MS Asymptotics For Stochastic Processes

Collection of results on
Stoch. processes:
elementary definitions;
important asymptotics

1. Elementary definitions

> The goal of this memory sheet is to compile a number of asymptotic results, akin to the standard versions of LLN & CLT for iid-sequences of random variables, for stochastic processes (in this cases sequences of r.v.s) of in principle arbitrary dependence structure.

> The way to do this here is as follows:

- Define important concepts rigorously
- Recap modes of convergence of stochastic sequences
- Asymptotic theory for Markov Processes
- Asymptotic theory for general Processes

useful mostly for theory (e.g. dynamic programming); heavy assumptions & heavy results

light assumptions & general but weak results; useful mostly for time series econometrics

> Let's go over some definitions; knowledge prerequisites:

- MBMT for σ-fields, measures, etc.
- MBI2 for Lebesgue-integration
- MSCE for conditional expectations

DEF 1.1 (Stoch. process, filtration). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a p-space and let $\mathcal{T}$ be some index set. A stochastic process $(X_t)_{t\in\mathcal{T}}$ is a family of random variables, $\forall t\in\mathcal{T}: X_t : (\Omega,\mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.
If $\mathcal{T}$ is a totally ordered set, we may define the sequence of σ-fields $(\mathcal{F}_t)_{t\in\mathcal{T}}$ with $\forall s\geq t, \mathcal{F}_s \supseteq \mathcal{F}_t$ and $\mathcal{F}_t \subseteq \mathcal{A}$ $\forall t$; we call $(\mathcal{F}_t)_{t\in\mathcal{T}}$ a filtration on $(\Omega, \mathcal{A}, \mathbb{P})$.
From now on, wlog concentrate on $d=1$
If we have $\sigma(X_t) \equiv X_t^{-1}(\mathcal{B}(\mathbb{R})) \subseteq \mathcal{F}_t$ $\forall t\in\mathcal{T}$, we call $(X_t)_{t\in\mathcal{T}}$ adapted to $(\mathcal{F}_t)_{t\in\mathcal{T}}$.

Often, given a process $(X_t)_{t\in\mathbb{Z}}$ ($\mathcal{T}$) on $(\Omega,\mathcal{A},\mathbb{P})$, we define
$\mathcal{F}_{t+n,t+m} := \sigma(\{X_{t+i}\}_{m\geq i\geq n})$ $\left(:= \sigma\left(\bigcup_{i=n}^{m} X_{t+i}^{-1}(\mathcal{B}(\mathbb{R}))\right)\right)$, the smallest σ-Alg. s.th. $\forall i\in\{0,\ldots,m\}$, X_{t+i} is measurable wrt it)
for $t\in\mathbb{Z}$, $n\in\mathbb{N}\cup\{-\infty\}$, $m\in\{k\in\mathbb{N}: k\geq n\}\cup\{+\infty\}$,
and the canonical filtration
$\mathcal{F}_t := \sigma(\{X_{t-i}\}_{i\geq 0}) = \sigma(\bigcup_{i\geq 0} \mathcal{F}_{t-i,t})$.

Remarks:

- Usually, and always in our context here, $\mathcal{T} \equiv \mathbb{Z}$.
- $(\mathcal{F}_t)_{t\in\mathcal{T}}$ is meant to capture the flow of information; informally, it models what sort of events may be observed/known at any point $t\in\mathcal{T}$; wlog, we may set $\mathcal{A} := \sigma(\bigcup_{t\in\mathbb{Z}} \mathcal{F}_t) = \sigma(\{X_t\}_{t\in\mathbb{Z}})$ for $\mathcal{T}\equiv\mathbb{Z}$
- For notational convenience, we write $\mathcal{L}(X_t)$ in reference to $P_{X_t} \equiv \mathbb{P}\circ X_t^{-1}$; analogously for $\mathcal{L}((X_s)_{s\in S\subseteq\mathcal{T}})$

From now on, fix $\mathcal{T}\equiv\mathbb{Z}$, and some $(\Omega,\mathcal{A},\mathbb{P})$. (While not important here, we usually choose $(\Omega,\mathcal{A}) \equiv (\mathbb{R}^{\mathbb{Z}}, \mathcal{B}(\mathbb{R})^{\otimes\mathbb{Z}})$.)
product σ-field.

> A core concept for stochastic processes is that of stationarity (or roughly the distribution always staying in the same place)

Def 1.2 (Stationarity). Let (X_t) be a process. We call (X_t) strictly stationary if

$$\forall t \in \mathbb{Z}, \forall m, s \geq 1, \quad \mathcal{L}(X_t, \dots, X_{t+s}) = \mathcal{L}(X_{t+m}, \dots, X_{t+m+s}).$$

We call (X_t) weakly stationary (or "covariance stationary") if $X_t \in \mathcal{L}^2(\mathbb{P})$ $\forall t$ and if $\forall t \in \mathbb{Z}$ $\forall k \in \mathbb{N}_0$,

$$\begin{cases} \mathbb{E}(X_t) = \mathbb{E}(X_1) \\ \mathbb{E}(X_t X_{t-k}) = \mathbb{E}(X_1 X_{1-k}). \end{cases}$$

(first 2^{nd} moments exist and are time-invariant)

Generally, weak stat.y never implies strict stat.y and str. st.y implies w. stat.y iff $(X_t) \in \mathcal{L}^2(\mathbb{P})$ (shorthand for $X_t \in \mathcal{L}^2(\mathbb{P})$ $\forall t \in \mathbb{Z}$).

> Stationarity is usually required when handling with stoch. proc.s — another often employed, but even more unintuitive concept is that of ergodicity; roughly, an ergodic process is such that along $\mathbb{Z}$ it a.s. travels through its entire range — it never gets stuck on a particular subset, once it has entered such a set. ("restricting the memory of the process")

DEF 1.3 (Shift-invariance; Ergodicity). Let (X_t) be a process on $(\Omega, \mathcal{A}, \mathbb{P})$. We define the shift-operator for (X_t),

$$\tau_X : \Omega \to \Omega : \omega \mapsto \tau_X(\omega) \text{ s.th. } X_t(\tau_X(\omega)) = X_{t+1}(\omega) \ \forall t \in \mathbb{Z}.$$

(concisely: τ s.th. $\forall t$ $X_t \circ \tau_X = X_{t+1}$; τ_X shifts ω backwards s.th. X_t at $\tau_X(\omega)$ gives the value of X_{t+1} at original ω; most easily seen for 'usual' way of defining stoch. proc.s: $(\Omega, \mathcal{A}) = (\mathbb{R}^{\mathbb{Z}}, \mathcal{B}(\mathbb{R})^{\mathbb{Z}})$ with $X_t : \underbrace{\omega}_{\text{a sequence } \{\omega_t\}_{t=-\infty}^{\infty}} \mapsto [\omega]_t$)

We define the σ-field of shift-invariant events for (X_t),

$$\mathcal{I}_X := \{A \in \mathcal{A} : \tau_X^{-1}(A) = A\}.$$ (can show this is σ-field, more useful descr. below.)

We say (X_t) is ergodic (on $(\Omega, \mathcal{A}, \mathbb{P})$) if

(i) $\mathcal{I}_X$ is $\mathbb{P}$-trivial, that is $\forall A \in \mathcal{I}_X$, $\mathbb{P}(A) \in \{0, 1\}$. (either A always occurs or it never occurs)

(ii) τ_X is measure-preserving, i.e. $\forall A \in \mathcal{A}$, $\mathbb{P}(\tau^{-1}(A)) = \mathbb{P}(A)$.

<u>Remark 1:</u> For $\mathcal{K}$ (some transf. $\Omega \to \Omega$) meas.p. & injective, and Y m.able, the process $X_0(\omega) = Y(\omega)$,

here need injectivity! $\mathcal{K}^{-1}$ should be single-valued

$$X_t(\omega) \overset{(*)}{:=} X_0(\underbrace{\mathcal{K}^t}_{\underbrace{\mathcal{K} \circ \dots \circ \mathcal{K}}_{t \text{ times}}}(\omega)), \forall t \in \mathbb{Z},$$ is strictly stationary; conversely every strictly stationary process <u>has a $\mathcal{K}$ (namely τ_X) measure-p. & injective and can be written as $(*)$.</u>

$\Rightarrow$ Ergodicity implies strict stat.y, if τ_X is injective — this is, e.g., the case for the 'canonical process':
$(\Omega, \mathcal{A}, \mathbb{P}) = (\mathbb{R}^{\mathbb{Z}}, \mathcal{B}(\mathbb{R})^{\otimes \mathbb{Z}}, \mathbb{P})$, $\underline{\tau_X : \Omega \to \Omega : ([\omega]_n)_{n\in\mathbb{Z}} \mapsto ([\omega]_{n+1})_n}$ ← invertible!
$X_n(\omega) := X_0(\tau_X^n(\omega))$ $\forall n \in \mathbb{Z}$.

Remark 2: Intuition of Shift-invariance & ergodicity.

- Shift-invariance can be 'decoded' into a more intuitive statement. First observe

$$A \in \mathcal{I} \overset{\text{def}}{\Longleftrightarrow} \tau^{-1}(A) = A \overset{\text{def}}{\Longleftrightarrow} \{\omega \in \Omega : \tau(\omega) \in A\} = A \Longleftrightarrow (\omega \in A \Leftrightarrow \tau(\omega) \in A)$$

$$\Rightarrow \tau(A) \subseteq A \quad \Rightarrow \quad \forall n \in \mathbb{N} \ \tau^n(A) \subseteq A \ \text{(by induction)}$$

(but not "$\supseteq$" since could $\exists \omega' \in A : \tau^{-1}(\omega') = \emptyset$)

- Hence we see: $\omega \in A$ for $A \in \mathcal{I}$ implies $\forall n \in \mathbb{N}$ $\tau^n(\omega) \in A$.
- Thus if $A \in \mathcal{I}$, we have $\forall n \in \mathbb{N}$: $\forall \omega \in A$,

$$X_n(\omega) = X_{n-1}(\tau(\omega)) = \ldots = X_0(\tau^n(\omega)) \in X_0(A) := \{X_0(\omega) \mid \omega \in A\}$$

since $\tau^n(\omega) \in A$.

$\Rightarrow$ If (X_t) is evaluated at a $\omega \in A$, the values $X_t(\omega)$ <u>never</u> leave the set $X_0(A)$!
This is what we want to prevent under ergodicity: either such sets A never occur, or they already capture the <u>wole</u> set of values that trajectories $(X_t(\omega))_{t\in\mathbb{Z}}$ can pass through. Either way, it cannot happen that two drawn trajectories $(X_t(\omega))$, $(X_t(\omega'))$ evolve in sets completely separate from each other.

> It can be shown that a strictly stationary & ergodic process satisfies a law of large numbers:

$$\frac{1}{T} \sum_{t=1}^{T} X_t \xrightarrow[T\to\infty]{a.s.} E[X_1]$$

Cf. WT1, Satz 11.4.

Often, such a LLN-property is made the definition of ergodicity.

> There are alternative definitions for ergodicity; a popular, stronger than the above, option is this sufficient condition:

Definition 3.33. Let $(\Omega, \mathscr{F}, P)$ be a probability space. Let $\{Y_t\}$ be a stationary sequence and let Z be the measure-preserving & injed. transformation <u>inducing this sequence</u>. Then, $\{Y_t\}$ is ergodic if and only if:
i.e. $Y_t(\omega) = Y_0(Z^t\omega)$

$$\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} P(F \cap Z^t G) = P(F)P(G)$$

Source: Davidson, p. 201 (221 in PDF)

13.13 Theorem A measure-preserving shift transformation T is ergodic if and only if, for any pair of events $A, B \in \mathscr{F}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} P(T^k A \cap B) = P(A)P(B). \qquad (13.28)$$

"Average asymptotic independence"

THM 1.4 (Measurable mappings). Let (X_t) be a strictly stationary (resp. ergodic) process, and let $g : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be measurable. Then, $(g(X_t))_t$ is strictly stationary (resp. ergodic).

