

Exerciții de seminar 1

Estimare prin metoda verosimilității maxime

Obiectivul acestui seminar este de a prezenta câteva exerciții de calcul cu metode utile atunci când vrem să verificăm dacă eșantionul provine dintr-o populație normală.

1 Metoda verosimilității maxime și repartiția Geometrică

Ex. 1.1



Fie X_1, X_2, \dots, X_n un eșantion de talie n dintr-o populație Geometrică a cărei funcție de masă este dată de

$$f_{\theta}(x) = \mathbb{P}_{\theta}(X = x) = \theta(1 - \theta)^{x-1}, \quad \forall x \in \{1, 2, 3, \dots\}$$

unde $\theta \in (0, 1)$ este necunoscut. Presupunem că

$$\mathbb{E}[X] = \frac{1}{\theta}, \quad \text{Var}(X) = \frac{1 - \theta}{\theta^2}$$

- Scriveți logaritmul funcției de verosimilitate pentru eșantionul dat.
- Determinați estimatorul de verosimilitate maximă $\hat{\theta}_n$ pentru θ .
- Arătați că estimatorul de verosimilitate maximă este consistent.
- Folosind proprietățile asimptotice ale estimatorilor de verosimilitate maximă, derivați repartiția asimptotică a lui $\hat{\theta}_n$.
- Folosind *Teorema Limită Centrală* și *metoda Delta*, derivați repartiția asimptotică a lui $\hat{\theta}_n$.
- Determinați marginea Rao-Cramer.
- Generați un eșantion de talie $n = 1000$ dintr-o populație Geometrică de parametru $\theta = 0.345$. Estimați parametru θ prin metoda verosimilității maxime folosind funcția `optim()` (sau `optimize()`).

- a) Din definiția funcției de verosimilitate avem

$$L(\theta|\mathbf{x}) = \prod_{i=1}^n f_{\theta}(x_i) = \prod_{i=1}^n \theta(1 - \theta)^{x_i-1} = \theta^n (1 - \theta)^{\sum_{i=1}^n x_i - n}$$

de unde logaritmul funcției de verosimilitate este

$$l(\theta|\mathbf{x}) = \sum_{i=1}^n \log f_{\theta}(x_i) = n \log \theta + \left(\sum_{i=1}^n x_i - n \right) \log(1 - \theta).$$

- b) Estimatorul de verosimilitate maximă pentru θ este definit prin

$$\hat{\theta}_n = \arg \max_{0 < \theta < 1} L(\theta|\mathbf{x}) = \arg \max_{0 < \theta < 1} l(\theta|\mathbf{x})$$

iar pentru determinarea acestuia (sub anumite condiții de regularitate) trebuie să rezolvăm ecuația de verosimilitate $\frac{\partial l(\theta|\mathbf{x})}{\partial \theta} = 0$ (condiție de ordin unu). Trebuie remarcat că în cazul în care $\theta = (\theta_1, \theta_2, \dots, \theta_k)$ condiția se scrie sub forma

$$\nabla L(\theta|\mathbf{x}) = \frac{\partial L(\theta|\mathbf{x})}{\partial \theta} = \begin{pmatrix} \frac{\partial L(\theta|\mathbf{x})}{\partial \theta_1} \\ \dots \\ \frac{\partial L(\theta|\mathbf{x})}{\partial \theta_k} \end{pmatrix} = \begin{pmatrix} 0 \\ \dots \\ 0 \end{pmatrix}.$$

Soluțiile acestei ecuații ne dau punctele critice (din interiorul domeniului) iar pentru determinarea maximului este necesară verificarea unor condiții de ordin doi: matricea Hessiană

$$\frac{\partial^2 L(\theta|\mathbf{x})}{\partial \theta \partial \theta^\top} = \begin{pmatrix} \frac{\partial^2 L(\theta|\mathbf{x})}{\partial \theta_1^2} & \frac{\partial^2 L(\theta|\mathbf{x})}{\partial \theta_1 \partial \theta_2} & \dots & \frac{\partial^2 L(\theta|\mathbf{x})}{\partial \theta_1 \partial \theta_k} \\ \frac{\partial^2 L(\theta|\mathbf{x})}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 L(\theta|\mathbf{x})}{\partial \theta_2^2} & \dots & \frac{\partial^2 L(\theta|\mathbf{x})}{\partial \theta_2 \partial \theta_k} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 L(\theta|\mathbf{x})}{\partial \theta_k \partial \theta_1} & \frac{\partial^2 L(\theta|\mathbf{x})}{\partial \theta_k \partial \theta_2} & \dots & \frac{\partial^2 L(\theta|\mathbf{x})}{\partial \theta_k^2} \end{pmatrix}$$

evaluată în $\hat{\theta}_n$ trebuie să fie negativ definită, adică

$$\mathbf{x}^\top \frac{\partial^2 L(\theta|\mathbf{x})}{\partial \theta \partial \theta^\top} \mathbf{x} < 0, \quad \forall \mathbf{x} \in \mathbb{R}^n \setminus \{0\}.$$

În cazul problemei noastre obținem

$$\frac{\partial l(\theta|\mathbf{x})}{\partial \theta} = \frac{n}{\theta} - \left(\sum_{i=1}^n x_i - n \right) \frac{1}{1 - \theta}$$

și rezolvând ecuația $\frac{\partial l(\theta|\mathbf{x})}{\partial \theta} = 0$ găsim că

$$\frac{n}{\theta} - \left(\sum_{i=1}^n x_i - n \right) \frac{1}{1 - \theta} \iff \frac{1 - \theta}{\theta} = \frac{1}{n} \sum_{i=1}^n x_i - 1 \iff \frac{1}{\theta} = \frac{1}{n} \sum_{i=1}^n x_i$$

de unde $\hat{\theta}_n = \frac{1}{\bar{X}_n}$. Pentru a vedea că $\hat{\theta}_n$ este într-adevăr valoarea care maximizează funcția de verosimilitate, avem

$$\left. \frac{\partial^2 l(\theta|\mathbf{x})}{\partial \theta^2} \right|_{\hat{\theta}_n} = -\frac{n}{\hat{\theta}_n^2} - \left(\frac{1}{1 - \hat{\theta}_n} \right)^2 \left(\sum_{i=1}^n x_i - n \right)$$

