Reading group mathematical foundations of statistics

Stefan Franssen, Msc

September 9, 2019

Goals

- To understand the current knowledge of nonparametric statistics better

Goals

- To understand the current knowledge of nonparametric statistics better
- To translate this knowledge into new results in statistics

Getting to know the Audience

Who are you?

Set up

- Held biweekly.

Set up

- Held biweekly.
- Alternates with the causality seminar.

Set up

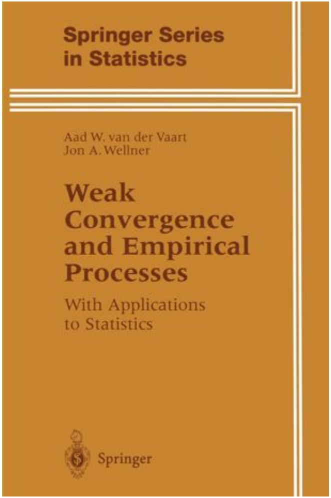
- Held biweekly.
- Alternates with the causality seminar.
- Amine, Bart, Geerten, Lasse and I will be the main speakers.

Set up

- Held biweekly.
- Alternates with the causality seminar.
- Amine, Bart, Geerten, Lasse and I will be the main speakers.
- First two talks will be held in Room 312, the rest of the semester will be in 176.

Set up

- Held biweekly.
- Alternates with the causality seminar.
- Amine, Bart, Geerten, Lasse and I will be the main speakers.
- First two talks will be held in Room 312, the rest of the semester will be in 176.
- Any questions so far?



The book we will start with.

You can get it from Springer link, where you can also order a softcover version for 25 euro.

I also have a pdf version of the second version of the book which I can share with the people here.

Topic of the book

The main object of interest is the empirical distribution

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

Topic of the book

The main object of interest is the empirical distribution

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

We will study its asymptotic normality, not just for a single function f but also uniformly over a class $\mathcal{F}$ of functions.

Topic of the book

The main object of interest is the empirical distribution

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

We will study its asymptotic normality, not just for a single function f but also uniformly over a class $\mathcal{F}$ of functions. We will develop tools for showing asymptotic normality in a uniform sense.

Outline of the book

- Part 1: Introduction of the basic definitions

Outline of the book

- Part 1: Introduction of the basic definitions
- Part 2: The theory on empirical processes

Outline of the book

- Part 1: Introduction of the basic definitions
- Part 2: The theory on empirical processes
- Part 3: Applications of the theory to statistical problems.

Part 1

Part 1 contains the basic theory of convergence of random variables in various settings.

Part 2

Part 2 contains the empirical process theory. The first 5 chapters are dealing with the main theory and the latter chapters deal mostly with examples.

Part 3

Part 3 contains a lot of statistical applications, for example the theory of M- and Z-estimators and the bootstrap is treated here.

Overview for today

We start the first two weeks with the first part of the book. Today I will do the following sections:

- Introduction
- Outer integrals and Measurable Majorants
- Weak convergence
- Product spaces

Introduction

The classical theory of weak convergence.

Introduction: weak convergence of probability measures

Definition

Let $(\mathbb{D}, d)$ be a metric space, and let $(P_n)_{n\in\mathbb{N}}$ and P be probability measures on the borel measurable space $(\mathbb{D}, \mathcal{D})$. Then P_n converges weakly to P if and only if

$$\int f \, dP_n \to \int f \, dP, \quad \forall f \in C_b(\mathbb{D})$$

Introduction: Weak convergence of random variables

Definition

In the same notation as before, let X_n, X be random variables taking values in $\mathbb{D}$. Then X_n converges weakly to X if and only if

$$\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)] \quad \forall f \in C_b(\mathbb{D}).$$

Introduction: The limitation of the classical theory

In the classical theory of weak convergence, we run into measurability issues, namely all the random variables X_n are measurable maps.

Introduction: The limitation of the classical theory

In the classical theory of weak convergence, we run into measurability issues, namely all the random variables X_n are measurable maps. However, classical examples of objects of interest are not measurable.

