Estimação

Ricardo Ehlers
ehlers@icmc.usp.br

Departamento de Matemática Aplicada e Estatística
Universidade de São Paulo

Para um processo estacionário $\{X_t\}$, $t = 1, \ldots, n$, a função de densidade conjunta de $X_1, \ldots, X_n$ pode ser fatorada como,

$$
\begin{aligned}
p(x_1, \ldots, x_n) &= p(x_n | x_{n-1}, \ldots, x_1) p(x_{n-1}, \ldots, x_1) \\
&= p(x_n | x_{n-1}, \ldots, x_1) p(x_{n-1} | x_{n-2}, \ldots, x_1) p(x_{n-2}, \ldots, x_1) \\
&\vdots \\
&= p(x_1) \prod_{t=2}^{n} p(x_t | x_{t-1}, \ldots, x_1).
\end{aligned}
$$

Para um modelo ARMA(p, q) com parâmetros,

$$\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \sigma_\epsilon^2),$$

$$\begin{aligned} p(x_1, \ldots, x_n|\boldsymbol{\theta}) &= p(x_1, \ldots, x_p|\boldsymbol{\theta}) \prod_{t=p+1}^{n} p(x_t|x_{t-1}, \ldots, x_1, \boldsymbol{\theta}) \\ &= p(x_1, \ldots, x_p|\boldsymbol{\theta}) \prod_{t=p+1}^{n} p(x_t|x_{t-1}, \ldots, x_{t-p}, \boldsymbol{\theta}) \end{aligned}$$

Tomando o logaritmo,

$$\log p(\mathbf{x}|\boldsymbol{\theta}) = \log p(x_1, \ldots, x_p|\boldsymbol{\theta}) + \sum_{t=p+1}^{n} \log p(x_t|x_{t-1}, \ldots, x_{t-p}, \boldsymbol{\theta})$$

- A última igualdade vem da estrutura Markoviana da componente autoregressiva.
- O segundo termo é a densidade condicional conjunta de $x_{p+1}, \ldots, x_n$ dados os valores iniciais $x_1, \ldots, x_p$ e define então uma função de verossimilhança condicional
- $p(x_1, \ldots, x_n|\boldsymbol{\theta})$ define a função de verossimilhança exata.

Se for atribuida uma distribuição de probabilidades conjunta para $\boldsymbol{\theta}$ então pelo Teorema de Bayes é possível obter sua distribuição atualizada após os dados serem observados (distribuição a posteriori),

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta})\ p(\boldsymbol{\theta})}{p(\mathbf{x})} \propto p(\mathbf{x}|\boldsymbol{\theta})\ p(\boldsymbol{\theta}).$$

Maximizando a verossimilhança condicional,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\theta} \sum_{t=p+1}^{n} \log p(x_t | x_{t-1}, \ldots, x_{t-p}, \boldsymbol{\theta})$$

Maximizando a verossimilhança completa,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\theta} \log p(x_1, \ldots, x_p | \boldsymbol{\theta}) + \sum_{t=p+1}^{n} \log p(x_t | x_{t-1}, \ldots, x_{t-p}, \boldsymbol{\theta})$$

Ajustando Processos AR

Para um processo AR(p) dado por,

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + \epsilon_t,$$

e dadas n observações $x_1, \ldots, x_n$, os parâmetros $\alpha_1, , \ldots, \alpha_p$ podem ser estimados pelo método de mínimos quadrados.

- Defina a soma de quadrados,

$$S(\boldsymbol{\alpha}) = \sum_{t=p+1}^{n} (x_t - \alpha_1 x_{t-1} - \cdots - \alpha_p x_{t-p})^2$$

- obtenha solução de minimos quadarados,

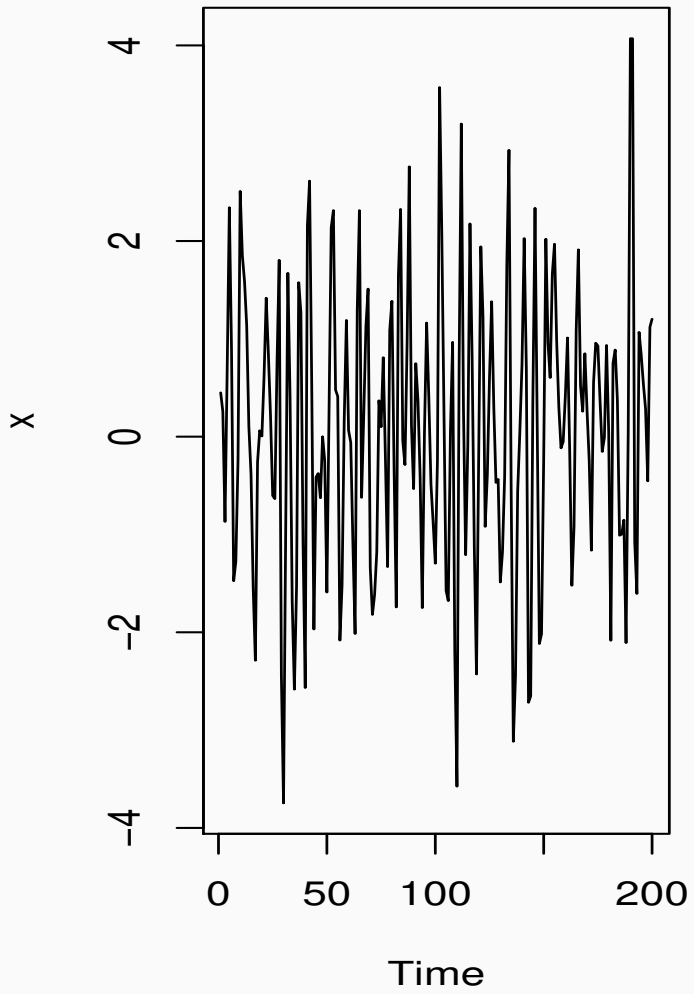
$$\hat{\boldsymbol{\alpha}} = \arg\min_{\alpha} S(\boldsymbol{\alpha}).$$

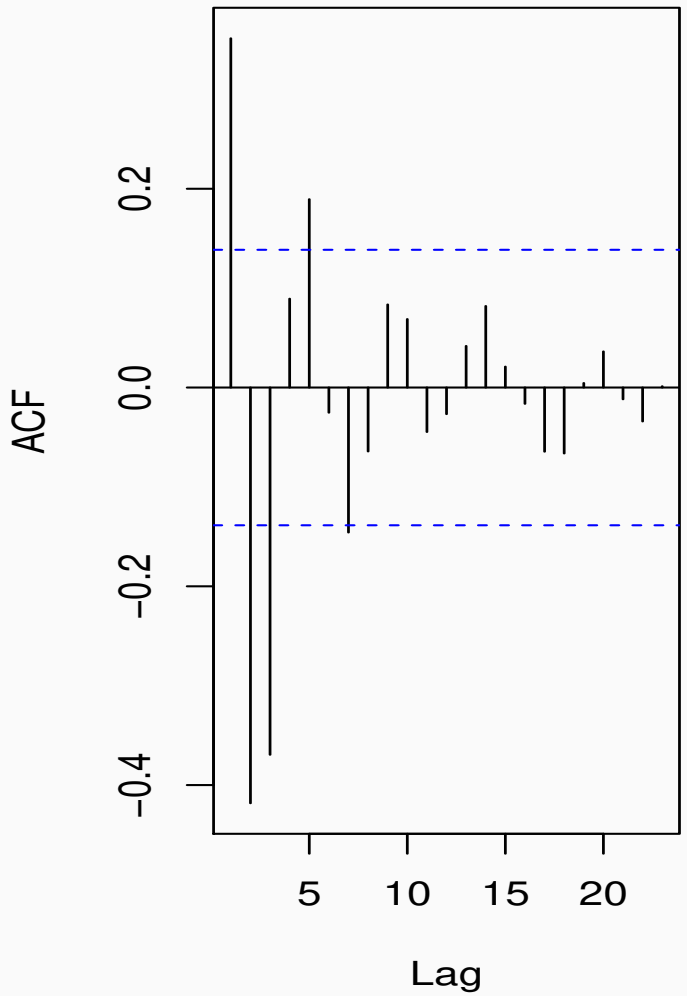
Usando as equações de Yule-Walker

- Estime as p primeiras autocorrelações, $r_1, \ldots, r_p$.
- Use as p primeiras equações de Yule-Walker, e resolva o sistema linear,

$$\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix} = \begin{bmatrix} 1 & r_1 & \ldots & r_{p-1} \\ r_1 & 1 & \ldots & r_{p-2} \\ \vdots & \vdots & & \vdots \\ r_{p-1} & r_{p-2} & \ldots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix}$$

Exemplo. Processo AR(3) simulado com coeficientes 0.6, -0.7 e 0.2. Usando as equações de Yule-Walker para estimar os coeficientes.





Use as 3 primeiras equações de Yule-Walker, e resolva o sistema linear,

$$\begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_1 \\ r_2 & r_1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

$r_1 = 0.3513$, $r_2 = -0.4182$, $r_3 = -0.3695$.

$\hat{\alpha}_1 = 0.6658$, $\hat{\alpha}_2 = -0.7075$, $\hat{\alpha}_3 = 0.1575$.

AR(p) como um modelo linear

Escrevendo o processo AR(p) a partir de $t = p + 1$ obtemos,

$$\begin{bmatrix} x_{p+1} \\ x_{p+2} \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_p & \cdots & x_1 \\ x_{p+1} & \cdots & x_2 \\ \vdots & & \vdots \\ x_{n-1} & \cdots & x_{n-p} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \epsilon_{p+1} \\ \epsilon_{p+2} \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Equivalentemente,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}.$$

sendo $E(\boldsymbol{\epsilon}) = 0$ e $Var(\boldsymbol{\epsilon}) = \sigma_\epsilon^2 I_{n-p}$.