și cum $\hat{\theta}_n = \frac{1}{\bar{X}_n}$ deducem că $\sum_{i=1}^n x_i - n = n \left(\frac{1}{\hat{\theta}_n} - 1 \right)$ iar

$$\begin{aligned} \left. \frac{\partial^2 l(\theta|\mathbf{x})}{\partial \theta^2} \right|_{\hat{\theta}_n} &= -\frac{n}{\hat{\theta}_n^2} - \left(\frac{1}{1 - \hat{\theta}_n} \right)^2 n \left(\frac{1}{\hat{\theta}_n} - 1 \right) = -n \left(\frac{1}{\hat{\theta}_n^2} + \frac{1}{\hat{\theta}_n(1 - \hat{\theta}_n)} \right) \\ &= -\frac{n}{\hat{\theta}_n^2(1 - \hat{\theta}_n)} < 0 \end{aligned}$$

ceea ce arată că $\hat{\theta}_n = \frac{1}{\bar{X}_n}$ este estimatorul de verosimilitate maximă.

c) Aplicând *Legea numerelor mari* (variantea slabă) avem că

$$\bar{X}_n \xrightarrow{\mathbb{P}} \mathbb{E}[X_1] = \frac{1}{\theta}$$

Cum $\hat{\theta}_n = \frac{1}{\bar{X}_n}$ putem aplica *Teorema aplicațiilor continue* pentru funcția $g(x) = \frac{1}{x}$, $0 < x < 1$ și găsim că

$$\hat{\theta}_n = g(\bar{X}_n) \xrightarrow{\mathbb{P}} g\left(\frac{1}{\theta}\right) = \theta$$

ceea ce arată că $\hat{\theta}_n$ este consistent.

d) Observăm că funcția de masă verifică condițiile de regularitate¹ prin urmare are loc

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, I_1^{-1}(\theta_0))$$

unde θ_0 este valoarea adevărată a parametrului iar $I_1^{-1}(\theta_0)$ este informația lui Fisher pentru o observație. În general *Informația lui Fisher* pentru eșantion este

$$\begin{aligned} I_n(\theta) &= \text{Var}_{\theta}(\nabla l(\theta|\mathbf{X})) = \text{Var}_{\theta}\left(\frac{\partial \log f_{\theta}(\mathbf{X})}{\partial \theta}\right) \\ &= \mathbb{E}_{\theta}\left[\frac{\partial \log f_{\theta}(\mathbf{X})}{\partial \theta} \times \frac{\partial \log f_{\theta}(\mathbf{X})}{\partial \theta}^{\top}\right] \\ &= \mathbb{E}_{\theta}\left[-\frac{\partial^2 \log f_{\theta}(\mathbf{X})}{\partial \theta \partial \theta^{\top}}\right]. \end{aligned}$$

Pentru cazul nostru găsim că informația lui Fisher este

$$I_1(\theta) = \mathbb{E}_{\theta}\left[-\frac{\partial^2 \log f_{\theta}(X_i)}{\partial \theta^2}\right] = \mathbb{E}_{\theta}\left[\frac{1}{\theta^2} - \left(\frac{1}{1-\theta}\right)^2 (X_i - 1)\right] = \frac{1}{\theta^2(1-\theta)}$$

și astfel

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \theta_0^2(1-\theta_0))$$

sau echivalent $\hat{\theta}_n \approx \mathcal{N}\left(\theta_0, \frac{\theta_0^2(1-\theta_0)}{n}\right)$.

e) Știind că $\mathbb{E}[X] = \frac{1}{\theta_0}$ și $\text{Var}(X) = \frac{1-\theta_0}{\theta_0^2}$ și aplicând *Teorema Limită Centrală* avem că

$$\sqrt{n}\left(\bar{X}_n - \frac{1}{\theta_0}\right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}\left(0, \frac{1-\theta_0}{\theta_0^2}\right).$$

Estimatorul de verosimilitate maximă este $\hat{\theta}_n = \frac{1}{\bar{X}_n}$ și considerând $g(x) = \frac{1}{x}$, $x \in (0, 1)$ (g este derivabilă cu derivata continuă) putem aplica metoda Delta care conduce la

$$\sqrt{n}\left(g(\bar{X}_n) - g\left(\frac{1}{\theta_0}\right)\right) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}\left(0, g'\left(\frac{1}{\theta_0}\right)^2 \frac{1-\theta_0}{\theta_0^2}\right)$$

¹e.g. Suportul $\{x \mid f_{\theta}(x) > 0\}$ nu depinde de θ ; $f_{\theta}(x)$ este de cel puțin 3 ori derivabilă în raport cu θ și derivatele sunt continue; Valoarea adevărată θ se află într-o mulțime compactă.

și cum $g'(x) = -\frac{1}{x^2}$ obținem același rezultat ca și în cazul punctului anterior

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow[n \rightarrow \infty]{d} \mathcal{N}(0, \theta_0^2(1 - \theta_0)).$$

f) Marginea inegalității Rao-Cramer (MIRC) este $I_n^{-1}(\theta_0)$ și cum

$$I_n(\theta_0) = \mathbb{E}_\theta \left[- \frac{\partial^2 \log f_\theta(\mathbf{X})}{\partial \theta \partial \theta^\top} \Big|_{\theta_0} \right] = n I_1(\theta_0) = \frac{n}{\theta_0^2(1 - \theta_0)}$$

găsim

$$MIRC = I_n^{-1}(\theta_0) = \frac{\theta_0^2(1 - \theta_0)}{n}.$$

g) Pentru a genera eșantionul X_1, X_2, \dots, X_n vom folosi funcția `rgeom()`. Atenție, această funcție permite generarea de observații repartizate Geometric de parametru θ , cu funcția de masă

$$\mathbb{P}_\theta(X = x) = \theta(1 - \theta)^x, \quad \forall x \in \{0, 1, 2, 3, \dots\}$$

deci trebuie să adăugăm 1 la fiecare observație pentru a fi în contextul din exercițiu.

```
theta = 0.345
n = 1000

x = rgeom(n, theta) + 1

# EVM gasit este
EVM = 1/mean(x)
EVM
[1] 0.3696858
```