Introduction: The limitation of the classical theory

In the classical theory of weak convergence, we run into measurability issues, namely all the random variables X_n are measurable maps. However, classical examples of objects of interest are not measurable. For example let X_i be iid uniformly distributed random variables. Then the following two maps are not measurable in the natural spaces (exercise 1.7.3):

$$\mathbb{F}_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \leq t}$$

$$\mathbb{G}_n(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \mathbb{1}_{X_i \leq t}$$

Introduction: Suggested solutions

Several people have suggested solutions, but the one that can cover the most is suggested by Hoffmann-Jørgensen. We will follow this approach.

The suggested solution

The suggested solution is to drop the measurability requirement and only require asymptotic measurability.

Introduction: Key differences

The four main differences are:

- The notion of (uniform) tightness of sequences needs a modification

Introduction: Key differences

The four main differences are:

- The notion of (uniform) tightness of sequences needs a modification
- Prohorov's theorem, asymptotic tightness implies relative compactness, must be modified by the addition of the requirement of asymptotic measurability

Introduction: Key differences

The four main differences are:

- The notion of (uniform) tightness of sequences needs a modification
- Prohorov's theorem, asymptotic tightness implies relative compactness, must be modified by the addition of the requirement of asymptotic measurability
- (Almost) sure convergence is meaningless without asymptotic measurability

Introduction: Key differences

The four main differences are:

- The notion of (uniform) tightness of sequences needs a modification
- Prohorov's theorem, asymptotic tightness implies relative compactness, must be modified by the addition of the requirement of asymptotic measurability
- (Almost) sure convergence is meaningless without asymptotic measurability
- No general Fubini's theorem.

Outer integrals and Majorants: The outer integral

Definition (p. 6)

Let $(\Omega, \mathcal{A}, \mathbb{P})$ *be an probability space and let* $T : \Omega \to \bar{\mathbb{R}}$ *be any (possibly nonmeasureable) function. The* **outer integral** *of* T *with respect to* P *is defined as:*

$$\mathbb{E}^* [T] = \inf \left\{ \mathbb{E} U : U \geq T, U : \Omega \to \bar{\mathbb{R}} \textit{ integrable} \right\}$$

Outer integrals and Majorants: The outer probability

Definition (p. 6)

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be an probability space and let B be any (possibly nonmeasureable) subset of Ω. The **outer probability** *of B with respect to P is defined as:*

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

Outer integrals and Majorants: Inner versions

The inner integral and probability are defined by reversing the direction of the inequalities.

Outer integrals and Majorants: Existance of a Majorant

Lemma (Lem 1.2.1)

For any map $T : \Omega \to \bar{\mathbb{R}}$ there exists a measurable function $T^* : \Omega \to \bar{\mathbb{R}}$ with

- $T^* \geq T$;

Outer integrals and Majorants: Existance of a Majorant

Lemma (Lem 1.2.1)

For any map $T : \Omega \to \bar{\mathbb{R}}$ *there exists a measurable function* $T^* : \Omega \to \bar{\mathbb{R}}$ *with*

- $T^* \geq T$;
- $T^* \leq U$ *almost surely, for any measurable* $U : \Omega \to \bar{\mathbb{R}}$ *with* $U \geq T$ *a.s.*

Outer integrals and Majorants: Existance of a Majorant

Lemma (Lem 1.2.1)

For any map $T : \Omega \to \bar{\mathbb{R}}$ *there exists a measurable function* $T^* : \Omega \to \bar{\mathbb{R}}$ *with*

- $T^* \geq T$;
- $T^* \leq U$ *almost surely, for any measurable* $U : \Omega \to \bar{\mathbb{R}}$ *with* $U \geq T$ *a.s.*

For any T^* *satisfying these requirements, it holds that* $\mathbb{E}^* T = \mathbb{E} T^*$, *provided that* $\mathbb{E} T^*$ *exists. The latter is certainly true if* $\mathbb{E}^* T < \infty$.