> Intuitively, ergodicity requires that our sequence will — in any realization w/o a negligible set — always explore the whole possible sample space; conversely, for a nonergodic process, there are starting values whose vicinity is never left, with probability one

(Ergodicity, as we saw, may be reformulated as 'average asymptotic independence'

> An even stronger notion than ergodicity is **mixing**:
Intuitively, mixing requires that sequence elements in (X_t) become independent of each other as they become farther apart

(For most mixing concepts employed, ergodicity follows from mixing
(For the cases where it doesn't the culprit is usually that the shift is not measure-preserving)

(Ergodicity does not imply mixing, since it doesn't rule out that a sequence is perfectly predictable from one point onwards (which clearly prevents asy. independence!); cf. Example 13.15 on p. 202 (222 in PDF) in Davidson

> There are many different concepts for mixing and they may differ even in the mathematical objects to which the definitions apply!
Here is a (non-comprehensive) summary:

DEF 1.5 (Mixing). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a p-space, $(X_t)_{t\in\mathbb{Z}}$ a process on it, T_X its shift-operator, $(\mathcal{F}_t)_{t\in\mathbb{Z}}$ the canonical filtration generated by (X_t), suppose $\mathcal{A} = \sigma(\bigcup_{t\in\mathbb{Z}} \mathcal{F}_t)$ and define

$$\bar{\mathcal{F}} := \bigcap_{t\in\mathbb{Z}} \mathcal{F}_{t,+\infty} = \bigcap_{t\in\mathbb{Z}} \underbrace{\mathcal{F}_{-\infty,t}}_{\equiv \sigma(\bigcup_{i\geq 0} X_{t-i}^{-1}(\mathcal{B}(\mathbb{R})))}$$

($\bar{\mathcal{F}}$ contains the events that are part of infinitely many $X_{t-i}^{-1}(\mathcal{B}(\mathbb{R}))$; cf. tail-$\sigma$-Alg. in WT1-script)

as the remote σ-field, or tail-σ-algebra.
Then, (X_t) is called

(i) (shift-) mixing if T_X is measure-preserving, injective and

$$\forall A, B \in \mathcal{A}, \quad \lim_{k\to\infty} P(T_X^k(A) \cap B) = P(A)\cdot P(B),$$

(ii) regular (-mixing) if $\bar{\mathcal{F}}$ is $\mathbb{P}$-trivial, i.e. $\forall A \in \bar{\mathcal{F}}, \mathbb{P}(A) \in \{0,1\}$; equivalently

$$\forall B \in \mathcal{A}, \quad \sup_{A\in\mathcal{F}_t} |P(A\cap B) - P(A)P(B)| \to 0 \text{ as } t \to -\infty$$

(informally: remote events are independent of events in $\mathcal{A}$)

(iii) ξ-mixing (for $\xi \in \{\alpha, \phi\}$) with size $-\varphi < 0$ if
"strong mixing" "uniform mixing"

$$\xi_m = O(m^{-\varphi-\varepsilon}) \text{ for some } \varepsilon > 0 \text{ and for}$$

— $\xi_m \to 0$ incrementally faster than $m^{-\varphi}$

$$\xi_m := \sup_{t\in\mathbb{Z}} \xi(\mathcal{F}_{-\infty,t}, \mathcal{F}_{t+m,\infty}) \text{ with (for } \mathcal{G}, \mathcal{H} \subseteq \mathcal{A} \text{ } \sigma\text{-alg.s)}$$

measure of dependence btwn $\mathcal{G}$ & $\mathcal{H}$

$$\alpha(\mathcal{G},\mathcal{H}) := \sup_{G\in\mathcal{G}, H\in\mathcal{H}} |P(G\cap H) - P(G)P(H)|, \quad \phi(\mathcal{G},\mathcal{H}) := \sup_{G, H \in \dots} |P(H|G) - P(H)|$$

Remarks:

- $\mathcal{I}_X \subseteq \mathcal{F}$ provided τ_X is injective (All shift-invariant events are remote) — or we go over from $\mathbb{Z}$ to $\mathbb{N}_0$
- regular-mixing and τ_X injective & measure-preserving implies equality (by $\mathcal{I}_X \subseteq \mathcal{F}$ and $\mathcal{F}$ $\mathbb{P}$-trivial)
- $\forall\, \mathcal{G}, \mathcal{H}$, $\alpha(\mathcal{G},\mathcal{H}) \leq \phi(\mathcal{G},\mathcal{H})$ so ϕ-mixing $\Rightarrow$ α-mixing $\Rightarrow$ regular-mixing

THM 1.6 (Mixing and measurable mappings). For $g: \mathbb{R}^{\tau+1} \to \mathbb{R}$ measurable, and (X_t) being ξ-mixing ($\xi \in \{\alpha, \phi\}$) of size $-\varphi < 0$, $Y_t := g(X_t, \dots, X_{t-\tau})$ is also.

Remark: The theorem does not hold for $\tau = +\infty$!

Mixingale, mixing (α, ϕ, β),
Thm 14.1 Davidson (m.able transf.s preserve α-/ϕ-mixing),
Relation Ergodicity & Mixing ?

Note: mixing process in $\mathcal{L}^p$ is also an $\mathcal{L}^p$-mixingale

Outline:

- Mixing
  - intuition: asy. independence
  - Filtrations, tail-σ-A, Kolmogorov 0-1
  - regularity-mixing: $\alpha(\mathcal{F}_{-\infty}^{t}, \mathcal{F}) \to 0$ as $t \to -\infty$
    - r-mixing & staty sequ. is ergodic
  - α-, ϕ-mixing coeff.s for $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$
  - α-, ϕ-mixing processes
    - $\Rightarrow$ regularity? Yes, I think for α-mixing
  - Mixing under m.able mappings
  - Word of caution: mixing may <u>not</u> survive infinite filtering!
  - Mixing inequ.s
- Mixingales

Davidson: Ch.s 13 & 14

2. Modes of convergence for stochastic sequences

> Let's explore the different notions of convergence for random variables; the standard reference for this is WT1-script, sect. I.5+7

> A more condensed exposition of the central concepts is in E703, II.1; this we simply recap here:

II.1 Recap I: important notions of convergence from p-theory

> Just as a reference (no lengthy explanations / intuitions), find here a summary of the definitions, properties and linkages of the most important notions of convergence of random variables

↳ for definition of here implicitly used concepts (like r.v.) or for general reference, cf. script WT1

> For a random vector $\underline{X} : (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $d \in \mathbb{N}$ and a sequence of r.v.s $(\underline{X}_n)_{n\in\mathbb{N}}$, we frequently use

applies to scalars as well!

$[\xrightarrow{a.s.}]$ $\underline{X}_n \xrightarrow{a.s.} \underline{X} \overset{def}{\iff} P(\limsup_{n\to\infty} \|\underline{X}_n - \underline{X}\| = 0) = 1$

← always Euclidean norm

useful for proofs (e.g. use Fatou) but not necess. here since $\|\cdot\| \geq 0$ ↳ lim exists

Intuition: $\underline{X}_n(\omega) \to \underline{X}(\omega)$ (deterministically) $\forall \omega \in \Omega \setminus N$ where $P(N) = 0$ ($\underline{X}_n$ converges for sure in all events w/ pos. probab.)

$\iff X_{j,n} \xrightarrow{a.s.} X_j \quad \forall j \in \{1, \dots, d\}$

$[\xrightarrow{p}]$ $\underline{X}_n \xrightarrow{p} \underline{X} \overset{def}{\iff} \limsup_{n\to\infty} P(\|\underline{X}_n - \underline{X}\| > \varepsilon) = 0 \quad \forall \varepsilon > 0$ ($\forall \varepsilon, \delta > 0\ \exists N: \forall n \geq N\ P(\ldots > \varepsilon) < \delta$)

$P(\cdot) \geq 0 \ldots$

Intuition: there might be nontrivial set where $\underline{X}_n$ does not converge, but the 'deviation probability' fades to 0 as $n \to \infty$ (it becomes increasingly improbable that $\underline{X}_n$ and $\underline{X}$ do not agree)

$\iff X_{j,n} \xrightarrow{p} X_j \quad \forall j$

„$\Rightarrow$": $P(|X_{j,n} - X_j| > \varepsilon) \leq P(\|\underline{X}_n - \underline{X}\| > \varepsilon) \to 0$ as $n \to \infty$
(by $\{\omega : |X_{j,n}(\omega) - X_j(\omega)| > \varepsilon\} \subseteq \{\omega : \sqrt{\sum_j |X_{j,n}(\omega) - X_j(\omega)|^2} > \varepsilon\}$)

„$\Leftarrow$": Suppose $\underline{X}_n \not\xrightarrow{p} \underline{X}$. Choose, then, $(\underline{X}_{n_k})_{k\in\mathbb{N}}$: $\exists \tilde{\varepsilon}, \tilde{\delta}$: $P(\|\underline{X}_{n_k} - \underline{X}\| > \tilde{\varepsilon}) > \tilde{\delta}$ $\forall k \in \mathbb{N}$. Since $X_{j,n} \xrightarrow{p} X_j$, can choose for each j a subsequence (of X_{j,n_k}!) that conv. to X_j a.s. (WT1, K5.8) wlog let X_{j,n_k} be such subs. for each j. Then, by the above, $\underline{X}_{n_k} \xrightarrow{a.s.} \underline{X}$. But then, $\underline{X}_{n_k} \xrightarrow{p} \underline{X} \Rightarrow \perp$.
(WT1, L5.6 (by Lemma of Fatou, limsup-part))

also have equivalence to statements about sequences being Cauchy (a.s. / in p)

$[\xrightarrow{\mathcal{L}^p}]$ $\underline{X}_n \xrightarrow{\mathcal{L}^p} \underline{X} \overset{def}{\iff} \underbrace{\mathbb{E}(\|\underline{X}_n - \underline{X}\|^p)^{(1/p)}}_{=: \|\underline{X}_n - \underline{X}\|_{\mathcal{L}^p}} \to 0$ as $n \to \infty$

← not nec. since $x \mapsto x^{1/p}$ contin.

Intuition: the expected Euclidean distance btwn $\underline{X}_n$ and $\underline{X}$ fades to 0

$\iff X_{j,n} \xrightarrow{\mathcal{L}^p} X_j \quad \forall j$

$\Rightarrow \mathbb{E}(\underline{X}_n) \to \mathbb{E}(\underline{X})$ as $n \to \infty$ (but not(!) „$\Leftarrow$")
(Pf.: for $d=1$ (use above) $0 \leq |\mathbb{E}(\underline{X}_n) - \mathbb{E}(\underline{X})| = |\mathbb{E}(\underline{X}_n - \underline{X})| \leq \mathbb{E}(|\underline{X}_n - \underline{X}|) \to 0$ as $n \to \infty$ since $\underline{X}_n \xrightarrow{\mathcal{L}^p} \underline{X} \Rightarrow \underline{X}_n \xrightarrow{\mathcal{L}^1} \underline{X}$ (WT1, MBI2, D3.2))

$[\xrightarrow{d}]$ $\underline{X}_n \xrightarrow{d} \underline{X} \overset{def}{\iff} \mathbb{E}(f(\underline{X}_n)) \to \mathbb{E}(f(\underline{X}))$ as $n \to \infty$ for all $f : \mathbb{R}^d \to \mathbb{R}$ bounded & continuous (note: $x \mapsto x$ is not bounded!)

