal{F}})$$

Proof. By the symmetrization Lemma 5, measurability of the class $\mathcal{F}$, and Fubini's theorem,

$$\begin{aligned} \mathrm{E}^* \|\mathbb{P}_n - P\|_{\mathcal{F}} &\leq 2\mathrm{E}_X\mathrm{E}_\epsilon \left\| \frac{1}{n}\sum_{i=1}^{n} \epsilon_i f(X_i) \right\|_{\mathcal{F}} \\ &\leq 2\mathrm{E}_X\mathrm{E}_\epsilon \left\| \frac{1}{n}\sum_{i=1}^{n} \epsilon_i f(X_i) \right\|_{\mathcal{F}_M} + 2P^* F\{F > M\} \end{aligned}$$

by the triangle inequality, for every $M > 0$. For sufficiently large M, the last term is arbitrarily small. To prove convergence in mean, it suffices to show that the first term converges to zero for fixed M. Fix $X_1, \ldots, X_n$. Let $\mathcal{G}$ be a set such that the $L_1(\mathbb{P}_n)$ ϵ-balls of the elements in $\mathcal{G}$ cover $\mathcal{F}_M$, then

$$\mathrm{E}_\epsilon \left\| \frac{1}{n}\sum_{i=1}^{n} \epsilon_i f(X_i) \right\|_{\mathcal{F}_M} \leq \mathrm{E}_\epsilon \left\| \frac{1}{n}\sum_{i=1}^{n} \epsilon_i f(X_i) \right\|_{\mathcal{G}} + \epsilon$$

Proof of Theorem 4, continued

Lemma 6

Let ψ be a convex, nondecreasing, nonzero function with $\psi(0) = 0$ and $\limsup_{x,y\to\infty} \psi(x)\psi(y)/\psi(cxy) < \infty$ for some constant c. Then, for any random variables $X_1, \ldots, X_m$,

$$\left\| \max_{1\le i\le m} X_i \right\|_\psi \le K\psi^{-1}(m) \max_i \|X_i\|_\psi$$

for a constant K depending only on ψ.

The cardinality of $\mathcal{G}$ can be chosen equal to $N(\epsilon, \mathcal{F}_M, L_1(\mathbb{P}_n))$. Bound the L_1-norm on the right by the Orlicz-norm for $\psi_2(x) = \exp(x^2) - 1$, and use the maximal inequality Lemma 6 to find that the last expression does not exceed a multiple of

$$\sqrt{1 + \log N(\epsilon, \mathcal{F}_M, L_1(\mathbb{P}_n))} \sup_{f\in\mathcal{G}} \left\| \frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i) \right\|_{\psi_2|X} + \epsilon$$

Proof of Theorem 4, continued

$$
\begin{aligned}
\mathrm{E}_\epsilon \left\| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i f\left(X_i\right) \right\|_{\mathcal{G}} &= \left\| \sup_{f \in \mathcal{G}} \left| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i f\left(X_i\right) \right| \right\|_{1|X} \\
&\leq K_1 \left\| \sup_{f \in \mathcal{G}} \left| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i f\left(X_i\right) \right| \right\|_{\psi_2|X} \\
&\leq K_2 \sqrt{\log(1 + N\left(\epsilon, \mathcal{F}_M, L_1\left(\mathbb{P}_n\right)\right))} \sup_{f \in \mathcal{G}} \left\| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i f\left(X_i\right) \right\|_{\psi_2|X} \\
&\leq K_3 \sqrt{1 + \log N\left(\epsilon, \mathcal{F}_M, L_1\left(\mathbb{P}_n\right)\right)} \sup_{f \in \mathcal{G}} \left\| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i f\left(X_i\right) \right\|_{\psi_2|X}
\end{aligned}
$$

Proof of Theorem 4, continued

By Hoeffding's inequality, the Orlicz norm can be bounded by $\sqrt{6/n}\left(\mathbb{P}_n f^2\right)^{1/2} \leq \sqrt{6/n}M$. Thus we have

$$\mathrm{E}_\epsilon \left\| \frac{1}{n} \sum_{i=1}^{n} \epsilon_i f\left(X_i\right) \right\|_{\mathcal{F}_M} \leq K_3 \sqrt{1 + \log N\left(\epsilon, \mathcal{F}_M, L_1\left(\mathbb{P}_n\right)\right)} \sqrt{\frac{6}{n}} M + \epsilon \xrightarrow{\mathrm{P}*} \epsilon$$

Since it is bounded by M, its expectation with respect to $X_1, \ldots, X_n$ converges to zero by the dominated convergence theorem.
This concludes the proof that $\|\mathbb{P}_n - P\|_{\mathcal{F}}^* \to 0$ in mean. It also converges almost surely follows from the fact that the sequence $\|\mathbb{P}_n - P\|_{\mathcal{F}}^*$ by the following lemma.