Outer integrals and Majorants: Relation between outer measure and probability

Lemma (Lem 1.2.3)

For any subset B of Ω,

- $P^*(B) = \mathbb{E}^* \mathbb{1}_B$; $\mathbb{P}_*(B) = \mathbb{E}(\mathbb{1}_B)_*$;

Outer integrals and Majorants: Relation between outer measure and probability

Lemma (Lem 1.2.3)

For any subset B of Ω,

- $P^*(B) = \mathbb{E}^* \mathbb{1}_B$; $\mathbb{P}_*(B) = \mathbb{E}(\mathbb{1}_B)_*$;
- *there exists a measurable set B^* containing B with $\mathbb{P}(B^*) = \mathbb{P}^*(B)$, for any such B^* it holds that $\mathbb{1}_{B^*} = (\mathbb{1}_B)^*$.*

Outer integrals and Majorants: Relation between outer measure and probability

Lemma (Lem 1.2.3)

For any subset B of Ω,

- $P^*(B) = \mathbb{E}^* \mathbb{1}_B$; $\mathbb{P}_*(B) = \mathbb{E}(\mathbb{1}_B)_*$;
- *there exists a measurable set B^* containing B with $\mathbb{P}(B^*) = \mathbb{P}^*(B)$, for any such B^* it holds that $\mathbb{1}_{B^*} = (\mathbb{1}_B)^*$.*
- $(\mathbb{1}_B)^* + (\mathbb{1}_{\Omega - B})_* = 1$.

Outer integrals and Majorants: Fubini

Lemma (Lemma 1.2.6)

Let T be defined on a product probability space. Then

$$\mathbb{E}_* T \leq \mathbb{E}_{1*}\mathbb{E}_{2*} T \leq \mathbb{E}_1^*\mathbb{E}_2^* T \leq \mathbb{E}^* T$$

Outer integrals and Majorants: Fubini

Lemma (Lemma 1.2.6)

Let T be defined on a product probability space. Then

$$\mathbb{E}_* T \leq \mathbb{E}_{1*}\mathbb{E}_{2*} T \leq \mathbb{E}_1^*\mathbb{E}_2^* T \leq \mathbb{E}^* T$$

Lemma (Lem 1.2.7)

Let $(\Omega_1, \mathcal{A}_1)$ be a separable metric space equipped with its Borel σ-field.

Outer integrals and Majorants: Fubini

Lemma (Lemma 1.2.6)

Let T be defined on a product probability space. Then

$$\mathbb{E}_* T \leq \mathbb{E}_{1*}\mathbb{E}_{2*} T \leq \mathbb{E}_1^*\mathbb{E}_2^* T \leq \mathbb{E}^* T$$

Lemma (Lem 1.2.7)

Let $(\Omega_1, \mathcal{A}_1)$ be a separable metric space equipped with its Borel σ-field. Suppose the map $T : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is defined on a product probability space

Outer integrals and Majorants: Fubini

Lemma (Lemma 1.2.6)

Let T be defined on a product probability space. Then

$$\mathbb{E}_* T \leq \mathbb{E}_{1*}\mathbb{E}_{2*} T \leq \mathbb{E}_1^*\mathbb{E}_2^* T \leq \mathbb{E}^* T$$

Lemma (Lem 1.2.7)

Let $(\Omega_1, \mathcal{A}_1)$ be a separable metric space equipped with its Borel σ-field. Suppose the map $T : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is defined on a product probability space and satisfies
$|T(\omega_1, \omega_2) - T(\omega_1', \omega_2)| \leq d(\omega_1, \omega_1')H(\omega_2)$,

Outer integrals and Majorants: Fubini

Lemma (Lemma 1.2.6)

Let T be defined on a product probability space. Then

$$\mathbb{E}_* T \leq \mathbb{E}_{1*}\mathbb{E}_{2*} T \leq \mathbb{E}_1^*\mathbb{E}_2^* T \leq \mathbb{E}^* T$$

Lemma (Lem 1.2.7)

Let $(\Omega_1, \mathcal{A}_1)$ be a separable metric space equipped with its Borel σ-field. Suppose the map $T : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is defined on a product probability space and satisfies $|T(\omega_1, \omega_2) - T(\omega_1', \omega_2)| \leq d(\omega_1, \omega_1')H(\omega_2)$, for a function H with $P_2^* H < \infty$ and for every sufficiently small $d(\omega_1, \omega_1')$.