Solução de minimos quadrados,

$$\begin{aligned}
\hat{\boldsymbol{\alpha}} &= \arg\min \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\
\hat{\boldsymbol{\epsilon}} &= \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\alpha}} \\
\hat{\sigma_\epsilon^2} &= \frac{1}{n-p}\sum_{t=p+1}^{n}\hat{\epsilon}_t^2 = \frac{1}{n-p}\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}}
\end{aligned}$$

Modelo AR(p) com erros normais

Se $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_{n-p})$, a função de verossimilhança condicional fica,

$$
\begin{aligned}
L(\boldsymbol{\alpha}, \sigma_\epsilon^2) &= \prod_{t=p+1}^{n} \frac{1}{(2\pi\sigma_\epsilon^2)^{1/2}} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} (x_t - \sum_{j=1}^{p} \alpha_j x_{t-j})^2 \right\} \\
&\propto (\sigma_\epsilon^2)^{-(n-p)/2} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}) \right\}.
\end{aligned}
$$

A solução de máxima verossimilhança é,

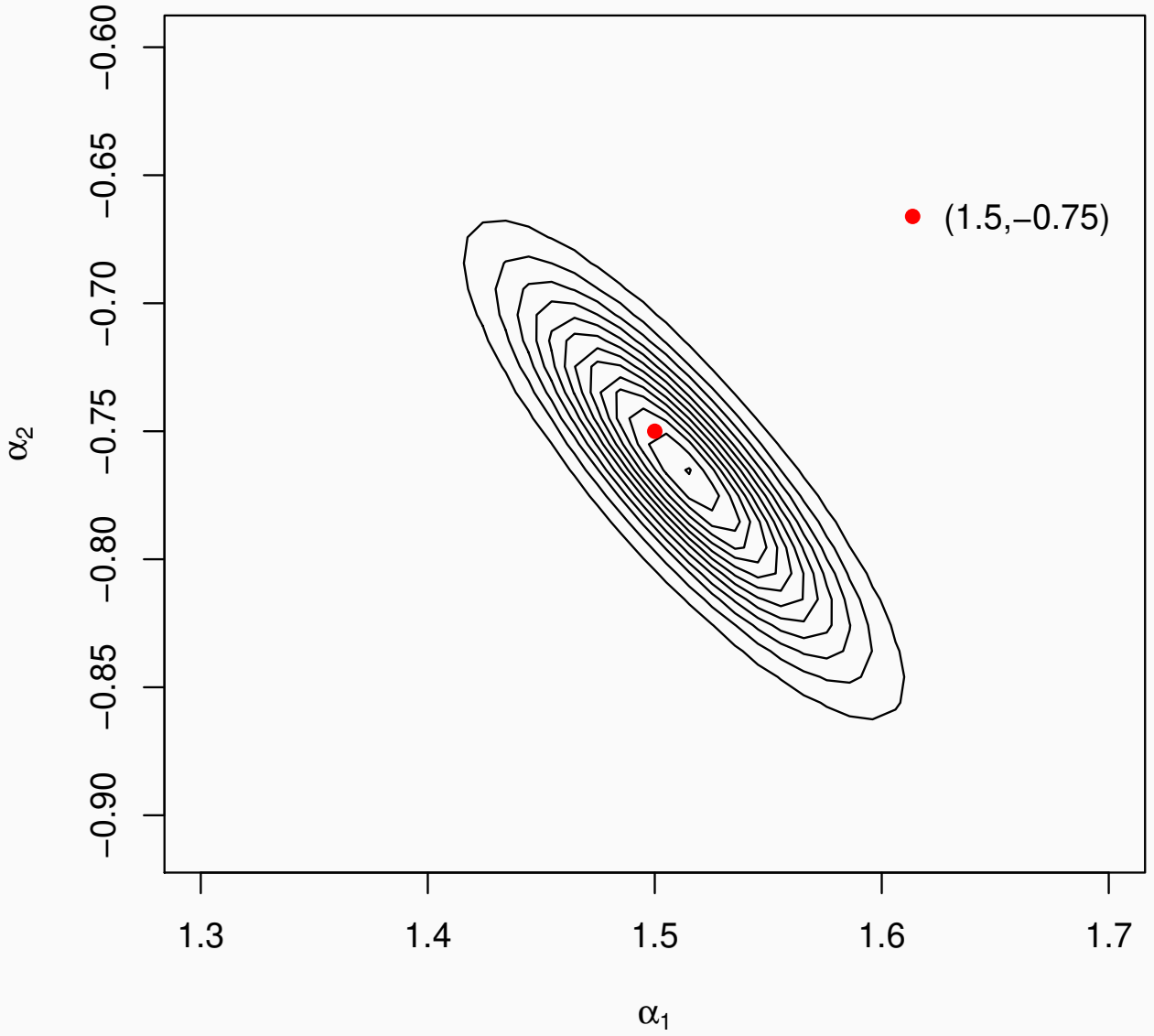
$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ \hat{\sigma^2_\epsilon} &= \frac{1}{n-p}(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}})'(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\alpha}}) = \frac{1}{n-p}\hat{\boldsymbol{\epsilon}}'\hat{\boldsymbol{\epsilon}} \end{aligned}$$

$$\begin{aligned}
\frac{\partial \log(L(\boldsymbol{\alpha}, \sigma_\epsilon^2))}{\partial \boldsymbol{\alpha}} &= -\frac{\sigma_\epsilon^{-2}}{2} \frac{\partial (\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})'(\mathbf{y} - \mathbf{X}\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \\
&= -\frac{\sigma_\epsilon^{-2}}{2} \frac{\partial (-2\boldsymbol{\alpha}'\mathbf{X}'\mathbf{y} + \boldsymbol{\alpha}'\mathbf{X}'\mathbf{X}\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \\
&= -\frac{\sigma_\epsilon^{-2}}{2} (-2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\alpha}).
\end{aligned}$$

$$\left.\frac{\partial \log(L(\boldsymbol{\alpha}, \sigma_\epsilon^2))}{\partial \boldsymbol{\alpha}}\right|_{\boldsymbol{\alpha}=\hat{\boldsymbol{\alpha}}} = \mathbf{0} \iff \hat{\boldsymbol{\alpha}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

Verifique para $\hat{\sigma}_\epsilon^2$.

Exemplo. Seja um processo AR(2) simulado com $\alpha_1 = 1.5$, $\alpha_2 = -0.75$ e $\sigma^2_\epsilon = 1$. A figura mostra a superficie da verossimilhança condicional.



Estimativas de máxima verossimilhança

```
> set.seed(90210)
> y = arima.sim(list(order=c(2,0,0),ar=c(1.5,-0.75)),n=200)
> reg = lm(y[3:n]~y[2:(n-1)]+y[1:(n-2)]-1)
```

	Estimate	Std. Error	t value	Pr(>|t|)
y[2:(n - 1)]	1.5128	0.0465	32.53	0.0000
y[1:(n - 2)]	-0.7650	0.0467	-16.37	0.0000

Exemplo. Para um modelo AR(1) com erros normais a matriz $\mathbf{X}$ tem somente uma coluna e não é difícil verificar que,

$$\mathbf{X}'\mathbf{X} = \sum_{t=2}^{n} x_{t-1}^2 \quad \text{e} \quad \mathbf{X}'\mathbf{y} = \sum_{t=2}^{n} x_t x_{t-1}.$$

Portanto, o EMV condicional é dado por

$$\hat{\alpha} = \frac{\sum_{t=2}^{n} x_t x_{t-1}}{\sum_{t=2}^{n} x_{t-1}^2} \quad \text{e} \quad \hat{\sigma_\epsilon^2} = \frac{1}{n-1} \sum_{t=2}^{n} (x_t - \hat{\alpha} x_{t-1})^2.$$

Modelo AR(1) com erros normais

A função de verossimilhança completa fica,

$$
\begin{aligned}
p(x_1, \ldots, x_n | \alpha, \sigma_\epsilon^2) &= p(x_1 | \alpha, \sigma_\epsilon^2) \prod_{t=2}^{n} p(x_t | x_{t-1}, \alpha, \sigma_\epsilon^2) \\
E(X_t) &= 0 \\
Var(X_t) &= \sigma_\epsilon^2 / (1 - \alpha^2) \\
\\
X_1 | \alpha, \sigma_\epsilon^2 &\sim N\left(0, \frac{\sigma_\epsilon^2}{1 - \alpha^2}\right) \\
\\
X_t | X_{t-1}, \alpha, \sigma_\epsilon^2 &\sim N(\alpha X_{t-1}, \sigma_\epsilon^2)
\end{aligned}
$$

