Random variable: measurable function $(\Omega, \mathcal{F}) \mapsto (\mathbb{R}^m, \mathcal{B})$

What is time series? Time series is a sequence of random variables

$(Y_1(w), Y_2(w), Y_3(w), \dots, Y_n(w)) \in \mathbb{R}^{m\times n}$

It can be extended to a doubly infinite sequence $(\cdots, Y_{t-1}, Y_t, Y_{t+1}, \cdots) \in \mathbb{R}^{m\times\infty}$

Here is discrete time series (vs. continuous time series).

* for each fixed w, the sequence is deterministic vector $(Y_1, Y_2, \dots, Y_n) \in \mathbb{R}^{m\times n}$
* for each fixed t, $Y_t(w)$ is a random vector in $\mathbb{R}^m$

In reality, we have only one realized sequence, but statistics needs repeated observations. We introduce the concept "stationarity".

* Def. (Y_t) is <u>covariance stationary</u> or <u>weakly stationary</u> if

  The mean $\mu = E(Y_t)$

  covariance $\Sigma = E[(Y_t - \mu)(Y_t - \mu)']$

  and autocovariance $\Gamma(k) = E[(Y_t - \mu)(Y_{t-k} - \mu)']$

  is indep of t

* if weakly stationary, $\Gamma(-k) = E[(Y_t - \mu)(Y_{t+k} - \mu)']$

$$= E[(Y_{t-k} - \mu)(Y_t - \mu)']$$

$$= \Gamma(k)'$$

e.g. To see asymmetry, consider $\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} u_t + v_{t-1} \\ v_t \end{pmatrix}$, where $\begin{pmatrix} u_t \\ v_t \end{pmatrix} \overset{iid}{\sim} N(0, I_2)$

$\Gamma(1) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$

$\Gamma(-1) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$

$\Sigma = \Gamma(0)$ is symmetric, but $\Gamma(k), k \neq 0$ is not symmetric in general.

* When $m = 1$, we use $\gamma(0), \gamma(1), \dots,$

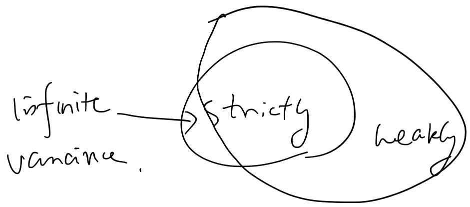
  autocorrelation $\rho(k) = \gamma(k)/\gamma(0) \in (-1, 1)$

  by Cauchy-Schwarz inequality

* (Y_t) is strictly stationary, if for each $\ell \in \mathbb{Z}^+$, the joint distributions of $(Y_t, Y_{t+1}, \dots, Y_{t+\ell})$ is indep. of t.

of $(Y_t, Y_{t+1}, \dots, Y_{t+\ell})$ is indep. of t.

* When mention "stationary", the default is "strictly stationary".



* if (Y_t) is i.i.d, then it is strictly stationary
* if (Y_t) is strictly stationary, it transformations $X_t \in \phi(Y_t, Y_{t-1}, \dots) \in R^q$ is also strictly stationary. In other words, strict stationarity is preserved by transformation.

Series: $X_t = \sum_{j=0}^{\infty} a_j Y_{t-j}$

* The infinite series X_t is convergent if the partial sum $\sum_{j=1}^{N} a_j Y_{t-j}$ has a finite limit as $N \to \infty$. almost surely.