Outer integrals and Majorants: Fubini

Lemma (Lemma 1.2.6)

Let T be defined on a product probability space. Then

$$\mathbb{E}_* T \leq \mathbb{E}_{1*}\mathbb{E}_{2*} T \leq \mathbb{E}_1^*\mathbb{E}_2^* T \leq \mathbb{E}^* T$$

Lemma (Lem 1.2.7)

Let $(\Omega_1, \mathcal{A}_1)$ be a separable metric space equipped with its Borel σ-field. Suppose the map $T : \Omega_1 \times \Omega_2 \to \mathbb{R}$ is defined on a product probability space and satisfies $|T(\omega_1, \omega_2) - T(\omega_1', \omega_2)| \leq d(\omega_1, \omega_1')H(\omega_2)$, for a function H with $P_2^* H < \infty$ and for every sufficiently small $d(\omega_1, \omega_1')$. Then $\mathbb{E}^* T = \mathbb{E}_1\mathbb{E}_2^* T$

The monotone convergence theorem

Lemma (Ex. 1.2.3)

Let T_n, T be maps on a probability space with $T_n \uparrow T$ pointwise on a set of probability one.

The monotone convergence theorem

Lemma (Ex. 1.2.3)

Let T_n, T be maps on a probability space with $T_n \uparrow T$ pointwise on a set of probability one. Then $T_n^* \uparrow T^*$ almost surely.

The monotone convergence theorem

Lemma (Ex. 1.2.3)

Let T_n, T be maps on a probability space with $T_n \uparrow T$ pointwise on a set of probability one. Then $T_n^* \uparrow T^*$ almost surely. If the maps are bounded from below, then $\mathbb{E}^* T_n \uparrow \mathbb{E}^* T$.

The dominated convergence theorem

Lemma (Ex 1.2.4)

Let T_n, T be maps on a probability space with $|T_n - T|^*$ converging to zero almost surely.

The dominated convergence theorem

Lemma (Ex 1.2.4)

Let T_n, T be maps on a probability space with $|T_n - T|^*$ converging to zero almost surely. Suppose $\mathbb{E}^*[\sup_n |T_n|] < \infty$, then $\mathbb{E}^* T_n \to \mathbb{E}^* T$.

Polish spaces

Definition (P. 17)

A topological space X is called **Polish** *if it is seperable and its topology can be generated by a complete metric.*

Tight and separable random variables

Definition (p. 16, 17)

- *A Borel probability measure P is* **tight** *if for every $\epsilon > 0$ there exists a compact K such that $P(K) \geq 1 - \epsilon$.*

Tight and separable random variables

Definition (p. 16, 17)

- *A Borel probability measure P is* **tight** *if for every $\epsilon > 0$ there exists a compact K such that $P(K) \geq 1 - \epsilon$.*
- *A Borel probability measure P is* **pre-tight** *if for every $\epsilon > 0$ there exists a totally bounded measurable K such that $P(K) \geq 1 - \epsilon$.*

Tight and separable random variables

Definition (p. 16, 17)

- *A Borel probability measure P is* **tight** *if for every $\epsilon > 0$ there exists a compact K such that $P(K) \geq 1 - \epsilon$.*
- *A Borel probability measure P is* **pre-tight** *if for every $\epsilon > 0$ there exists a totally bounded measurable K such that $P(K) \geq 1 - \epsilon$.*
- *We call a Borel probability measure* **separable** *if there exists a separable measurable set S with $P(S) = 1$.*

Tight and separable random variables

Definition (p. 16, 17)

- *A Borel probability measure P is* **tight** *if for every $\epsilon > 0$ there exists a compact K such that $P(K) \geq 1 - \epsilon$.*
- *A Borel probability measure P is* **pre-tight** *if for every $\epsilon > 0$ there exists a totally bounded measurable K such that $P(K) \geq 1 - \epsilon$.*
- *We call a Borel probability measure* **separable** *if there exists a separable measurable set S with $P(S) = 1$.*

Lemma (1.3.2)

On a complete metric space these concepts are equivalent.