Vom crea o funcție care să calculeze estimatorul de verosimilitate maximă plecând de la logaritmul funcției de verosimilitate (îi determinăm maximum cu ajutorul funcției `optimize()`):

```
EVM_geom = function(theta, n, init = 0.5, seed = NULL){

  if (!is.null(seed)){
    set.seed(seed)
  }

  x = rgeom(n, theta)+1

  loglik_geom = function(param){

    l = n*log(param) + (sum(x) - n)*log(1-param)
    # intoarcem -l pentru ca vrem maximumul
    return(-l)
  }

  # folosim functia optimize
  # a se vedea ?optimize
  return(optimize(loglik_geom, c(0,1))[[1]])
}

# exemple
```

```
EVM_geom(0.345, 100000)
[1] 0.3448424
EVM_geom(0.478, 1000)
[1] 0.4847251
EVM_geom(0.222, 1000)
[1] 0.2200082
```

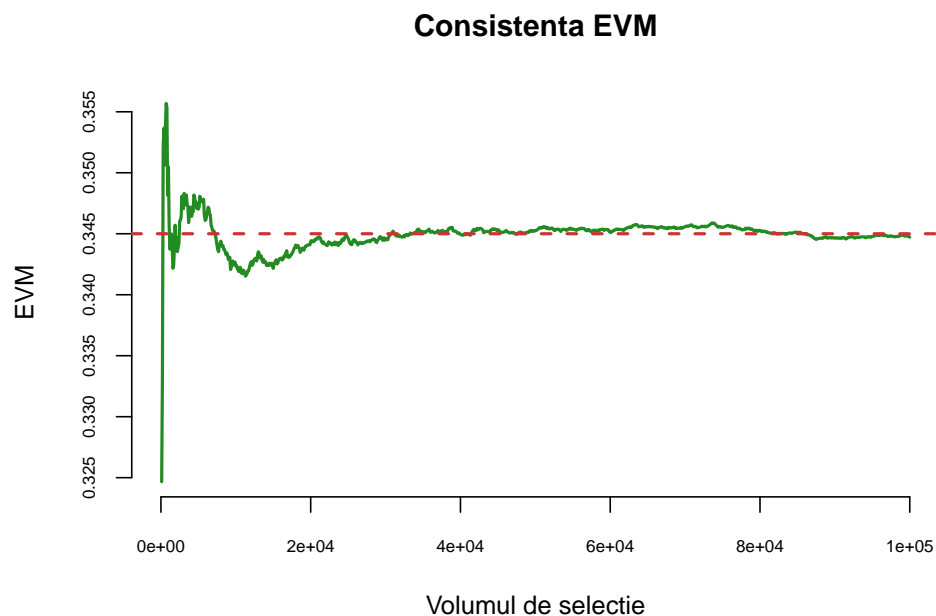
În figura de mai jos este ilustrată proprietatea de consistență a estimatorului de verosimilitate maximă, pentru $\theta = 0.345$:

```
theta = 0.345
t = seq(100, 100000, 100)

y = sapply(t, function(x){
  r = EVM_geom(theta, x, seed = 2018)
  return(r)
})

plot(t, y, type = "l",
     col = "forestgreen",
     xlab = "Volumul de selectie",
     ylab = "EVM",
     main = "Consistentă EVM",
     bty = "n",
     cex.axis = 0.7,
     lwd = 2)

abline(h = theta, col = "brown3",
       lty = 2, lwd = 2)
```



2 Exemplu de EVM determinat prin soluții numerice

Ex. 2.1



Fie X_1, X_2, \dots, X_n un eșantion de talie n dintr-o populație logistică a cărei densitate este dată de formula

$$f_{\theta}(x) = \frac{e^{-(x-\theta)}}{(1 + e^{-(x-\theta)})^2}, \quad x \in \mathbb{R}, \theta \in \mathbb{R}$$

Determinați estimatorul de verosimilitate maximă $\hat{\theta}_n$ pentru θ .

Densitatea de repartiție și funcția de repartiție a repartiției logistice sunt ilustrate mai jos (în R se folosesc funcțiile: `rlogis`, `dlogis`, `plogis` și respectiv `qlogis`):

```
# Generam graficele
pars = c(2, 4, 6, 9)

x = seq(-8, 15, length.out = 250)

set.seed(1234)
cols = sample(colors(), length(pars))

par(mfrow = c(1, 2))
# densitatile
plot(x, dlogis(x, location = pars[1]),
     xlab = "x",
     ylab = TeX("$f_{\\theta}(x)$"),
     # ylim = c(0,1),
     col = "brown3",
     lwd = 2, type = "l",
     bty = "n",
     main = "Densitatea")

for (i in seq(length(pars)-1)){
  location = pars[i+1]

  y = dlogis(x, location = location)

  lines(x, y, lwd = 2,
        col = cols[i])
}

legend("topright",
      legend = TeX(paste0("$\\theta = ", pars, "$")),
      col = cols,
      lwd = rep(2, length(pars)),
      bty = "n",
      cex = 0.7,
      seg.len = 1.5)

# functiile de repartitie
plot(x, plogis(x, location = pars[1]),
     xlab = "x",
```

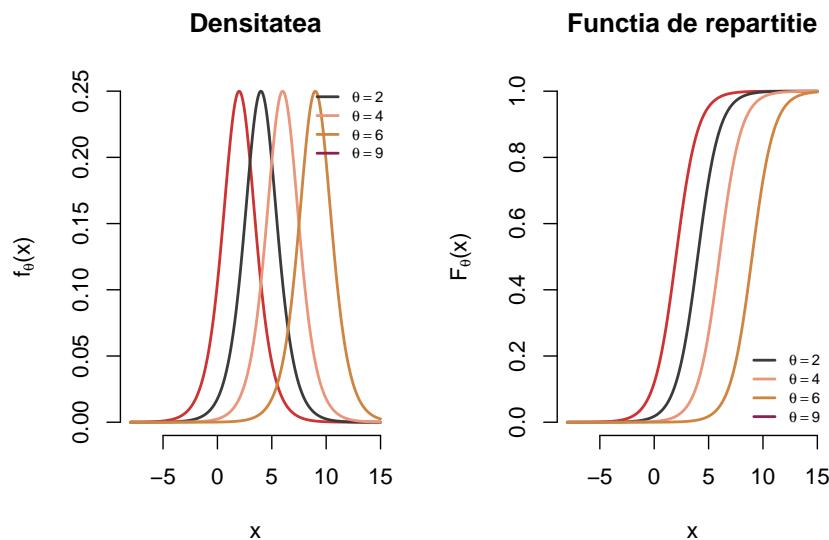
```
ylab = TeX("$F_{\\theta}(x)$"),
ylim = c(0,1),
col = "brown3",
lwd = 2, type = "l",
bty = "n",
main = "Funcția de repartiție")

for (i in seq(length(pars)-1)){
  location = pars[i+1]

  y = plogis(x, location = location)

  lines(x, y, lwd = 2,
        col = cols[i])
}

legend("bottomright",
  legend = TeX(paste0("$\\theta = ", pars, "$")),
  col = cols,
  lwd = rep(2, length(pars)),
  bty = "n",
  cex = 0.7,
  seg.len = 1.5)
```