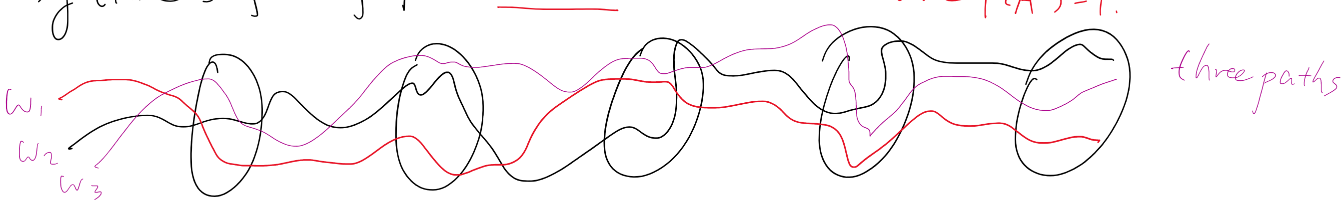
* If Y_t is strictly stationary, $E\|Y\| < \infty$ and $\sum_{j=0}^{\infty} |a_j| < \infty$ (absolutely summable), then X_t is convergent and strictly stationary.

Ergodicity

"invariant" mean the sequence of r.v. gets stuck somewhere.

* $\{Y_t\}$ is ergodic if all invariant events are trivial. Any event unaffected by time shift is of prob 0 or 1.

Such invariant cannot happen (if $P(A)=0$), and in the mean time $P(A^c)=1$.



* Ergodicity is preserved by transformation

  if $\{Y_t\}$ is stationary and ergodic, the same is for $X_t = \phi(Y_t, Y_{t-1}, \dots)$

  (function with infinite terms)

  e.g. $X_t = \sum_{j=1}^{\infty} a_j Y_{t-j}$ if convergent

(Cesaro means) If $a_j \to a$ as $j \to \infty$ then $\frac{1}{n}\sum a_j \to a$ as $n \to \infty$

(Cesaro means) if $a_j \to a$ as $j \to \infty$, then $\frac{1}{n}\sum_{j=1}^{n} a_j \to a$ as $n \to \infty$

* Thm: if $Y_t \in \mathbb{R}^m$ is stationary and ergodic, and $var(Y_t) < \infty$, then

$$\frac{1}{n}\sum_{\ell=1}^{n} Cov(Y_t, Y_{t+\ell}) \longrightarrow 0 \quad \text{as } n \to \infty$$

Text book provides a proof for discrete r.v. Continuous r.v. is more complicated.

* Formal definition:

Let $\tilde{Y}_t = (\cdots, Y_{t-1}, Y_t, Y_{t+1}, \cdots)$. an event $A = \{\tilde{Y}_t \in G\}$ for some $G \subseteq \mathbb{R}^{m\infty}$.

The ℓ-th time shift is $\tilde{Y}_{t+\ell} = (\cdots, Y_{t-1+\ell}, Y_{t+\ell}, Y_{t+\ell+1}, \cdots)$ and a time shift of the event is $A_\ell = \{\tilde{Y}_{t+\ell} \in G\}$.

An event is invariant if $A_\ell = A$

An event is trivial if $P(A) = 0$ or $P(A) = 1$

Thm: A stationary $\{Y_t\}$ is ergodic iff for all events A and B,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{\ell=1}^{n} P(A_\ell \cap B) = P(A)P(B)$$

Let $B = A$, and then we solve $P(A) = [P(A)]^2 \Rightarrow P(A) = 0$ or 1.

A "sufficient" condition for ergodicity is $P(A_\ell \cap B) \longrightarrow P(A)P(B)$ as $\ell \to \infty$, according to Cesaro means. This sufficient condition is called "mixing".

* Mixing says that separate events (any A and B) are asy. indep when A is shifted to A_ℓ as $\ell \to \infty$. (week dependence)

* Ergodicity is slightly weaker than mixing, in the sense that the indp. is "on averge" in the form of $\frac{1}{n}\sum_{\ell=1}^{n} P(A_\ell \cap B)$.

* Ergodic Theorem:

$Y_t \in \mathbb{R}^m$ is stationary & ergodic, and $E\|Y\| < \infty$, then

$E\|\bar{Y} - \mu\| \to 0$ and $\bar{Y} \xrightarrow{p} \mu$.

Interpretation: Convergence in the 1st mean implies $\xrightarrow{p}$.