Nets

Definition

Let A be a directed set with order relation $\geq$ and X be a topological space. A map $f : A \to X$ is called a net.

Example

- Sequences

Weak convergence of random variables

Definition (1.3.3)

Let $(\Omega_\alpha, \mathcal{A}_\alpha P_\alpha)$ be a net of probability spaces and let $X_\alpha : \Omega_\alpha \to \mathbb{D}$ be arbitrary maps. The net X_α **converges weakly** *to a borel measure L if*

$$\mathbb{E}^* f(X_\alpha) \to \int f \, dL, \qquad \forall f \in C_b(\mathbb{D}).$$

We denote this convergence by $X_\alpha \rightsquigarrow L$.

Portmanteau lemma

Lemma (1.3.4)

The following statements are equivalent.

- $X_\alpha \rightsquigarrow L$;

Portmanteau lemma

Lemma (1.3.4)

The following statements are equivalent.

- $X_\alpha \rightsquigarrow L$;
- $\liminf P_*(X_\alpha \in G) \geq L(G)$ *for every open* G;

Portmanteau lemma

Lemma (1.3.4)

The following statements are equivalent.

- $X_\alpha \rightsquigarrow L$;
- $\liminf P_*(X_\alpha \in G) \geq L(G)$ *for every open* G;
- $\limsup P^*(X_\alpha \in F) \leq L(F)$ *for every closed* F;

Portmanteau lemma

Lemma (1.3.4)

The following statements are equivalent.

- $X_\alpha \rightsquigarrow L$;
- $\liminf P_*(X_\alpha \in G) \geq L(G)$ *for every open* G;
- $\limsup P^*(X_\alpha \in F) \leq L(F)$ *for every closed* F;
- $\liminf \mathbb{E}_* f(X_\alpha) \geq \int f \, dL$ *for every lower semicontinuous* f *that is bounded below;*

Portmanteau lemma

Lemma (1.3.4)

The following statements are equivalent.

- $X_\alpha \rightsquigarrow L$;
- $\liminf P_*(X_\alpha \in G) \geq L(G)$ *for every open* G*;*
- $\limsup P^*(X_\alpha \in F) \leq L(F)$ *for every closed* F*;*
- $\liminf \mathbb{E}_* f(X_\alpha) \geq \int f\, dL$ *for every lower semicontinuous* f *that is bounded below;*
- $\limsup \mathbb{E}^* f(X_\alpha) \leq \int f\, dL$ *for every upper semicontinuous* f *that is bounded above;*

Portmanteau lemma

Lemma (1.3.4)

The following statements are equivalent.

- $X_\alpha \rightsquigarrow L$;
- $\liminf P_*(X_\alpha \in G) \geq L(G)$ *for every open* G;
- $\limsup P^*(X_\alpha \in F) \leq L(F)$ *for every closed* F;
- $\liminf \mathbb{E}_* f(X_\alpha) \geq \int f \, dL$ *for every lower semicontinuous* f *that is bounded below;*
- $\limsup \mathbb{E}^* f(X_\alpha) \leq \int f \, dL$ *for every upper semicontinuous* f *that is bounded above;*
- $\lim P^*(X_\alpha \in B) = \lim P_*(X_\alpha \in B) = L(B)$ *for every Borel set* B *with* $L(\delta B) = 0$;

Portmanteau lemma

Lemma (1.3.4)

The following statements are equivalent.

- $X_\alpha \rightsquigarrow L$;
- $\liminf P_*(X_\alpha \in G) \geq L(G)$ *for every open* G;
- $\limsup P^*(X_\alpha \in F) \leq L(F)$ *for every closed* F;
- $\liminf \mathbb{E}_* f(X_\alpha) \geq \int f\, dL$ *for every lower semicontinuous* f *that is bounded below;*
- $\limsup \mathbb{E}^* f(X_\alpha) \leq \int f\, dL$ *for every upper semicontinuous* f *that is bounded above;*
- $\lim P^*(X_\alpha \in B) = \lim P_*(X_\alpha \in B) = L(B)$ *for every Borel set* B *with* $L(\delta B) = 0$;
- $\liminf \mathbb{E}_* f(X_\alpha) \geq \int f\, dL$ *for every bounded, Lipschitz continuous, nonnegative* f.

