Reading group mathematical foundations of statistics

Chapter 2.9: Multiplier Central Limit Theorems

Stefan Franssen, Msc

May 11, 2020

The ultimate recall

If we have data $X_1, \cdots, X_n$, we can form the empirical process

$$\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}.$$

The ultimate recall

If we have data $X_1, \cdots, X_n$, we can form the empirical process

$$\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}.$$

This will become random noise around the true value:

$$\sqrt{n}\,(\mathbb{P}_n - P) \rightsquigarrow G_P.$$

The ultimate recall

If we have data $X_1, \cdots, X_n$, we can form the empirical process

$$\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}.$$

This will become random noise around the true value:

$$\sqrt{n} (\mathbb{P}_n - P) \rightsquigarrow G_P.$$

When this convergence is uniform on a class of functions $\mathcal{F}$ we call this class Donsker.

Reminder of the last chapters

Since it has been a while since we have been together, let us remind ourselves what we did in the previous chapters.

Maximal inequalities

We introduced the Orlitz norm

$$||X||_\phi = \inf\{C > 0 : E[\phi(\frac{|X|}{C})] \le 1\}.$$

Maximal inequalities

We introduced the Orlitz norm

$$||X||_\phi = \inf\{C > 0 : E[\phi(\frac{|X|}{C})] \le 1\}.$$

We used the Orlitz norm to control the maxima of random variables using the Orlitz norm.

Covering numbers

We introduced Covering numbers and packing numbers, the number of ϵ balls needed to cover a space and the maximal number of ϵ separated points respectively.

Covering numbers

We introduced Covering numbers and packing numbers, the number of ϵ balls needed to cover a space and the maximal number of ϵ separated points respectively. These play a very important role in the coming chapters, but in the first application we used them for was controlling maxima of stochastic processes.

Symmetrization

For a given distribution P we want to understand the empirical process and how close it gets to the true distribution P, so we want to study

$$\frac{1}{n}\sum_{i=1}^{n}(f(X_i) - Pf),$$

Symmetrization

For a given distribution P we want to understand the empirical process and how close it gets to the true distribution P, so we want to study

$$\frac{1}{n}\sum_{i=1}^{n}(f(X_i) - Pf),$$

This might be a bit complicated so we instead study

$$\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i)$$

with ϵ_i iid Rademacher RV.

Symmetrization continued

It turns out the symmetrized process can be used to control the behaviour of deviations of the empirical process. As main result it allows us to give equivalent statements for classes of functions being P-Donsker.

In the next chapter we provide theorems to show when the empirical process almost surely converges to the true distribution.

Glivenko-Cantelli Theorems

In the next chapter we provide theorems to show when the empirical process almost surely converges to the true distribution. In particular, we use that when the entropy numbers with the $L_1(P)$-distance are finite for all $\epsilon > 0$ these classes are Glivenko-Cantelli.

Donsker Theorems

In this chapter we use the covering and bracketing numbers to derive theorems which show that a class of functions is Donsker.

Uniform entropy numbers

In this chapter we study Vapnik Cervonenkis classes of functions. These are Donsker for any probability distribution as long as there exists a integrable envelope function.

Bracketing Numbers

In this chapter we bound the bracketing numbers of common classes of functions coming from approximation theory.

Uniformity in the Underlying Distribution

We extend the results of the previous chapter on Glivenko-Cantelli and Donkser Theorems so that the convergence is true uniformly for all true distributions.

End of recap

Rewriting

Let $Z_i = \delta_{X_i} - P$. Then the Central limit theorem asserts

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i \rightsquigarrow G_p.$$

Random weights

Let ξ_i be mean zero variance 1 i.i.d. random variables.

Random weights

Let ξ_i be mean zero variance 1 i.i.d. random variables. We can now wonder what the behaviour is of

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i Z_i.$$

The multiplier central limit theorem

Under some moment conditions, pointwise

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i Z_i \rightsquigarrow \mathbb{G}.$$

Equivalance

If ξ_i are i.i.d. mean zero variance one and some finite higher absolute moment, then

Equivalance

If ξ_i are i.i.d. mean zero variance one and some finite higher absolute moment, then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i Z_i \rightsquigarrow \mathbb{G}$$

Equivalance

If ξ_i are i.i.d. mean zero variance one and some finite higher absolute moment, then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i Z_i \rightsquigarrow \mathbb{G}$$

if and only if $\mathcal{F}$ is Donsker.

The conditional multiplier central limit theorem

Under the requirements before, conditional on $Z_1, Z_2, \cdots$, for almost every sequence $Z_1, Z_2, \cdots$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i Z_i \rightsquigarrow \mathbb{G}.$$

The $2,1$ norm

It turns out that one measure of distance that is relevant is the $2,1$ norm.

Definition

$$\|\xi\|_{2,1} = \int_0^\infty \sqrt{\mathbb{P}(|\xi| > x)}\,\mathrm{d}\,x.$$

The $2,1$ norm

It turns out that one measure of distance that is relevant is the $2,1$ norm.

Definition

$$\|\xi\|_{2,1} = \int_0^\infty \sqrt{\mathbb{P}(|\xi| > x)}\,\mathrm{d}\,x.$$

Note that this is not a norm, but the topology can be generated by a norm. Finiteness is implied by existence of a $2+\epsilon$ absolute moment for some $\epsilon > 0$.