Information set

* for a univariate time series, def. $E_{t-1}[Y_t] = E[Y_t \mid Y_{t-1}, Y_{t-2}, \cdots]$

* for a univariate time series, def. $E_{t-1}[Y_t] = E[Y_t \mid Y_{t-1}, Y_{t-2}, \dots]$ as the conditional expectation of Y_t given the past history $(Y_{t-1}, Y_{t-2}, \dots)$
* More generally, we write $\mathcal{F}_t$ as the smallest σ-field generated by the information up to time t. $\mathcal{F}_t$ is called an "information set".

$$E(Y_t \mid \mathcal{F}_{t-1}] = E_{t-1}[Y_t]$$

* Information sets are nested. $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t \subseteq \mathcal{F}_{t+1}, \dots$
* Depends on the definition, when multiple r.v. are involved,

$$\sigma(Y_t, Y_{t-1}, \dots) \neq \sigma(Y_t, X_t, Y_{t-1}, X_{t-1}, \dots)$$

Martingale difference sequence (MDS)

* let $\{e_t\}$ be a time series, and $\mathcal{F}_t$ be an information set.

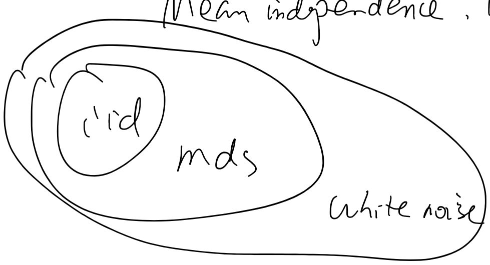
$\{e_t\}$ is adapted to $\mathcal{F}_t$ if $E(e_t \mid \mathcal{F}_t) = e_t$ ($\mathcal{F}_t$ contain the complete information of e_t. A natural filteration is $\mathcal{F}_t = \sigma(e_t, e_{t-1}, \dots)$

* MDS: a process $\{e_t, \mathcal{F}_t\}$ is mds if

    1. e_t is adapted to $\mathcal{F}_t$
    2. $E|e_t| < \infty$
    3. $E(e_t \mid \mathcal{F}_{t-1}] = 0$

Interpretation: unforeseeable.

Mean independence. But it doesn't rule out predictability in other moments.



eg. $e_t = u_t u_{t-1}$, $u_t \sim$ iid $N(0,1)$

the fiteration here is $\mathcal{F}_t = \sigma(u_t, u_{t-1}, \dots)$, which subsumes $\sigma(e_t, e_{t-1}, \dots)$

e_t is mds, but not iid. the covariance of e_t^2 and e_{t-1}^2 is not 0.

$$\mathrm{cov}(e_t, e_{t-k}) = E(e_t e_{t-k}) = E[E(e_t e_{t-k} \mid \mathcal{F}_{t-1})]$$
$$= E[e_{t-k} E(e_t \mid \mathcal{F}_{t-1})] = 0$$

* A mds $(e_t, \mathcal{F}_t)$ is a homoskedastic martingale diff. sequence if $E[e_t^2 \mid \mathcal{F}_{t-1}] = \sigma^2$.

* A mds $(e_t, \mathcal{F}_t)$ is a homoskedastic martingale diff. sequence if $E(e_t^2 | \mathcal{F}_{t-1}) = \sigma^2$.

  $e_t = u_t u_{t-1}$ is mds, but not homoskedastic

* CLT for mds

  Thm: if $\{u_t\}$ is strictly stationary, ergodic and mds, then

  $$S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u_t \xrightarrow{d} N(0, \Sigma). \quad \text{where } \Sigma = E(u_t u_t')$$

  This is the t.s. counterpart of the Lindeberg-Levy CLT.

Mixing.

We will loose the restriction of mds. The price is stronger assumptions on the dependence than ergodicity.

* $\alpha(A, B) = |P(AB) - P(A)P(B)|$

* Let two σ-fields be $\mathcal{F}_{-\infty}^{t} = \sigma(\ldots, Y_{t-1}, Y_t)$.

  $$\mathcal{F}_{t}^{\infty} = \sigma(Y_t, Y_{t+1}, \ldots)$$

* Strong mixing coef. $\alpha(\ell) = \sup_t \sup_{A \in \mathcal{F}_{-\infty}^{t-\ell},\, B \in \mathcal{F}_t^{\infty}} \alpha(A, B)$.