The continuous mapping theorem for weak convergence

Lemma (1.3.6)

Let $g : \mathbb{D} \to \mathbb{E}$ *be continous at every point of set* $\mathbb{D}_0 \subset \mathbb{D}$*.*

The continuous mapping theorem for weak convergence

Lemma (1.3.6)

Let $g : \mathbb{D} \to \mathbb{E}$ *be continous at every point of set* $\mathbb{D}_0 \subset \mathbb{D}$*. If* $X_\alpha \rightsquigarrow X$ *and* X *takes its values in* $\mathbb{D}_0$*,*

The continuous mapping theorem for weak convergence

Lemma (1.3.6)

Let $g : \mathbb{D} \to \mathbb{E}$ *be continous at every point of set* $\mathbb{D}_0 \subset \mathbb{D}$. *If* $X_\alpha \rightsquigarrow X$ *and* X *takes its values in* $\mathbb{D}_0$, *then* $g(X_\alpha) \rightsquigarrow g(X)$.

Asymptotically measurability and asymptotically tight

Definition (1.3.7)

The net of maps X_α is:

- *asymptotically measurable iff*

$$\mathbb{E}^* f(X_\alpha) - \mathbb{E}_* f(X_\alpha) \to 0, \qquad \forall f \in C_b(\mathbb{D}).$$

Asymptotically measurability and asymptotically tight

Definition (1.3.7)

The net of maps X_α is:

- *asymptotically measurable iff*

$$\mathbb{E}^* f(X_\alpha) - \mathbb{E}_* f(X_\alpha) \to 0, \qquad \forall f \in C_b(\mathbb{D}).$$

- *Asymptitically tight if for every $\epsilon > 0$ there exists a compact K such that*

$$\liminf P_*(X_\alpha \in K^\delta) \geq 1 - \epsilon, \qquad \forall \delta > 0.$$

Here $K^\delta = \{y \in \mathbb{D} : d(y, K) < \delta\}$ the δ-enlargement around K.

Prohorov's theorem 1: The forward statement

Lemma (1.3.8)

- *If $X_\alpha \rightsquigarrow X$, then X_α is asymptotically measurable.*

Prohorov's theorem 1: The forward statement

Lemma (1.3.8)

- *If* $X_\alpha \rightsquigarrow X$, *then* X_α *is asymptotically measurable.*
- *If* $X_\alpha \rightsquigarrow X$, *then* X_α *is asymptotically tight if and only if* X *is tight.*

Prohorov's theorem 2: The reverse statement

Lemma (1.3.9)

- *If the net X_α is asymptotically tight and asymptotically measurable, then it has a subnet $X_{\alpha(\beta)}$ that converges in law to a tight Borel law.*

Prohorov's theorem 2: The reverse statement

Lemma (1.3.9)

- *If the net* X_α *is asymptotically tight and asymptotically measurable, then it has a subnet* $X_{\alpha(\beta)}$ *that converges in law to a tight Borel law.*
- *If the sequence* X_n *is asymptotically tight and asymptotically measurable, then it has a subsequence* X_{n_j} *that converges weakly to a tight Borel law.*

Prohorov's theorem 2: The reverse statement

Lemma (1.3.9)

- *If the net X_α is asymptotically tight and asymptotically measurable, then it has a subnet $X_{\alpha(\beta)}$ that converges in law to a tight Borel law.*
- *If the sequence X_n is asymptotically tight and asymptotically measurable, then it has a subsequence X_{n_j} that converges weakly to a tight Borel law.*

https://math.stackexchange.com/questions/2150490/
trying-to-understand-how-a-subnet-of-a-sequence-differs-fr
https://math.stackexchange.com/questions/1209341/example-of-
converging-subnet-when-there-is-no-converging-subsequence?
noredirect=1&lq=1

The end

Any questions?