Quick break

Multiplier inequalities

Lemma

let $Z_1, \cdots, Z_n$ *be i.i.d. stochastic processes*

Multiplier inequalities

Lemma

let $Z_1, \cdots, Z_n$ *be i.i.d. stochastic processes with* $\mathbb{E}^*\|Z_i\|_{\mathcal{F}} < \infty$

Multiplier inequalities

Lemma

let $Z_1, \cdots, Z_n$ be i.i.d. stochastic processes with $\mathbb{E}^* \|Z_i\|_{\mathcal{F}} < \infty$ independent of Rademacher variables $\epsilon_1, \cdots, \epsilon_n$.

Multiplier inequalities

Lemma

let $Z_1, \cdots, Z_n$ *be i.i.d. stochastic processes with* $\mathbb{E}^*\|Z_i\|_{\mathcal{F}} < \infty$ *independent of Rademacher variables* $\epsilon_1, \cdots, \epsilon_n$*. Then for every i.i.d. sample* $\xi_1, \cdots, \xi_n$ *of mean-zero random variables independent of* $Z_1, \cdots, Z_n$ *and any* $1 \leq n_0 \leq n$*,*

Multiplier inequalities

Lemma

let $Z_1, \cdots, Z_n$ be i.i.d. stochastic processes with $\mathbb{E}^* \|Z_i\|_{\mathcal{F}} < \infty$ independent of Rademacher variables $\epsilon_1, \cdots, \epsilon_n$. Then for every i.i.d. sample $\xi_1, \cdots, \xi_n$ of mean-zero random variables independent of $Z_1, \cdots, Z_n$ and any $1 \leq n_0 \leq n$,

$$\frac{1}{2} \|\xi\|_1 \mathbb{E}^* \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i Z_i \|_{\mathcal{F}} \leq \mathbb{E}^* \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i Z_i \|_{\mathcal{F}}$$

Multiplier inequalities

Lemma

let $Z_1, \cdots, Z_n$ be i.i.d. stochastic processes with $\mathbb{E}^*\|Z_i\|_{\mathcal{F}} < \infty$ independent of Rademacher variables $\epsilon_1, \cdots, \epsilon_n$. Then for every i.i.d. sample $\xi_1, \cdots, \xi_n$ of mean-zero random variables independent of $Z_1, \cdots, Z_n$ and any $1 \le n_0 \le n$,

$$\begin{aligned}
\frac{1}{2}\|\xi\|_1 \mathbb{E}^*\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i Z_i\|_{\mathcal{F}} &\le \mathbb{E}^*\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i Z_i\|_{\mathcal{F}} \\
&\le 2(n_0-1)\mathbb{E}^*\|Z_1\|_{\mathcal{F}}\mathbb{E}\max_{1\le i\le n}\frac{|\xi_i|}{\sqrt{n}} \\
&\quad + 2\sqrt{2}\|\xi\|_{2,1}\max_{n_0\le k\le n}\mathbb{E}^*\|\frac{1}{\sqrt{k}}\sum_{i=n_0}^{k}\epsilon_i Z_i\|_{\mathcal{F}}
\end{aligned}$$

Remark

For symmetric random variables ξ *we can replace each the constants* $\frac{1}{2}$, 2 *and* $2\sqrt{2}$ *by* 1.

Step 1: First inequality symmetric case

For the first inequality, first suppose the distribution of the ξ random variables is symmetric around 0.

Step 1: First inequality symmetric case

For the first inequality, first suppose the distribution of the ξ random variables is symmetric around 0. Let $\epsilon_1, \cdots, \epsilon_n$ be Rademacher random variables independent of $Z_1, \cdots, Z_n$ and $\xi_1, \cdots, \xi_n$.

Step 1: First inequality symmetric case

For the first inequality, first suppose the distribution of the ξ random variables is symmetric around 0. Let $\epsilon_1, \cdots, \epsilon_n$ be Rademacher random variables independent of $Z_1, \cdots, Z_n$ and $\xi_1, \cdots, \xi_n$. Note that $\epsilon_i |\xi_i|$ has the same distribution as ξ_i.

Step 1: First inequality symmetric case

For the first inequality, first suppose the distribution of the ξ random variables is symmetric around 0. Let $\epsilon_1, \cdots, \epsilon_n$ be Rademacher random variables independent of $Z_1, \cdots, Z_n$ and $\xi_1, \cdots, \xi_n$. Note that $\epsilon_i|\xi_i|$ has the same distribution as ξ_i. Now, due to Jensen's inequality

$$\mathbb{E}^*\|\sum_{i=1}^n \epsilon_i \mathbb{E}_\xi|\xi_i|Z_i\|_{\mathcal{F}} \leq \mathbb{E}^*\|\sum_{i=1}^n \epsilon_i|\xi_i|Z_i\|_{\mathcal{F}}$$

Step 2: First inequality asymmetric case

In the general case, let $\eta_1, \cdots, \eta_n$ be an independent copy of $\xi_1, \cdots, \xi_n$.

Step 2: First inequality asymmetric case

In the general case, let $\eta_1, \cdots, \eta_n$ be an independent copy of $\xi_1, \cdots, \xi_n$. Then

$$\|\xi\|_1 = \mathbb{E}|\xi_i - \mathbb{E}\eta_i| \leq \|\xi_i - \eta_i\|_1.$$

Step 2: First inequality asymmetric case

In the general case, let $\eta_1, \cdots, \eta_n$ be an independent copy of $\xi_1, \cdots, \xi_n$. Then

$$\|\xi\|_1 = \mathbb{E}|\xi_i - \mathbb{E}\eta_i| \leq \|\xi_i - \eta_i\|_1.$$

Now by the triangle inequality it follows that

$$\mathbb{E}^* \| \sum_{i=1}^{n} (\xi_i - \eta_i) Z_i \|_{\mathcal{F}} \leq 2\mathbb{E}^* \| \sum_{i=1}^{n} \xi_i Z_i \|_{\mathcal{F}}$$

Step 3: Second inequality symmetric case: Notation

Suppose again that ξ_i are symmetric distributed around zero.