  In general, the α-coef should have a sup over t.

  Y_t is strong mixing if $\alpha(\ell) \to 0$ as $\ell \to \infty$.

* A mixing process is ergodic.

* Absolute regularity (β-mixing)

  $$\beta(\ell) = \sup_{A \in \mathcal{F}_t^{\infty}} |P(A | \mathcal{F}_{-\infty}^{t-\ell}) - P(A)|$$

  β mixing is stronger than α mixing.

* Strong mixing is preserved by finite transformation

  Thm: Y_t has mixing coef $\alpha_Y(\ell)$. $X_t = \sigma(Y_t, Y_{t-1}, \ldots, Y_{t-q})$.

  Then $\alpha_X(\ell) \le \alpha_Y(\ell - q)$ for $\ell \ge q$.

  The α-coefs satisfy the same rate and summation properties.

* Rate condition $\alpha(\ell) = O(\ell^{-r})$

  Summation restriction $\sum^{\infty} \alpha(\ell)^{r} < \infty$, $\sum^{\infty}$

* Rate condition $\alpha(\ell) = O(\ell^{-r})$

Summation restriction $\sum_{\ell=0}^{\infty} \alpha(\ell)^r < \infty$ or $\sum_{\ell=0}^{\infty} \ell^s \alpha(\ell)^r < \infty$.

Thm 14.13 bounds covariances with functions of α-coef.

CLT for correlated variables

$$\mathrm{var}(S_n) = \mathrm{var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} Y_t\right)$$

$$= \frac{1}{n} 1_N' E[Y Y'] 1_N$$

$$= \frac{1}{n} 1_N' \begin{bmatrix} \sigma^2 & \gamma(1) & \gamma(2) & \cdots & \\ \gamma(1) & \sigma^2 & \gamma(1) & \ddots & \\ \gamma(2) & \gamma(1) & \ddots & \ddots & \ddots \\ & \ddots & \ddots & \ddots & \\ & & & & \sigma^2 \end{bmatrix} 1_N$$

$$= \frac{1}{n}\left(n\sigma^2 + 2(n-1)\gamma(1) + 2(n-2)\gamma(2) + \cdots + 2\cdot\gamma(n-1) + 2\times 0 \times \gamma(n)\right]$$

$$= \sigma^2 + 2\sum_{\ell=1}^{n}\left(1-\frac{\ell}{n}\right)\gamma(\ell)$$

As $\gamma(-\ell) = \gamma(\ell)$, $\mathrm{var}(S_n) = \sum_{\ell=-n}^{n}\left(1-\frac{|\ell|}{n}\right)\gamma(\ell)$

In vector case, similarly we have

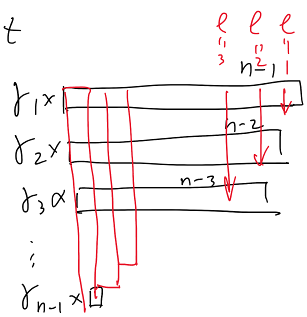
$$\mathrm{var}(S_n) = \Gamma(0) + \sum_{\ell=1}^{n}\left(1-\frac{\ell}{n}\right)\left(\Gamma(\ell)+\Gamma(\ell)'\right) = \sum_{\ell=-n}^{n}\left(1-\frac{|\ell|}{n}\right)\Gamma(\ell)$$

* For CLT to work, $\mathrm{var}(S_n)$ must be convergent in the limit

$$\sum_{\ell=1}^{n}\left(1-\frac{\ell}{n}\right)\gamma(\ell) = \frac{1}{n}\sum_{\ell=1}^{n}(n-\ell)\gamma(\ell)$$

$$= \frac{1}{n}\sum_{\ell=1}^{n-1}\sum_{j=1}^{\ell}\gamma(j)$$

$$\longrightarrow \sum_{j=1}^{\infty}\gamma(j) = \sum_{\ell=1}^{\infty}\gamma(\ell)$$



by the Theorem of Cesaro means if $\sum_{\ell=1}^{\infty}\gamma(\ell)$ is convergent

