Séries de Tempo

Aula 1 - Conceitos Introdutórios

Regis A. Ely

Departamento de Economia
Universidade Federal de Pelotas

09 de julho de 2020

Conteúdo

Como instalar o R

Durante este curso usaremos o R para exercícios e exemplos[1]

- O R é uma software estatístico gratuito e de código aberto
- O R pode ser instalado em `http://cran.r-project.org/`
- É recomendável também instalar o RStudio Desktop em `https://rstudio.com/products/rstudio/`
- O RStudio é uma interface gráfica que facilita o uso do R
- Você pode encontrar muito material sobre o R online, um bom exemplo é o curso em `https://r4ds.had.co.nz/`

[1] *Curiosidade*: Todo este curso, incluindo os slides, são feitos no R.

Instalar e carregar pacotes no R

Após instalar o **R** e o **RStudio** você deverá instalar e ler os pacotes que serão utilizados durante a aula

- A função `install.packages()` instala pacotes no R
- A função `library()` lê pacotes no ambiente do R

Durante o curso, você sempre deve garantir que os pacotes utilizados através da função `library` estão devidamente instalados em seu `R`

Definição de processo estocástico

- Um processo estocástico é uma **família de variáveis aleatórias**
- Ao observarmos uma série de tempo, $\{y_1, y_2, \ldots, y_T\}$, cada y_t representa uma variável aleatória
- Assim, uma série de tempo é apenas uma realização do processo estocástico que gerou os dados

Definição de processo estocástico

Vamos gerar um processo estocástico no R com 100 realizações para cada mês de 2019[2]:

```r
library(tidyverse) # Para manipular dados e graficos
library(lubridate) # Para manipular datas
set.seed(9817) # Fixar semente de dados aleatorios
stoc <- tibble(
  y = rnorm(1200) + rep(1:12, each = 100),
  mes = rep(
    seq.Date(ymd("2019-01-01"), ymd("2019-12-01"), by = "month"),
    each = 100
  )
)
```

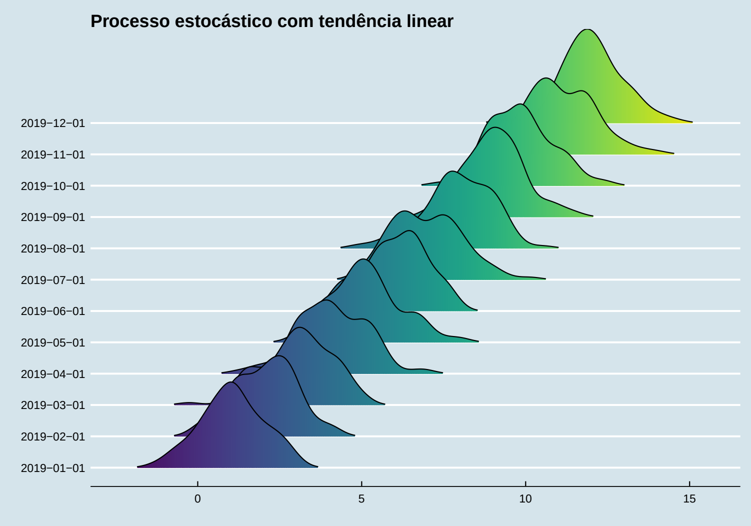
[2] A média do processo estocástico irá mudar ao longo do tempo pois adicionamos uma tendência linear aos dados.

Definição de processo estocástico

Vamos plotar a densidade de cada variável aleatória que compõe o processo estocástico:

```r
library(ggridges) # Para plotar densidades de variaveis aleatorias
library(ggthemes) # Temas para graficos
library(viridis) # Cores para graficos
ggplot(stoc, aes(x = y, y = as.factor(mes), fill = stat(x))) +
  geom_density_ridges_gradient(scale = 3, rel_min_height = 0.01) +
  scale_fill_viridis(option = "D")+
  theme_economist() +
  theme(legend.position="none") +
  labs(
    x = "", y = "",
    title = "Processo estocástico com tendência linear"
  )
```

Definição de processo estocástico



Definição de processo estocástico

- Em uma série de tempo, observamos apenas um valor da distribuição de probabilidade para cada unidade de tempo
- Podemos chamar o conjunto de valores possíveis para cada unidade de tempo de **espaço de estados**, e o valor observado de **estado**
- Um processo estocástico está completamente especificado apenas se conhecermos todas as funções densidade de probabilidade de y para cada unidade de tempo

Esperança de um processo estocástico

O valor esperado da t-ésima observação de uma série de tempo é conhecido como *ensemble average*:

- $E(y_t) = \text{plim}_{N \to \infty} \frac{\sum_{i=1}^{N} y_t^{(i)}}{N}$, onde N é o número de observações em cada período de tempo

Já a média temporal de uma série de tempo é dada por:

- $\bar{Y}_t = \frac{\sum_{t=1}^{T} y_t}{T}$, onde T é o número de observações no tempo

Esperança de um processo estocástico

Utilizando o processo estocástico criado no exemplo anterior, podemos calcular a *ensemble average* para cada mês no `R`:

```
stoc %>%
  group_by(mes) %>%
  summarise(mean(y), .groups = "drop")
```

Esperança de um processo estocástico

```
## # A tibble: 12 x 2
##    mes          `mean(y)`
##    <date>           <dbl>
##  1 2019-01-01        1.04
##  2 2019-02-01        2.04
##  3 2019-03-01        3.02
##  4 2019-04-01        4.08
##  5 2019-05-01        5.11
##  6 2019-06-01        6.04
##  7 2019-07-01        6.98
##  8 2019-08-01        7.93
##  9 2019-09-01        9.05
## 10 2019-10-01        9.84
## 11 2019-11-01       11.1
## 12 2019-12-01       12.0
```

Esperança de um processo estocástico

Vamos selecionar apenas uma realização de y para cada mês, de maneira aleatória, e então calcular a média temporal:

```
obs <- sample(1:100, 12) + seq(0, 1100, by = 100)
yt <- stoc[obs, ]
mean(yt$y)
```

```
## [1] 6.543201
```

Note como a média temporal, mesmo selecionando as observações de maneira aleatória, pode divergir significativamente da *ensemble average* para alguns meses

Variância e Autocovariância

A autocovariância de ordem j de um processo estocástico é dada por:

- $\gamma_{jt} = E(y_t - \mu_t)(y_{t-j} - \mu_{t-j})$, onde $\mu_t = E(y_t)$

Quando $j = 0$, temos a variância do processo estocástico:

- $\gamma_{0t} = E(y_t - \mu_t)^2$

Em termos da *ensemble average*, temos:

- $\gamma_{jt} = \text{plim}_{N\to\infty} \frac{\sum_{i=1}^{N}(y_t^{(i)} - \mu_t)(y_{t-j}^{(i)} - \mu_{t-j})}{N}$

Já utilizando os dados observados e a média temporal:

- $\gamma_{jt} = \frac{\sum_{t=1}^{T-j}(y_t - \mu_t)(y_{t-j} - \mu_{t-j})}{T}$, para $j = 0, 1, \ldots, T$

Variância e Autocovariância

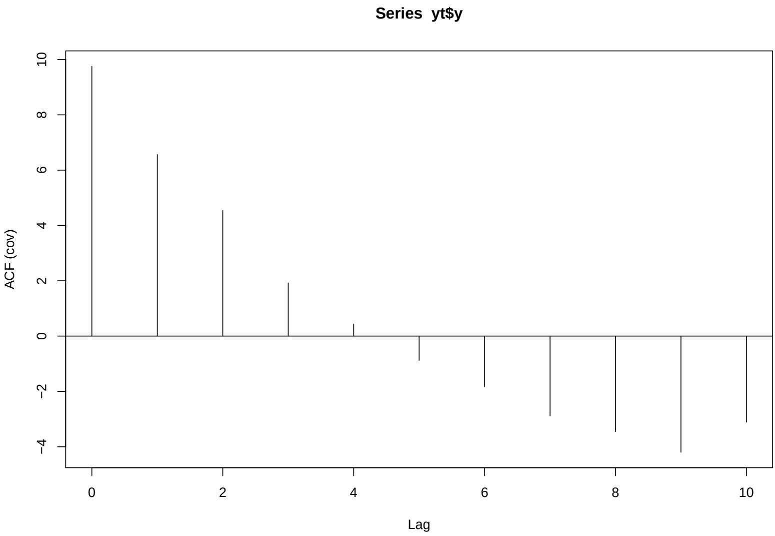
Podemos plotar a variância e as autocovariâncias da série `yt` com a função `acf` no `R`:

```
acf(yt$y, type = "covariance")
```

Observações:

- O primeiro valor corresponde a variância ($j = 0$)
- Para obter os valores da autocovariância, basta salvar o resultado do comando `acf` em uma variável
- A autocovariância não é uma medida padronizada para comparar séries diferentes pois ela depende da magnitude dos valores.