Step 3: Second inequality symmetric case: Notation

Suppose again that ξ_i are symmetric distributed around zero. Denote $\tilde{\xi}_i$ the reversed order statistics of $|\xi_1|, \cdots, |\xi_n|$.

Step 3: Second inequality symmetric case: Sketch part 1

We first rewrite and use Fubini to see that

$$\mathbb{E}^*\|\sum_{i=1}^{n}\xi_i Z_i\|_{\mathcal{F}} = \mathbb{E}_{\xi,\epsilon}\mathbb{E}_Z^*\|\sum_{i=1}^{n}\epsilon_i\tilde{\xi}_i Z_i\|_{\mathcal{F}}$$

Step 3: Second inequality symmetric case: Sketch part 1

We first rewrite and use Fubini to see that

$$\mathbb{E}^*\|\sum_{i=1}^{n}\xi_i Z_i\|_{\mathcal{F}} = \mathbb{E}_{\xi,\epsilon}\mathbb{E}_Z^*\|\sum_{i=1}^{n}\epsilon_i\tilde{\xi}_i Z_i\|_{\mathcal{F}}$$

Now apply the triangle inequality to the first $n_0 - 1$ terms

Step 3: Second inequality symmetric case: Sketch part 1

We first rewrite and use Fubini to see that

$$\mathbb{E}^*\|\sum_{i=1}^{n}\xi_i Z_i\|_{\mathcal{F}} = \mathbb{E}_{\xi,\epsilon}\mathbb{E}_Z^*\|\sum_{i=1}^{n}\epsilon_i\tilde{\xi}_i Z_i\|_{\mathcal{F}}$$

Now apply the triangle inequality to the first $n_0 - 1$ terms

$$\leq (n_0 - 1)\mathbb{E}\tilde{\xi}_1\mathbb{E}^*\|Z_1\|_{\mathcal{F}} + \mathbb{E}^*\|\sum_{i=n_0}^{n}\epsilon_i\tilde{\xi}_i Z_i\|_{\mathcal{F}}$$

Step 3: Second inequality symmetric case: Sketch part 2

We continue by rewriting $\tilde{\xi}_i = \sum_{k=i}^{n} \tilde{\xi}_k - \tilde{\xi}_{k+1}$ with $\tilde{\xi}_{n+1} = 0$.

Step 3: Second inequality symmetric case: Sketch part 2

We continue by rewriting $\tilde{\xi}_i = \sum_{k=i}^{n} \tilde{\xi}_k - \tilde{\xi}_{k+1}$ with $\tilde{\xi}_{n+1} = 0$. Plugging this in yields

$$\begin{aligned}
\mathbb{E}^* \| \sum_{i=n_0}^{n} \epsilon_i \tilde{\xi}_i Z_i \|_{\mathcal{F}} &= \mathbb{E}^* \| \sum_{k=n_0}^{n} (\tilde{\xi}_k - \tilde{\xi}_{k+1}) \sum_{i=n_0}^{k} \epsilon_i Z_i \|_{\mathcal{F}} \\
&= \mathbb{E}^* \| \sum_{k=n_0}^{n} \sqrt{k} (\tilde{\xi}_k - \tilde{\xi}_{k+1}) \frac{1}{\sqrt{k}} \sum_{i=n_0}^{k} \epsilon_i Z_i \|_{\mathcal{F}}
\end{aligned}$$

Step 3: Second inequality symmetric case: Sketch part 2

We continue by rewriting $\tilde{\xi}_i = \sum_{k=i}^{n} \tilde{\xi}_k - \tilde{\xi}_{k+1}$ with $\tilde{\xi}_{n+1} = 0$. Plugging this in yields

$$\begin{aligned}\mathbb{E}^*\|\sum_{i=n_0}^{n} \epsilon_i \tilde{\xi}_i Z_i\|_{\mathcal{F}} &= \mathbb{E}^*\|\sum_{k=n_0}^{n} (\tilde{\xi}_k - \tilde{\xi}_{k+1}) \sum_{i=n_0}^{k} \epsilon_i Z_i\|_{\mathcal{F}} \\ &= \mathbb{E}^*\|\sum_{k=n_0}^{n} \sqrt{k}(\tilde{\xi}_k - \tilde{\xi}_{k+1}) \frac{1}{\sqrt{k}} \sum_{i=n_0}^{k} \epsilon_i Z_i\|_{\mathcal{F}}\end{aligned}$$

Fubini + Independence + Bounding + positivity yields

Step 3: Second inequality symmetric case: Sketch part 2

We continue by rewriting $\tilde{\xi}_i = \sum_{k=i}^{n} \tilde{\xi}_k - \tilde{\xi}_{k+1}$ with $\tilde{\xi}_{n+1} = 0$. Plugging this in yields

$$\mathbb{E}^* \| \sum_{i=n_0}^{n} \epsilon_i \tilde{\xi}_i Z_i \|_{\mathcal{F}} = \mathbb{E}^* \| \sum_{k=n_0}^{n} (\tilde{\xi}_k - \tilde{\xi}_{k+1}) \sum_{i=n_0}^{k} \epsilon_i Z_i \|_{\mathcal{F}}$$

$$= \mathbb{E}^* \| \sum_{k=n_0}^{n} \sqrt{k} (\tilde{\xi}_k - \tilde{\xi}_{k+1}) \frac{1}{\sqrt{k}} \sum_{i=n_0}^{k} \epsilon_i Z_i \|_{\mathcal{F}}$$