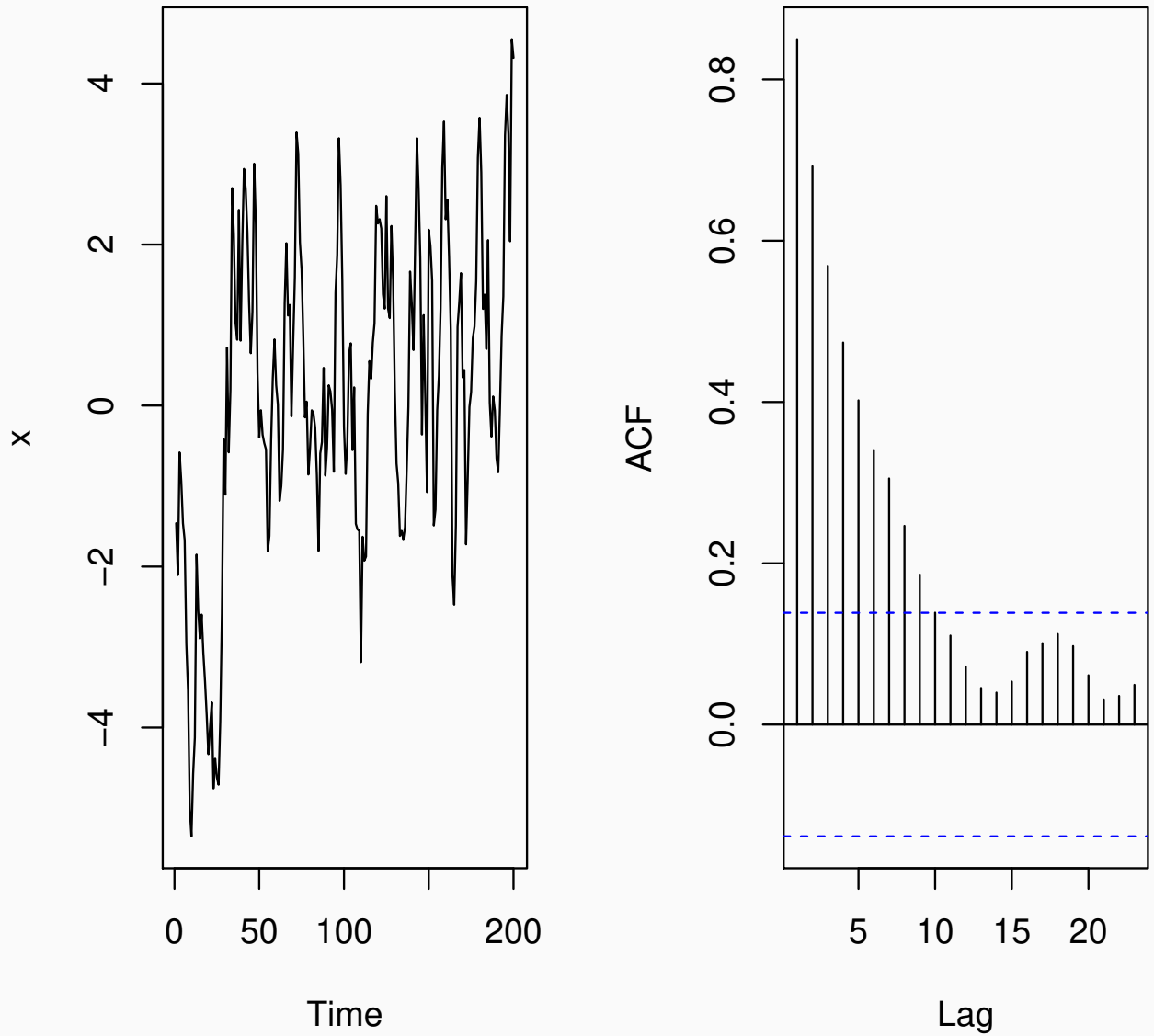
$$
\begin{aligned}
& L(\alpha, \sigma_\epsilon^2) \propto \left(\frac{\sigma_\epsilon^2}{1-\alpha^2}\right)^{-1/2} \\
\times \quad & \exp\left\{-\frac{1-\alpha^2}{2\sigma_\epsilon^2}\, x_1^2\right\} \prod_{t=2}^{n} (\sigma_\epsilon^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}(x_t - \alpha x_{t-1})^2\right\} \\
& \propto \left(\frac{\sigma_\epsilon^2}{1-\alpha^2}\right)^{-1/2} (\sigma_\epsilon^2)^{-(n-1)/2} \exp\left\{-\frac{1-\alpha^2}{2\sigma_\epsilon^2}\, x_1^2\right\} \\
\times \quad & \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\sum_{t=2}^{n}(x_t - \alpha x_{t-1})^2\right\} \\
\propto \quad & (1-\alpha^2)^{1/2}(\sigma_\epsilon^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left((1-\alpha^2)x_1^2 + \sum_{t=2}^{n}(x_t - \alpha x_{t-1})^2\right)\right\}.
\end{aligned}
$$

- Com erros normais e $E(X_t) = 0$ e $Var(X_t) = \sigma_\epsilon^2/(1-\alpha^2)$ foi razoável assumir que,

$$X_1|\alpha, \sigma_\epsilon^2 \sim N\left(0, \frac{\sigma_\epsilon^2}{1-\alpha^2}\right)$$

- As equações de verossimilhança não tem solução analitica e a maximização requer métodos numéricos.

Exemplo. Foram gerados 200 valores de um processo AR(1) com parâmetros $\alpha = 0.8$ e $\sigma^2_\epsilon = 1$.



No R podemos criar uma função com o logaritmo da verossimilhança e usar o pacote optim para fazer a maximização.

```
> args(optim)

function (par, fn, gr = NULL, ..., method = c("Nelder-Mead",
    "BFGS", "CG", "L-BFGS-B", "SANN", "Brent"), lower = -Inf,
    upper = Inf, control = list(), hessian = FALSE)
NULL
```

As estimativas de máxima verossimilhança incondicional são, $\hat{\alpha} = 0.8676$ e $\hat{\sigma}_\epsilon^2 = 1.0534$.

```
> fun <- function(theta,x) {
+  s2 = theta[1]
+  alpha = theta[2]
+  if (abs(alpha)>=1) return(-Inf)
+  n = length(x)
+  e = x[2:n] - alpha * x[1:(n-1)]
+  Q = (1-alpha^2)*x[1]^2 + sum(e^2)
+  return(-0.5*(n*log(s2) -log(1-alpha^2) + Q/s2))
+ }

>  init=c(1,0.5)
>  out=optim(init,fn=fun,control=list(fnscale=-1),x=x)
```

Exemplo. Para os dados do exemplo anterior, sejam as seguintes distribuições para os parâmetros,

$$\begin{aligned} \alpha &\sim N(0,1) \\ \sigma_\epsilon^2 &\sim \mathit{Gama} - \mathit{Inversa}(1,1) \end{aligned}$$

Então,

$$\begin{aligned} p(\alpha) &\propto \exp\left(-\frac{\alpha^2}{2}\right) \\ p(\sigma_\epsilon^2) &\propto (\sigma_\epsilon^2)^{-2} \exp\left(-\frac{1}{\sigma_\epsilon^2}\right) \end{aligned}$$

α e σ_ϵ^2 são independentes a priori.

Pelo Teorema de Bayes temos a densidade conjunta atualizada de α e σ_ϵ^2,

$$
\begin{aligned}
p(\alpha, \sigma_\epsilon^2|\mathbf{x}) &\propto p(\mathbf{x}|\alpha, \sigma_\epsilon^2)\ p(\alpha, \sigma_\epsilon^2) \\
&\propto \exp\left(-\frac{\alpha^2}{2}\right)(\sigma_\epsilon^2)^{-2}\exp\left(-\frac{1}{\sigma_\epsilon^2}\right) \times \\
&\quad (1-\alpha^2)^{1/2}(\sigma_\epsilon^2)^{-n/2} \times \\
&\quad \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left((1-\alpha^2)x_1^2 + \sum_{t=2}^{n}(x_t - \alpha x_{t-1})^2\right)\right\}
\end{aligned}
$$

Podemos criar uma função no R com o logaritmo desta densidade (a menos de uma constante) e usar o pacote optim para fazer a maximização.

```
> prior <- function(theta) {
+  s2 = theta[1]
+  alpha = theta[2]
+  return(-alpha^2/2 - 1/s2 -2*log(s2))
+ }

> post <- function(theta,x) fun(theta,x) + prior(theta)

> out = optim(init,fn=post, control=list(fnscale=-1),x=x)
> out$par

[1] 1.0425583 0.8665942
```

A moda de $p(\alpha, \sigma^2_\epsilon|\mathbf{x})$ é, $\hat{\alpha} = 0.8666$ e $\hat{\sigma}^2_\epsilon = 1.0426$.

Exemplo. Estimando um processo AR(1) com verossimilhança condicional e priori não informativa.

Defina $\phi = \sigma_\epsilon^{-2}$,

$$p(\mathbf{x}|\alpha,\phi) \propto \phi^{(n-1)/2} \exp\left\{-\frac{\phi}{2}\sum_{t=2}^{n}(x_t - \alpha x_{t-1})^2\right\}$$

$$p(\alpha,\phi) \propto 1/\phi$$

Pelo Teorema de Bayes temos a densidade conjunta atualizada de α e ϕ,

$$p(\alpha,\phi|\mathbf{x}) \propto \phi^{(n-1)/2-1} \exp\left\{-\frac{\phi}{2}\sum_{t=2}^{n}(x_t - \alpha x_{t-1})^2\right\},$$

Obtemos a densidade condicional de α dado ϕ e a densidade marginal de ϕ,

$$
\begin{aligned}
p(\alpha|\mathbf{x},\phi) &\propto \exp\left\{-\frac{\phi}{2}\sum_{t=2}^{n}(x_t-\alpha x_{t-1})^2\right\} \\
p(\phi|\mathbf{x}) &= \int p(\alpha,\phi|\mathbf{x})d\alpha \\
&\propto \phi^{(n-1)/2-1}\int \exp\left\{-\frac{\phi}{2}\sum_{t=2}^{n}(x_t-\alpha x_{t-1})^2\right\}d\alpha
\end{aligned}
$$

Após algum algebrismo obtemos que,

$$\begin{aligned}
\alpha|\mathbf{x} &\sim t_{n-1}\left(\frac{\sum x_t x_{t-1}}{\sum x_{t-1}^2}, \frac{1}{\sum x_{t-1}^2}\right) \\
\phi|\mathbf{x} &\sim \mathit{Gama}\left(\frac{n-1}{2}, \frac{1}{2}\sum (x_t - \hat{\alpha} x_{t-1})^2\right) \\
\hat{\alpha} &= \frac{\sum x_t x_{t-1}}{\sum x_{t-1}^2}
\end{aligned}$$

Estimativas: $\hat{\alpha} = 0.8699$ e $\hat{\sigma}_\epsilon^2 = 1.0667$.

Exemplo. No Exemplo com verossimilhança completa com distribuições a priori,

$$\begin{aligned} \alpha &\sim N(0,1) \\ \sigma_\epsilon^2 &\sim \textit{Gama} - \textit{Inversa}(1,1) \end{aligned}$$

podemos obter a distribuição *condicional completa* de σ_ϵ^2 como,

$$p(\sigma_\epsilon^2|\alpha,\mathbf{x}) \propto p(\alpha,\sigma_\epsilon^2|\mathbf{x}) \propto p(\mathbf{x}|\alpha,\sigma_\epsilon^2)p(\sigma_\epsilon^2)$$

Usando o Teorema de Bayes segue que,

$$\begin{aligned} p(\sigma_\epsilon^2|\alpha, \mathbf{x}) &\propto (\sigma_\epsilon^2)^{-2} \exp\left(-\frac{1}{\sigma_\epsilon^2}\right) (\sigma_\epsilon^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma_\epsilon^2} Q(\alpha)\right\} \\ &\propto (\sigma_\epsilon^2)^{-(n/2+2)} \exp\left\{-\frac{1}{\sigma_\epsilon^2}\left[1 + \frac{1}{2} Q(\alpha)\right]\right\} \end{aligned}$$

sendo $Q(\alpha) = (1-\alpha^2)x_1^2 + \sum_{t=2}^{n}(x_t - \alpha x_{t-1})^2$.

