HW3_Yi_Xiong

Yi Xiong

2021/2/18

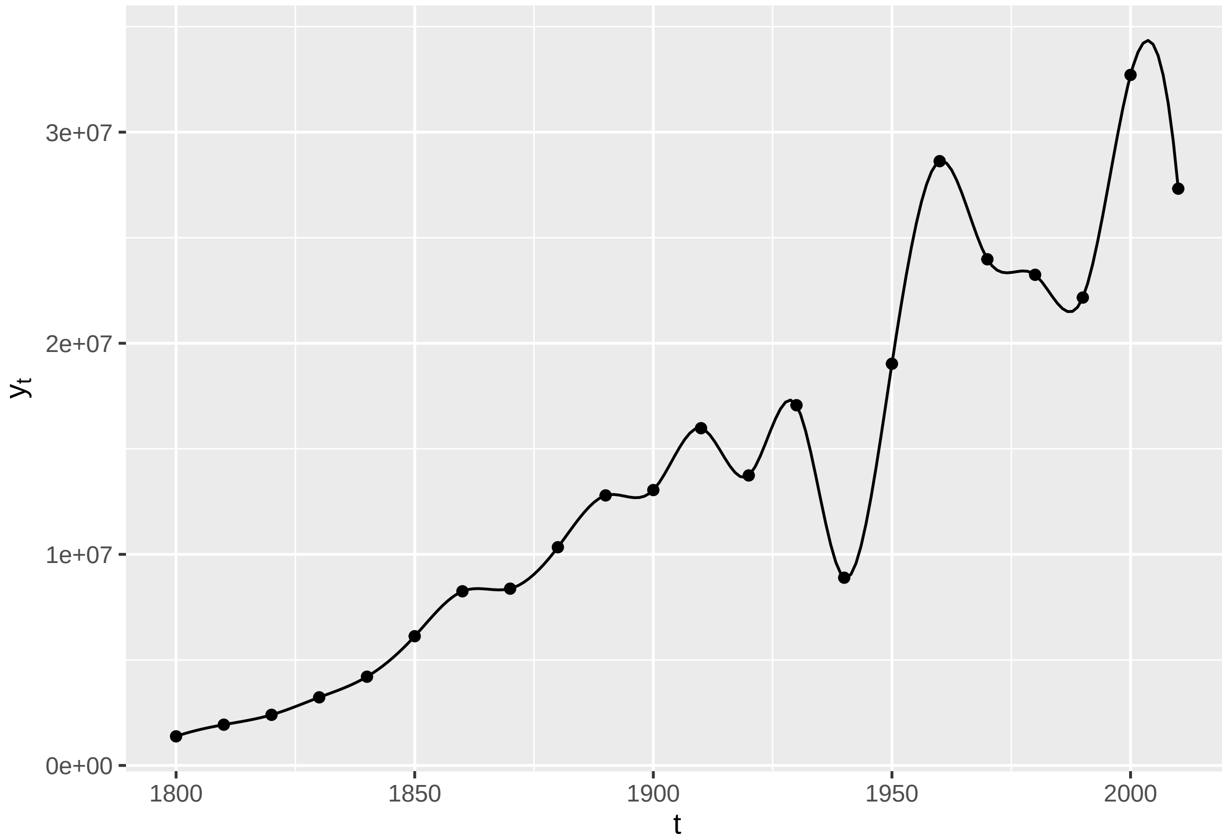
Contents

1) Consider the US population growth data $\mathbf{x_t}$ from Module 1.2

a) Define $y_t = \nabla x_t - Lx_t = x_t - x_{t-1}$ .compute and plot $(y_t\, Vs.t)$. Do you see a trend from this plot? Explain.

```
library(ggplot2)
po <- "
Year USPopulation
1790 3,929,214
1800 5,308,483
1810 7,239,881
1820 9,638,453
1830 12,866,020
1840 17,069,453
1850 23,191,876
1860 31,443,321
1870 39,818,449
1880 50,155,783
1890 62,947,714
1900 75,994,575
1910 91,972,266
```