Fubini + Independence + Bounding + positivity yields

$$\leq \mathbb{E} \sum_{k=n_0}^{n} \sqrt{k} (\tilde{\xi}_k - \tilde{\xi}_{k+1}) \max_{n_0 \leq l \leq n} \mathbb{E}^* \| \frac{1}{\sqrt{l}} \sum_{i=n_0}^{l} \epsilon_i Z_i \|_{\mathcal{F}}$$

Step 3: Second inequality symmetric case: Sketch part 3

For $\tilde{\xi}_{k+1} < t < \tilde{\xi}_k$ we see that k is the number of ξ_i such that $|\xi_i| \geq t$. This allows us to do the following bound

$$\mathbb{E} \sum_{k=n_0}^{n} \int_{\tilde{\xi}_{k+1}}^{\tilde{\xi}_k} \leq \int_0^\infty \mathbb{E}\sqrt{\#(i : \xi_i| \geq t)}\, \mathrm{d}\, t \leq \int_0^\infty \sqrt{nP(|\xi_i| \geq t}\, \mathrm{d}\, t$$

The last step is Jensen's inequality.

Step 4: Second inequality general case

Instead of looking at ξ_i we look at $\xi_i - \nu_i$, where ν_i is an independent copy of the ξ_i random variables.

Step 4: Second inequality general case

Instead of looking at ξ_i we look at $\xi_i - \nu_i$, where ν_i is an independent copy of the ξ_i random variables. Note that

$$\mathbb{E}^* \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i Z_i \|_{\mathcal{F}} \leq \mathbb{E}^* \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\xi_i - \nu_i) Z_i \|_{\mathcal{F}}$$

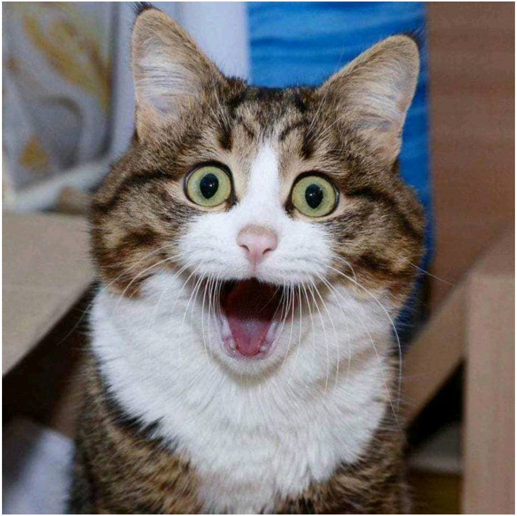
Step 4: Second inequality general case

Instead of looking at ξ_i we look at $\xi_i - \nu_i$, where ν_i is an independent copy of the ξ_i random variables. Note that

$$\mathbb{E}^* \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i Z_i \|_{\mathcal{F}} \leq \mathbb{E}^* \| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\xi_i - \nu_i) Z_i \|_{\mathcal{F}}$$

and use that $\|\xi - \nu\|_{2,1} \leq 2\sqrt{2}\|\xi\|$ (exercise 2.9.2). Now apply the result for the symmetric case.

Taking a breather



Theorem

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

Theorem

Let $\mathcal{F}$ be a class of measurable functions. Let $\xi_1, \cdots, \xi_n$ be i.i.d. random variables with mean zero, variance 1 and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$.

Theorem

Let $\mathcal{F}$ be a class of measurable functions. Let $\xi_1, \cdots, \xi_n$ be i.i.d. random variables with mean zero, variance 1 and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$. Then the sequence $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i(\delta_{X_i} - P)$ converges to a tight limit process in $\ell^\infty(\mathcal{F})$ if and only if $\mathcal{F}$ is Donsker.

Theorem

Let $\mathcal{F}$ be a class of measurable functions. Let $\xi_1, \cdots, \xi_n$ be i.i.d. random variables with mean zero, variance 1 and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$. Then the sequence $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i(\delta_{X_i} - P)$ converges to a tight limit process in $\ell^\infty(\mathcal{F})$ if and only if $\mathcal{F}$ is Donsker. In that case, the limit process is a P-Brownian bridge.

Proof: Preparation

Marginal convergence for both process is equivalent to $\mathcal{F} \subset \mathcal{L}_2(P)$, hence they are equivalent to each other.

Proof: Preparation

Marginal convergence for both process is equivalent to $\mathcal{F} \subset \mathcal{L}_2(P)$, hence they are equivalent to each other. So what is left to show is the asymptotic equicontinuity conditions are equivalent.

- If $\mathcal{F}$ is Donsker, then by lemma 2.3.9 $P^*(F > x) = o(x^{-2})$ in both cases.

Proof: Preparation

Marginal convergence for both process is equivalent to $\mathcal{F} \subset \mathcal{L}_2(P)$, hence they are equivalent to each other. So what is left to show is the asymptotic equicontinuity conditions are equivalent.

- If $\mathcal{F}$ is Donsker, then by lemma 2.3.9 $P^*(F > x) = o(x^{-2})$ in both cases.
- Since $\|\xi\|_{2,1}$ implies existence of a second moment, we have that $\mathbb{E}^* \max_{1 \leq i \leq n} |\xi_i| / \sqrt{n} \to 0$.

Proof: Translating.

Using Lemma 2.9.1 combined with previous results we get

$$\frac{1}{2}\limsup_{n\to\infty}\mathbb{E}^*\|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i Z_i\|_{\mathcal{F}_\delta} \leq \limsup_{n\to\infty}\mathbb{E}^*\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i Z_i\|_{\mathcal{F}_\delta}$$

$$\leq 2\sqrt{2}\|\xi\|_{2,1}\sup_{n_0\leq k}\mathbb{E}^*\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\epsilon_i Z_i\|_{\mathcal{F}_\delta}$$

for every n_0 and δ.