Observăm că funcția de verosimilitate este dată de

$$L(\theta|\mathbf{x}) = \prod_{i=1}^n f_{\theta}(x_i) = \prod_{i=1}^n \frac{e^{-(x_i-\theta)}}{(1 + e^{-(x_i-\theta)})^2}$$

iar logaritmul funcției de verosimilitate este

$$l(\theta|\mathbf{x}) = \sum_{i=1}^n \log f_{\theta}(x_i) = n\theta - n\bar{x}_n - 2 \sum_{i=1}^n \log \left(1 + e^{-(x_i-\theta)} \right).$$

Pentru a găsi valoarea lui θ care maximizează logaritmul funcției de verosimilitate și prin urmare a funcției de verosimilitate trebuie să rezolvăm ecuația $l'(\theta|\mathbf{x}) = 0$, unde derivata lui $l(\theta|\mathbf{x})$ este

$$l'(\theta|\mathbf{x}) = n - 2 \sum_{i=1}^n \frac{e^{-(x_i-\theta)}}{1 + e^{-(x_i-\theta)}}$$

ceea ce conduce la ecuația

$$\sum_{i=1}^n \frac{e^{-(x_i-\theta)}}{1 + e^{-(x_i-\theta)}} = \frac{n}{2} \quad (\star)$$

Chiar dacă această ecuație nu se simplifică, se poate arăta că această ecuația admite soluție unică. Observăm că derivata parțială a membrului drept în (\star) devine

$$\frac{\partial}{\partial \theta} \sum_{i=1}^n \frac{e^{-(x_i-\theta)}}{1 + e^{-(x_i-\theta)}} = \sum_{i=1}^n \frac{e^{-(x_i-\theta)}}{(1 + e^{-(x_i-\theta)})^2} > 0$$

ceea ce arată că membrul stâng este o funcție strict crescătoare în θ . Cum membrul stâng în (\star) tinde spre 0 atunci când $\theta \rightarrow -\infty$ și spre n pentru $\theta \rightarrow \infty$ deducem că ecuația (\star) admite soluție unică (vezi graficul de mai jos).

```
set.seed(112)
n = 20
x = rlogis(n, location = 7.5)

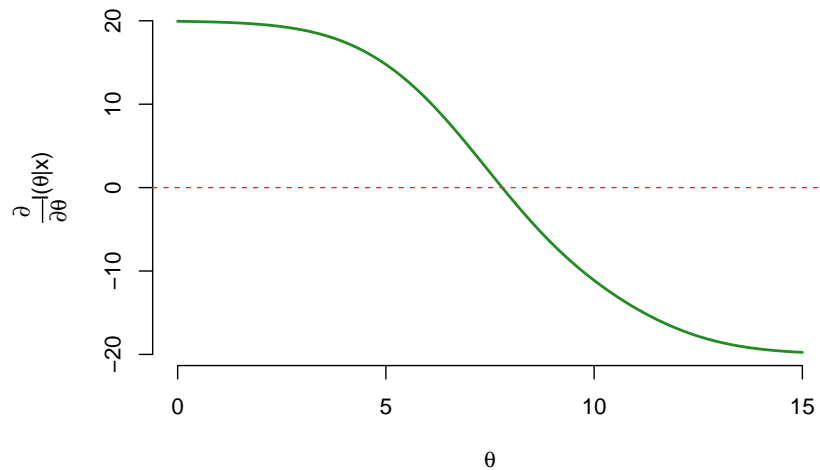
# derivata logaritmului functiei de verosimilitate
dLogLogistic = function(n, x, theta){
  sapply(theta, function(t){
    y = exp(-(x - t))
    n - 2*sum(y/(1+y))
  })
}

theta = seq(0, 15, length.out = 250)

mar.default <- c(5,4,4,2) + 0.1
par(mar = mar.default + c(0, 1.2, 0, 0))

plot(theta, dLogLogistic(n, x, theta), type = "l",
     col = "forestgreen", lwd = 2,
     bty = "n",
     xlab = TeX("$\\theta$"),
     ylab = TeX("$\\frac{\\partial}{\\partial \\theta} l(\\theta | x)$"))

abline(h = 0, col = "brown3",
       lty = 2)
```

Cum nu putem găsi o soluție a ecuației $l'(\theta|\mathbf{x}) = 0$ sub formă compactă, este necesar să apelăm la metode numerice. O astfel de metodă numerică este binecunoscuta [metodă a lui Newton-Raphson](#). Metoda presupune să începem cu o valoare (soluție) inițială $\hat{\theta}^{(0)}$ și să alegem, plecând de la aceasta, o nouă valoare $\hat{\theta}^{(1)}$ definită prin

$$\hat{\theta}^{(1)} = \hat{\theta}^{(0)} - \frac{l'(\hat{\theta}^{(0)})}{l''(\hat{\theta}^{(0)})},$$

adică $\hat{\theta}^{(1)}$ este intersecția cu axa absciselor a tangentei în punctul $(\hat{\theta}^{(0)}, l'(\hat{\theta}^{(0)}))$ la graficul funcției $l'(\theta)$. Ideea este de a itera procesul până când soluția converge, cu alte cuvinte pornind de la o valoare *rezonabilă* de start $\hat{\theta}^{(0)}$ la pasul $k + 1$ avem

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - \frac{l'(\hat{\theta}^{(k)})}{l''(\hat{\theta}^{(k)})}$$