Proof of Theorem 4, continued

Lemma 7

Let $\mathcal{F}$ be a class of measurable functions with envelope F such that $P^*F < \infty$. Define a filtration by letting Σ_n be the σ-field generated by all measurable functions $h : \mathcal{X}^\infty \mapsto \mathbb{R}$ that are permutationsymmetric in their first n arguments. Then

$$\mathrm{E}\left(\|\mathbb{P}_n - P\|_{\mathcal{F}}^* \mid \Sigma_{n+1}\right) \geq \|\mathbb{P}_{n+1} - P\|_{\mathcal{F}}^*, \qquad a.s.$$

Furthermore, there exist versions of the measurable cover functions $\|\mathbb{P}_n - P\|_{\mathcal{F}}^*$ that are adapted to the filtration. Any such versions form a reverse submartingale and converge almost surely to a constant.

Proof. See Lemma 2.4.5 of VW. From this Lemma , we know that there exists a version of $\|\mathbb{P}_n - P\|_{\mathcal{F}}^*$ that converges almost surely to a constant. Since we already know that $\|\mathbb{P}_n - P\|_{\mathcal{F}}^* \xrightarrow{\mathrm{P}*} 0$, this constant must be zero.

Theorem 8

Let $\mathcal{F}$ be a P measurable class of measurable functions with envelope F and

$$\sup_{Q} N\left(\epsilon\|F\|_{Q,1}, \mathcal{F}, L_1(Q)\right) < \infty$$

for every $\epsilon > 0$, where the supremum is taken over all finite probability measures Q with $\|F\|_{Q,1} > 0$. If $P^*F < \infty$, then $\mathcal{F}$ is Glivenko-Cantelli.

Proof. Assume $P^*F > 0$. Thus there exists $\delta, \eta > 0$ such that, with probability 1, $\delta < \mathbb{P}_n F < \eta$ for all n large enough. Fix $\epsilon > 0$. There is a $K < \infty$ such that $1\left\{\mathbb{P}_n F > 0\right\} \log N\left(\epsilon \mathbb{P}_n F, \mathcal{F}, L_1\left(\mathbb{P}_n\right)\right) \leq K$ almost surely. Hence, with probability 1, $\log N\left(\epsilon\eta, \mathcal{F}, L_1\left(\mathbb{P}_n\right)\right) \leq K$ for all n large enough. We have

$$\log N\left(\epsilon, \mathcal{F}_M, L_1\left(\mathbb{P}_n\right)\right) \leq \log N\left(\epsilon, \mathcal{F}, L_1\left(\mathbb{P}_n\right)\right) = O_P^*(1)$$

for all $\epsilon > 0$ and $M < \infty$.

Donsker results without bracketing

Theorem 9

Let $\mathcal{F}$ be a class of measurable functions that satisfies the uniform entropy bound

$$\int_0^\infty \sup_Q \sqrt{\log N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\epsilon < \infty$$

Let the classes $\mathcal{F}_\delta = \{f - g : f, g \in \mathcal{F}, \|f - g\|_{P,2} < \delta\}$ and $\mathcal{F}_\infty^2$ be P -measurable for every $\delta > 0$. If $P^* F^2 < \infty$, then $\mathcal{F}$ is P-Donsker.

Theorem 10

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.

By Theorem 10, we only need to show $\mathbb{G}_n$ is asymptotically uniformly equicontinuous in probability and $\mathcal{F}$ is totally bounded in $L_2(P)$.