Proof: Translating.

Using Lemma 2.9.1 combined with previous results we get

$$
\begin{aligned}
\frac{1}{2}\limsup_{n\to\infty}\mathbb{E}^*\|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i Z_i\|_{\mathcal{F}_\delta} &\leq \limsup_{n\to\infty}\mathbb{E}^*\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i Z_i\|_{\mathcal{F}_\delta} \\
&\leq 2\sqrt{2}\|\xi\|_{2,1}\sup_{n_0\leq k}\mathbb{E}^*\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\epsilon_i Z_i\|_{\mathcal{F}_\delta}
\end{aligned}
$$

for every n_0 and δ. If we now apply Lemma 2.3.6 we can get rid of the Rademacher for the price of extra factors $\frac{1}{2}$ and 2.

Proof: Translating.

Using Lemma 2.9.1 combined with previous results we get

$$\frac{1}{2}\limsup_{n\to\infty}\mathbb{E}^*\|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i Z_i\|_{\mathcal{F}_\delta} \leq \limsup_{n\to\infty}\mathbb{E}^*\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i Z_i\|_{\mathcal{F}_\delta}$$

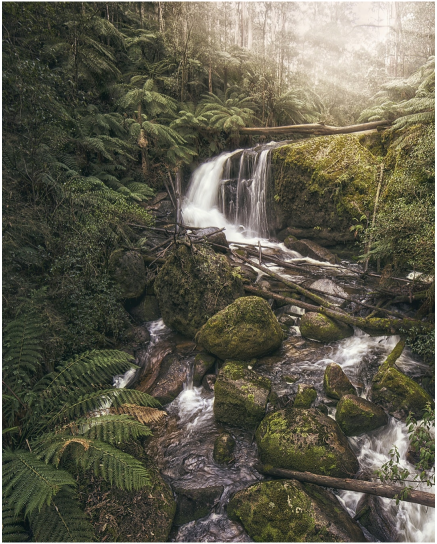
$$\leq 2\sqrt{2}\|\xi\|_{2,1}\sup_{n_0\leq k}\mathbb{E}^*\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k}\epsilon_i Z_i\|_{\mathcal{F}_\delta}$$

for every n_0 and δ. If we now apply Lemma 2.3.6 we can get rid of the Rademacher for the price of extra factors $\frac{1}{2}$ and 2. Now consider what happens when $n_0 \to \infty$.

Proof: Finalisation

From the previous slide we conclude that $\mathbb{E}^*\|\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i\|_{\mathcal{F}_{\delta_n}} \to 0$ if and only if $\mathbb{E}^*\|\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i Z_i\|_{\mathcal{F}_{\delta_n}}$. By lemma 2.3.11 these mean versions of the asymptotic equicontinuity conditions are equivalent to the probability versions, which shows tightness.

More Eyebleach



Joint convergence

Corollary

Under the conditions of the previous theorem

$$\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\delta_{X_i} - P), \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i(\delta_{X_i} - P)\right)$$

Converges in $\ell^\infty(\mathcal{F}) \times \ell^\infty(\mathcal{F})$ *to a vector* $(\mathbb{G}, \mathbb{G}')$ *of independent tight brownian bridges* $\mathbb{G}$ *and* $\mathbb{G}'$.

Corollary

Let $\mathcal{F}$ be Donsker with $\|P\|_{\mathcal{F}} < \infty$. Let $\xi_1, \cdots, X_n$ be i.i.d. random variables with mean μ, variance σ^2 and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$.

Corollary

Let $\mathcal{F}$ be Donsker with $\|P\|_{\mathcal{F}} < \infty$. Let $\xi_1, \cdots, X_n$ be i.i.d. random variables with mean μ, variance σ^2 and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$. Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\xi_i \delta_{X_i} - \mu P) \rightsquigarrow \mu \mathbb{G} + \sigma \mathbb{G}' + \sigma Z P$$

Corollary

Let $\mathcal{F}$ be Donsker with $\|P\|_{\mathcal{F}} < \infty$. Let $\xi_1, \cdots, X_n$ be i.i.d. random variables with mean μ, variance σ^2 and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$. Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\xi_i \delta_{X_i} - \mu P) \rightsquigarrow \mu \mathbb{G} + \sigma \mathbb{G}' + \sigma Z P$$

where $\mathbb{G}$ and $\mathbb{G}'$ are independent Brownian bridges and Z is a standard normal random variable.

Corollary

Let $\mathcal{F}$ be Donsker with $\|P\|_{\mathcal{F}} < \infty$. Let $\xi_1, \cdots, X_n$ be i.i.d. random variables with mean μ, variance σ^2 and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$. Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\xi_i \delta_{X_i} - \mu P) \rightsquigarrow \mu \mathbb{G} + \sigma \mathbb{G}' + \sigma Z P$$

where $\mathbb{G}$ and $\mathbb{G}'$ are independent Brownian bridges and Z is a standard normal random variable. The limit process $\mu \mathbb{G} + \sigma \mathbb{G}' + \sigma Z P$ is a Gaussian Process with zero mean and covariance function $(\sigma^2 + \mu^2) Pgf - \mu^2 PfPg$.

Questions so far?

Conditional multiplier process

Now consider the process

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i Z_i$$

Given $Z_1, Z_2, \cdots$.