$\not\Leftrightarrow (X_{j,n} \xrightarrow{d} X_j\ \forall j)$!

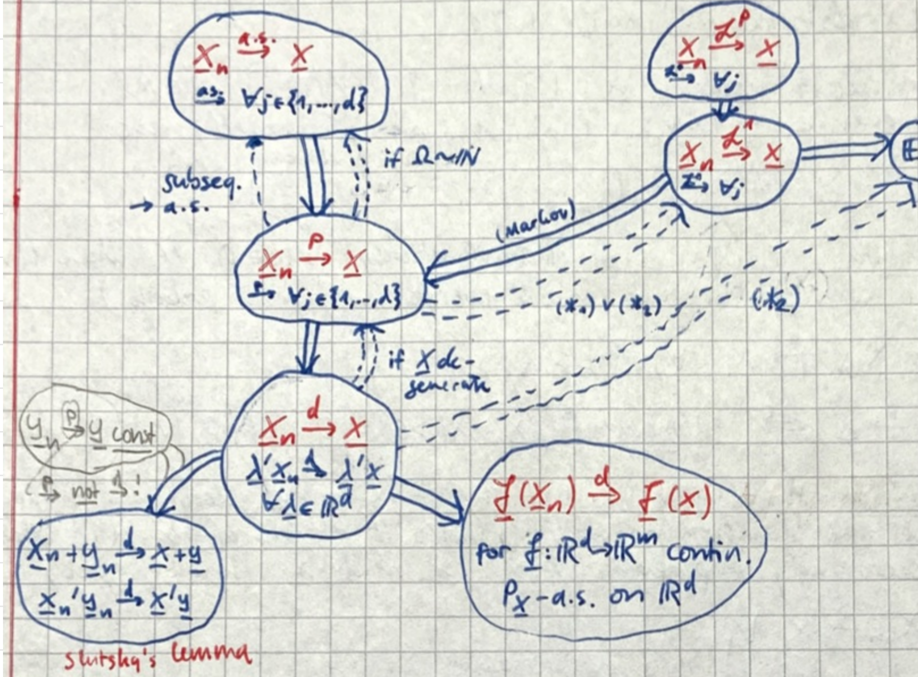
Intuition: the integrals converge for some functions (not for all) — it can be shown that this def. is eqvt. to a more intuitive one which (roughly) states that the distributions should become more 'similar' (cf. WT1, sect. 7, i.p. Satz 7.16, for this)

$\Leftrightarrow \quad \underbrace{F_{\underline{X}_n}(\underline{t})}_{\equiv P(\underline{X}_n \leq \underline{t}) \equiv P_{\underline{X}_n}((-\infty, t_1] \times \dots \times (-\infty, t_d])} \longrightarrow F_{\underline{X}}(\underline{t})$ as $n \to \infty$ pointwise at all $\underline{t}$: $F(\cdot)$ contin. at $\underline{t}$

$\Leftrightarrow \quad \varphi_{\underline{X}_n}(\underline{t}) \longrightarrow \varphi_{\underline{X}}(\underline{t})$ pointwise as $n \to \infty$ for $\varphi_{\underline{X}}(\underline{t}) \equiv \mathbb{E}(\exp(i \cdot \underline{t}'\underline{X}))$ the char. function

$\Leftrightarrow \quad \underbrace{\underline{\lambda}'\underline{X}_n}_{\in \mathbb{R}!} \xrightarrow{d} \underline{\lambda}'\underline{X} \quad \forall \underline{\lambda} \in \mathbb{R}^d$ (Cramér-Wold-Device)

> And we can use the following relations:



dominated conveg.

$(*_1)\ \exists \underline{g}: (\Omega, \mathcal{A}) \to (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ s.th.:
$\underline{g} \geq \underline{X}_n \ \forall n$ a.s.
$\wedge\ \underline{g} \in \mathcal{L}^1 \ (\Leftrightarrow \mathbb{E}(\|\underline{g}\|) < +\infty)$

$(*_2)\ \lim_{c \to 0} \sup_{n \geq 1} \mathbb{E}[\|\underline{X}_n\| \mathbb{1}\{\|\underline{X}_n\| > c\}] = 0$

uniform integrability; act. $\underline{X}_n \xrightarrow{\mathcal{L}^1} \underline{X} \Leftrightarrow \underline{X}_n \xrightarrow{p} \underline{X}$ and $(\underline{X}_n)$ unif. int.able

> It is convenient to denote convergence in the analogue of the order notation known from deterministic sequences:

Let $\underline{(a_n)_{n \in \mathbb{N}}}$ be a convergent deterministic sequence (to $a \in \mathbb{R}$ (i.p. 0) or to $+\infty$)

We define for a sequence of r.v.s, $(\underline{X}_n)_{n \in \mathbb{N}}$, to be

- Bounded in p (probability): if $\forall \varepsilon > 0, \exists M \in \mathbb{R}, N \in \mathbb{N}: \forall n \geq N \ P(\|\underline{X}_n\| \leq M) \geq 1 - \varepsilon$
  (As we consider higher sequence elements, p-mass does not 'wander off to ∞' but stays in finite regions around $\underline{0}$)
  this is actually equvt. to $(P_{\underline{X}_n})_{n \in \mathbb{N}}$ being tight!
- $o_p(a_n)$: if $\frac{1}{a_n} \cdot \underline{X}_n \xrightarrow{p} \underline{0}$ (Note: $\underline{X}_n \xrightarrow{p} \underline{0} \Leftrightarrow \underline{X}_n = o_p(1)$)
- $O_p(a_n)$: if $\frac{1}{a_n} \underline{X}_n$ is bounded in p $(\forall \varepsilon > 0\ \exists M, N: P(\|\underline{X}_n\| \leq a_n \cdot M) \geq 1 - \varepsilon \ \forall n \geq N)$
  helpful heuristic notation: $\underline{X}_n \in O_p(a_n) \Leftrightarrow$ as $n \to \infty$, $\underline{X}_n \leq a_n \cdot \underline{X}$ for some r.v. $\underline{X}$

slight abuse of "=" sign, since $\underline{X}_n = O_p(1) \wedge \underline{Y}_n = o_p(1) \not\Rightarrow \underline{X}_n = \underline{Y}_n$!

- We have a number of rules: $\underline{X}_n = o_p(a_n) \Leftrightarrow a_n^{-1}\underline{X}_n = o_p(1)$, $\underline{X}_n = O_p(a_n) \Leftrightarrow \underline{X}_n = a_n O_p(1)$

Naturally $o_p(1) \in O_p(1)$

- $b_n \cdot O_p(a_n) = O_p(b_n a_n)$ (same for $o_p(\cdot)$),

cf. PS 3 for proofs.

- $o_p(1) + o_p(1) = o_p(1)$, $O_p(1) + O_p(1) = O_p(1)$, $o_p(1) + O_p(1) = O_p(1)$,
- $o_p(1) \cdot o_p(1) = o_p(1)$, $O_p(1) \cdot O_p(1) = O_p(1)$, $o_p(1) \cdot O_p(1) = o_p(1)$

- For $a_n \to +\infty$: $O_p(a_n^{-\alpha-h}) \in O_p(a_n^{-\alpha})$ $\forall h \geq 0, \forall \alpha \in \mathbb{R}$

(e.g. $O_p(n^{-1/2}) \in O_p(n^{-1/4})$)

reminder: this means $\underline{x}_n = O_p(a_n^{-\alpha-h}) \Rightarrow \underline{x}_n = O_p(a_n^{-\alpha})$

Proof: $O_p(a_n^{-\alpha-h}) = a_n^{-\alpha-h} \cdot O_p(1) = a_n^{-\alpha} \cdot \underbrace{a_n^{-h} O_p(1)}_{= o_p(1) O_p(1) = o_p(1)}$

$\Rightarrow \; O_p(a_n^{-\alpha-h}) = O_p(a_n^{-\alpha})$

$\square$

We define, with help of these concepts, for O-p-convergent sequences the **rate of convergence**: let

$\underline{x}_n \xrightarrow{P} \underline{0}$, specifically $\underline{x}_n = O_p(a_n^{-1}) \;(= a_n^{-1} \cdot O_p(1))$ for $a_n \to +\infty$ ($a_n^{-1} = o_p(1)$ ($= o(1)$ in fact))

then: "a_n is rate of convergence of $\underline{x}_n$" $\overset{\text{def}}{\Longleftrightarrow}$ $\nexists (b_n)_n$: $b_n \to +\infty$, $\underline{x}_n = O_p(b_n^{-1})$, $b_n \to \infty$ faster than $\to$ $b_n / a_n \to +\infty$

$a_n \to \infty$; hence can 'inflate' $\underline{x}_n$ by even more than a_n and still get p-boundedness.

> Almost directly tied to these notions are several 'classic' limit theorems for sequences of random variables that, unfortunately for stochastic processes, are often required to be independent:

II.2 Recap II: limit theorems

> As might be apparent, the limit notions from II.1 (as with any limit notion) are particularly useful in approximating objects, here random variables (or rather sequences), that converge to some well-behaved limit

> The first class of limit theorems are "laws of large numbers": under the right conditions, averages ($\frac{1}{n}\sum_n \ldots$) of random variables converge (in $\mathcal{L}^1$/P/a.s.) to the (deterministic) expectation (if they don't have the same expectation, the centered r.v.s, averaged, go to 0)

> The second class are "central limit theorems": appropriately scaled averages of centered r.v.s converge (under conditions) in distribution to the standard normal distribution

> There are many more limit theorems, and even the ones from the classes presented here are not an exhaustive representation! We just look at the most important theorems.

> **Laws of large numbers:**

└ for $d=1$ (for $d>1$, analogues hold - just use component-wise!)

(standard) SLLN: $(X_n)_{n\in\mathbb{N}}$ i.i.d. in $\mathcal{L}^2(P) \Rightarrow \frac{1}{n}\sum_{i=1}^n X_i \xrightarrow{a.s.} \mathbb{E}(X_1)$

$\mathcal{L}^2$-LLN for $\neg\perp\!\!\!\perp$, $\neg$ident.: (X_n) in $\mathcal{L}^2(P)$ p.w. uncorr with $V[\frac{1}{n}\sum_i X_i] = \frac{1}{n^2}\sum_i V[X_i] \to 0 \Rightarrow \frac{1}{n}\sum_i (X_i - \mathbb{E}(X_i)) \xrightarrow{\mathcal{L}^2} 0$ (This result looks super powerful - but since $V \to 0$ is very hard to verify in practice, it's use is limited)

$\mathcal{L}^2$-LLN for $\neg\perp\!\!\!\perp$, ident (Khintchine): (X_n) in $\mathcal{L}^2(P)$ p.w. uncorr and ident. dist. $\Rightarrow P(|\frac{1}{n}\sum_i X_i - \mathbb{E}X_1| \geq \varepsilon) \leq \frac{1}{\varepsilon^2 n} V[X_1] \Rightarrow \frac{1}{n}\sum_i X_i \xrightarrow{\mathcal{L}^2} \mathbb{E}(X_1)$

SLLN for p.w. $\perp\!\!\!\perp$, ident (Etemadi): (X_n) in $\mathcal{L}^2(P)$ p.w. $\perp\!\!\!\perp$ and ident. dist. $\Rightarrow \frac{1}{n}\sum_i X_i \xrightarrow{a.s.} \mathbb{E}(X_1)$

> **Central limit theorems:** first for $d=1$, for $d>1$ use Cramér-Wold-device! (transfer premises on $X_{j,n}$ to $\underline{\lambda}'\underline{x}_n$!)