Necessary condition: $\gamma(\ell) \to 0$ as $\ell \to \infty$

Sufficient: $\sum_{\ell=1}^{\infty}|\gamma(\ell)| < \infty$

It can be shown if $E\|u_t\|^r < \infty$ and $\sum_{\ell=0}^{\infty}\alpha(\ell)^{1-2/r} < \infty$ for some $r > 2$,

then $\sum_{\ell=0}^{\infty}\|\Gamma(\ell)\| < \infty$ is absolutely convergent.

* Thm (CLT) If Y_t is strictly stationary with α-mixing coef $\sum_{\ell=0}^{\infty} \alpha(\ell)^{1-2/r} < \infty$ and $E\|u\|^r < \infty$ for some $r > 2$, $E(u_t) = 0$, then
$S_n \xrightarrow{d} N(0, \Omega)$ where $\Omega = \sum_{\ell=-\infty}^{\infty} \Gamma(\ell)$ is the long-run variance.

---

linear projection

* In regression problems, $P(Y|X) = X\beta^* = X'(EXX')^{-1}EXY$
* Extend to a projection to the infinite past history $\tilde{Y}_{t-1} = (Y_{t-1}, Y_{t-2}, \cdots)$

Denote $P_{t-1}(Y_t) = P[Y_t | \tilde{Y}_{t-1}]$, and the projection error $e_t := Y_t - P_{t-1}(Y_t)$.

* Projection Thm:

If $Y_t \in \mathbb{R}$ is covariance stationary, then the projection error satisfies

① $E(e_t) = 0$ ② $\sigma^2 = E(e_t^2) \leq E(Y_t^2)$

② $E(e_t e_{t-j}) = 0$ for all $j \geq 1$.

In other words, $\{e_t\}$ is a white noise.

If $\{Y_t\}$ is strictly stationary, then $\{e_t\}$ is strictly stationary.

* Def: a time series is a white noise if it is covariance stationary with 0 autocovariance.

It is helpful to image the projection as a linear combination

$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + e_t$

The nature of projection ensures e_t is uncorrelated with all regressors.

e_{t-j} is a linear combination $Y_{t-j} - \alpha_1 Y_{t-j-1} - \alpha_2 Y_{t-j-2} - \cdots$.

Then e_t is uncorrelated with e_{t-j}.

Wold decomposition

* If Y_t is covariance stationary, and the linear projection error has $\sigma^2 > 0$, then

$$Y_t = \mu_t + \sum_{j=0}^{\infty} b_j e_{t-j}, \quad b_0 = 1, \quad \sum_{j=0}^{\infty} b_j^2 < \infty, \text{ and } \mu_t = \lim_{m\to\infty} P_{t-m}(Y_t)$$

Project Y_t onto the orthogonal elements $e_t, e_{t-1}, e_{t-2} \ldots$

For simplicity, we can consider the case $\mu_t = \mu$

Project Y_t onto the orthogonal elements $e_t, e_{t-1}, e_{t-2} \ldots$

For simplicity, we can consider the case $\mu_t = \mu$

* Def: lag operator: $LY_t = Y_{t-1}$

$$L^2 Y_t = L(LY_t) = LY_{t-1} = Y_{t-2}$$, and so on.

$$Y_t = \mu + \sum_{j=1}^{\infty} b_j e_{t-j} = \mu + (b_0 + b_1 L + b_2 L^2 + \cdots) e_t = \mu + b(L) e_t$$

where $b(L)$ is an infinite-order polynomial.

* Autoregressive Wold Representation

If Y_t is covariance stationary with $Y_t = \mu + b(L) e_t$, then with some additional technical restrictions, $Y_t = \mu + \sum_{j=1}^{\infty} a_j Y_{t-j} + e_j$.