Lemma

Let $Z_1, Z_2, \cdots$ be i.i.d. euclidean random vectors with $\mathbb{E}Z_i = 0$ and $\mathbb{E}\|Z_i\|_2^2 < \infty$ independent of the i.i.d. sequence $\xi_1, \xi_2, \cdots$ with $\mathbb{E}\xi_i = 0$ and $\mathbb{E}\xi_i^2 = 1$. Then, conditionally on $Z_1, Z_2, \cdots$,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i Z_i \rightsquigarrow N(0, covZ_1),$$

For almost every sequence $Z_1, Z_2, \cdots$.

Proof

By the strong law of large numbers, $\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i^T \to \mathrm{cov} Z_1$ almost surely.

Proof

By the strong law of large numbers, $\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i^T \to \mathrm{cov} Z_1$ almost surely. Since $\mathbb{E}\|Z_i\|_2^2 < \infty$, we know that $\max_{1\le i\le n} \|Z_i\|_2/\sqrt{n} \to 0$ almost surely.

Proof

By the strong law of large numbers, $\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i^T \to \mathrm{cov} Z_1$ almost surely. Since $\mathbb{E}\|Z_i\|_2^2 < \infty$, we know that $\max_{1\le i\le n} \|Z_i\|_2/\sqrt{n} \to 0$ almost surely. Combined with the Strong law of large numbers we get $\frac{1}{n}\sum_{i=1}^{n} \|Z_i\|_2^2 \mathbb{E}_\xi \xi_i^2 \mathbb{1}_{|\xi_i|\|Z_i\|>\epsilon\sqrt{n}} \to 0$ almost surely.

Proof

By the strong law of large numbers, $\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i^T \to \mathrm{cov} Z_1$ almost surely. Since $\mathbb{E}\|Z_i\|_2^2 < \infty$, we know that $\max_{1\leq i\leq n}\|Z_i\|_2/\sqrt{n} \to 0$ almost surely. Combined with the Strong law of large numbers we get $\frac{1}{n}\sum_{i=1}^{n}\|Z_i\|_2^2 \mathbb{E}_\xi \xi_i^2 \mathbb{1}_{|\xi_i|\|Z_i\|>\epsilon\sqrt{n}} \to 0$ almost surely. Now we can apply the Lindeberg central limit theorem and conclude.

More Eyebleach

Bounded lipschitz metric

From the results of chapter 1.12 it follows that weak convergence to separable limits is metrizable.

Bounded lipschitz metric

From the results of chapter 1.12 it follows that weak convergence to separable limits is metrizable. Define

$$\mathrm{BL}_1 = \{f | f \text{ is bounded and 1 Lipschitz}\}$$

Bounded lipschitz metric

From the results of chapter 1.12 it follows that weak convergence to separable limits is metrizable. Define

$$\mathrm{BL}_1 = \{f | f \text{ is bounded and 1 Lipschitz}\}$$

Then $X_\alpha \rightsquigarrow X$, where X is Borel measurable and separable, if and only if

Bounded lipschitz metric

From the results of chapter 1.12 it follows that weak convergence to separable limits is metrizable. Define

$$\mathrm{BL}_1 = \{f | f \text{ is bounded and 1 Lipschitz}\}$$

Then $X_\alpha \rightsquigarrow X$, where X is Borel measurable and separable, if and only if

$$\sup_{f \in \mathrm{BL}_1} |\mathbb{E}^* f(X_\alpha) - \mathbb{E} f(X)| \to 0.$$

Break

Conditional multiplier central limit theorem

Theorem

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

Conditional multiplier central limit theorem

Theorem

Let $\mathcal{F}$ be a class of measurable functions. Let $\xi_1, \cdots, \xi_n$ be i.i.d. random variables with mean zero, variance 1, and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$.

Conditional multiplier central limit theorem

Theorem

Let $\mathcal{F}$ be a class of measurable functions. Let $\xi_1, \cdots, \xi_n$ be i.i.d. random variables with mean zero, variance 1, and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$. Let $\mathbb{G}'_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i(\delta_{X_i} - P)$.

Conditional multiplier central limit theorem

Theorem

Let $\mathcal{F}$ be a class of measurable functions. Let $\xi_1, \cdots, \xi_n$ be i.i.d. random variables with mean zero, variance 1, and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$. Let $\mathbb{G}'_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i(\delta_{X_i} - P)$. Then the following assertions are equivalent:

Conditional multiplier central limit theorem

Theorem

Let $\mathcal{F}$ be a class of measurable functions. Let $\xi_1, \cdots, \xi_n$ be i.i.d. random variables with mean zero, variance 1, and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$. Let $\mathbb{G}'_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i(\delta_{X_i} - P)$. Then the following assertions are equivalent:

- *$\mathcal{F}$ is Donsker.*

Conditional multiplier central limit theorem

Theorem

Let $\mathcal{F}$ be a class of measurable functions. Let $\xi_1, \cdots, \xi_n$ be i.i.d. random variables with mean zero, variance 1, and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$. Let $\mathbb{G}'_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i(\delta_{X_i} - P)$. Then the following assertions are equivalent:

- *$\mathcal{F}$ is Donsker.*
- $\sup_{h \in BL_1} |\mathbb{E}_\xi h(\mathbb{G}'_n) - \mathbb{E} h(G)| \to 0$ *in outer probability, and the sequence $\mathbb{G}'_n$ is asymptotically measurable.*

proof

We begin with the forward statement of the theorem. We split it in 6 steps.

Step 1: asymptotic measurability

For a Donsker class $\mathcal{F}$, from the unconditional multiplier central limiet theorem, it follows that G_n' converges in distribution to a Brownian bridge process. Thus it is asymptotically measurable.

Step 2: Delta nets and projections

By 2.3.12 we see that a Donsker class is totally bounded in ρ_P.