DeMoivre-Laplace / Lindeberg-Lévy: (X_n) iid in $\mathcal{L}^2(P) \Rightarrow \sqrt{n}\,\frac{\frac{1}{n}\sum_i X_i - \mathbb{E}(X_i)}{\sqrt{V(X_i)}} \xrightarrow{d} \mathcal{N}(0,1)$

Béry-Esseen (speed o. conv.): under stated conditions,

$$\forall n \in \mathbb{N}, \quad \sup_{t \in \mathbb{R}} \left| P\left(\sqrt{n} \frac{\frac{1}{n}\sum_i x_i - \mathbb{E}x_i}{\sqrt{V(x_i)}} \leq t\right) - \Phi(t) \right| < \frac{C}{\sqrt{n}} \cdot \frac{\mathbb{E}(|x_i - \mathbb{E}x_i|^3)}{\sqrt{Var(x_i)}^3}$$

for some $C \in \mathbb{R}$ (we know $C \in [0.4097, 0.4748]$ but not more)

Lyapunov: (X_n) ⫫ in $\mathcal{L}^2(P)$. For $\bar{X}_n \equiv \frac{1}{n}\sum_{i=1}^n X_i$, $s_n^2 \equiv \sum_{i=1}^n Var(X_i)$, if $\exists \delta > 0$ (in practice often use $\delta = 1$): $\frac{1}{s_n^{2+\delta}} \cdot \sum_{i \leq n} \mathbb{E}(|X_i - \mathbb{E}X_i|^{2+\delta}) \to 0$

then $\frac{1}{s_n} \sum_{i=1}^n (X_i - \mathbb{E}X_i) \xrightarrow{d} \mathcal{N}(0,1)$ — Lyapunov-cond $\Leftarrow \sum_{i=1}^n \frac{\mathbb{E}(|X_i - \mathbb{E}X_i|^3)}{s_n^3} \to 0$

Lindeberg-Feller: (X_n) ⫫ in $\mathcal{L}^2(P)$. For $\bar{X}_n, s_n^2$ as above, if $L_n(\varepsilon) \to 0$ $\forall \varepsilon > 0$ for

$$L_n(\varepsilon) \equiv \frac{1}{s_n^2} \sum_{i \leq n} \mathbb{E}\left((X_i - \mathbb{E}X_i)^2 \cdot \mathbb{1}\{|X_i - \mathbb{E}X_i| \geq \varepsilon \cdot s_n^2\}\right),$$

we have $\frac{1}{s_n} \sum_{i=1}^n (X_i - \mathbb{E}X_i) \xrightarrow{d} \mathcal{N}(0,1)$

Notice: Lyapunov-cond $\Rightarrow$ Lindeberg-condition.

Not a CLT itself, but very practical:

Δ-method: For $a_n(\underline{X}_n - \underline{\mu}) \xrightarrow{d} \underline{X}$ for some $\underline{\mu}, \underline{X}$, $a_n \to \infty$ and for $g: \mathbb{R}^d \to \mathbb{R} \in \mathcal{C}^1(U(\underline{x}))$ with $\nabla g(\underline{\mu}) \neq \underline{0}$,

$$a_n(g(\underline{X}_n) - g(\underline{\mu})) \xrightarrow{d} \nabla g(\underline{\mu})' \underline{X}$$

Proof: by 1OTA, sandwich theorem, CMT and Slutsky's lemma; here we see we can also use k^{th} OTA under the right conditions!

$$a_n(g(\underline{X}_n) - g(\underline{\mu})) = a_n \nabla g(\underbrace{\underline{\xi}_n}_{\text{interm. values}})'(\underline{X}_n - \underline{\mu}).$$

□

Looks weird? Suppose (X_n) iid. Then $\mathbb{E}X_i = \mu \ \forall i$, $V[X_i] = \sigma^2 \ \forall i$. Then

$\sum_{i=1}^n (X_i - \mathbb{E}X_i) = \sum_i X_i - n\mathbb{E}X_i$, $\quad S_n = \sqrt{\sum_i V[X_i]} = \sqrt{n}\sqrt{V[X_i]}$

$$\Rightarrow \frac{\sum_i (X_i - \mathbb{E}X_i)}{S_n} = \frac{n\left(\frac{1}{n}\sum_i X_i - \mathbb{E}X_i\right)}{\sqrt{n}\sqrt{V[X_i]}} = \sqrt{n} \frac{\left(\frac{1}{n}\sum_i X_i - \mathbb{E}X_i\right)}{\sqrt{V[X_i]}}$$

$\Rightarrow$ It's always about dividing sums of deviations by ($\sim$) sums of standard deviations!

> These results are useful as a benchmark, but in severely limited scope for stochastic processes; the next section covers limit results explicitly for Markov processes

4. Limit results for Markov-Processes

(This section is taken from RecMeth, II.5+6.)

- sect. II.5 clarifies what Markov-Processes are & some of their basic properties
- sect. II.6 builds on this & presents important limit results for Markov-Processes

Besides giving LLNs, II.6 is actually mainly concerned with characterizing $\mathcal{L}(X_t)$ for $t \to \infty$; this is useful in itself for later since most processes with which we will work can be represented as Markov processes!

5. Markov processes for Dynamic programming I: Definitions & Foundations

↳ Rd. MBMT + MBI2 + MSCE before this

> MP.s are an integral ingredient to stochastic dynamic programming: in this section, we define key building blocks of Markov process (trans. fnct. & Markov operators) and then sketch how to construct a Markov process; finally, we show which objects from dynamic programming give rise to transition functions & potentially a Markov process.

> Most basic definition needed for Markov process in discrete time is that of <u>transition function</u>; let $(Z, \mathcal{Z})$ be meas. space

$$\boxed{\begin{aligned} &Q: Z \times \mathcal{Z} \to [0,1] \text{ s.th. } (a)\ \forall z,\ Q(z,\cdot) \text{ is p-measure on } \mathcal{Z}, \\ &\text{close to Markov kernel, see MSCE, D2.1 } (b)\ \forall A,\ Q(\cdot, A) \text{ is } \mathcal{Z}\text{-measurable} \end{aligned}}$$

We have that $\forall n \in \mathbb{N}$, $Q^{n+1}(z,A) \equiv \int Q^n(z',A)\,Q(z,dz')$ is a trans. function

↳ way of reading: start at z; then $z \rightsquigarrow dz'$, then $z' \overset{Q^n}{\rightsquigarrow} A$

> And asctd with this transition function we define

Markov operator of Q:

$\sim \mathbb{E}(f|z) \to$ $$T: \underbrace{\mathcal{M}(Z,\mathcal{Z})}_{\mathcal{Z}\text{-measurable functions on } Z} \to \underbrace{\overline{\mathbb{R}}(Z)}_{\overline{\mathbb{R}}\text{-valued functions on } Z} : f(z) \mapsto (Tf)(z) \equiv \int f(z')\,Q(z,dz')$$

$\sim \mathcal{L}(z'|z) \to$ Adjoint of T:

law/distribution

$$T^*: \underbrace{\mathcal{P}(Z,\mathcal{Z})}_{\text{probability measures on } (Z,\mathcal{Z})} \to \underbrace{\mathcal{M}(Z,\mathcal{Z})}_{\text{measures on } (Z,\mathcal{Z})} : \nu(A) \mapsto (T^*\nu)(A) \equiv \int Q(z,A)\,\nu(dz)$$

<u>Intuition</u>: Tf is exp. value of f after transition started at z;
~~$T^*\nu$ is prob. that next state is in A after trans. start at z~~

not quite. $T^*\nu$ is the distrib. over $(Z,\mathcal{Z})$ that is induced when a r.v. following ν is subjected to the Markov-transition T^*

and to work with them, one can show the following properties (see SLP, Ch 8.1 for proofs)

Thm II.5.1 (Properties of Q's markov operator and its adjoint)

Let Q be a transition on $(Z,\mathcal{Z})$ some meas. space; then:

(i) $T: B(Z,\mathcal{Z}) \to B(Z,\mathcal{Z})$ — bounded $\mathcal{Z}$-measurable functions

(ii) $T^*: \mathcal{P}(Z,\mathcal{Z}) \to \mathcal{P}(Z,\mathcal{Z})$

"input stability"

(iii) <u>Linearity</u>: for $f, g \in B(Z,\mathcal{Z})$, $\alpha, \beta \in \mathbb{R}$,
$T(\alpha f + \beta g) = \alpha Tf + \beta Tg$
for $\tau, \varsigma \in \mathcal{P}(Z,\mathcal{Z})$, $\gamma \in (0,1)$,
$T^*(\gamma\tau + (1-\gamma)\varsigma) = \gamma T^*\tau + (1-\gamma)T^*\varsigma$

(iv) <u>Exchangeability</u>: for $\langle\cdot,\cdot\rangle: B(Z,\mathcal{Z}) \times \mathcal{P}(Z,\mathcal{Z}) \to \mathbb{R}: (f,\nu) \mapsto \int_Z f(z)\,\nu(dz)$
we can say
$$\langle Tf, \nu\rangle = \langle f, T^*\nu\rangle$$
or more lucidly, $\forall f \in B(Z,\mathcal{Z})$, $\forall \nu \in \mathcal{P}(Z,\mathcal{Z})$,
$$\int_Z (Tf)(z)\,\nu(dz) = \int_Z f(z)\,(T^*\nu)(dz)$$
$$\Leftrightarrow \boxed{\int_Z \int_Z f(z')\,Q(z,dz')\,\nu(dz) = \int_Z f(z') \int_Z Q(z,dz')\,\nu(dz)}$$

(v) Iterateability the operators T^n, T^{*n} corresponding to the TF $Q^n(z,A) \equiv \int Q^{n-1}(z',A)\,Q(z,dz')$ have the property that we can obtain them by stacking the Q^1-operators; more generally: $T_{Q^n} = T_Q^n$, $T_{Q^n}^* = T_Q^{*n}$ and

$$T^{n+m} = T^n \circ T^m, \quad T^{*(n+m)} = T^{*n} \circ T^{*m}$$

(in sect 2, will want to show under which conditions $\{T^{*n}\}_{n\in\mathbb{N}}$ converges to a limiting measure $\nu^* \in \mathcal{P}(Z,\mathcal{Z})$)

Proof remark: most of these properties can be shown using the 'standard machinery' of integration theory:

1) Show it holds for an indicator function
2) using 1) and linearity, show it holds for an elementary function
3) using 2) and $\exists$ of isotonically conv. sequ. of elem. functions and def. of integral, show it holds for nonnegative m.able functions
4) using 3) and $f = f^+ - f^-$ and linearity, show it holds for arb. m.able function.

may have to use monotone convergence thm. for expressions in double integral

> If we can presume that $(Z,\mathcal{Z})$ has a structure (usually the case with (Z,ρ) being metric space) we can define the following properties for T, T^*:

Feller property: Q has FP if $T_Q: \mathcal{C}_b(Z) \to \mathcal{C}_b(Z)$

bounded continuous functions on Z wrt ρ

Monotonicity: Q is monotone if $T_Q: N(Z) \to N(Z)$

non-decreasing functions on Z where Z needs to have some pre-order (like $\geq_n$ on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$)

condexp (under Markov chain) of continuous bounded function is cont. & bound.; condexp (under MC) preserves monotonicity function

> Now we can use Q to construct a measure on $\otimes_{n=1}^\infty (Z,\mathcal{Z})$ which will give us the means to define the actual Markov process!