și oprim procesul atunco când k este suficient de mare și/sau $|\hat{\theta}^{(k+1)} - \hat{\theta}^{(k)}|$ este suficient de mic. Următorul grafic ilustrează grafic algoritmul lui Newton:

```
set.seed(112)
n = 20
x = rlogis(n, location = 7.5)

# derivata logaritmului functiei de verosimilitate
dLogLogistic = function(n, x, theta){
  sapply(theta, function(t){
    y = exp(-(x - t))
    n - 2*sum(y/(1+y))
  })
}

theta = seq(0, 15, length.out = 250)
```

```
mar.default <- c(5,4,4,2) + 0.1
par(mar = mar.default + c(0, 1.2, 0, 0))

plot(theta, dLogLogistic(n, x, theta), type = "l",
     col = "forestgreen", lwd = 2,
     bty = "n",
     xlab = TeX("$\\theta$"),
     ylab = TeX("$\\frac{\\partial}{\\partial \\theta} l(\\theta | x)$"))

abline(h = 0, col = "brown3",
       lty = 2)

# ilustrarea metodei Newton

dl = function(theta) n - 2 * sum(exp(theta - x) / (1 + exp(theta - x)))
ddl = function(theta) {-2 * sum(exp(theta - x) / (1 + exp(theta - x))^2)}

x0 = 5 # punctul de start

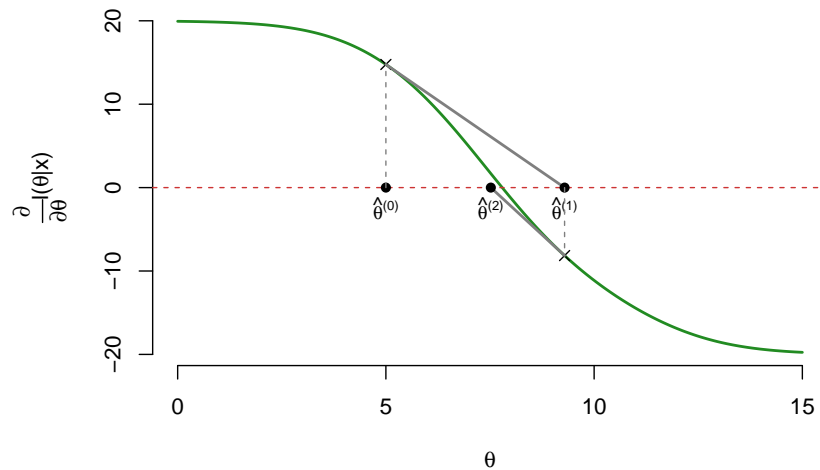
points(x0, 0, pch = 16, col = "black")
text(x0, 0, labels = TeX("$\\hat{\\theta}^{(0)}$"), pos = 1, cex = 0.8)
segments(x0, 0, x0, dl(x0), lty = 2, col = "grey50")
points(x0, dl(x0), pch = 4)

x1 = x0 - dl(x0)/ddl(x0)

segments(x0, dl(x0), x1, 0, lty = 1, lwd = 2, col = "grey50")
points(x1, 0, pch = 16, col = "black")
text(x1, 0, labels = TeX("$\\hat{\\theta}^{(1)}$"), pos = 1, cex = 0.8)
segments(x1, 0, x1, dl(x1), lty = 2, col = "grey50")
points(x1, dl(x1), pch = 4)

x2 = x1 - dl(x1)/ddl(x1)

segments(x1, dl(x1), x2, 0, lty = 1, lwd = 2, col = "grey50")
points(x2, 0, pch = 16, col = "black")
text(x2, 0, labels = TeX("$\\hat{\\theta}^{(2)}$"), pos = 1, cex = 0.8)
```



Obs: Singurul lucru care se schimbă atunci când trecem de la scalar la vector, este funcția $l(\theta)$ care acum este o funcție de $p > 1$ variabile, $\theta = (\theta_1, \theta_2, \dots, \theta_p)^T \in \mathbb{R}^p$. În acest context $l'(\theta)$ este un vector de derivate parțiale iar $l''(\theta)$ este o matrice de derivate parțiale de ordin doi. Prin urmare iterațiile din metoda lui Newton sunt

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} - [l''(\hat{\theta}^{(k)})]^{-1} l'(\hat{\theta}^{(k)})$$

unde $[\cdot]^{-1}$ este [pseudoinversa](#) unei matrici.

Funcția de mai jos implementează metoda lui Newton pentru cazul multidimensional:

```
# Metoda lui Newton

newton <- function(f, df, x0, eps=1e-08, maxiter=1000, ...) {
  # in caz ca nu e incarcat pachetul sa putem accesa pseudoinversa
  if(!exists("ginv")) library(MASS)

  x <- x0
  k <- 0

  repeat {
    k <- k + 1

    x.new <- x - as.numeric(ginv(df(x, ...)) %*% f(x, ...))

    if(mean(abs(x.new - x)) < eps | k >= maxiter) {
      if(k >= maxiter) warning("S-a atins numarul maxim de iteratii!")
      break
    }
    x <- x.new
  }
  out <- list(solution = x.new, value = f(x.new, ...), iter = k)

  return(out)
}
```

Să presupunem că am observat următorul eșantion de talie 20 din repartiția logistică:

```
[1] 6.996304 9.970107 12.304991 11.259549 6.326912 5.378941 4.299639  
[8] 8.484635 5.601117 7.094335 6.324731 6.868456 9.753360 8.042095  
[15] 8.227830 10.977982 7.743096 7.722159 8.562884 6.968356
```

```
set.seed(112)  
x = rlogis(20, location = 7.5)  
  
n = length(x)  
dl = function(theta) n - 2 * sum(exp(theta - x) / (1 + exp(theta - x)))  
ddl = function(theta) {-2 * sum(exp(theta - x) / (1 + exp(theta - x))^2)}  
  
logis.newton = newton(dl, ddl, median(x))
```

și aplicând metoda lui Newton găsim estimatorul de verosimilitate maximă $\hat{\theta}_n = 7.7933$ după numai 3 iterații (datele au fost simulate folosind $\theta = 7.5$).

3 Metoda verosimilității maxime și procese autoregresive $AR(r)$

Ex. 3.1



Se numește proces autoregresiv de ordin 1 $AR(1)$, un proces Gaussian staționar definit prin

$$Y_t = c + \rho Y_{t-1} + \epsilon_t$$

cu ϵ_t variabile aleatoare i.i.d. repartizate $\mathcal{N}(0, \sigma^2)$ și $|\rho| < 1$.

Observăm că din condiția de staționaritate² rezultă că

$$\mathbb{E}[Y_t] = \frac{c}{1 - \rho}, \quad \text{Var}[Y_t] = \frac{\sigma^2}{1 - \rho^2}.$$

Ex. 3.2



Fie $\theta = (c, \rho, \sigma^2)^\top$ vectorul parametrilor modelului. Scrieți funcția de verosimilitate și logaritmul funcției de verosimilitate pentru o observație, y_1 .

