Asymptotic statistics

Chapter 1 and 2: Introduction and weak convergence

Stefan Franssen, Msc

October 5, 2021

Organisation

- Introduction
- Overal set up
- Rota for presenting
- Discussion

One by one

Set up

- Goal

Set up

- Goal
- History

Set up

- Goal
- History
- Weekly presentation

Rota

Rota Link

Discussion

Questions?

Asymptotic statistics

- Chapter 1: Introduction
- Chapter 2: Stochastic convergence

Introduction

- Approximate statistical procedures;
- Asymptotic optimality theory;
- Limitations.

Frequentist statistics

We study the statistical tools from an (empirical) frequentist point of view.

Chapter 2: Stochastic convergence

Almost sure convergence

Definition

We say a sequence of random variables X_n converges almost surely to X if

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1.$$

This is denoted by

$$X_n \xrightarrow{\text{as}} X.$$

Convergence in probability

Definition

We say a sequence of random variables X_n taking values in a metric space converges in probability to X if for all $\epsilon > 0$

$$\mathbb{P}\left(d(X_n, X) > \epsilon\right) \to 0.$$

This is denoted by

$$X_n \xrightarrow{P} X.$$

Weak convergence: Definition 1

Definition

We say a sequence of real valued random variables X_n converges weakly to X if

$$\mathbb{P}(X_n \le x) \to \mathbb{P}(X \le x)$$

at every point where $x \mapsto \mathbb{P}(X \le x)$ is continuous. We denote this by

$$X_n \rightsquigarrow X.$$

Weak convergence: Definition 1

Definition

We say a sequence of real valued random variables X_n converges weakly to X if

$$\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$$

at every point where $x \mapsto \mathbb{P}(X \leq x)$ is continuous. We denote this by

$$X_n \rightsquigarrow X.$$

Remark

Note that we can extend this to vector valued random variables by defining that a vector $v \leq u$ iff $v_i \leq u_i$ for all i.

Portmanteau Lemma

Lemma (Portmanteau)

For random vectors X_n and X the following are equivalent

- X_n *converges weakly to* X;
- $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ *for all bounded continuous functions* f;
- $\mathbb{E}[f(X_n)] \to \mathbb{E}[f(X)]$ *for all bounded lipschitz functions* f;
- $\liminf \mathbb{E}[f(X_n)] \geq \mathbb{E}[f(X)]$ *for all nonnegative, continuous functions* f;
- $\liminf \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$ *for every open set* G;
- $\limsup \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F)$ *for every closed set* F;
- $\mathbb{P}(X_n \in B) \to \mathbb{P}(X \in B)$ *for all Borel sets* B *with* $\mathbb{P}(X \in \delta B) = 0$, *where* δB *is the boundary of* B.

proof sketch, step 1 $\mathrm{i} \Rightarrow \mathrm{ii}$

Claim

There exists a set C such that C is dense and rectangles I with corners in C are continuity sets.

proof sketch, step 1 i $\Rightarrow$ ii

Claim

There exists a set C such that C is dense and rectangles I with corners in C are continuity sets.

Let $\epsilon > 0$. Pick a rectangle I with all their corners in C such that $\mathbb{P}(X \in I) > 1 - \epsilon$. Since I is compact, f is uniformly continuous on I, and hence we can partition I into finitely many sets I_j such that f varies at most ϵ on each I_j, and each I_j has corners in C. Pick x_j arbitrary points in I_j, then $|f(x) - f(x_j)| \leq 2\epsilon$ for all $x \in I_j$. Define $f_\epsilon = \sum_j f(x_j)\mathbb{1}_{I_j}$. Then

$$\|\mathbb{E}f(X_n) - \mathbb{E}f_\epsilon(X_n)\| \leq 2\epsilon + \mathbb{P}(X_n \notin I)$$
$$\|\mathbb{E}f(X) - \mathbb{E}f_\epsilon(X)\| \leq 2\epsilon + \mathbb{P}(X \notin I) \leq 2\epsilon$$

proof cont'd

$$\|\mathbb{E}f(X_n) - \mathbb{E}f_\epsilon(X_n)\| \leq 2\epsilon + \mathbb{P}(X_n \notin I)$$
$$\|\mathbb{E}f(X) - \mathbb{E}f_\epsilon(X_n)\| \leq 2\epsilon + \mathbb{P}(X \notin I) \leq 3\epsilon$$