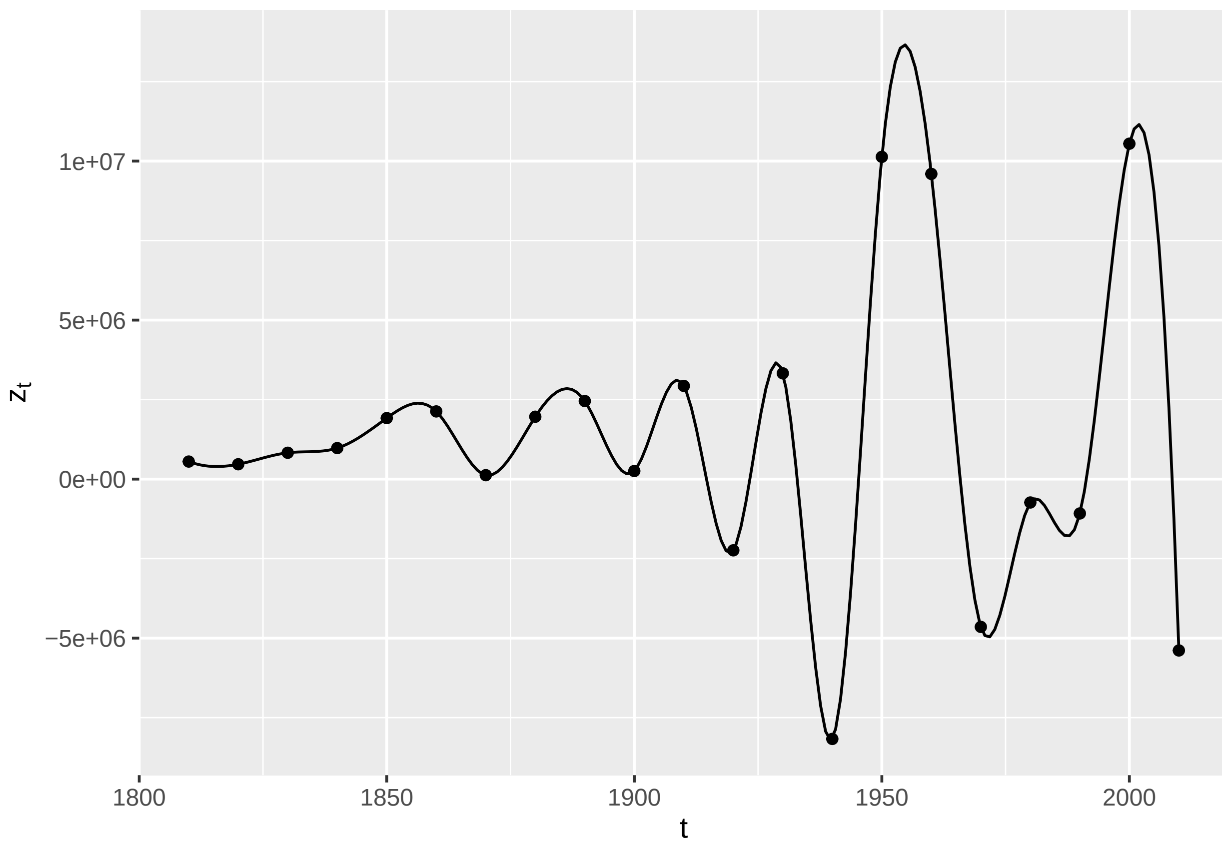
```
1920 105,710,620
1930 122,775,046
1940 131,669,275
1950 150,697,361
1960 179,323,175
1970 203,302,031
1980 226,545,805
1990 248,709,873
2000 281,421,906
2010 308,745,538"
po <- read.table(
  text = po,
  header = T,
  sep = " ",
  fill = T
)
po$USPopulation <- as.numeric(gsub(",", "", po$USPopulation))
yt <- c()
x <- po$USPopulation
t <- 2:nrow(po)
for (i in t) {
  yt <- c(yt, x[i] - x[i - 1])
}
df <- data.frame(x = po$Year[-1], y = yt)
df2 <- as.data.frame(spline(df$x, df$y, n = 200))
p <- ggplot(df, aes(x = x, y = y)) +
  geom_point() +
  xlab(expression(t)) +
  ylab(expression(y[t])) +
  geom_line(data = df2, aes(x = x, y = y))
p
```



The population growth rate increased steadily in past 200 years. On the plot the influence of the World War II and the baby boom after it is easily identified.

b) Define $z_t = \nabla y_t - L y_t = y_t - y_{t-1}$. Compute and plot $(z_t\, Vs. t)$. Do you see a trend from this plot? Explain.

```
zt <- c()
t <- 2:length(yt)
for (i in t) {
  zt <- c(zt, yt[i] - yt[i - 1])
}
df <- data.frame(x = po$Year[c(-1, -2)], y = zt)
df2 <- as.data.frame(spline(df$x, df$y, n = 200))
p <- ggplot(df, aes(x = x, y = y)) +
  geom_point() +
  xlab(expression(t)) +
  ylab(expression(z[t])) +
  geom_line(data = df2, aes(x = x, y = y))
p
```



The change rate of population growth rate is less stable after 1900

c) Compute the mean, variance and auto-correlation ρ_k for z_t and plot $(\rho_k \, Vs. \, k)$

```r
# mean
(ztMean <- mean(zt))
```