Cum variabila aleatoare Y_1 are media și varianța date de

$$\mathbb{E}[Y_1] = \frac{c}{1 - \rho}, \quad \text{Var}[Y_1] = \frac{\sigma^2}{1 - \rho^2}.$$

iar ϵ_t sunt i.i.d. repartizate $\mathcal{N}(0, \sigma^2)$, deducem că Y_1 este repartizată tot normal, cu $Y_1 \sim \mathcal{N}\left(\frac{c}{1 - \rho}, \frac{\sigma^2}{1 - \rho^2}\right)$. Astfel funcția de verosimilitate pentru y_1 este

$$L(\theta; y_1) = \frac{1}{\sqrt{2\pi} \sqrt{\frac{\sigma^2}{1 - \rho^2}}} e^{-\frac{1}{2} \frac{(y_1 - \frac{c}{1 - \rho})^2}{\frac{\sigma^2}{1 - \rho^2}}}$$

iar logaritmul funcției de verosimilitate pentru y_1 este

²Aici ne referim la proprietatea de staționaritate în sens larg (wide-sense stationary) care presupune că $\forall t_1, t_2 \in \mathbb{N}$ și $\forall \tau \in \mathbb{N}$ avem $\mathbb{E}[Y_{t_1}] = \mathbb{E}[Y_{t_2}]$ și $\mathbb{E}[Y_{t_1} Y_{t_2}] = \mathbb{E}[Y_{t_1 + \tau} Y_{t_2 + \tau}]$.

$$l(\theta; y_1) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log\left(\frac{\sigma^2}{1-\rho^2}\right) - \frac{1}{2} \frac{\left(y_1 - \frac{c}{1-\rho}\right)^2}{\frac{\sigma^2}{1-\rho^2}}.$$

Ex. 3.3



Care este repartiția condiționată a lui Y_2 la $Y_1 = y_1$? Scrieți funcția de verosimilitate și logaritmul funcției de verosimilitate (condiționată) pentru a doua observație y_2 .

Observăm că pentru $t = 2$ avem

$$Y_2 = c + \rho Y_1 + \epsilon_2,$$

unde $\epsilon_2 \sim \mathcal{N}(0, \sigma^2)$. Prin urmare repartiția condiționată a lui Y_2 dat fiind $Y_1 = y_1$ este

$$Y_2|Y_1 = y_1 \sim \mathcal{N}(c + \rho y_1, \sigma^2)$$

de unde funcția de verosimilitate (condiționată) pentru y_2 este

$$L(\theta; y_2|y_1) = f_{Y_2|Y_1}(y_2|y_1; \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(y_2 - c - \rho y_1)^2}{\sigma^2}}$$

iar logaritmul funcției de verosimilitate (condiționată) pentru y_2 este

$$l(\theta; y_2|y_1) = \log f_{Y_2|Y_1}(y_2|y_1; \theta) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{1}{2} \frac{(y_2 - c - \rho y_1)^2}{\sigma^2}.$$

Ex. 3.4



Considerați eșantionul $\{y_1, y_2\}$ de talie 2. Scrieți funcția de verosimilitate (completă) și logaritmul funcției de verosimilitate a modelului $AR(1)$ pentru acest eșantion. Extindeți rezultatul pentru un eșantion y_1, y_2, \dots, y_T de talie T .

Reamintim că dacă avem două variabile aleatoare continue (absolut continue) X și Y atunci densitatea cuplului (X, Y) este

$$f_{(X,Y)}(x, y) = f_{Y|X}(y|x) f_X(x),$$

prin urmare funcția de verosimilitate (completă) pentru eșantionul $\{y_1, y_2\}$ este

$$L(\theta; y_1, y_2) = f_{(Y_1, Y_2)}(y_1, y_2; \theta) = f_{Y_2|Y_1}(y_2|y_1; \theta) f_{Y_1}(y_1; \theta)$$

sau echivalent

$$L(\theta; y_1, y_2) = L(\theta; y_2|y_1) L(\theta; y_1) = \frac{\sqrt{1-\rho^2}}{2\pi\sigma^2} e^{-\frac{1}{2} \frac{(1-\rho^2)(y_1 - \frac{c}{1-\rho})^2}{\sigma^2} - \frac{1}{2} \frac{(y_2 - c - \rho y_1)^2}{\sigma^2}}.$$

În mod similar, logaritmul funcției de verosimilitate este

$$l(\theta; y_1, y_2) = l(\theta; y_2|y_1) + l(\theta; y_1) = \frac{1}{2} \log(1-\rho^2) - \log(2\pi\sigma^2) - \frac{1}{2} \frac{(1-\rho^2)(y_1 - \frac{c}{1-\rho})^2}{\sigma^2} - \frac{1}{2} \frac{(y_2 - c - \rho y_1)^2}{\sigma^2}.$$

Observăm că densitatea lui Y_3 condiționată la primele două variabile este

$$f_{Y_3|Y_2, Y_1}(y_3|y_2, y_1; \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(y_3 - c - \rho y_2)^2}{\sigma^2}}$$

de unde

$$\begin{aligned} f_{Y_3, Y_2, Y_1}(y_3, y_2, y_1; \theta) &= f_{Y_3|Y_2, Y_1}(y_3|y_2, y_1; \theta) f_{Y_2, Y_1}(y_2, y_1; \theta) \\ &= f_{Y_3|Y_2, Y_1}(y_3|y_2, y_1; \theta) f_{Y_2|Y_1}(y_2|y_1; \theta) f_{Y_1}(y_1; \theta). \end{aligned}$$

În general, valoarea lui Y_1, Y_2, \dots, Y_{t-1} influențează valoarea lui Y_t doar prin valoarea lui Y_{t-1} ceea ce arată că densitatea lui Y_t condiționată la celelalte $t-1$ variabile este

$$f_{Y_t|Y_{t-1}, Y_{t-2}, \dots, Y_1}(y_t|y_{t-1}, y_{t-2}, \dots, y_1; \theta) = f_{Y_t|Y_{t-1}}(y_t|y_{t-1}; \theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(y_t - c - \rho y_{t-1})^2}{\sigma^2}}.$$