Proof of Theorem 9

Proof. To prove asymptotically uniformly equicontinuity, we only need to prove that for arbitrary sequence $\delta_n \downarrow 0$ and $x > 0$, $\mathrm{P}^*(\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} > x) \to 0$. By Markov's inequality and the symmetrization Lemma 5,

$$\mathrm{P}^*\left(\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} > x\right) \leq \frac{1}{x}\mathrm{E}^*\sqrt{n}\,\|\mathbb{P}_n - P\|_{\mathcal{F}_{\delta_n}} \leq \frac{2}{x}\mathrm{E}^*\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i f(X_i)\right\|_{\mathcal{F}_{\delta_n}}$$

Since the supremum in the right-hand side is measurable by assumption, Fubini's theorem applies and the outer expectation can be calculated as $\mathrm{E}_X\mathrm{E}_\epsilon$. Fix $X_1, \ldots, X_n$. By Hoeffding's inequality, the stochastic process $f \mapsto \left\{n^{-1/2}\sum_{i=1}^{n}\epsilon_i f(X_i)\right\}$ is sub-Gaussian for the $L_2(\mathbb{P}_n)$-seminorm

$$\|f\|_n = \sqrt{\frac{1}{n}\sum_{i=1}^{n} f^2(X_i)}.$$

Proof of Theorem 9, continued

Lemma 11

Let $\{X_t : t \in T\}$ be a separable sub-Gaussian process. Then for every $\delta > 0$ and any t_0,

$$\mathrm{E} \sup_t |X_t| \le \mathrm{E} |X_{t_0}| + K \int_0^\infty \sqrt{\log D(\epsilon, d)} d\epsilon$$

for a universal constant K.

Use the above Lemma 11 to find that

$$\mathrm{E}_\epsilon \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i f(X_i) \right\|_{\mathcal{F}_{\delta_n}} \lesssim \int_0^\infty \sqrt{\log N(\epsilon, \mathcal{F}_{\delta_n}, L_2(\mathbb{P}_n))} d\epsilon$$

For large values of ϵ the set $\mathcal{F}_{\delta_n}$ fits in a single ball of radius ϵ around the origin, in which case the integrand is zero. This is certainly the case for values of ϵ larger than θ_n, where
$\theta_n^2 = \sup_{f \in \mathcal{F}_{\delta_n}} \|f\|_n^2 = \left\| \frac{1}{n} \sum_{i=1}^n f^2(X_i) \right\|_{\mathcal{F}_{\delta_n}}$.

Proof of Theorem 9, continued

Furthermore, we have

$$N\left(\epsilon, \mathcal{F}_{\delta_n}, L_2(Q)\right) \leq N\left(\epsilon, \mathcal{F}_{\infty}, L_2(Q)\right) \leq N^2\left(\epsilon/2, \mathcal{F}, L_2(Q)\right)$$

for every measure Q. The first inequality holds because $\mathcal{F}_{\delta_n} \subset \mathcal{F}_{\infty}$. For the second inequality, fix ϵ and choose $f_1, \ldots, f_N \in \mathcal{F}$ such that their $\epsilon/2$-balls cover $\mathcal{F}$. Then consider the set $\{f_i - f_j : i, j = 1, \ldots, N\}$, which is a subset of $\mathcal{F}_{\infty}$ and contains at most N^2 elements. For any $h \in \mathcal{F}_{\infty}$, it can be written as $h = f - g$ where $f, g \in \mathcal{F}$. By the construction of $f_1, \ldots, f_N$, there exist i_0 and j_0 such that $f \in B(f_{i_0}; \epsilon/2)$ and $g \in B(f_{j_0}; \epsilon/2)$. Thus, we must have $h = f - g \in B(f_{i_0} - f_{j_0}; \epsilon)$.

Proof of Theorem 9, continued

Then we can limit the integral to the interval $(0, \theta_n)$, make a change of variables, and bound the integrand to obtain the bound

$$\int_0^{\theta_n/\|F\|_n} \sup_Q \sqrt{\log N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right)} d\epsilon \|F\|_n$$

Now we take expectation on X and apply Cauchy-Schwarz inequality to the abouve quantity,

$$\left\{\mathrm{E}_X\left(\int_0^{\theta_n/\|F\|_n} \sup_Q \sqrt{\log N(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q))} \mathrm{d}\epsilon\right)^2\right\}^{1/2} \left\{\mathrm{E}_X(\|F\|_n^2)\right\}^{1/2}$$

Since $\|F\|_n = O_p(1)$, we can conclude the above quantity converges to zero if we can show $\theta_n = \left\|\frac{1}{n}\sum_{i=1}^n f^2\left(X_i\right)\right\|_{\mathcal{F}_{\delta_n}}^{1/2} \xrightarrow{\mathrm{P}*} 0$.