Step 2: Delta nets and projections

By 2.3.12 we see that a Donsker class is totally bounded in ρ_P. Therefore there is a maximal family of points which are δ much separated.

Step 2: Delta nets and projections

By 2.3.12 we see that a Donsker class is totally bounded in ρ_P. Therefore there is a maximal family of points which are δ much separated. Denote $\Pi_\delta f$ the projection of f onto the closest element in a fixed δ-net for $\mathcal{F}$.

Step 3: Using continuity of G

Since G is a Brownian bridge process, it is continuous. Thus $G\Pi_\delta f \to Gf$ when $\delta \to 0$, and even stronger, $G \circ \pi \to G$ almost surely when $\delta \to 0$.

Step 4: Using the pointwise convergence

Note that in the previous lemma we have shown the pointwise convergence for any fixed f, and therefore for finitely many points, of $\mathbb{G}'_n$ to G (in distribution). Thus, for every fixed $\delta > 0$:

Step 4: Using the pointwise convergence

Note that in the previous lemma we have shown the pointwise convergence for any fixed f, and therefore for finitely many points, of $\mathbb{G}_n'$ to G (in distribution). Thus, for every fixed $\delta > 0$:

$$\sup_{h \in \mathrm{BL}_1} \left| \mathbb{E}_\xi h(\mathbb{G}_n' \circ \Pi_\delta) - \mathbb{E} h(G \circ \Pi_\delta) \right| \to 0,$$

for almost every sequence $X_1, X_2, \cdots$.

Step 5: Bounding the error between $\mathbb{G}_n\Pi_\delta$ and $\mathbb{G}_n$

$$
\begin{aligned}
\sup_{f\in \mathsf{BL}_1} |\mathbb{E}_\xi h(\mathbb{G}_n' \circ \Pi_\delta) - \mathbb{E}_\xi h(\mathbb{G}_n')| &\leq \sup_{h\in \mathsf{BL}_1} \mathbb{E}|h(\mathbb{G}_n' \circ \Pi_\delta) - h(\mathbb{G}_n')| \\
&\leq \mathbb{E}_\xi \|\mathbb{G}_n' \circ |pi_\delta - \mathbb{G}_n'\|_{\mathbb{F}}^* \\
&\leq \mathbb{E}_\xi \|\mathbb{G}_n'\|_{\mathcal{F}_\delta}^*
\end{aligned}
$$

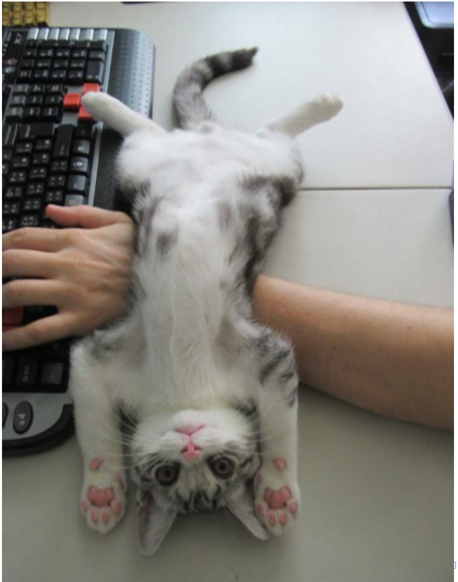
Step 6: finishing up

We use the multiplier central limit theorem to control the last term, same as in the theorem 2.9.2. Then we let $n \to \infty$ and then $\delta \downarrow 0$. This concludes one direction of the proof.

Other direction

This direction is much less interesting, it follows from the triangle inequality, Fubini, dominated convergence theorem and then theorem 2.9.2.

Yet another kitty picture



Right notion of almost sure convergence

We want to strengthen the statement from outer probability to an almost sure statement.

Right notion of almost sure convergence

We want to strengthen the statement from outer probability to an almost sure statement. However, almost sure convergence itself does not mean much without the proper measurability requirements, so we have to decide how to proceed.

Right notion of almost sure convergence

We want to strengthen the statement from outer probability to an almost sure statement. However, almost sure convergence itself does not mean much without the proper measurability requirements, so we have to decide how to proceed. The option chosen in the book is to go for outer almost sure convergence.

Consequences of our choices

We will show that

$$\sup_{h \in \mathrm{BL}_1} |\mathbb{E}_\xi h(\mathbb{G}'_n) - \mathbb{E}h(\mathbb{G})| \overset{\mathrm{as}*}{\to} 0.$$

This implies that $\mathbb{E}_\xi h(G'_n) \to \mathbb{E}h(G)$ for every $h \in \mathrm{BL}_1$, for almost every sequence $X_1, X_2, \cdots$. By the portmanteau lemma this is then also true for every continuous, bounded h.

Almost sure conditional multiplier central limit theorem

Theorem

Let $\mathcal{F}$ *be a class of measurable functions. Let* $\xi_1, \cdots, \xi_n$ *be i.i.d. random variables with mean zero, variance* 1 *and* $\|\xi\|_{2,1} < \infty$*, independent of* $X_1, \cdots, X_n$*.*

Almost sure conditional multiplier central limit theorem

Theorem

Let $\mathcal{F}$ be a class of measurable functions. Let $\xi_1, \cdots, \xi_n$ be i.i.d. random variables with mean zero, variance 1 and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$. . Define $\mathbb{G}'_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i (\delta_{X_i} - P)$.