First, we use Q to define a measure on $\otimes_{n=1}^\infty (Z,\mathcal{Z})$;

1. Step: consider the finite-product space $(Z^T, \mathcal{Z}^{\otimes T}) \equiv \otimes_{n=1}^T (Z^n, \mathcal{Z}^n)$. now define the set function

$$\mu^T(z_0,\cdot): \mathcal{R}^T \to [0,1]: R \mapsto \int_{A_1} \cdots \int_{A_{T-1}} \int_{A_T} Q(z_{T-1}, dz_T)\, Q(z_{T-2}, dz_{T-1}) \cdots Q(z_0, dz_1)$$

here we 'start' in the fixed value z_0, but we could also 'start' in some $\nu \in \mathcal{P}(Z,\mathcal{Z})$

on the set of measurable rectangles in Z^T $\mathcal{R}^T := \{ X_{n=1}^T A_n \mid A_n \in \mathcal{Z} \}$.

Then one can check that μ^T is finite and σ-additive on $\mathcal{R}^T$ which is $\cap$-stable and generates $\mathcal{Z}^T$, hence μ^T has a unique extension to $\mathcal{Z}^T$ (see MBMT, S1.2.3, "Carathéodory")

2. Step: Now for all $T \in \mathbb{N}$, define the set of finite measurable rectangles $\bar{\mathcal{R}}^T := \{ X_{n=1}^T A_n \times (X_{n=T+1}^\infty Z) \mid A_n \in \mathcal{Z} \}$ and define the set function $\bar{\mu}^T: B \in \bar{\mathcal{R}}^T \mapsto \int_{A_1} \cdots \int_{A_T} Q(z_{T-1}, dz_T) \cdots Q(z_0, dz_1)$. Again, $\bar{\mu}^T$ is a finite pre-measure on $\bar{\mathcal{R}}^T$ and hence has a unique extension to

$$\underbrace{\sigma(\bar{\mathcal{R}}^T)}_{=:\mathcal{F}_T} \equiv \{ B_T \times (X_{n=T+1}^\infty Z) \mid B_T \in \mathcal{Z}^T \}$$ which we again call $\bar{\mu}^T$

Now $\mathcal{F}_T$ is an incr. sequence of σ-a.s and hence $\mathcal{F} := \cup_{T\in\mathbb{N}} \mathcal{F}_T$ is an algebra with the property $\sigma(\mathcal{F}) = \mathcal{Z}^\infty$

3. Step: now define the set function $\overline{\mu}$ on $\mathcal{F}$ by

$$\overline{\mu}: \mathcal{F} \to [0,1]: F \mapsto \mu^T(z_0, F) \text{ for } F \in \mathcal{F}_T$$

is a tricky proof...

again, one can check that this is a finite, σ-additive set function on the algebra (thus, ring) $\mathcal{F}$ and hence has the unique extension μ to $\sigma(\mathcal{F}) = \mathcal{Z}^T$.

$\Rightarrow$ $(\times_{n=1}^{\infty} Z, \otimes_{n=1}^{\infty} \mathcal{Z}, \mu)$ is a measure space where μ maps all finite rectangles just as described in step 1.

We say the TF Q generates μ on $(\times_{n=1}^{\infty} Z, \otimes_{n=1}^{\infty} \mathcal{Z})$

Def & Prop. II.5.2 (first order Markov process)

A Markov process is a stochastic process

"Filtration"

sequence of σ-A.s on Ω: $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots \subseteq \mathcal{A}$

increas.

SP on $(\Omega, \mathcal{A}, P)$ is collection $((Z, \mathcal{Z}), (\mathcal{F}_t)_{t\in\mathbb{N}}, (X_t)_t)$

m. able space

sequence of mappings $X_t: (\Omega, \mathcal{F}_t) \to (Z, \mathcal{Z})$

with the property $\forall t \in \mathbb{N}\ \forall n \in \mathbb{N}$

$$\text{(MP)}\quad P((X_{t+1}, \ldots, X_{t+n}) \in C \mid X_t = z_t, \ldots, X_1 = z_1)$$
$$= P((X_{t+1}, \ldots, X_{t+n}) \in C \mid X_t = z_t) \quad \forall C \in \mathcal{Z}^{\otimes n}.$$

Short-hand, we just write $(X_t)_{t\geq 1}$ instead of $((Z,\mathcal{Z}), (\mathcal{F}_t)_t, (X_t)_t)$

If the probabilities in (MP) are independent of t, we say (X_t) is "time homogeneous".

For $(Z, \mathcal{Z})$ and a TF Q given, one can construct the canonical markov process as

sequence projection: $p_t: \{z_i\}_{i\in\mathbb{N}} \mapsto z_t$

Important: this is what makes the process Markovian!

$$((Z,\mathcal{Z}), (\mathcal{F}_t)_t, (X_t)_t) = ((Z,\mathcal{Z}), (\sigma(\mathcal{R}^T))_{T\geq 1}, (p_t)_{t\geq 1})$$

on $(\times_{n=1}^{\infty} Z, \otimes_{n=1}^{\infty} \mathcal{Z}, \mu)$ where μ generated by Q.

Remark:

actually slightly more involved proof needed, see SLP, Ch 8.3

- By construction μ^T has the property that for any $\mathcal{Z}^T$-measurable function $F: Z^T \to \overline{\mathbb{R}}$, we have

$$\int_{Z^T} F(z^T)\, \mu^T(z_0, dz^T) = \int_{Z^{T-1}} \int_Z F(z^{T-1}, z_T)\, Q(z_{T-1}, dz_T)\, \mu^{T-1}(z_0, dz^{T-1})$$
$$= \int_Z \int_{Z^{T-1}} F(z_1, (z_2, \ldots, z_T))\, \mu^{T-1}(z_1, d(z_2, \ldots, z_T))\, Q(z_0, dz_1)$$

> Lastly, we want to show that – building on the stochastic difference equation $z_{t+1} = g(z_t, w_t)$ for (w_t) a stochastic process – we can construct a transition function for z_t and show that it is a Markov chain!

↳ this is an integral result, since in DP we will most often work directly with AR-processes!

↳ Intuitively, Q should be such that for z_t fixed, it gives the transition probability for $z_{t+1} \in A$ by the law of w_t through giving the mass of the $w_t : g(z_t, w_t) \in A$; this is indeed the case (however the proof for measurability of $Q(\cdot, A)$ is more involved bcs it uses monotone class lemma – here more of a practical exercise – for DP only result is important)

Thm II.5.3 (from stoch. diff. equ. to Markov process)

Let $(W, \mathcal{W}, \mu)$ be a p-space and let $(W_t)_{t\geq 0}$ be a stochastic process on $(W, \mathcal{W}, \mu)^{\otimes \mathbb{N}}$, s.th. $W_t = \pi_t \; \forall t$.
— projection on tth coordinate

(i.e. state space is $(W, \mathcal{W})$, $(\sigma(\pi_t))_t$ is filtration – note that (W_t) is just iid-process on $(W, \mathcal{W})$!)

Let $(Z, \mathcal{Z})$ be a m.able space and define the process (z_t) recursively by

$$\boxed{z_{t+1} = g(z_t, w_t) \quad \forall t \in \mathbb{N}}$$ — e.g. AR(1) for z_t with $(w_t) \overset{iid}{\sim} \mathcal{N}(0,1)$

where $g : (Z \times W, \mathcal{Z} \otimes \mathcal{W}) \to (Z, \mathcal{Z})$. Then, we can define the transition function for (z_t): $\forall z \in Z, \forall A \in \mathcal{Z}$

$$Q(z, A) := \mu\left(g_z^{-1}(A)\right)$$ ← Q says: given $z_t = z$, proba to get $z_{t+1} \in A$ is equ. to getting $w_t : g(z_t, w_t) \in A$!

all outputs are $\mathcal{W}$-measurable! see proof

where $g_{(\cdot)}^{-1} : Z \times \mathcal{Z} \overset{\text{correspondence}}{\rightrightarrows} W : (z, A) \mapsto \{w : g(z, w) \in A\}$ (and $g_{(\cdot)}^{-1} = f : Z \times \mathcal{Z} \to \mathcal{W} : \ldots$)

And by Thm 5.2 we can construct (z_t) as a Markov process.

↳ i.e. can construct MC that behaves exactly like $\{z_t\}$, i.e. $\{z_t\}$ is a MC.
— by D&P II.5.2
— bcs transition prob.s are the same by def. of Q!

SLP, p.220 –222 in book

Proof: For the course of proof, define

$$\Gamma : \mathcal{Z} \rightrightarrows Z \times W : A \mapsto \{(z, w) : g(z, w) \in A\}$$

and $\forall C \in \mathcal{Z} \times \mathcal{W}$, $C_z := \{w \in W : (z, w) \in C\}$ the z-section of C.
Then, $\forall z \forall A$,

$$\boxed{g_z^{-1}(A) \equiv (\Gamma(A))_z}$$

1) Q is well-defined: $g_{(\cdot)}^{-1}(\cdot)$ is w.d. by the above, so suffices to show that g^{-1} maps to $\mathcal{W}$.
Note that by g being $\mathcal{Z} \otimes \mathcal{W}$-$\mathcal{Z}$-measurable, Γ maps to $\mathcal{Z} \otimes \mathcal{W}$ and because sections can be obtained as limits of countable intersections $(\Gamma(A))_z \in \mathcal{W}$.

2) $\forall z \in Z$, $Q(z, \cdot)$ is p-measure on $(Z, \mathcal{Z})$: $Q(z, \emptyset) = 0$, $Q(z, Z) = 1$ are obvious; remains σ-additivity: fix $\{A_n\} \subseteq \mathcal{Z}$ disjoint, then:

$$Q(z, \uplus_n A_n) \equiv \mu\left((\Gamma(\uplus_n A_n))_z\right) \overset{\text{by properties of inverse and z-section}}{=} \mu\left(\uplus_n \Gamma(A_n))_z\right)$$

$$= \Sigma_n \, \mu\left((\Gamma(A_n))_z\right) = \Sigma_n \, Q(z, A_n)$$

3) $\forall A \in \mathcal{Z}$, $Q(\cdot, A)$ is $\mathcal{Z}$-measurable this is the tricky part;

we will use the technique "principle of the good sets" on basis of the monotone class lemma (MBMT, S0.2.4)

Define

$$\mathcal{E} := \{ C \in \mathcal{Z} \otimes \mathcal{W} : z \mapsto \mu(C_z) \text{ is a } \mathcal{Z}\text{-measurable function} \}$$

we now show:

(a) $\mathcal{R} \subseteq \mathcal{Z} \times \mathcal{W}$ the system of meas. rectangles is in $\mathcal{E}$

Pf.: let $R \in \mathcal{R}$, then $\forall z$,

$$\mu(R_z) = \mu((\underbrace{A}_{\in \mathcal{Z}} \times \overbrace{B}^{\in \mathcal{W}})_z) = \mathbb{1}_A(z) \cdot \mu(B)$$

the set of inverse images to this, $\{A, A^c\}$ is in $\mathcal{Z}$ bcs A is, and hence $\sigma(\{A, A^c\}) \subseteq \mathcal{Z}$ which gives m.ability.