Proof of Theorem 9, continued

Since $\sup\left\{Pf^2 : f \in \mathcal{F}_{\delta_n}\right\} \to 0$ and $\mathcal{F}_{\delta_n} \subset \mathcal{F}_\infty$, it is certainly enough to prove that

$$\left\|\mathbb{P}_n f^2 - Pf^2\right\|_{\mathcal{F}_\infty} \stackrel{\mathrm{P}_*}{\to} 0$$

This is a uniform law of large numbers for the class $\mathcal{F}_\infty^2$. This class has integrable envelope $(2F)^2$ and is measurable by assumption. For any pair f, g of functions in $\mathcal{F}_\infty$,

$$\mathbb{P}_n\left|f^2 - g^2\right| \leq \mathbb{P}_n|f-g|4F \leq \|f-g\|_n\|4F\|_n$$

Thus,

$$\log N\left(\epsilon\|2F\|_n^2, \mathcal{F}_\infty^2, L_1\left(\mathbb{P}_n\right)\right) \leq \log N\left(\epsilon\|F\|_n, \mathcal{F}_\infty, L_2\left(\mathbb{P}_n\right)\right) = o_p^*(n)$$

By Theorem 4, $\mathcal{F}_\infty^2$ is Glivenko-Cantelli. This completes the proof of asymptotic equicontinuity.

Proof of Theorem 9, continued

Finally, we show that $\mathcal{F}$ is totally bounded in $L_2(P)$. By the result of the last slide, there exists a sequence of discrete measures P_n with

$$\left\|(P_n - P) f^2\right\|_{\mathcal{F}_\infty} \to 0$$

Take n sufficiently large so that the supremum is bounded by ϵ^2. By assumption, $N(\epsilon, \mathcal{F}, L_2(P_n))$ is finite. For any $f, g \in \mathcal{F}$ with $\|f - g\|_{P_n,2} < \epsilon$,

$$P(f-g)^2 \leq P_n(f-g)^2 + \left|(P_n - P)(f-g)^2\right| \leq 2\epsilon^2$$

Any ϵ-net for $\mathcal{F}$ in $L_2(P_n)$ is a $\sqrt{2}\epsilon$-net in $L_2(P)$. Hence $\mathcal{F}$ is totally bounded in $L_2(P)$ since ϵ was arbitrary.

Donsker results with bracketing

The second main empirical central limit theorem uses bracketing entropy rather than uniform entropy.

Theorem 12

Suppose that $\mathcal{F}$ is a class of measurable functions satisfying

$$\int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P)) \mathrm{d}\epsilon} < \infty$$

Then $\mathcal{F}$ is P-Donsker.

Unlike the uniform entropy condition, this bracketing integral involves only the true underlying measure P. However, part of this gain is offset by the fact that bracketing numbers can be larger than covering numbers. As a result, the two sufficient conditions for a class to be Donsker are not comparable.

Donsker results with bracketing

Define $L_{2,\infty}$-norm as $\|f\|_{P,2,\infty} = \sup_{x>0} \left(x^2 P(|f| > x)\right)^{1/2}$. Note that $\|f\|_{P,2,\infty} \le \|f\|_{P,2}$, so that the bracketing numbers relative to $L_{2,\infty}(P)$ are smaller. We can prove a more general theorem.

Theorem 13

Let $\mathcal{F}$ be a class of measurable functions such that

$$\int_0^\infty \sqrt{\log N_{[]}\left(\epsilon, \mathcal{F}, L_{2,\infty}(P)\right)} d\epsilon + \int_0^\infty \sqrt{\log N\left(\epsilon, \mathcal{F}, L_2(P)\right)} d\epsilon < \infty$$

Moreover, assume that the envelope function F of $\mathcal{F}$ possesses a weak second moment, i.e., $x^2 P^*(F(X) > x) \to 0$ as $x \to \infty$. Then $\mathcal{F}$ is P-Donsker.

Proof of Theorem 13

Proof. To prove Theorem 13, we need to use the following theorem.

Theorem 14

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, for all $\epsilon, \eta > 0$, there exists a finite partition $T = \cup_{i=1}^k T_i$ such that

$$\limsup_\alpha \mathrm{P}^* \left(\sup_i \sup_{s,t \in T_i} |X_\alpha(s) - X_\alpha(t)| > \epsilon \right) < \eta$$

holds for every $\epsilon, \eta > 0$.