Note that for large enough n, we can make $\mathbb{P}(X_n \notin I) < 2\epsilon$. Moreover

$$\|\mathbb{E}f_\epsilon(X_n) - \mathbb{E}f_\epsilon(X)\| \leq \sum_j \|\mathbb{P}(X_n \in I_j) - \mathbb{P}(X \in I_j)\||f(x_j)| \to 0$$

Hence, $\|\mathbb{E}f(X_n) - \mathbb{E}f(X)\| \leq 7\epsilon$ for n large enough, for all $\epsilon > 0$. The rest of the proof is left to read in the book.

Weak convergence: Definition 2

Definition

We say a sequence of random variables X_n converges weakly to X if for every bounded continuous real valued function

$$\mathbb{E}f(X_n) \to \mathbb{E}f(X).$$

Continuous mapping

Lemma

Let $g : \mathbb{R}^k \mapsto \mathbb{R}^m$ *be continuous at every point of a set* C *sch that* $\mathbb{P}(X \in C) = 1$.

- *If* $X_n \rightsquigarrow X$, *then* $g(X_n) \rightsquigarrow g(X)$;
- *If* $X_n \xrightarrow{P} X$, *then* $g(X_n) \xrightarrow{P} G(X)$;
- *If* $X_n \xrightarrow{as} X$, *then* $g(X) \xrightarrow{as} G(X)$.

Tightness and uniform tightness

Tightness is the corresponding concept of being bounded for a random variable

Definition

We say a random variable X is tight if for every $\epsilon > 0$ there exists a compact set C such that

$$\mathbb{P}(X \notin C) < \epsilon$$

Tightness and uniform tightness

Tightness is the corresponding concept of being bounded for a random variable

Definition

We say a random variable X is tight if for every $\epsilon > 0$ there exists a compact set C such that

$$\mathbb{P}(X \notin C) < \epsilon$$

Every random vector is tight.

Tightness and uniform tightness

Tightness is the corresponding concept of being bounded for a random variable

Definition

We say a random variable X is tight if for every $\epsilon > 0$ there exists a compact set C such that

$$\mathbb{P}(X \notin C) < \epsilon$$

Every random vector is tight.

Definition

A set of random variables $\{X_\alpha : \alpha \in \mathcal{A}\}$ is called uniformly tight if for all $\epsilon > 0$ there exists a compact set C such that

$$\mathbb{P}(X_\alpha \notin C) < \epsilon \forall \alpha \in \mathcal{A}.$$

Heine-Borel for random variables

Theorem (Prohorov's theorem)

Let X_n be random vectors in $\mathbb{R}^k$.

- *If $X_n \rightsquigarrow X$ for some X, then $\{X_n : n \in \mathbb{N}\}$ is uniformly tight;*
- *If X_n is uniformly tight, then there exists a subsequence of $X_{n_j} \rightsquigarrow X$ as $j \to \infty$, for some X.*

Relations between convergence

Theorem

Let X_n, X, Y_n, Y be random vectors and c a constant. Then

- $X_n \stackrel{as}{\to} X$ *implies* $X_n \stackrel{P}{\to} X$;

Relations between convergence

Theorem

Let X_n, X, Y_n, Y be random vectors and c a constant. Then

- $X_n \stackrel{as}{\to} X$ *implies* $X_n \stackrel{P}{\to} X$;
- $X_n \stackrel{P}{\to} X$ *implies* $X_n \rightsquigarrow X$;