Variância e Autocovariância



Autocorrelação

A função de autocorrelação de um processo estocástico está sempre entre -1 e 1, e é descrita por[3]:

- $\rho_{jt} = \frac{\gamma_{jt}}{\gamma_{0t}}$

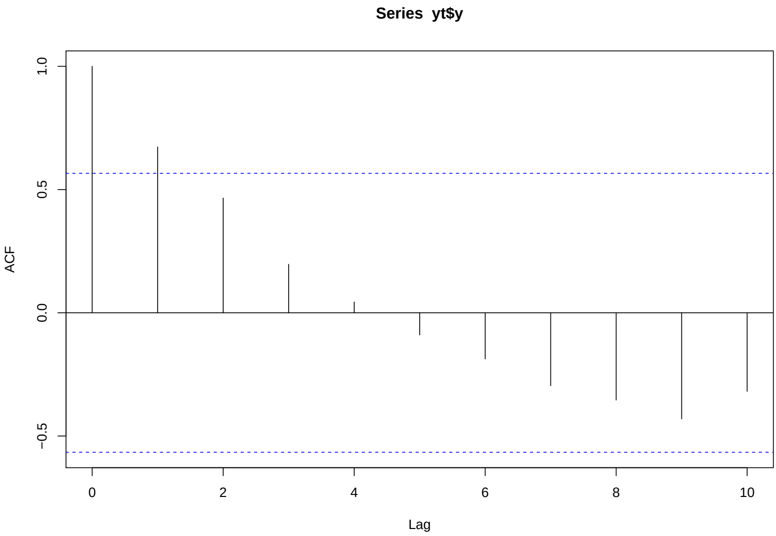
Para plotar as autocorrelações no R basta remover o argumento *covariance* da função `acf`:

```
acf(yt$y)
```

A linha tracejada indica significância estatística

[3] Note que $\rho_{0t} = 1$ para todo t

Autocorrelação



Estacionariedade

Um processo estocástico é **estacionário em covariância** (ou fracamente estacionário) se a média e a autocovariância não dependerem da unidade de tempo t

- $E(Y_t) = \mu$ para todo t
- $E(Y_t - \mu)(Y_{t-j} - \mu) = \gamma_j$ para todo t e qualquer j

Observação:

- Note que $\gamma_j = \gamma_{-j}$ para qualquer j

Estacionariedade

Um processo estocástico é **estritamente estacionário** se para quaisquer valores $j_1, j_2, \ldots, j_n$, a distribuição conjunta de $(y_t, y_{t+j_1}, \ldots, y_{t+j_n})$ depende apenas de $j_1, j_2, \ldots, j_n$ e não da unidade de tempo t

- Nem todo processo estacionário em covariância é estritamente estacionário
- Um processo estritamente estacionário com segundos momentos finitos é também estacionário em covariância
- Um processo estacionário gaussiano sempre é estritamente estacionário

Estacionariedade

Exercícios:

- O processo estocástico criado na variável `stoc` é estacionário em covariância?
- Este processo estocástico é gaussiano?
- A partir da variável `stoc`, como você criaria um novo processo estacionário gaussiano?
- Este novo processo estocástico é estritamente estacionário?

Ergodicidade

Sob quais condições a média temporal é uma boa aproximação da *ensemble average*?

Ergodicidade

- *Ensemble averages* normalmente são impossíveis de calcular, pois temos apenas uma realização do processo estocástico
- Se o processo estocástico for **ergódico para a média**, então a média temporal irá convergir para a *ensemble average* quando o número de observações for grande o suficiente:
  - $\bar{Y}_t \xrightarrow{p} E(Y_t)$ quando $T \to \infty$
- Na maior parte deste curso iremos supor que as séries de tempo são ergódicas

Ergodicidade

- Uma **condição suficiente para ergodicidade na média** é $\sum_{j=0}^{\infty} |\gamma_j| < \infty$
  - Se y_t for um processo estacionário gaussiano, esta condição garante a ergodicidade em todos os momentos
- Um processo estocástico estacionário em covariância não necessariamente é ergódico
- Um processo estocástico ergódico (em todos os momentos) é necessariamente estacionário em covariância
- Logo, um processo não-estacionário é necessariamente não-ergódico

Ergodicidade

Podemos checar se a soma dos valores absolutos das autocovariâncias da série de tempo `yt` está se aproximando de um valor finito:

```
autocov <- acf(yt$y, type = "covariance")
cumsum(abs(autocov$acf))
```

```
##  [1]  9.753141 16.315166 20.856001 22.775924 23.201596 24.078678 25.906582
##  [8] 28.792365 32.245030 36.443645 39.552414
```

Parece que esta soma está aumentando a taxas não decrescentes, o que indica que a condição suficiente de ergodicidade na média não é satisfeita

Ruído branco

- O ruído branco é o principal componente aleatório dos modelos de séries de tempo
- Um ruído branco é definido por uma sequência $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ tal que:
    - $E(\varepsilon_t) = 0$
    - $E(\varepsilon_t^2) = \sigma^2$
    - $E(\varepsilon_t \varepsilon_\tau) = 0$ para $t \neq \tau$
- Se supormos independência entre ε_t e ε_τ para qualquer $t \neq \tau$, então teremos um **ruído branco independente**
- Se supormos $\varepsilon_t \sim N(0, \sigma^2)$ então teremos um **ruído branco gaussiano**

Ruído branco

```r
ts.plot(rnorm(100)) # Ruído branco gaussiano
```