The first condition is easy by the CLT theorem and the weak second moment of F. To verify the second condition, it is sufficient to show that for any $\eta > 0$, we can find a finite partition of $\mathcal{F} : \mathcal{F}_1, \ldots, \mathcal{F}_m$ such that $\sup_{f,q \in \mathcal{F}_i} |\mathbb{G}_n(f - g)| = o_p(1)$.

Proof of Theorem 13, continued

The proof of is based on a chaining argument. We first state that, for each natural number q, there exists a partition $\mathcal{F} = \cup_{i=1}^{N_q} \mathcal{F}_{qi}$ of $\mathcal{F}$ into N_q disjoint subsets such that

$$\sum 2^{-q} \sqrt{\log N_q} < \infty$$

$$\|(\sup_{f,g \in \mathcal{F}_{qi}} |f - g|)^* \|_{P,2,\infty} < 2^{-q}$$

$$\sup_{f,g \in \mathcal{F}_{qi}} \|f - g\|_{P,2} < 2^{-q}$$

To see this, first cover $\mathcal{F}$ separately with minimal numbers of $L_2(P)$-balls and $L_{2,\infty}(P)$-brackets of size 2^{-q}, disjointify, and take the intersection of the two partitions. The total number of sets will be $\bar{N}_q = \bar{N}_q^1 \bar{N}_q^2$ if $\bar{N}_q^i$ are the number of sets in the two separate partitions. The logarithm turns the product into a sum, and the first condition is satisfied if it is satisfied for both $\bar{N}_q^i$, which can be satisfied by the discrete version of the assumption.

Proof of Theorem 13, continued

By successive refinements, we can reconstruct the sequence of partitions. We take the partition at stage q to consist of all intersections of the form $\cap_{p=1}^{q} \mathcal{F}_{p,i_p}$. This gives new partitions into $N_q = \bar{N}_1 \cdots \bar{N}_q$ sets. Using the inequality

$$\left(\log \prod \bar{N}_p\right)^{1/2} \leq \sum \left(\log \bar{N}_p\right)^{1/2}$$

we can prove the first of the three displayed conditions is still satisfied.

Proof of Theorem 13, continued

Choose for each q a fixed element f_{qi} from each partitioning set $\mathcal{F}_{qi}$ and set

$$\begin{aligned} \pi_q f &= f_{qi} && \text{if } f \in \mathcal{F}_{qi} \\ \Delta_q f &= \sup_{f,g \in \mathcal{F}_{qi}} |f - g|^*, && \text{if } f \in \mathcal{F}_{qi} \end{aligned}$$

In other words, $\pi_q f$ is the projection of f in the q-th partition and $\Delta_q f$ is the maximal possible difference between f and its projection in q-th partition. In view of Theorem 14, it suffices to show that

$$\|\mathbb{G}_n(f - \pi_{q_0} f)\|_{\mathcal{F}} \xrightarrow{\mathrm{P}*} 0$$

as $n \to \infty$ followed by $q_0 \to \infty$.

Proof of Theorem 13, continued

The chaining method works as follows: for each f, we obtain $\pi_{q_0} f$ then $\pi_{q_0+1} f$ and so on. Therefore, there is a chain corresponding to $f : \pi_{q_0} f \to \pi_{q_0+1} f \to \pi_{q_0+2} f \to \ldots$. However, the chain stops once $\Delta_q f$ is larger than $\sqrt{n} a_q$, where $a_q = 2^{-q}/\sqrt{\ln N_{q+1}}$. Mathematically, we obtain

$$f - \pi_{q_0} f = (f - \pi_{q_0} f)\, B_{q_0} f + \sum_{q_0+1}^{\infty} (f - \pi_q f)\, B_q f + \sum_{q_0+1}^{\infty} (\pi_q f - \pi_{q-1} f)\, A_{q-1} f$$

where

$$\begin{aligned} A_{q-1} f &= 1\left\{\Delta_{q_0} f \le \sqrt{n} a_{q_0}, \ldots, \Delta_{q-1} f \le \sqrt{n} a_{q-1}\right\} \\ B_q f &= 1\left\{\Delta_{q_0} f \le \sqrt{n} a_{q_0}, \ldots, \Delta_{q-1} f \le \sqrt{n} a_{q-1}, \Delta_q f > \sqrt{n} a_q\right\} \\ B_{q_0} f &= 1\left\{\Delta_{q_0} f > \sqrt{n} a_{q_0}\right\} \end{aligned}$$

Noticing that either all $B_q f$ are zero or there is a unique q_1 with $B_{q_1} f = 1$, we can easily prove this equation.