Relations between convergence

Theorem

Let X_n, X, Y_n, Y be random vectors and c a constant. Then

- $X_n \stackrel{as}{\to} X$ *implies* $X_n \stackrel{P}{\to} X$;
- $X_n \stackrel{P}{\to} X$ *implies* $X_n \rightsquigarrow X$;
- $X_n \stackrel{P}{\to} c$ *if and only if* $X_n \rightsquigarrow c$;

Relations between convergence

Theorem

Let X_n, X, Y_n, Y be random vectors and c a constant. Then

- $X_n \xrightarrow{as} X$ *implies* $X_n \xrightarrow{P} X$;
- $X_n \xrightarrow{P} X$ *implies* $X_n \rightsquigarrow X$;
- $X_n \xrightarrow{P} c$ *if and only if* $X_n \rightsquigarrow c$;
- *if* $X_n \rightsquigarrow X$ *and* $d(X_n, Y_n) \xrightarrow{P} 0$, *then* $Y_n \rightsquigarrow X$;

Relations between convergence

Theorem

Let X_n, X, Y_n, Y be random vectors and c a constant. Then

- $X_n \xrightarrow{as} X$ *implies* $X_n \xrightarrow{P} X$;
- $X_n \xrightarrow{P} X$ *implies* $X_n \rightsquigarrow X$;
- $X_n \xrightarrow{P} c$ *if and only if* $X_n \rightsquigarrow c$;
- *if* $X_n \rightsquigarrow X$ *and* $d(X_n, Y_n) \xrightarrow{P} 0$, *then* $Y_n \rightsquigarrow X$;
- *if* $X_n \rightsquigarrow X$ *and* $Y_n \xrightarrow{P} c$, *then* $(X_n, Y_n) \rightsquigarrow (X, c)$;

Relations between convergence

Theorem

Let X_n, X, Y_n, Y be random vectors and c a constant. Then

- $X_n \xrightarrow{as} X$ *implies* $X_n \xrightarrow{P} X$;
- $X_n \xrightarrow{P} X$ *implies* $X_n \rightsquigarrow X$;
- $X_n \xrightarrow{P} c$ *if and only if* $X_n \rightsquigarrow c$;
- *if* $X_n \rightsquigarrow X$ *and* $d(X_n, Y_n) \xrightarrow{P} 0$, *then* $Y_n \rightsquigarrow X$;
- *if* $X_n \rightsquigarrow X$ *and* $Y_n \xrightarrow{P} c$, *then* $(X_n, Y_n) \rightsquigarrow (X, c)$;
- *if* $X_n \xrightarrow{P} X$ *and* $Y_n \xrightarrow{P} Y$, *then* $(X_n, Y_n) \xrightarrow{P} (X, Y)$.

Slutsky

Lemma

Let X_n, X, Y_n be random variables. If $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$ for a constant c, then

- $X_n + Y_n \rightsquigarrow X + c$;
- $Y_n X_n \rightsquigarrow cX$;
- $Y_n^{-1} X_n \rightsquigarrow c^{-1} X$ *provided that c is invertible.*

Slutsky

Lemma

Let X_n, X, Y_n be random variables. If $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow c$ for a constant c, then

- $X_n + Y_n \rightsquigarrow X + c$;
- $Y_n X_n \rightsquigarrow cX$;
- $Y_n^{-1} X_n \rightsquigarrow c^{-1} X$ *provided that c is invertible.*

Proof.

Previous theorem implies $(X_n, Y_n) \rightsquigarrow (X, c)$. Now apply continuous mapping to each of the maps and note that they are continuous. □

Stochastic o and O Symbols

$$X_n = o_P(R_n) \quad \text{means} \quad X_n = Y_n R_n \quad \text{and} \, Y_n \xrightarrow{P} 0;$$
$$X_n = O_P(R_n) \quad \text{means} \quad X_n = Y_n R_n \quad \text{and} \, Y_n \text{ bounded in probability}$$