Portanto,

$$(\sigma_\epsilon^2|\alpha, \mathbf{x}) \sim \mathit{Gama - Inversa}\left(\frac{n}{2} + 1, 1 + \frac{1}{2} Q(\alpha)\right).$$

Um procedimento MCMC consiste em usar o amostrador de Gibbs para σ_ϵ^2 e propor valores $\alpha^* \sim N(\alpha, c\sigma_\epsilon^2)$,

1. especifique valores iniciais para $(\alpha, \sigma_\epsilon^2)$,
2. para $i = 2, \ldots, N$

   2.1 gere

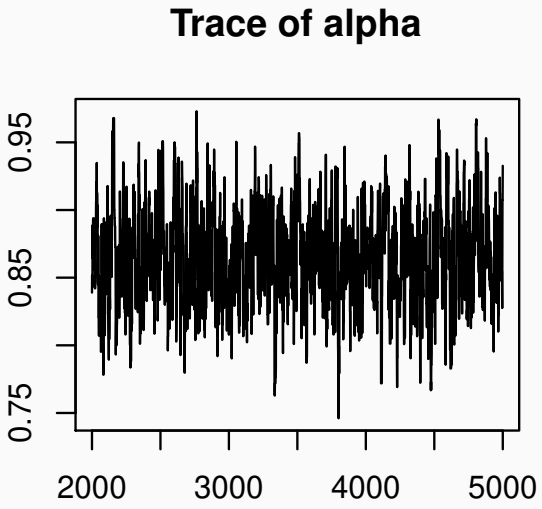
$$(\sigma_\epsilon^{2(i)} | \alpha^{(i-1)}, \mathbf{x}) \sim \textit{Gama} - \textit{Inversa}\left(\frac{n}{2} + 1, 1 + \frac{1}{2} Q(\alpha^{(i-1)})\right)$$

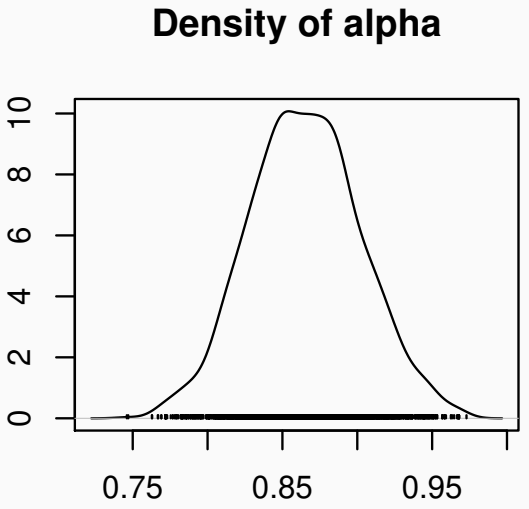
   2.2 gere $\alpha^* \sim N(\alpha^{(i-1)}, c\sigma_\epsilon^{2(i)})$,
   2.3 faça $\alpha^{(i)} = \alpha^*$ com probabilidade,

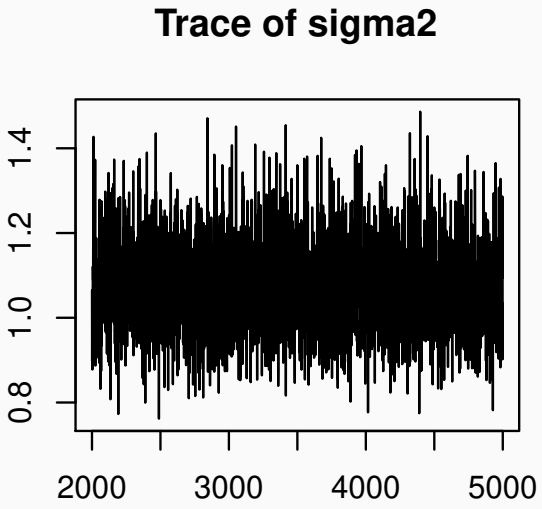
$$\min\left\{1, \frac{p(\alpha^*, \sigma^{2(i)}|\mathbf{x})}{p(\alpha^{(i-1)}, \sigma^{2(i)}|\mathbf{x})}\right\}.$$

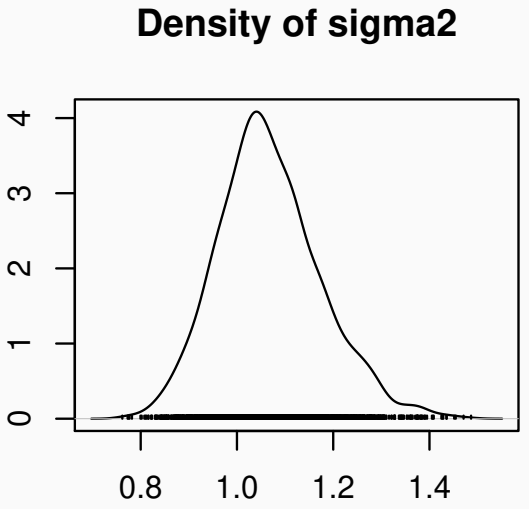
```
Iterations = 2000:5000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 3001

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

          Mean      SD Naive SE Time-series SE
alpha  0.8651 0.03703 0.000676       0.002140
sigma2 1.0670 0.10713 0.001956       0.001956

2. Quantiles for each variable:

          2.5%    25%    50%    75%  97.5%
alpha  0.7928 0.8395 0.8646 0.8894 0.9408
sigma2 0.8760 0.9947 1.0574 1.1293 1.2966
```

	alpha	sigma2
Mean	0.8651	1.0670
SD	0.0370	0.1071
Naive SE	0.0007	0.0020
Time-series SE	0.0021	0.0020
2.5%	0.7928	0.8760
25%	0.8395	0.9947
50%	0.8646	1.0574
75%	0.8894	1.1293
97.5%	0.9408	1.2966


Trace of alpha
0.75
0.85
0.95
2000
3000
4000
5000
Iterations

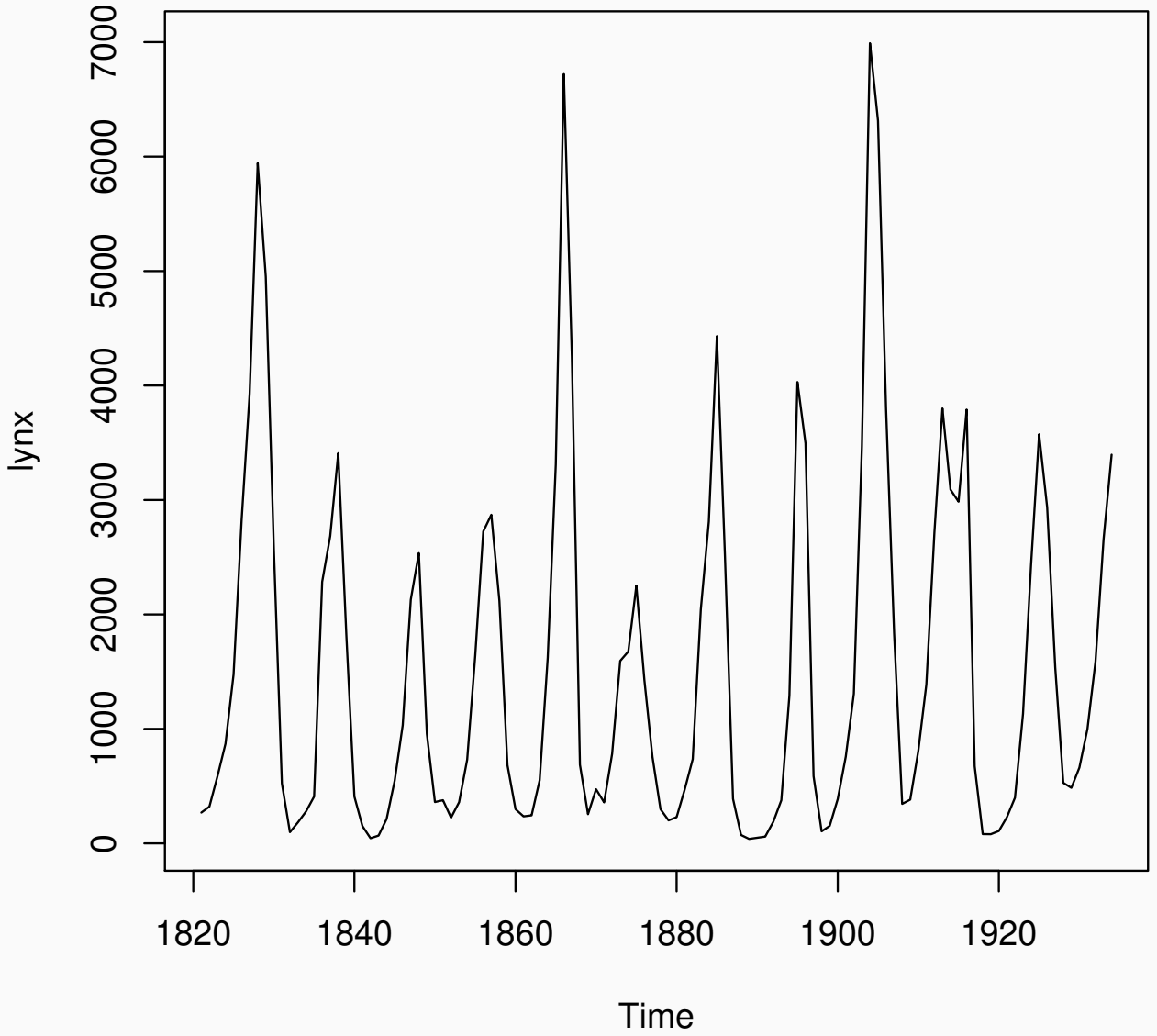

Density of alpha
0
2
4
6
8
10
0.75
0.85
0.95
N = 3001 Bandwidth = 0.007915


Trace of sigma2
0.8
1.0
1.2
1.4
2000
3000
4000
5000
Iterations


Density of sigma2
0
1
2
3
4
0.8
1.0
1.2
1.4
N = 3001 Bandwidth = 0.02147

- As estimativas pontuais nos exemplos são bastante similares.
- Nenhuma restrição de estacionariedade foi imposta na distribuição a priori.
- é possível fazer otimização restrita ou impor esta restrição a priori.
- No AR(1) poderiamos atribuir uma distribuição normal truncada ou uma distribuição uniforme em (-1,1) para o parâmetro α.