(b) $\overline{\mathcal{R}} \subseteq \mathcal{Z} \otimes \mathcal{W}$ the system of finite unions of rectangles is in $\mathcal{E}$

Pf.: let $\overline{R}_N = \bigcup_{n=1}^N R_n$ for $R_n \in \mathcal{R}$ and $N \in \mathbb{N}$. For $N=1$ is given by (a). Suppose it's given for $N-1$, then:

$$\mu((\overline{R}_N)_z) = \mu(\textstyle\bigcup_{n=1}^N (R_n)_z) = \mu(\bigcup_{n=1}^{N-1} (R_n)_z \cup (R_N)_z)$$

$$= \mu(\textstyle\bigcup_{n=1}^{N-1} (R_n)_z) + \mu((R_N)_z) - \mu(\underbrace{\bigcup_{n=1}^{N-1} (R_n)_z \cap (R_N)_z}_{= \bigcup_{n=1}^{N-1} (R_n \cap R_N)_z})$$

(using MBMT, S 1.1.1, (a))

and all three functions in the sum are $\mathcal{Z}$-m.able by (a), hence so is the sum. Thus, by induction, $\overline{R}_N \in \mathcal{E}$ $\forall N \in \mathbb{N}$ so $\overline{\mathcal{R}} \in \mathcal{E}$.

(c) $\overline{\mathcal{R}}$ generates $\mathcal{Z} \otimes \mathcal{W}$

by choice of $\mathcal{Z} \otimes \mathcal{W}$.

(d) $\mathcal{E}$ is a δ-system

(i) $Z \times W \in \mathcal{E}$ (trivial by $\mu(W) = 1$ $\forall z$ and constant is m.able)

(ii) $C \in \mathcal{E} \Rightarrow C^c \in \mathcal{E}$

$\mu((C^c)_z) = \mu((C)_z^c) = 1 - \mu(C_z)$ which is $\mathcal{Z}$-m.able

(iii) for $\{D_n\} \subseteq \mathcal{E}$ pw. disjoint is $\bigcup_n D_n \in \mathcal{E}$

$\mu((\bigcup_n D_n)_z) = \mu(\biguplus_n (D_n)_z) = \sum_n \mu((D_n)_z)$ which is $\mathcal{Z}$-m.able

which, together (ie. $\cap$-stable generator of $\mathcal{Z} \otimes \mathcal{W}$ in $\mathcal{E}$ and $\mathcal{E}$ a δ-system) means, by MCL (MBMT S0.2.4), that

$\mathcal{Z} \otimes \mathcal{W} = \mathcal{E}$ (notice $\mathcal{E} \subseteq \mathcal{Z} \otimes \mathcal{W}$ by construction)

ie. $\forall A \in \mathcal{Z}$, $Q(\cdot, A)$ is a $\mathcal{Z}$-m.able function.

$\Rightarrow Q(\cdot,\cdot)$ as defined is a transition function. We can construct a Markov chain with it (see D&D II.5.2).

□

6. Markov processes for Dynamic programming II: Convergence of Markov processes

2Rd. MBMT, MBI2, MSCE, WT1 ▸ 1-3, 5-8, 10.①, II.5 before this

> As with stability analysis in deterministic DP setting, it is also for stochastic DP interesting to talk about the long-run properties of the system that is implicitly defined by a SDP-setting

> In this section, we will see the most important theorems when it comes to analyzing the convergence properties of a MP — which is especially interesting before the background that the optimal transitions as implied by a SDP-problem can be viewed as a Markov process!

> First, we will clarify some terminology around Markov processes, then we will have a look at the collection of assumptions we will need to invoke, finally we will list the results from SLP89 (but not prove them) to see which kinds of results are possible

Note: in this primer we focus on results for Markov processes on an infinite state space — they will naturally hold also for MP.s on a finite (countable) state space (i.e. for Markov Chains) but for MC.s one can get results more easily — see DynMo-Script, section I. for a light introduction!

> In general throughout this section let

- $(S,\mathcal{S})$: $S \equiv \underbrace{[\underline{a},\underline{b}]}_{\equiv [a_1,b_1]\times\ldots\times[a_l,b_l]} \subsetneq \mathbb{R}^l$ compact with $\mathcal{S} \equiv \mathcal{B}([\underline{a},\underline{b}])$
  ↳ this is clearly restrictive, e.g. if we want S to be the state space in a SDP-setting (see I.4 beginning), but note that there we can sometimes go from X to a compact set WLOG

- $P: S\times\mathcal{S} \to [0,1]$ a transition function on $(S,\mathcal{S})$ with associated operators T, T^*, P^n as the n-step-trans. function

- $\mu^t(\underline{s}_0,\cdot)$ and $\mu(\underline{s}_0,\cdot)$ the measures on $(S,\mathcal{S})^{\otimes t}$ and $(S,\mathcal{S})^{\otimes\mathbb{N}}$ generated by P

- $((S,\mathcal{S}), (\sigma(\overline{\mathcal{R}}^t))_{t\geq1}, (p_t)_{t\geq1})$ the canonical Markov process on $(S^{\mathbb{N}}, \mathcal{S}^{\otimes\mathbb{N}}, \mu)$ with $p_t: (S^\infty, \sigma(\overline{\mathcal{R}}^t)) \to (S,\mathcal{S}): s \mapsto p_t(s) = \underline{s}_t$, with $s \equiv \{\underline{s}_i\}_{i\geq1}$ typical element from S^∞

↳ in short: $(\underline{s}_t)_{t\geq1}$ is a Markov process with transition P on $(S,\mathcal{S})$.

> Before starting, we clarify some concepts:

[Ergodic set] Let $\underline{s}\in S$. If for $E\in\mathcal{S}$, $P^n(\underline{s},E)=1$ $\forall n\in\mathbb{N}$ we call E consequent set of $\underline{s}$; if E is consequent to every $\underline{s}\in E$, then E is called invariant; We call E ergodic if $\nexists E' \subsetneq E$: E' is invariant; We call $S\setminus(\bigcup_i E_i)$ transient
(ergodic sets are the smallest sets that are never left, once entered)
($\bigcup_i E_i$: union of all ergodic sets; transient: a set that is left & never returned to w/ pos. prob.)

[Invariant measure] Let $\mathcal{P}(S,\mathcal{S})$ be the set of p-measures on $(S,\mathcal{S})$, then $\rho\in\mathcal{P}(S,\mathcal{S})$ is called invariant wrt the Markov process $(\underline{s}_t)_{t\geq1}$ or wrt P if $T^*\rho=\rho$ (ie ρ is fixed point of T^*)

[Dominance for (p-)measures] Let ς, τ be two (signed) measures. We say ς dominates τ, $\varsigma \geq \tau$, if

bounded, measurable, w. increasing

$$\langle f, \varsigma \rangle \geq \langle f, \tau \rangle \quad \forall f \in B_{\geq}(S, \mathfrak{S})$$

where $\langle \cdot, \cdot \rangle : \mathcal{F}_m(S,\mathfrak{S}) \times \Sigma(S,\mathfrak{S}) \to \overline{\mathbb{R}} : (f, \mu) \mapsto \int_S f \, d\mu$

measurable functions on $(S,\mathfrak{S})$; signed measures on $(S,\mathfrak{S})$

It holds that $\geq$ is a partial order on $\mathcal{P}(S,\mathfrak{S})$

[Strong convergence of measures] Let $\Sigma(S,\mathfrak{S})$ be the signed measures on $(S,\mathfrak{S})$. A sequence $(\varsigma_n)_{n\in\mathbb{N}} \subseteq \Sigma(S,\mathfrak{S})$ is said to converge strongly to some $\varsigma \in \Sigma(S,\mathfrak{S})$, $\varsigma_n \xrightarrow{s} \varsigma$, if

see SLP, p. 177ff. (in PDF)

$$\lim_{n\to\infty} \int f \, d\varsigma_n = \int f \, d\varsigma \quad \forall f \in \mathcal{C}_b(S,\mathfrak{S})$$

(this is weak convergence)

and the convergence is uniform over

$$\{f \in \mathcal{C}_b(S,\mathfrak{S}) : \|f\|_\infty \leq 1\}.$$

It can be shown that on $\Sigma(S,\mathfrak{S})$, the total variation distance

$$d_{TV}(\cdot,\cdot) : \Sigma(S,\mathfrak{S})^2 \to \mathbb{R}_+ : (\mu, \varsigma) \mapsto 2 \sup_{A\in\mathfrak{S}} |\mu(A) - \varsigma(A)|$$

metrises strong convergence and that strong convergence implies w. conv. (is obvious from def.)

conv. in d_{TV} is eqvt to uniform convergence on $\mathfrak{S}$ (directly visible from definition: $d_{TV}(\mu_n, \mu) \to 0$ $\Rightarrow$ $\mu_n(A) \to \mu(A)$ $\forall A \in \mathfrak{S}$ uniformly) bcs of sup

Furthermore, $(\mathcal{P}(S,\mathfrak{S}), d_{TV})$ is a complete metric space (this can be used for CMT-arguments, will see later) and d_{TV} induces the norm (and is induced by)

$$\|\cdot\|_{TV} : \Sigma(S,\mathfrak{S}) \to \mathbb{R}_+ : \mu \mapsto \sup_{\{A_n\} \in Part(S,\mathfrak{S})} \sum_n |\mu(A_n)|$$

set of measurable disjoint partitions of S

and $(\Sigma(S,\mathfrak{S}), \|\cdot\|_{TV})$ is a normed space.

> Now we have all the terminology we need; we will now look at the assumptions potentially needed – these are sometimes not really intuitive-looking, but esp. the latter ones are not too hard to check in practice

Ass. II.6.1 (Assumptions on Markov processes for convergence)

finite measures on $(S,\mathfrak{S})$

(A1) Doeblin's condition: $\exists \phi \in M_f(S,\mathfrak{S}),\ N \in \mathbb{N},\ \varepsilon > 0 : \ \forall A \in \mathfrak{S}, \forall \underline{s} \in S,$

$$\phi(A) \leq \varepsilon \Rightarrow P^N(\underline{s}, A) \leq 1 - \varepsilon$$

(we can always find finite measure ϕ, an integer N and an ε s.th. a small ϕ measure of a set tells us that an N-step ($\Rightarrow \geq N$-step) transition into this set is not completely certain, regardless of the starting point)

(A2) Condition M: $\exists \varepsilon > 0, N \in \mathbb{N} : \forall A \in \mathfrak{S}, \forall \underline{s} \in S,\ P^N(\underline{s}, A) \geq \varepsilon \vee P^N(\underline{s}, A^c) \geq \varepsilon$

(despite appearance, Cond. M is not always satisfied: imagine

$P^N(\underline{s}, A) = P^N(\underline{s}, A^c) = 0.5 \wedge \varepsilon = 0.9$) from cond. M

(this condition will be needed to establish that T^{*N} is a contraction on $(\mathcal{P}(S,\mathcal{S}), d_{TV})$)

suffic. to give $\xrightarrow{w}$ and LLN! together with $\boxed{S \subseteq \mathbb{R}^l \text{ compact}}$

(A3) P has the Feller property: $T: \mathcal{C}_b(S) \to \mathcal{C}_b(S)$
This is familiar from I.5; what's new is the triplet of equivalent criteria for checking the Feller property
We have $(1) \Leftrightarrow (2) \Leftrightarrow (3)$ for:

(1) $T: \mathcal{F}_m(S,\mathcal{S}) \to \mathcal{F}_m(S,\mathcal{S}): f(\underline{s}) \mapsto (Tf)(\underline{s}) := \int_S f(\underline{s}') P(\underline{s}, d\underline{s}')$ maps $\mathcal{C}_b(S)$ into $\mathcal{C}_b(S)$, sometimes written $T(\mathcal{C}_b(S)) \subseteq \mathcal{C}_b(S)$

(2) $\underline{s}_n \to \underline{s} \Rightarrow P(\underline{s}_n, \cdot) \xrightarrow{w} P(\underline{s}, \cdot)$ (directly from continuity of Tf and definition of $\xrightarrow{w}$)

(3) $\nu_n \xrightarrow{w} \nu \Rightarrow T^*\nu_n \xrightarrow{w} T^*\nu$ (by Thm II.5.1, (iv))

(A4) P is monotone: $T(\mathcal{F}_{m,\geq}(S,\mathcal{S})) \subseteq \mathcal{F}_{m,\geq}(S,\mathcal{S})$ (1)
non-decreasing functions

$\Leftrightarrow$ (2) for $\rho, \mu \in \mathcal{P}(S,\mathcal{S})$: $\mu \geq \rho$ we have $T^*\mu \geq T^*\rho$ (i.e. $\langle f, T^*\mu\rangle \geq \langle f, T^*\rho\rangle$
$\Leftrightarrow \langle Tf, \mu\rangle \geq \langle Tf, \rho\rangle$)

$\Leftrightarrow$ (3) $\forall \underline{s}, \underline{s}' \in S: \underline{s} \geq \underline{s}'$, $P(\underline{s}, \cdot) \geq P(\underline{s}', \cdot)$

(A5) P has Mixing-property: $\exists \underline{c} \in S \equiv [\underline{a}, \underline{b}]$, $\varepsilon > 0$, $N \in \mathbb{N}$:

$$P^N(\underline{a}, [\underline{c}, \underline{b}]) \geq \varepsilon \wedge P^N(\underline{b}, [\underline{a}, \underline{c}]) \geq \varepsilon$$

(intuitively, mixing ensures that there is enough mobility in the stochastic transitions, guarantees uniqueness of ergodic set)

again: all relevant objects are defined at beginning of chapter!