The next thing we need to do is to apply the empirical process $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$ to each of the three terms separately, and take suprema over $f \in \mathcal{F}$. It will be shown that the resulting three variables converge to zero in probability as $n \to \infty$ followed by $q_0 \to \infty$.

Proof of Theorem 13, continued

First, since $|f - \pi_{q_0} f| \, B_{q_0} f \leq 2F1\left\{2F > \sqrt{n} a_{q_0}\right\}$, one has

$$\mathrm{E}^* \left\|\mathbb{G}_n \left(f - \pi_{q_0} f\right) B_{q_0} f\right\|_{\mathcal{F}} \leq 4\sqrt{n} P^* F \left\{2F > \sqrt{n} a_{q_0}\right\}$$

The right side converges to zero as $n \to \infty$, for each fixed q_0, by the assumption that F has a weak second moment.

Proof of Theorem 13, continued

Second, by the inequality $\sup_{t>0} t\mathrm{E}X\{X > t\} \leq 2\|X\|_{2,\infty}^2$,

$$\sqrt{n}a_q P\Delta_q f B_q f \leq \sqrt{n}a_q P\Delta_q f\{\Delta_q f > \sqrt{n}a_q\} \leq 2\left\|\Delta_q f\right\|_{P,2,\infty}^2 \leq 2 \times 2^{-2q}$$

Since $\Delta_q f B_q f \leq \Delta_{q-1} f B_q f \leq \sqrt{n}a_{q-1}$ for $q > q_0$, we obtain

$$P\left(\Delta_q f B_q f\right)^2 \leq \sqrt{n}a_{q-1} P\Delta_q f\left\{\Delta_q f > \sqrt{n}a_q\right\} \leq 2\frac{a_{q-1}}{a_q}2^{-2q}$$

Proof of Theorem 13, continued

By Bernstein's inequality and lemma 8.3,

$$\mathrm{E}\left\|\mathbb{G}_n\right\|_{\mathcal{F}} \lesssim \max_f \frac{\|f\|_\infty}{\sqrt{n}} \log|\mathcal{F}| + \max_f \|f\|_{P,2}\sqrt{\log|\mathcal{F}|} \tag{1}$$

Apply the triangle inequality and inequality (1) to find

$$\begin{aligned}
&\mathrm{E}^*\left\|\sum_{q_0+1}^{\infty} \mathbb{G}_n\left(f-\pi_q f\right) B_q f\right\|_{\mathcal{F}} \\
&\leq \sum_{q_0+1}^{\infty} \mathrm{E}^*\left\|\mathbb{G}_n \Delta_q f B_q f\right\|_{\mathcal{F}} + \sum_{q_0+1}^{\infty} 2\sqrt{n}\left\|P\Delta_q f B_q f\right\|_{\mathcal{F}} \\
&\lesssim \sum_{q_0+1}^{\infty}\left[a_{q-1}\log N_q + \sqrt{\tfrac{a_{q-1}}{a_q}} 2^{-q}\sqrt{\log N_q} + \tfrac{4}{a_q} 2^{-2q}\right]
\end{aligned}$$

Since a_q is decreasing, the quotient a_{q-1}/a_q can be replaced by its square. Then in view of the definition of a_q, the series on the right can be bounded by a multiple of $\sum_{q_0+1}^{\infty} 2^{-q}\sqrt{\log N_q}$. This upper bound is independent of n and converges to zero as $q_0 \to \infty$.

Proof of Theorem 13, continued

Third, there are at most N_q functions $\pi_q f - \pi_{q-1} f$ and at most N_{q-1} functions $A_{q-1} f$. Since the partitions are nested, the function $|\pi_q f - \pi_{q-1} f| A_{q-1} f$ is bounded by $\Delta_{q-1} f A_{q-1} f \leq \sqrt{n} a_{q-1}$. The $L_2(P)$-norm of $|\pi_q f - \pi_{q-1} f|$ is bounded by 2^{-q+1}. Apply inequality (1) to find

$$\mathrm{E}^* \left\| \sum_{q_0+1}^{\infty} \mathbb{G}_n \left(\pi_q f - \pi_{q-1} f \right) A_{q-1} f \right\|_{\mathcal{F}} \lesssim \sum_{q_0+1}^{\infty} \left[a_{q-1} \log N_q + 2^{-q} \sqrt{\log N_q} \right]$$