Exemplo. Totais anuais de linces canadenses capturados em armadilhas entre 1821 e 1934.



- Estes dados têm sido modelados na literatura após uma transformação que consiste em tomar o logaritmo na base 10 e subtrair a média dos dados transformados.
- Vamos ajustar modelos AR(p) com p variando de 1 até 5 e calcular os critérios de informação e os respectivos pesos para cada modelo.

	p	AIC	pesos AIC	BIC	pesos BIC
1	1	-242.3913	0.0000	-234.9189	0.0000
2	2	-333.0988	0.1057	-321.8902	0.8137
3	3	-332.7283	0.0878	-317.7835	0.1044
4	4	-335.6596	0.3802	-316.9786	0.0698
5	5	-335.8881	0.4263	-313.4709	0.0121

Table 1: Critérios de informação AIC e BIC e respectivos pesos para modelos AR(p) ajustados a série Lynx.

- Há falta de concordância entre os critérios de informação quanto ao melhor modelo.
- Isto pode ser uma indicação de que na verdade há 2 modelos descrevendo bem os dados.
- AIC seleciona um modelo com o valor máximo de p e isto pode indicar a necessidade de considerar mais termos autoregressivos.

	p	AIC	pesos AIC	BIC	pesos BIC
1	1	-242.3913	0.0000	-234.9189	0.0000
2	2	-333.0988	0.0000	-321.8902	0.8100
3	3	-332.7283	0.0000	-317.7835	0.1039
4	4	-335.6596	0.0000	-316.9786	0.0695
5	5	-335.8881	0.0000	-313.4709	0.0120
6	6	-334.4484	0.0000	-308.2950	0.0009
7	7	-338.8427	0.0001	-308.9531	0.0013
8	8	-338.8505	0.0001	-305.2247	0.0002
9	9	-338.3849	0.0001	-301.0229	0.0000
10	10	-341.8678	0.0006	-300.7696	0.0000
11	11	-354.5690	0.3581	-309.7346	0.0019
12	12	-354.7117	0.3846	-306.1411	0.0003
13	13	-353.0609	0.1685	-300.7541	0.0000
14	14	-351.0895	0.0629	-295.0465	0.0000
15	15	-349.2335	0.0249	-289.4543	0.0000

Modelo AR(12) ajustados a série Lynx

	Coef	SE
ar1	1.1159	0.0089
ar2	-0.5143	0.0196
ar3	0.2875	0.0216
ar4	-0.3123	0.0219
ar5	0.1613	0.0225
ar6	-0.1648	0.0225
ar7	0.0759	0.0227
ar8	-0.0699	0.0225
ar9	0.1701	0.0215
ar10	0.1385	0.0210
ar11	-0.1903	0.0193
ar12	-0.1338	0.0094

Ajustando Processos Médias Móveis

- estimação dos parâmetros em modelos MA é bem mais complicado do que em modelos AR.
- Os erros ϵ_t são agora funções não lineares complicadas dos parâmetros $\beta_1, \ldots, \beta_q$
- expressões analíticas para os estimadores não podem ser obtidas.
- métodos computacionais iterativos precisam ser utilizados para minimizar a soma de quadrados residual.

Dado um modelo MA(q)

$$X_t = \mu + \epsilon_t + \beta_1 \epsilon_{t-1} + \cdots + \beta_q \epsilon_{t-q}$$

e uma série observada $x_1, \ldots, x_n$ o procedimento iterativo consiste em fixar os valores de $\mu, \beta_1, \ldots, \beta_q$ e calcular os resíduos

$$e_t = x_t - \mu - \beta_1 \epsilon_{t-1} - \cdots - \beta_q \epsilon_{t-q}$$

sequencialmente para $t = 1, \ldots, n$ assumindo que $\epsilon_0 = \epsilon_{-1} = \cdots = \epsilon_{-q+1} = 0$ e substituindo $\epsilon_{t-1}, \ldots, \epsilon_{t-q}$ pelos residuos calculados. Assim,

$$\begin{aligned} e_1 &= x_1 - \mu \\ e_2 &= x_2 - \mu - \beta_1 e_1 = x_2 - \mu - \beta_1 x_1 + \beta_1 \mu \\ e_3 &= x_3 - \mu - \beta_1 e_2 - \beta_2 e_1 \\ &\vdots \end{aligned}$$

- Dados estes resíduos calcule a soma de quadrados residual $S(\mu, \boldsymbol{\beta}) = \sum_{t=1}^{n} e_t^2$.
- Repita o procedimento para $\mu, \beta_1, \ldots, \beta_q$ variando em uma grade de pontos.
- Escolha os valores que minimizam a soma de quadrados.
- Este procedimento requer o uso de algoritmos eficientes de otimização numérica e nada garante a sua convergência para um mínimo global.

Se $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ a função de verossimilhança fica,

$$\begin{aligned} L(\mu, \boldsymbol{\beta}, \sigma_\epsilon^2) &= \prod_{t=1}^{n} (2\pi\sigma_\epsilon^2)^{-1/2} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} e_t^2 \right\} \\ &\propto (\sigma_\epsilon^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} \sum_{t=1}^{n} e_t^2 \right\}. \end{aligned}$$

e os valores de e_t são calculados como anteriormente.

Portanto $L(\mu, \boldsymbol{\beta}, \sigma_\epsilon^2)$ é uma função não linear dos parâmetros.

- Em termos práticos, se o procedimento de otimização utilizado levar muitas iterações para convergir ou não convergir deve-se "desconfiar" das estimativas.

- Neste caso as estimativas podem ser instáveis no sentido de que adicionando-se ou removendo-se uma ou duas observações pode-se obter valores muito diferentes.

- pode ser computacionalmente mais vantajoso ajustar um modelo AR aos dados mesmo que o modelo resultante tenha mais parâmetros do que o modelo MA sugerido pela função de autocorrelação.

Ajustando Processos ARMA

- Os problemas de estimação para modelos ARMA são similares aqueles para modelos MA.
- um procedimento iterativo precisa ser utilizado.
- Isto ocorre porque os erros $\{\epsilon_t\}$ são funções não lineares complicadas de todos os coeficientes $\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q$.
- os mesmos comentários são válidos para procedimentos que levam muitas iterações para convergir, i.e deve-se "desconfiar" das estimativas.
- Os residuos são calculados de forma análoga ao modelo MA.

Cancelamento de raízes

Seja o modelo ARMA(2,1),

$$X_t = 2\theta X_{t-1} - \theta^2 X_{t-2} - \phi\epsilon_{t-1} + \epsilon_t$$

que pode ser reescrito em termos do operador de retardo como

$$(1 - \theta B)^2 X_t = (1 - \phi B)\epsilon_t.$$

Note como $\theta = \phi$ implica em um modelo AR(1) $X_t = \theta X_{t-1} + \epsilon_t$, ou seja ambos os modelos implicam exatamento no mesmo comportamento para a série temporal X_t.

- Este é um problema de identificação que fica ainda mais complicado em modelos de ordem mais alta.

- Na prática é difícil identificar o cancelamento de raízes (o procedimento iterativo deverá ter convergência lenta).

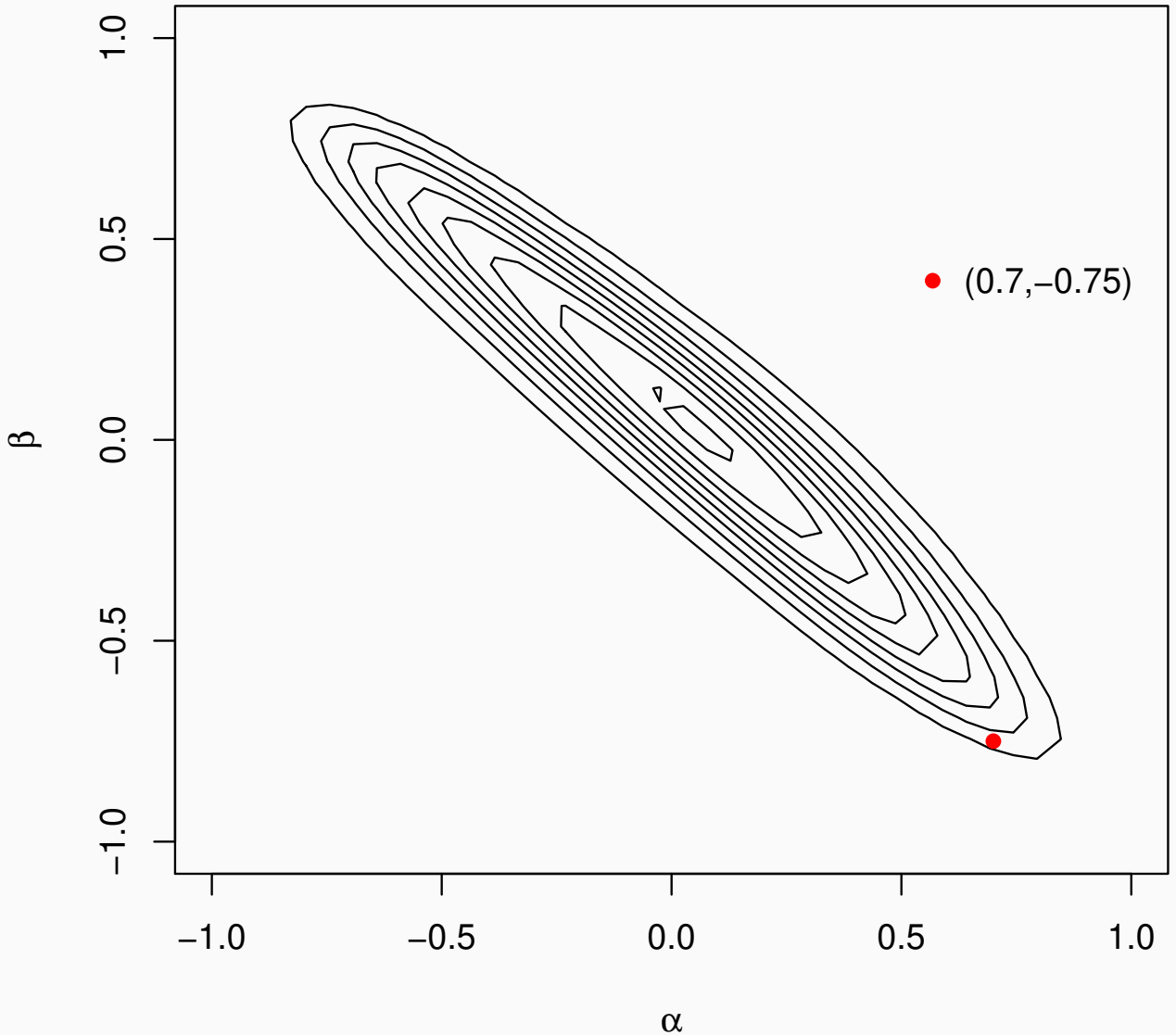
- para tentar minimizar o problema não incluir muitos parâmetros no modelo ARMA(p, q)