Astfel, pentru un eșantion y_1, y_2, \dots, y_T de talie T avem

$$\begin{aligned} L(\theta; y_1, y_2, \dots, y_T) &= L(\theta; y_1) \times \prod_{t=2}^T L(\theta; y_t|y_{t-1}) \\ l(\theta; y_1, y_2, \dots, y_T) &= l(\theta; y_1) + \sum_{t=2}^T l(\theta; y_t|y_{t-1}) \end{aligned}$$

ceea ce conduce la

$$L(\theta; y_1, y_2, \dots, y_T) = \frac{1}{\sqrt{2\pi} \sqrt{\frac{\sigma^2}{1-\rho^2}}} e^{-\frac{1}{2} \frac{\left(y_1 - \frac{c}{1-\rho}\right)^2}{\frac{\sigma^2}{1-\rho^2}}} \times \prod_{t=2}^T \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(y_t - c - \rho y_{t-1})^2}{\sigma^2}}$$

și respectiv la

$$\begin{aligned} l(\theta; y_1, y_2, \dots, y_T) &= -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log\left(\frac{\sigma^2}{1-\rho^2}\right) - \frac{1}{2} \frac{\left(y_1 - \frac{c}{1-\rho}\right)^2}{\frac{\sigma^2}{1-\rho^2}} \\ &\quad + \sum_{t=2}^T \left(-\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{1}{2} \frac{(y_t - c - \rho y_{t-1})^2}{\sigma^2} \right) \\ &= -\frac{T}{2} \log(2\pi) - \frac{T}{2} \log(\sigma^2) + \frac{1}{2} \log(1-\rho^2) + \\ &\quad + \frac{1}{2\sigma^2} \left[(1-\rho^2) \left(y_1 - \frac{c}{1-\rho}\right)^2 + \sum_{t=2}^T (y_t - c - \rho y_{t-1})^2 \right] \end{aligned}$$

Ex. 3.5



crieți o funcție în R care să permită generarea unui eșantion dintr-un proces $AR(1)$. Ilustrați grafic traiectoriile procesului $AR(1)$ pentru diverse seturi de parametrii.

Avem următoarea funcție care generează procesul autoregresiv $AR(1)$:

```
genAR1 = function(n, c, rho, sigma){
  # n - marimea esantionului
  # c - termenul constant
  # rho - parametrul autoregresiv
  # sigma - abaterea standard a erorii

  # generam Y_1 repartizat normal
  y1 = rnorm(1, mean = c/(1-rho), sd = sqrt(sigma^2/(1-rho^2)))

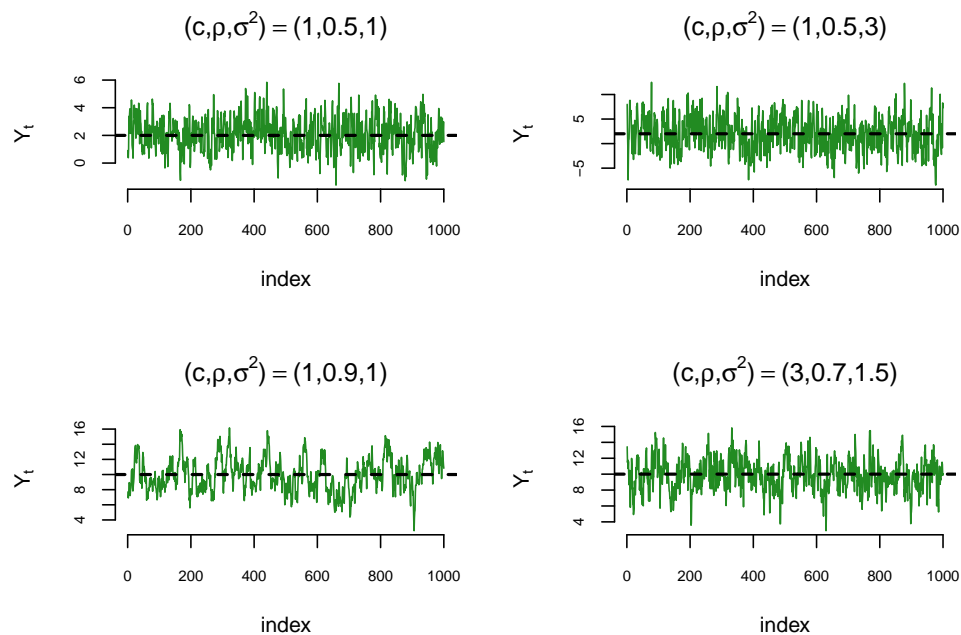
  # nitializam
  y = rep(1, n)*y1

  # vectorul de erori
  epsilon = rnorm(n-1, 0, sigma)

  for (i in 2:n){
    y[i] = c + rho*y[i-1] + epsilon[i-1]
  }

  return(y)
}
```

Ilustrăm grafic traiectoriile procesului $AR(1)$ pentru diverse seturi de parametri (c, ρ, σ^2) :



Ex. 3.6



Acum considerăm că prima observație y_1 este dată (deterministă) și avem $f_{Y_1}(y_1; \theta) = 1$. Scrieți logaritmul funcției de verosimilitate condiționat a modelului $AR(1)$ pentru eșantionul y_1, y_2, \dots, y_T .

Funcția de verosimilitate condiționată este definită prin

$$L(\boldsymbol{\theta}; y_2, \dots, y_T | y_1) = \prod_{t=2}^T f_{Y_t | Y_{t-1}, Y_1=y_1}(y_t | y_{t-1}, y_1; \boldsymbol{\theta}) \times \underbrace{f_{Y_1}(y_1; \boldsymbol{\theta})}_{=1} = \prod_{t=2}^T f_{Y_t | Y_{t-1}}(y_t | y_{t-1}; \boldsymbol{\theta})$$

iar logaritmul funcției de verosimilitate condiționată devine

$$l(\boldsymbol{\theta}; y_2, \dots, y_T | y_1) = \sum_{t=2}^T l_t(\boldsymbol{\theta}; y_t | y_{t-1})$$

cu $l_t(\boldsymbol{\theta}; y_t | y_{t-1}) = \log(f_{Y_t | Y_{t-1}}(y_t | y_{t-1}; \boldsymbol{\theta}))$. Găsim că

$$l(\boldsymbol{\theta}; y_2, \dots, y_T | y_1) = -\frac{T-1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=2}^T (y_t - c - \rho y_{t-1})^2.$$

Ex. 3.7



Scrieți ecuațiile de verosimilitate asociate modelului condiționat de mai sus.