> Find here now the compiled list of results on the convergence of Markov processes from SLP89, all w/o proof:

Thm II.6.2 (Existence of invariant measures; conveg. of avg. prob.s)
T11.9 p. 182 SLP

Suppose P satisfies Ass II.6.1, (A1) (Doeblin's condition) for some (ϕ, N, ε); Then we have:

(a) S can be partitioned into a transient set and $1 \leq M \leq \phi(S)/\varepsilon$ ergodic sets

(b) $\forall \nu_0 \in \mathcal{P}(S,\mathcal{S})$, $\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^N T^{*n}\nu_0$ exists and is an invariant measure of T^*
(i.e. average probabilities converge strongly to some inv. measure)

(c) There is one invar. m. for each ergodic set and every inv. m. of T^* can be written as convex combination of these

Thm II.6.3 (Ex. & uniqueness of inv. measure; conv. of avg. prob.s)
T11.10 p. 183 SLP

Assume Ass. II.6.1, (A1) holds for some (ϕ, N, ε) and additionally, if $A: \phi(A) > 0$ then $\forall \underline{s} \exists n \in \mathbb{N}: P^n(\underline{s}, A) > 0$ (this is close in spirit to irreducibility, see DynMo, I.4, D4.4); Then:

(a) $\exists!$ ergodic set $E \subseteq S$

(b) $\exists!$ inv. meas. to T^*, call it ν^*

(c) $\lim_N \frac{1}{N} \sum_{n=1}^N T^{*n}\nu_0 = \nu^*$ strongly $\forall \nu_0 \in \mathcal{P}(S,\mathcal{S})$

Thm II.6.4 (T^* as contraction; strong conv. of measures)

L11.11+T11.12, p 183 SLP

Assume Ass. II.6.1, (A2) holds (P satisfies condition M) for some $N \in \mathbb{N}$, $\varepsilon > 0$; then,

T^{*N} is a contraction on $(\mathcal{P}(S,\mathcal{S}), d_{TV})$ of modulus $(1-\varepsilon)$ and $\forall \nu_0 \in \mathcal{P}(S,\mathcal{S})$, $\nu_k \equiv T^{*Nk}\nu_0$ converges strongly ($\Rightarrow$ weakly) to the unique invariant measure ν^*

The converse (T^{*N} a contraction $\Rightarrow$ Cond M for N, ε) holds also.
Note: this is a quite nice result: we have both uniqueness of ν^* and convergence (strongly!) to it, regardless of the starting measure)
cannot say anything about a LLN at this stage

broadening of thm 6.2 →

Thm II.6.5 (Existence of inv. m. under weaker condition)

T12.10, p197 SLP

Assume Ass. II.6.1, (A3) (P has Feller prop) holds; then for $S \subseteq \mathbb{R}^l$ compact (Thm.s 6.2-6.4 hold for arbitrary meas. space $(S,\mathcal{S})$)

$$\exists \nu^* \in \mathcal{P}(S,\mathcal{S}) : T^*\nu^* = \nu^*$$

From here on, $S \subseteq \mathbb{R}^l$ compact.

broadening of thm 6.4 →

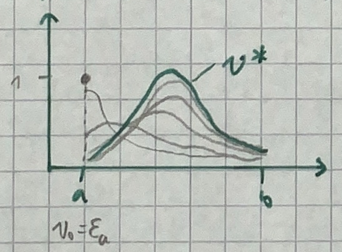
Thm II.6.6 (Ex. & uniqueness + weak convergence under weaker conditions)

T12.12, p200 SLP

Assume $S \subseteq \mathbb{R}^l$ compact and Ass. II.6.1, (A3), (A4), (A5) (P has Feller, is monotone and satisfies mixing) hold; Then:

(a) $\exists!\ \nu^* \in \mathcal{P}(S,\mathcal{S}) : T^*\nu^* = \nu^*$ and
(b) $\forall \nu_0 \in \mathcal{P}(S,\mathcal{S})$, $T^{*n}\nu_0 \xrightarrow{w} \nu^*$
weak convergence

Graphically (only sketch!)



ν^* can be interpreted as the long-run unconditional distribution of s_t that P produces

Thm II.6.7 (Law of large numbers for $(s_t)_{t\geq 1}$)

T14.7, p221 SLP

Assume $S \subseteq \mathbb{R}^l$ compact and Ass. II.6.1, (A3) (Feller) holds and $\exists \nu^*$ invariant s.th. $\forall \nu_0 \in \mathcal{P}(S,\mathcal{S})$, $\lim_N \frac{1}{N}\sum_n T^{*n}\nu_0 = \nu^*$ weakly (e.g. (A4)+(A5) hold add. so that T II.6.6 holds);
Then

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} f(\sigma_t(s)) = \int f \, d\nu^* \quad \mu(s_0,\cdot)\text{-a.s. } \forall s_0 \in S \ \forall f \in \mathcal{C}(S)$$

continuous functions $S \to \mathbb{R}$

or more commonly

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} f(s_t) = \mathbb{E}_{\nu^*}[f] \quad \text{a.s.}$$

5. Limit results for general processes

Source: Hamilton, Ch. 7

> The first result we already encountered in section 1:

THM 5.0 (Ergodic theorem). Consider (X_t) strictly stationary and ergodic in $\mathcal{L}^1(\mathbb{P})$. Then,

$$\frac{1}{T}\sum_{t=1}^{T} X_t \xrightarrow[T\to\infty]{a.s.} \mathbb{E}[X_1]$$

Proof: Cf. WT1, Satz 11.4. or Klenke "Probability theory", Ch 20.

> This theorem is useful in conjunction with THM 1.4: measurable transformations preserve strict stationarity & ergodicity

> The next result is very useful, since it is very general.

THM 5.1 (LLN for weakly stationary processes). Let $(X_t)_{t\in\mathbb{Z}}$ be weakly stationary, and define

$$\mathbb{E}(X_t) \equiv \mu, \quad \mathbb{E}[(X_t-\mu)(X_{t-j}-\mu)] \equiv \gamma_j.$$

Provided we have $\sum_{j\geq 0}|\gamma_j| < +\infty$, it is

actually, $|\gamma_j| \to 0$ sufft.!

$$\frac{1}{T}\sum_{t=1}^{T} X_t \xrightarrow{\mathcal{L}^2} \mu \quad \left(\text{and } \operatorname{Avar}\left(\tfrac{1}{T}\textstyle\sum_t X_t\right) := \lim_{T\to\infty} \operatorname{Var}\left(\sqrt{T}\,\tfrac{1}{T}\textstyle\sum_t X_t\right) = \displaystyle\sum_{j\in\mathbb{Z}} \gamma_j.\right)$$

Proof. We can show the result by verifying that

$$\mathbb{E}\left[\left|\tfrac{1}{T}\textstyle\sum_t X_t - \mu\right|^2\right] \to 0 \text{ as } T\to\infty.$$

To this end, compute

$$\mathbb{E}\left[\left|\tfrac{1}{T}\textstyle\sum_t X_t - \mu\right|^2\right] = \mathbb{E}\left[\left(\tfrac{1}{T}\textstyle\sum_t (X_t-\mu)\right)^2\right] = \frac{1}{T^2}\mathbb{E}\left[\left(\textstyle\sum_t X_t-\mu\right)\left(\textstyle\sum_t X_t-\mu\right)\right]$$

$$= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{T}\underbrace{\mathbb{E}[(X_t-\mu)\cdot(X_j-\mu)]}_{=t-(t-j)} = \frac{1}{T^2}\sum_{t=1}^{T}\sum_{j=1}^{T}\gamma_{t-j}$$

$t=1$: $\gamma_0+\gamma_1+\ldots+\gamma_{T-1}$
$t=2$: $\gamma_1+\gamma_0+\gamma_1+\ldots+\gamma_{T-2}$
$t=3$: $\gamma_2+\gamma_1+\gamma_0+\gamma_1+\ldots+\gamma_{T-3}$
$\vdots$
$t=T$: $\gamma_{T-1}+\ldots+\gamma_0$

$$= \frac{1}{T^2}\left(T\cdot\gamma_0 + 2\sum_{j=1}^{T-1}(T-j)\gamma_j\right).$$

note: $\gamma_{-h}=\gamma_h \ \forall h\in\mathbb{N}$

using only $|\gamma_j|\to 0$:

$$\operatorname{Var}\left(\tfrac{1}{T}\textstyle\sum_t X_t\right) \leq \frac{\gamma_0}{T} + \frac{2}{T}\sum_{j=1}^{T}|\gamma_j| \to 0$$

using $a_j\to 0 \Rightarrow \frac{1}{T}\sum_{j=1}^{T} a_j \to 0$

Pf: $\left|\frac{1}{T}\sum_t a_t\right| < \frac{1}{T}\sum_{t=1}^{N_\varepsilon}|a_t| + \frac{1}{T}\sum_{t=N_\varepsilon}^{T}\varepsilon/2 \leq \underbrace{\frac{N_\varepsilon}{T}}_{\to 0}\sup_{t\leq N_\varepsilon}|a_t| + \underbrace{\frac{T-N_\varepsilon}{T}}_{\to 1}\cdot\varepsilon/2 \xrightarrow[T\to\infty]{} \varepsilon/2.$ □

$\forall j>N_\varepsilon,\ |a_j|<\varepsilon/2$ — exists by $a_j\to 0$

Now observe:

$$T\cdot\mathbb{E}\left[\left(\tfrac{1}{T}\textstyle\sum_t X_t-\mu\right)^2\right] = \gamma_0 + 2\sum_{j=1}^{T-1}\underbrace{\frac{T-j}{T}}_{\leq 1}\gamma_j \leq 2\sum_{j\geq 0}|\gamma_j| < +\infty \quad \text{by premise}$$

$\forall T\in\mathbb{N}$.

Hence, $\mathbb{E}\left[\left(\tfrac{1}{T}\sum_t X_t-\mu\right)^2\right] = O(T)$ and the claim follows. □

Note: see Hamilton, p.187 (101 in PDF) for explicit limit $T\cdot\mathbb{E}\left[\left(\tfrac{1}{T}\sum_t X_t-\mu\right)^2\right] \to \sum_{j\in\mathbb{Z}}\gamma_j$. □

Or Cf. E806—Exam 2018, Problem 3

> Since mostly we operate with covariance-stationary processes, this LLN is mostly sufficient for our purposes; it is good to note, though, that there exist LLNs even when we take away weak stationarity.