Almost sure conditional multiplier central limit theorem

Theorem

Let $\mathcal{F}$ be a class of measurable functions. Let $\xi_1, \cdots, \xi_n$ be i.i.d. random variables with mean zero, variance 1 and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$. . Define $\mathbb{G}'_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i (\delta_{X_i} - P)$. Then the follow assertions are equivalent:

Almost sure conditional multiplier central limit theorem

Theorem

Let $\mathcal{F}$ be a class of measurable functions. Let $\xi_1, \cdots, \xi_n$ be i.i.d. random variables with mean zero, variance 1 and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$. . Define $\mathbb{G}'_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i(\delta_{X_i} - P)$. Then the follow assertions are equivalent:

- *$\mathcal{F}$ is Donsker and $P^*\|f - Pf\|_{\mathcal{F}}^2 < \infty$.*

Almost sure conditional multiplier central limit theorem

Theorem

Let $\mathcal{F}$ be a class of measurable functions. Let $\xi_1, \cdots, \xi_n$ be i.i.d. random variables with mean zero, variance 1 and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$. . Define $\mathbb{G}'_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i(\delta_{X_i} - P)$. Then the follow assertions are equivalent:

- *$\mathcal{F}$ is Donsker and $P^*\|f - Pf\|_{\mathcal{F}}^2 < \infty$.*
- $\sup_{h \in BL_1} |\mathbb{E}_\xi h(\mathbb{G}'_n) - Eh(G)| \to 0$ *outer almost surely, and the sequence $\mathbb{E}\xi h(\mathbb{G}'_n)^* - E\xi h(\mathbb{G}'_n)_*$ converges out almost surely to zero for every $h \in BL_1$.*

Almost sure conditional multiplier central limit theorem

Theorem

Let $\mathcal{F}$ be a class of measurable functions. Let $\xi_1, \cdots, \xi_n$ be i.i.d. random variables with mean zero, variance 1 and $\|\xi\|_{2,1} < \infty$, independent of $X_1, \cdots, X_n$. . Define $\mathbb{G}'_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i(\delta_{X_i} - P)$. Then the follow assertions are equivalent:

- $\mathcal{F}$ *is Donsker and* $P^*\|f - Pf\|_{\mathcal{F}}^2 < \infty$.
- $\sup_{h \in BL_1} |\mathbb{E}_\xi h(\mathbb{G}'_n) - Eh(G)| \to 0$ *outer almost surely, and the sequence* $\mathbb{E}\xi h(\mathbb{G}'_n)^* - E\xi h(\mathbb{G}'_n)_*$ *converges out almost surely to zero for every* $h \in BL_1$.

Here $h(\mathbb{G}'_n)^*$ *and* $h(\mathbb{G}'_n)_*$ *denote measurable majorants and minorants with respect to* $(\xi_1, \cdots, \xi_n, X_1, \cdots, X_n)$ *jointly.*

proof overview

The proof is more or less the same as the previous proof, except that all the arguments have to be applied to outer almost sure convergence instead of outer probability. This does not change much, only the final step of the forward implication changes. We need to control

$$\limsup_{n\to\infty} \mathbb{E}_\xi \|\mathbb{G}'_n\|^*_{\mathcal{F}_\delta}.$$

An application of corollary 2.9.9

By Corollary 2.9.9 (possible thanks to the extra assumption)

$$\limsup_{n\to\infty} \mathbb{E}_\xi \|\mathbb{G}'_n\|^*_{\mathcal{F}_\delta} \leq 6\sqrt{2} \limsup_{n\to\infty} \mathbb{E}^* \|\mathbb{G}'_n\|_{\mathcal{F}_\delta}, \qquad a.s.$$

Now take the limit as $n \to \infty$ and then $\delta \downarrow 0$, using theorem 2.9.2.

An application of corollary 2.9.9

By Corollary 2.9.9 (possible thanks to the extra assumption)

$$\limsup_{n\to\infty} \mathbb{E}_\xi \|\mathbb{G}'_n\|^*_{\mathcal{F}_\delta} \leq 6\sqrt{2} \limsup_{n\to\infty} \mathbb{E}^* \|\mathbb{G}'_n\|_{\mathcal{F}_\delta}, \qquad a.s.$$

Now take the limit as $n \to \infty$ and then $\delta \downarrow 0$, using theorem 2.9.2.
We omit the proof of the reverse implication.

Deviation bounds for conditional multiplier central limits

Let $Z_1, Z_2, \cdots$ be i.i.d. stochastic processes such that $\mathbb{E}^* \|Z_1\|_{\mathcal{F}}^2 < \infty$.

Deviation bounds for conditional multiplier central limits

Let $Z_1, Z_2, \cdots$ be i.i.d. stochastic processes such that $\mathbb{E}^* \|Z_1\|_{\mathcal{F}}^2 < \infty$. Let $\xi_1, \xi_2, \cdots$ be i.i.d. random variables with mean zero, independent of $Z_1, Z_2, \cdots$. Then

$$\limsup_{n\to\infty} \mathbb{E}_\xi \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i Z_i\|_{\mathcal{F}}^* \le 6\sqrt{2} \limsup_{n\to\infty} \mathbb{E}^* \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i Z_i\|_{\mathcal{F}}, \text{ a.s.}$$

$$\mathbb{E} \limsup_{n\to\infty} \mathbb{E}_\xi \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i Z_i\|_{\mathcal{F}}^* \le C \limsup_{n\to\infty} \mathbb{E}^* \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i Z_i\|_{\mathcal{F}} + C\sqrt{\mathbb{E}^* \|Z_1\|_{\mathcal{F}}^2}$$

For some universal constant C.

Where to find the proofs

The proof of this lemma is way too long to present, it is spread over 1 page after the statements, and 6 pages of appendix A.4.

Now you are as educated as this smart quokka

Sources for pictures

The pictures are coming from posts on reddit from the /r/Awww, /r/eyebleach and /r/EarthPorn subreddits.