Estimatorul de verosimilitate maximă $\hat{\boldsymbol{\theta}} = (\hat{c}, \hat{\rho}, \hat{\sigma}^2)^\top$ a lui $\boldsymbol{\theta}$ este definit prin

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Theta} l_T(\boldsymbol{\theta}; y_1, \dots, y_T)$$

iar ecuațiile de verosimilitate se scriu

$$\left. \frac{\partial l_T(\boldsymbol{\theta}; y)}{\partial \boldsymbol{\theta}} \right|_{\hat{\boldsymbol{\theta}}} = \begin{pmatrix} \left. \frac{\partial l_T(\boldsymbol{\theta}; y)}{\partial c} \right|_{\hat{\boldsymbol{\theta}}} \\ \left. \frac{\partial l_T(\boldsymbol{\theta}; y)}{\partial \rho} \right|_{\hat{\boldsymbol{\theta}}} \\ \left. \frac{\partial l_T(\boldsymbol{\theta}; y)}{\partial \sigma^2} \right|_{\hat{\boldsymbol{\theta}}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

și înlocuind cu expresia logaritmului funcției de verosimilitate condiționată $l_T(\boldsymbol{\theta}; y) = l(\boldsymbol{\theta}; y_2, \dots, y_T | y_1)$ găsim

$$\begin{aligned} \left. \frac{\partial l_T(\boldsymbol{\theta}; y)}{\partial c} \right|_{\hat{\boldsymbol{\theta}}} &= \frac{1}{\hat{\sigma}^2} \sum_{t=2}^T (y_t - \hat{c} - \hat{\rho} y_{t-1}) = 0 \\ \left. \frac{\partial l_T(\boldsymbol{\theta}; y)}{\partial \rho} \right|_{\hat{\boldsymbol{\theta}}} &= \frac{1}{\hat{\sigma}^2} \sum_{t=2}^T (y_t - \hat{c} - \hat{\rho} y_{t-1}) y_{t-1} = 0 \\ \left. \frac{\partial l_T(\boldsymbol{\theta}; y)}{\partial \sigma^2} \right|_{\hat{\boldsymbol{\theta}}} &= -\frac{(T-1)}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{t=2}^T (y_t - \hat{c} - \hat{\rho} y_{t-1})^2 = 0 \end{aligned}$$

Ex. 3.8



Determinați estimatorul de verosimilitate maximă $\hat{\boldsymbol{\theta}}$ pentru $\boldsymbol{\theta}$.

Problema maximizării logaritmului funcției de verosimilitate $l_T(\boldsymbol{\theta}; y)$ în raport cu c și ρ

$$l(\boldsymbol{\theta}; y_2, \dots, y_T | y_1) = -\frac{T-1}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=2}^T (y_t - c - \rho y_{t-1})^2$$

este echivalentă cu problema minimizării termenului

$$\sum_{t=2}^T (y_t - c - \rho y_{t-1})^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

unde $\mathbf{y} = (y_2, \dots, y_T)^\top$, $\boldsymbol{\beta} = (c, \rho)^\top$ și $\mathbf{X} = \begin{pmatrix} 1 & y_1 \\ \dots & \dots \\ 1 & y_{T-1} \end{pmatrix}$. Observăm că estimatorii de verosimilitate maximă pentru c și ρ sunt echivalenți cu estimatorii obținuți prin metoda celor mai mici pătrate în problema de regresie:

$$\begin{pmatrix} \hat{c} \\ \hat{\rho} \end{pmatrix} = \left(\sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}_i y_i \right) = \begin{pmatrix} T-1 & \sum_{t=2}^T y_{t-1} \\ \sum_{t=2}^T y_{t-1} & \sum_{t=2}^T y_{t-1}^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{t=2}^T y_t \\ \sum_{t=2}^T y_{t-1} y_t \end{pmatrix}.$$

Estimatorul de verosimilitate maximă pentru σ^2 se obține prin rezolvarea ecuației

$$\left. \frac{\partial l_T(\boldsymbol{\theta}; y)}{\partial \sigma^2} \right|_{\hat{\boldsymbol{\theta}}} = -\frac{(T-1)}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4} \sum_{t=2}^T (y_t - \hat{c} - \hat{\rho} y_{t-1})^2 = 0$$

de unde găsim

$$\hat{\sigma}^2 = \frac{1}{T-1} \sum_{t=2}^T (y_t - \hat{c} - \hat{\rho} y_{t-1})^2.$$

Ex. 3.9



Funcția de verosimilitate este o funcție neliniară în parametrii $\boldsymbol{\theta}$, prin urmare estimatorul de verosimilitate maximă $\hat{\boldsymbol{\theta}} = (\hat{c}, \hat{\rho}, \hat{\sigma}^2)^\top$ va fi determinat prin metode numerice. Pentru $c = 1$, $\rho = 0.5$ și $\sigma^2 = 1$ generați un eșantion de talie $T = 1000$, calculați estimatorul de verosimilitate maximă și comparați valorile obținute pentru \hat{c} și $\hat{\rho}$ cu cele obținute prin metoda celor mai mici pătrate.

Considerăm setul de parametri $(c, \rho, \sigma^2) = (1, 0.5, 1)$ și calculăm estimatorul de verosimilitate maximă plecând de la logaritmul funcției de verosimilitate (utilizăm funcția `optim()`):

```
N = 1000
y = genAR1(N, 1, 0.5, 1)

loglik_AR1 = function(param){
  # pentru a folosi functia optim trebuie sa avem un singur argument
  # parametrii
  c = param[1]
  rho = param[2]
  sigma = param[3]

  # esantionul
  ly = length(y) # talia esantionului

  # prima observatie
  ll = log(dnorm(y[1], mean = c/(1-rho), sd = sqrt(sigma^2/(1-rho^2))))

  # celelalte observatii
  dif = y[2:ly] - c - rho*y[1:(ly-1)]
```

```
l2 = log(dnorm(dif, 0, sigma))

# logaritmul verosimilitatii
l = l1 + sum(l2)

# intoarcem -l pentru ca vrem maximul
return(-l)
}

# determinam MLE
param = c(0.6, 0.6, 0.6)
MLE = optim(param, loglik_AR1)$par

# OLS
X = cbind(rep(1, N-1), y[-N])
Y = y[-1]

XtX = t(X) %*% X
XtY = t(X) %*% Y
beta_ols = solve(XtX,XtY)
```

Obținem următoarele rezultate

	Theta	MLE	OLS
c	1.0	0.9261	0.9251
rho	0.5	0.5212	0.5217
sigma	1.0	1.0104	

care sunt apropiate de valorile reale.