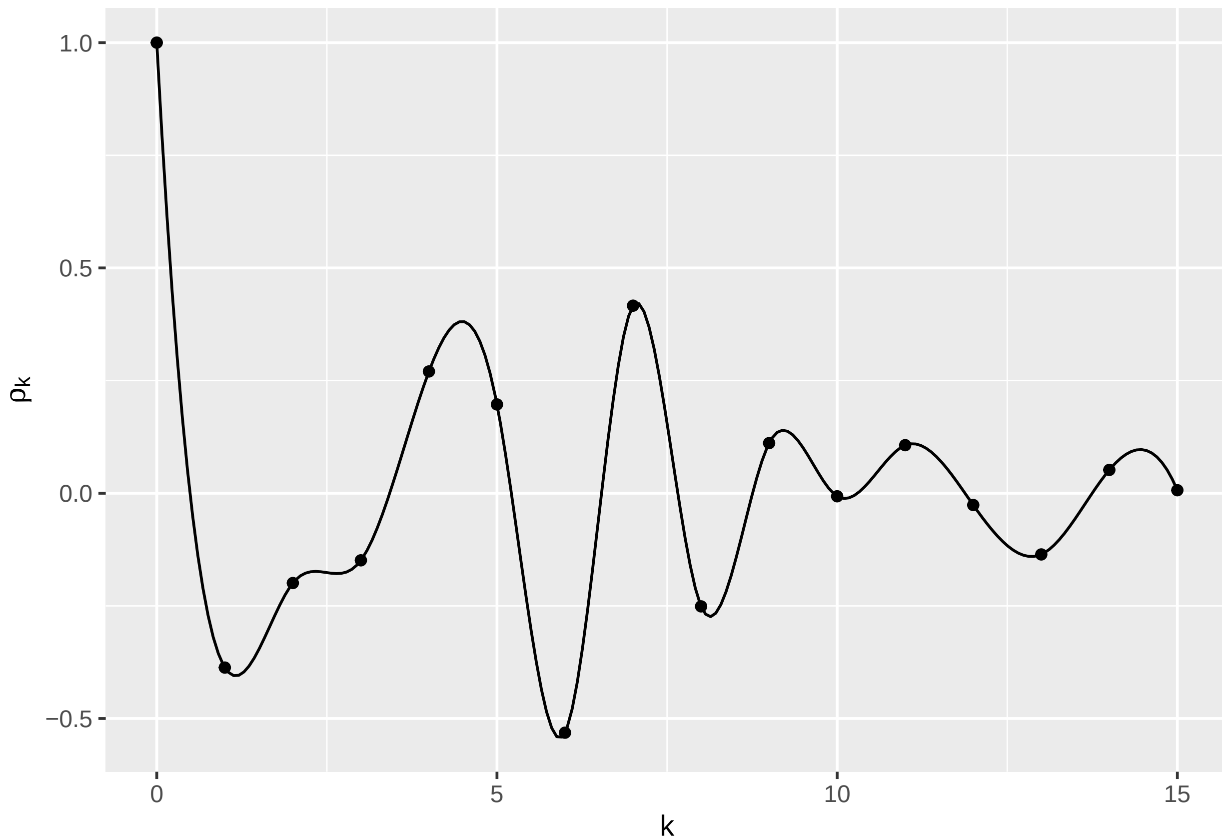
```
## [1] 1235446
```

```r
# variance
(ztVar <- var(zt))
```

```
## [1] 2.190531e+13
```

```r
# autocorrelation
n <- length(zt)
getCov <- function(k) {
  x1 <- zt[1:(n - k)]
  x2 <- zt[(1 + k):(n)]
  cov(x1, x2)
}
k <- 0:floor(3 / 4 * n)
cov1 <- sapply(k, getCov)
```

```r
p <- cov1 / cov1[1]
dfP <- data.frame(x = k, y = p)
dfP2 <- as.data.frame(spline(dfP$x, dfP$y, n = 200))
p <- ggplot(dfP, aes(x = x, y = y)) +
  geom_point() +
  geom_line(data = dfP2, aes(x = x, y = y)) +
  xlab(expression(k)) +
  ylab(expression(rho[k]))
p
```