Exemplo. Processo ARMA(1,1) simulado com raízes similares ($\alpha = 0.70$ e $\beta = -0.75$). Obteve-se as seguintes estimativas,

```
Call:
arima(x = x, order = c(1, 0, 1), include.mean = F)

Coefficients:
         ar1     ma1
      0.0632  0.0212
s.e.  0.5205  0.5183

sigma^2 estimated as 1.018:  log likelihood = -285.54,  aic = 57
```

Note como as estimativas dos coeficientes estão muito diferentes dos valores verdadeiros e os erros padrões estão enormes!

Verossimilhança condicional do ARMA(1,1) simulado



Modelos ARIMA Sazonais

- Componente sazonal se repete a cada s observações ($s > 1$).
- Com dados mensais e $s = 12$ tipicamente espera-se que X_t dependa de X_{t-12} e talvez de X_{t-24} além de $X_{t-1}, X_{t-2}, \ldots$.
- Tomar a primeira diferença $x_t - x_{t-1}$ não é suficiente para tornar a série (aproximadamente) estacionária.
- Diferenças Sazonais,

$$\nabla_s x_t = (1 - B^s)x_t = x_t - x_{t-s}$$

  sendo s o periodo sazonal.
- A D-ésima diferença sazonal é denotada por ∇_s^D.
- Combinando-se diferenciação simples e sazonais obtem-se o operador $\nabla^d \nabla_s^D$.

Modelo $SARIMA(p, d, q) \times (P, D, Q)_s$

$$\phi(B)\ \Phi(B^s)\ W_t = \theta(B)\ \Theta(B^s)\ \epsilon_t \tag{1}$$

onde

$$\begin{aligned}
\phi(B) &= (1 - \alpha_1 B - \cdots - \alpha_p B^p) \\
\Phi(B^s) &= (1 - \phi_1 B^s - \cdots - \phi_P B^{Ps})
\end{aligned}$$

$$W_t = \nabla^d\ \nabla_s^D\ X_t = (1 - B)^d\ (1 - B^s)^D\ X_t$$

$$\begin{aligned}
\theta(B) &= (1 + \beta_1 B + \cdots + \beta_q B^q) \\
\Theta(B^s) &= (1 + \theta_1 B^s + \cdots + \theta_Q B^{Qs}).
\end{aligned}$$

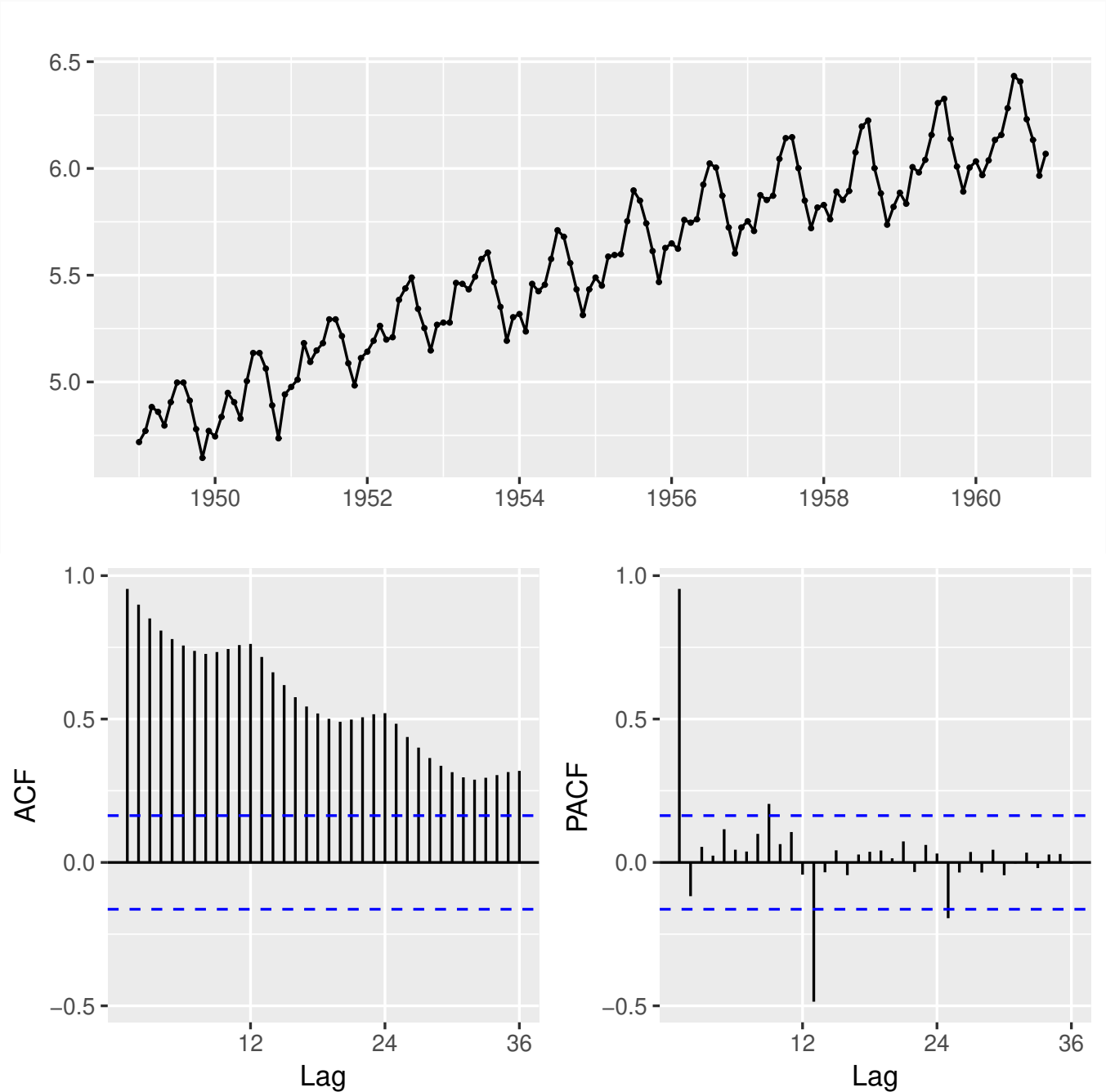
Exemplo. Série mensal com 1 diferença simples e 1 sazonal,

$$\begin{aligned} \nabla\nabla_{12}x_t &= (1-B)(1-B^{12})x_t \\ &= (1-B-B^{12}+B^{13})x_t \\ &= x_t - x_{t-1} - x_{t-12} + x_{t-13} \end{aligned}$$

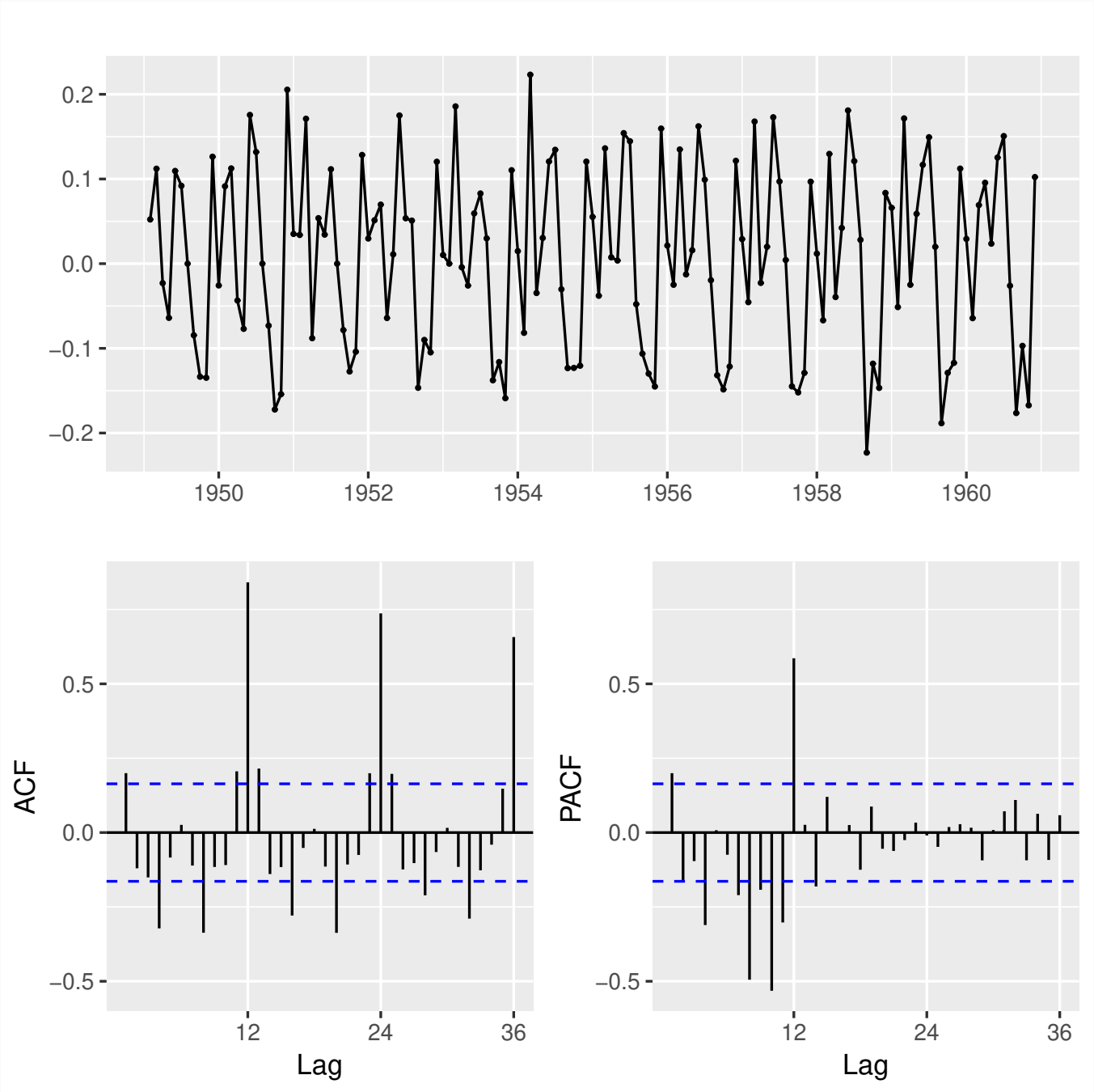
Exemplo. Modelo SARIMA(1,0,0)$\times$ $(0,1,1)_{12}$ para dados mensais.