DEF & THM 5.2 ($\mathcal{L}^1$-Mixingales & LLN). Consider a process $(X_t) \in \mathcal{L}^1(\mathbb{P})$ with $E(X_t) = 0\ \forall t$ and a filtration $(\mathcal{F}_t)_t$ to which (X_t) is adapted. We call (X_t) an $\mathcal{L}^1$-Mixingale if $\exists (c_t)_{t\geq 1}, (\xi_m)_{m\geq 0} \in \mathbb{R}^{\mathbb{N}_0}$ ← not for c..., with $\xi_m \to 0$ s.th. $\forall t \geq 1, m \geq 0$:

$$\mathbb{E}\left|\mathbb{E}(X_t|\mathcal{F}_{t-m})\right| \leq c_t \xi_m. \quad (\to 0 \text{ as } m\to\infty)$$

(Intuitively: the forecasts of X_t using info in $t-m$ converge to 0, the uncondit. expect., in $\mathcal{L}^1$ as $m\to\infty$; e.g. a stable AR(p) started at the nonergodic distrib., but with zero mean is an $\mathcal{L}^1$-Mixingale)

Now if

(a) $(X_t)_{t\in\mathbb{N}_0}$ is uniformly int.able $\left(\lim_{\alpha\to\infty} \sup_{t\in\mathbb{N}_0} E(|X_t| \mathbb{1}\{|X_t|\geq\alpha\}) = 0\right.$

Sufficient conditions for uniform integrability (Prop. 7.7)

(a) Suppose there exist an $r>1$ and an $M'<\infty$ such that $E(|y_t|^r) < M'$ for all t. Then, $\{y_t\}$ is uniformly integrable.

(b) Suppose there exist an $r>1$ and an $M'<\infty$ such that $E(|x_t|^r) < M'$ for all t. If $y_t = \sum_{j=-\infty}^{\infty} h_j x_{t-j}$ with $\sum_{j=-\infty}^{\infty}|h_j| < \infty$, then $\{y_t\}$ is uniformly integrable.

$\left.\right)$

↑ P7.7 in Hamilton

and

(b) $\exists (c_t)_{t\geq 1}$ for the above s.th. $\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^T c_t < +\infty$

Then $\frac{1}{T}\sum_{t=1}^T X_t \xrightarrow{P} 0.\ (= EX_t)$

Remark: For $X_t \equiv \sum_{j\geq 0} \psi_j \varepsilon_{t-j}$, (ε_t) iid in $\mathcal{L}^r(\mathbb{P})$, $r>2$, $\sum_{j\geq 0}|\psi_j| < +\infty$, we can show that $\forall k \in \mathbb{N}_0$, the process $(X_t X_{t-k} - \mathbb{E}(X_t X_{t-k}))$ satisfies the above conditions and thus

(for (ψ_j) abs. summable & (ε_t) w.stat., this series exists as $\mathcal{L}^2$-limit, cf. MS: ARMA.)

$$\frac{1}{T}\sum_t X_t X_{t-k} \xrightarrow{P} \mathbb{E}(X_t X_{t-k}).$$

Cf. Hamilton, p. 192 (104 in PDF).

> As usual, LLNs are only one part of asymptotic analysis — the other part are CLTs; again, there is quite a number of CLTs for stoch. processes, here we cover just a few that are quite useful.

THM 5.3 (CLT for MDS). Let (X_t) be a martingale difference sequence wrt some $(\mathcal{F}_t)$. Suppose that

In particular, $(\mathcal{F}_t)$ adapt. to (X_t): $\sigma(X_{t-s}, s\geq 0) \subseteq \mathcal{F}_t\ \forall t$

(a) $(X_t) \in \mathcal{L}^r(\mathbb{P})$, for some $r > 2$, and

(b) $\frac{1}{T}\sum_t \mathbb{E}(X_t^2) \longrightarrow \sigma^2 > 0$ for some σ^2, and

(c) $\frac{1}{T}\sum_t X_t^2 \xrightarrow{P} \sigma^2$.

Then, $\sqrt{T}\cdot\frac{1}{T}\sum_{t=1}^T X_t \xrightarrow{d} \mathcal{N}(0,\sigma^2)$.

E823-Ch2-sl.17

- Let $\{\boldsymbol{g}_t, \mathscr{G}_t\}$ be a n-dimensional MDS with
  - (a) $E(\boldsymbol{g}_t\boldsymbol{g}_t') = \Sigma_t$, where Σ_t is a positive definite matrix with $(1/T)\sum_{t=1}^T \Sigma_t \overset{T\to\infty}{\to} \Sigma$, a positive definite matrix
  - (b) $E|g_{it}g_{jt}g_{lt}g_{mt}| < \infty\ \forall t$ and all i,j,l, and m, where g_{it} is the i-th element of the vector $\boldsymbol{g}_t$
  - (c) $(1/T)\sum_{t=1}^T \boldsymbol{g}_t\boldsymbol{g}_t' \xrightarrow{p} \Sigma$
- Then,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^T \boldsymbol{g}_t \xrightarrow{d} N(0,\Sigma)$$

$\Leftarrow \underline{g}_t \in \mathcal{L}^4(\mathbb{P})$ $\overset{\text{def}}{\Leftrightarrow} \|\underline{g}_t\|^4_{\mathcal{L}^4} := \mathbb{E}(\|\underline{g}_t\|_4^4) = \mathbb{E}(\sum_i [\underline{g}_t]_i^4) < +\infty$ iff $\|[\underline{g}_t]_i\|^4_{\mathcal{L}^4} < +\infty$

Hölder: $\leq (\mathbb{E}|g_{it}|^4)^{1/4} \cdots (\mathbb{E}|g_{mt}|^4)^{1/4}$

Proof: White (1984): "Asymptotic Theory for Econometricians.", p. 130.

Remark: This result generalizes to a vector-valued $(\underline{X}_t)$ MDS [by showing that $\forall \underline{\lambda}$, $(\underline{\lambda}^T \underline{X}_t)$ satisfies (a)-(c) above, given vector equivalents to (a)-(c) for $(\underline{X}_t)$, and using the Cramér-Wold-device (see above, sect. 2)].

$$\sqrt{T}\,\frac{1}{T}\sum_{t=1}^T \underline{X}_t \xrightarrow{d} \mathcal{N}(\underline{0}, \underset{\sim}{\Sigma}),\quad \underset{\sim}{\Sigma} := \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T \mathbb{E}\,\underline{X}_t\underline{X}_t'.$$

THM 5.4 (CLT for iid-driven MA(∞)). Let (ε_t) iid $\in \mathcal{L}^2(\mathbb{P})$ and let $(\psi_j) \in \mathbb{R}^{\mathbb{N}_0}$ s.th. $\sum_{j\geq 0} |\psi_j| < +\infty$, and define $\forall t \in \mathbb{Z}$,

$$X_t = \underbrace{\mu}_{\in \mathbb{R}} + \sum_{j\geq 0} \psi_j \varepsilon_{t-j}. \qquad (\Rightarrow (X_t) \text{ is w. stat., can apply THM 5.1})$$

Then, $\sqrt{T}\left(\frac{1}{n}\sum_t X_t - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sum_{j\in\mathbb{Z}} \gamma_j\right)$ where $\forall j$ $\gamma_j = \mathbb{E}(X_t X_{t-j})$.

Proof: Anderson (1971): "The Statistical Analysis of Time series." p. 429.

THM 5.5 (Gordin's CLT for str. stationary & ergodic processes). Consider a process $(X_t)_{t\in\mathbb{Z}}$ that is strictly stationary and ergodic, and suppose it satisfies

GORDIN'S CONDITION

(a) $(X_t) \in \mathcal{L}^2(\mathbb{P})$

(b) $\mathbb{E}_{t-j}(X_t) := \mathbb{E}(X_t \mid (X_{t-h})_{h\geq j}) \xrightarrow{\mathcal{L}^2} 0$ as $j \to \infty$

(forecasts disappear in $\mathcal{L}^2$ as horizon increases)

(c) $\sum_{j\geq 0} \|r_{t,j}\|_{\mathcal{L}^2} < +\infty$ $\forall t \in \mathbb{Z}$, where $r_{t,j} := (\mathbb{E}_{t-j} - \mathbb{E}_{t-j-1})(X_t)$

($\Rightarrow \forall t$, $\sum_{j\geq 0} r_{t,j} = \sum_{j\geq 0}(\mathbb{E}_{t-j}X_t - \mathbb{E}_{t-j-1}X_t) \overset{\text{by rearrangement of } (\mathcal{L}^2\text{-}) \text{ convergent series.}}{=} X_t$)

(forecast updates are $\mathcal{L}^2$-summable.)

Then, we have:

(i) $\mathbb{E}X_t = 0$

(ii) $\sum_{h\in\mathbb{Z}} |\gamma_h| < +\infty$ for $\gamma_h := \mathbb{E}(X_t X_{t-h})$

(iii) $\sqrt{T}\,\frac{1}{T}\sum_{t=1}^{T} X_t \xrightarrow{d} \mathcal{N}\left(0, \underline{\sum_{h\in\mathbb{Z}} \gamma_h}\right)$ cf. remark 2

Remark: using Cramér-Wold-device, it directly follows that we get corresponding CLT for vector-valued process $(\underline{X}_t)$; (a)-(c) now read:

(a) $(\underline{X}_t) \in \mathcal{L}^2(\mathbb{P})$ $\left(\overset{\text{def}}{\Leftrightarrow} \forall t \left(\mathbb{E}\|\underline{X}_t\|^2\right)^{1/2} < +\infty\right)$,
Eucl. norm, $\|\underline{x}\| = (\underline{x}^T\underline{x})^{1/2}$

(b) $\mathbb{E}_{t-j}\underline{X}_t \xrightarrow{\mathcal{L}^2} \underline{0}$ for $j \to \infty$ $\forall t \in \mathbb{Z}$,

(c) $\sum_{j\geq 0} \|\underline{r}_{t,j}\|_{\mathcal{L}^2} < +\infty$,

then $\sqrt{T} \cdot \frac{1}{T}\sum_{t=1}^{T} \underline{X}_t \xrightarrow{d} \mathcal{N}\left(\underline{0}, \sum_{h\in\mathbb{Z}} \underline{\Gamma}_h\right)$, $\underline{\Gamma}_h := \mathbb{E}(\underline{X}_t \underline{X}_{t-h}^T)$.

THM 5.6 (CLT composition). (cf. Brockwell & Davis Prop. 6.3.9, pp. 221 in PDF)

For $(\underline{X}_t)_{n\in\mathbb{N}}$, $(\underline{Y}_{n,t})_{n,t\in\mathbb{N}}$, $\underline{X}$ random vectors,

$$\left.\begin{array}{l} \text{(i) } \underline{Y}_{n,t} \xrightarrow[t\to\infty]{d} \underline{Y}_n \ \forall n \in \mathbb{N} \\ \text{(ii) } \underline{Y}_n \xrightarrow[n\to\infty]{d} \underline{X} \\ \text{(iii) } \lim_{n\to\infty} \limsup_{t\to\infty} \mathbb{P}(|\underline{X}_t - \underline{Y}_{n,t}| > \varepsilon) = 0 \ \forall \varepsilon > 0 \end{array}\right\} \Rightarrow \underline{X}_t \xrightarrow[t\to\infty]{d} \underline{X}.$$