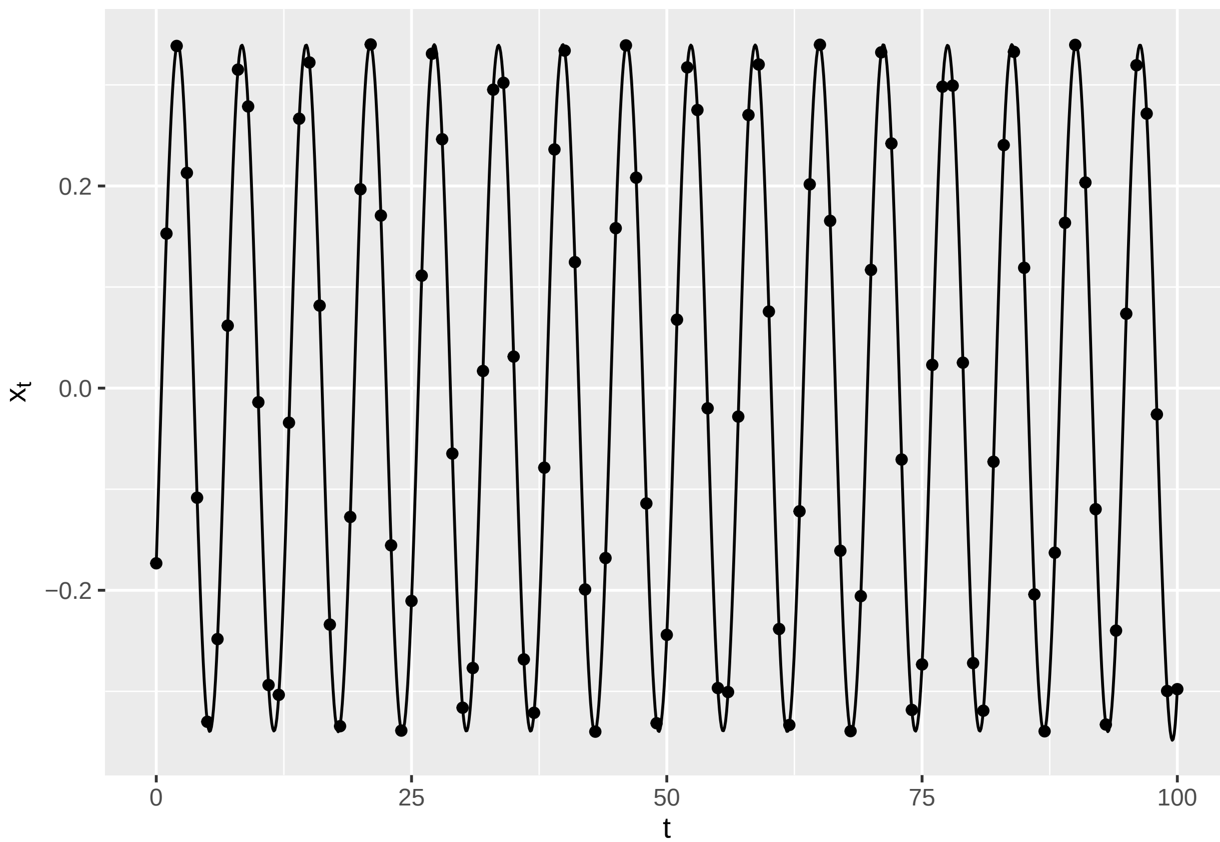
The absolute value of auto-correlation decreases when k increases.

2) (10 points) Define $x_t = a\cos(t) + b\sin(t)$ for $0 \leqslant t \leqslant 100$ where a and b are from $N(0,1)$ and are independent.

a) Draw the pair a and b from $N(0,1)$ and plot $(x_t\, Vs.\, t)$ for $0 \leqslant t \leqslant 100$

```r
set.seed(6666)
a = rnorm(1)
set.seed(6667)
b = rnorm(1)
t <- 0:100
```

```
xt <- c()
for (i in t) {
  xt <- c(xt, a * cos(i) + b * sin(i))
}
df <- data.frame(x = t, y = xt)
df2 <- as.data.frame(spline(df$x, df$y, n = 1000))
p <- ggplot(df, aes(x = x, y = y)) +
  geom_point() +
  xlab(expression(t)) +
  ylab(expression(x[t])) +
  geom_line(data = df2, aes(x = x, y = y))
p
```



b) Repeat this experiment for 20 different pairs: a and b from $N(0, 1)$ and plot all in the same plot.

```
suppressMessages(library(tidyverse))
getXt <- function(a, b) {
  t <- 0:100
  xt <- c()
  for (i in t) {
    xt <- c(xt, a * cos(i) + b * sin(i))
  }
```

```
  return(xt)
}
set.seed(6666)
a = rnorm(20)
set.seed(6667)
b = rnorm(20)
xt <- mapply(getXt, a, b)
colnames(xt) <- paste0("pair", 1:20)
xt <- cbind(t = t, xt)
xt <- as.data.frame(xt)
xt <- pivot_longer(xt, -t, names_to = "Pair", values_to = "xt")
xt$Pair <- factor(xt$Pair, levels = paste0("pair", 1:20))
p <- ggplot(xt, aes(x = t, y = xt, color = Pair)) +
  geom_point() +
  xlab(expression(t)) +
  ylab(expression(x[t]))
p + theme(legend.position = "bottom")
```