$$\begin{aligned} (1-\alpha B)(1-B^{12})X_t &= (1+\theta B^{12})\epsilon_t \\ X_t &= X_{t-12} + \alpha(X_{t-1} - X_{t-13}) + \epsilon_t + \theta\epsilon_{t-12}. \end{aligned}$$

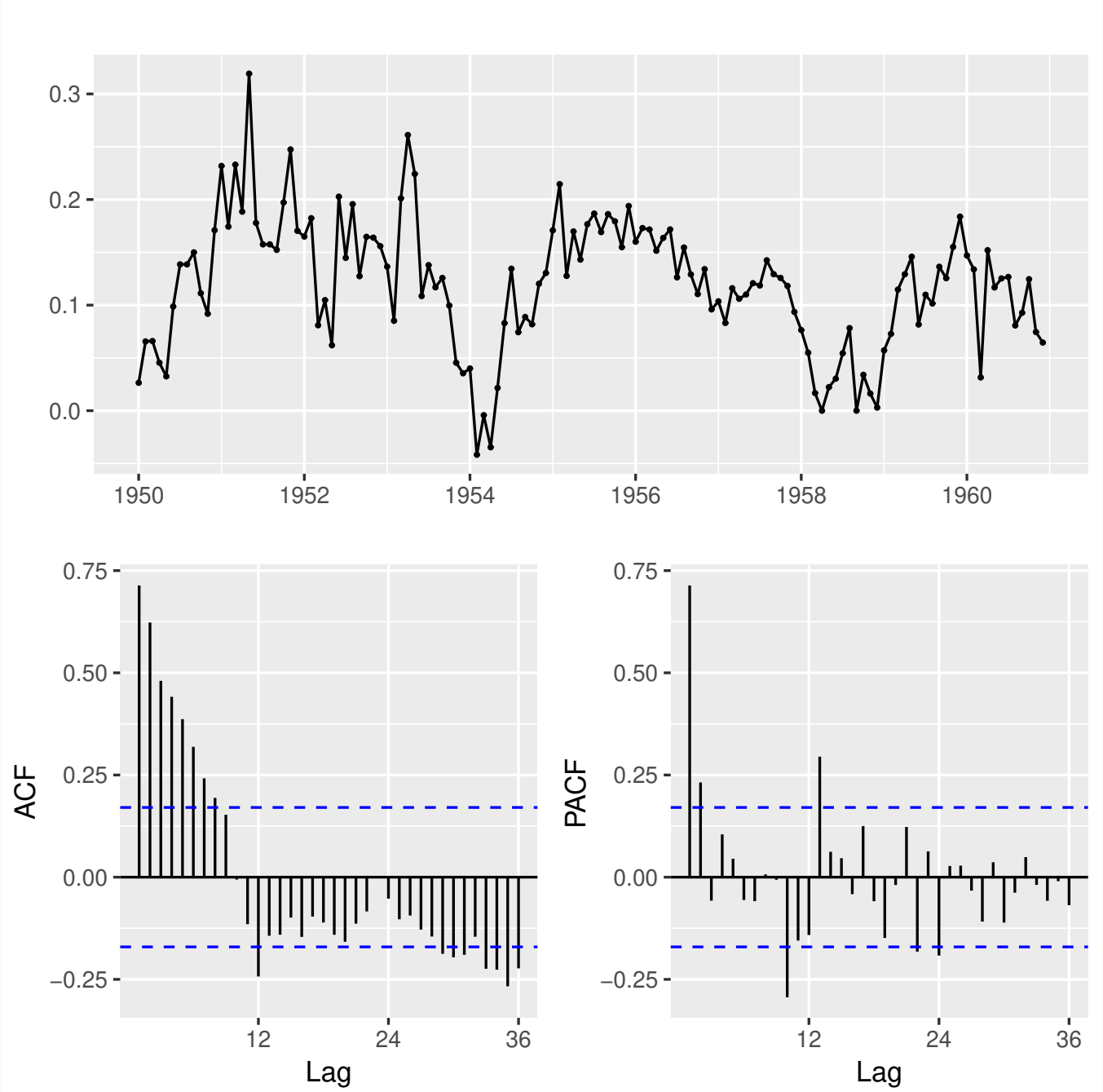
Exemplo. Série com os totais mensais de passageiros em linhas aéreas internacionais nos EUA.



Primeira diferença da série.



Primeira diferença sazonal da série.



- Na prática os valores de d e D em geral não serão maiores do que 1 e um número pequeno de coeficientes será suficiente.

- Especificar os valores de d e D que tornam a série (aproximadamente) estacionária e remove a maior parte da sazonalidade.

- Os valores de p, P, q e Q devem ser especificados com base nas funções de autocorrelação e autocorrelação parcial da série diferenciada.

- Os valores de P e Q são especificados basicamente olhando-se para as defasagens $k = s, 2s, \ldots$.

Adequação do Modelo

Todos os modelos são errados mas alguns são úteis (George Box)

- Após identificar a ordem e estimar os parâmetros de um modelo é necessário verificar sua adequação antes de utilizá-lo por exemplo para fazer previsões.
- Pode-se fazer testes de sobreajustamento (incluir parâmetros extras no modelo e verificar sua significância estatística).
- Em modelos ARMA deve-se incluir um parâmetro de cada vez para evitar o problema de cancelamento de raízes.

Análise dos Resíduos

$$\text{residuo} = \text{observação - valor ajustado.}$$

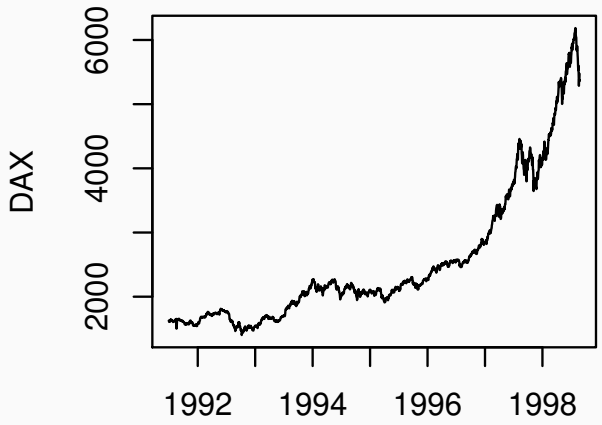
Por exemplo, em um modelo AR(1) se $\hat{\alpha}$ é a estimativa do coeficiente autoregressivo então o valor ajustado no tempo t é $\hat{\alpha}x_{t-1}$ e o resíduo correspondente é

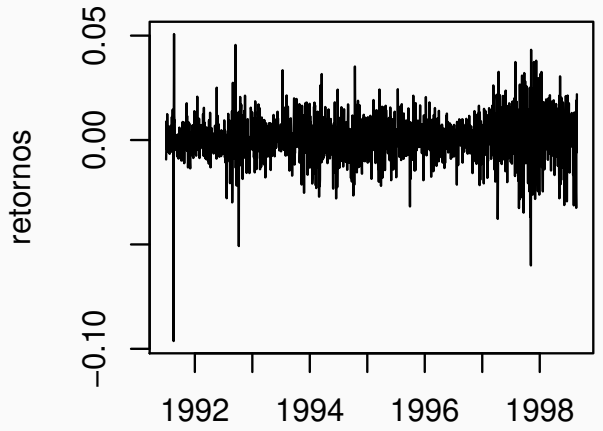
$$e_t = x_t - \hat{\alpha}x_{t-1}.$$

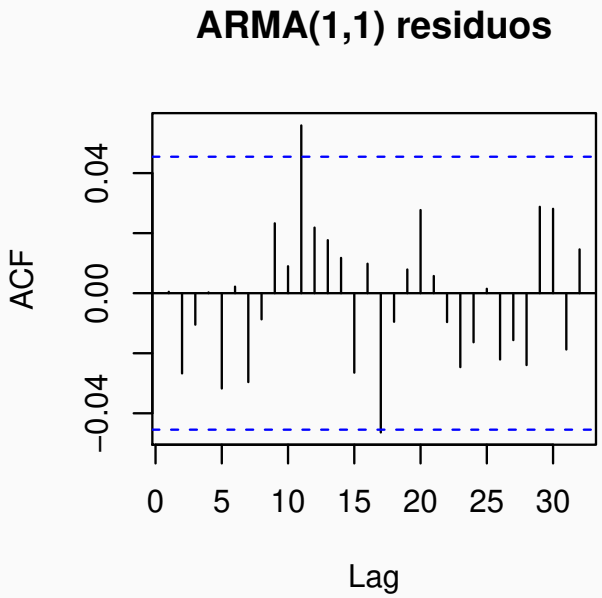
Para o modelo ARMA(p, q),

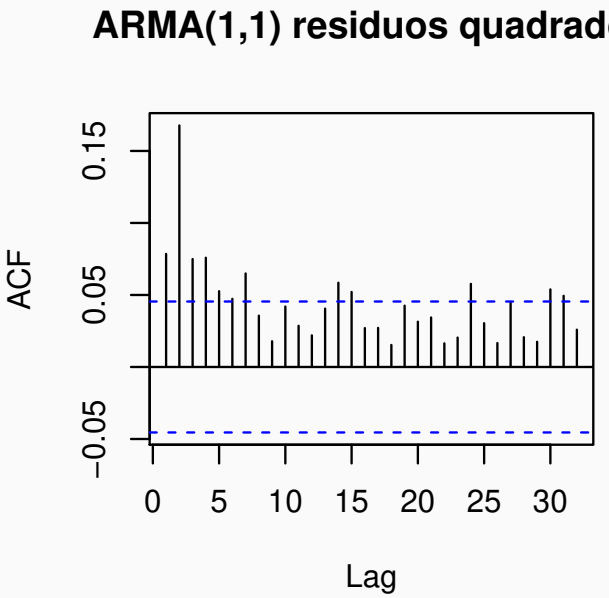
$$e_t = x_t - \hat{\alpha}_1 x_{t-1} - \cdots - \hat{\alpha}_p x_{t-p} - \hat{\beta}_1 e_{t-1} - \cdots - \hat{\beta}_q e_{t-q}$$

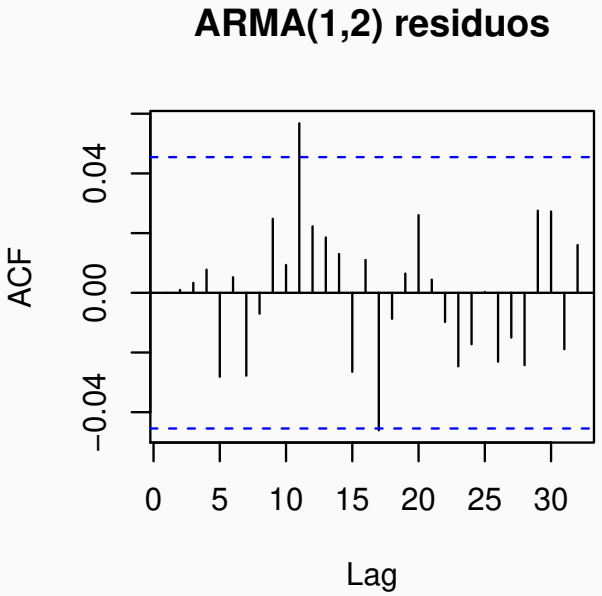
Indice de preços diários de fechamento da bolsa de Frankfurt (DAX).

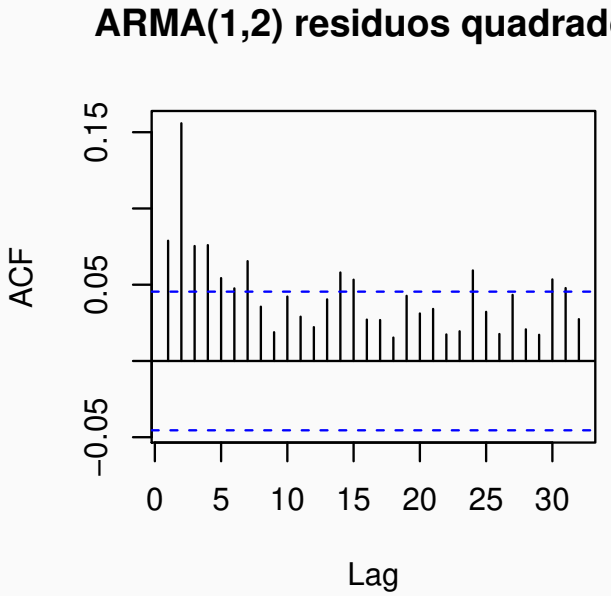


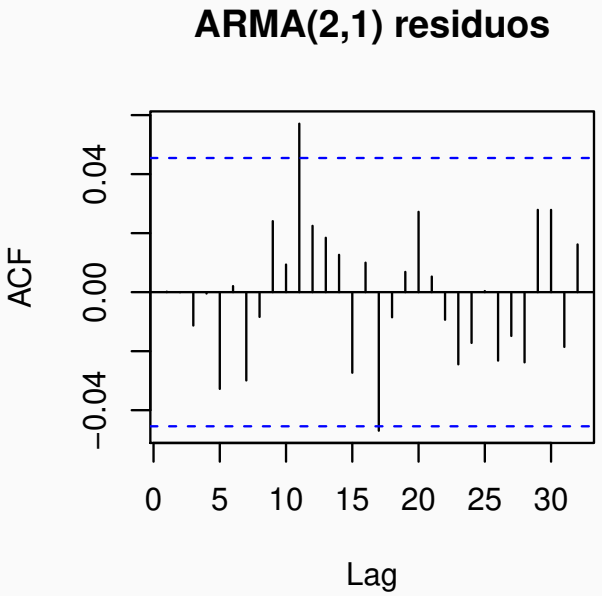


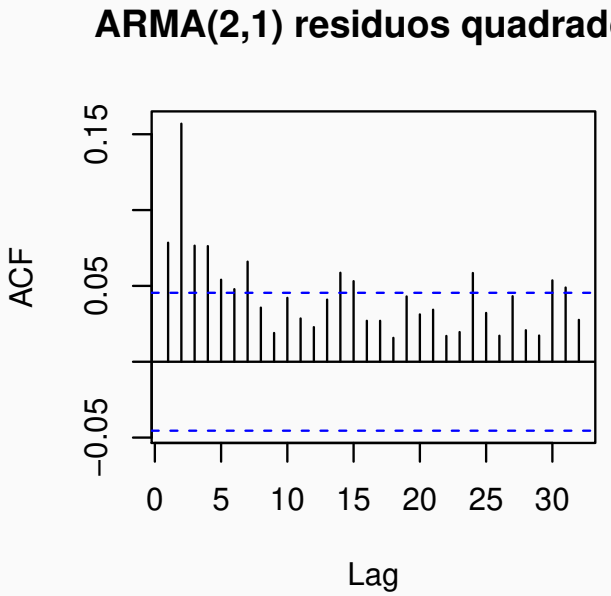


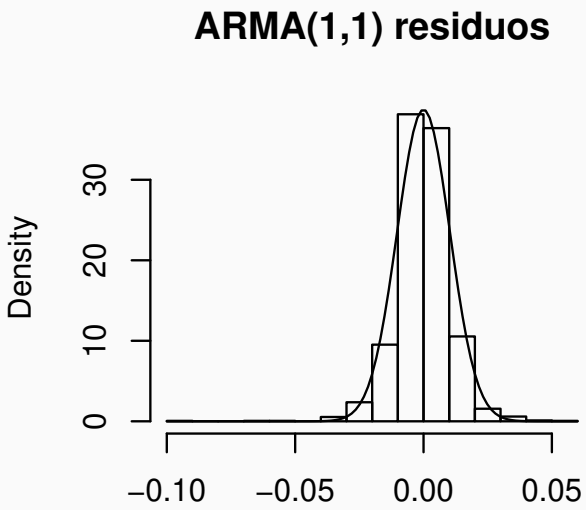


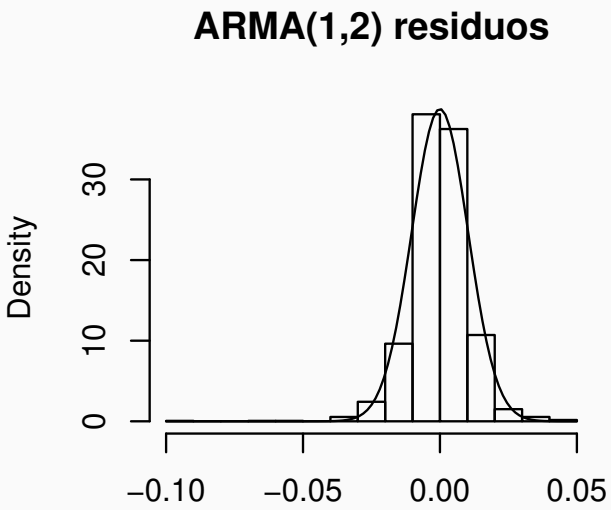

ARMA(1,2) residuos
ACF
0.04
0.00
−0.04
0
5
10
15
20
25
30
Lag

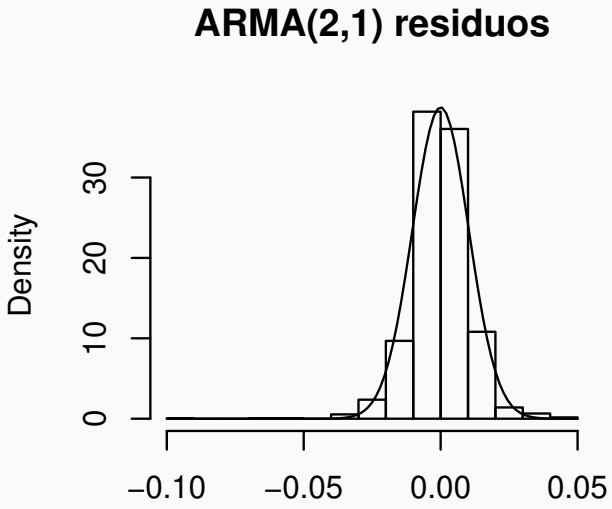

ARMA(1,2) residuos quadrados
ACF
0.15
0.05
−0.05
0
5
10
15
20
25
30
Lag


ARMA(2,1) residuos
ACF
0.04
0.00
−0.04
0
5
10
15
20
25
30
Lag


ARMA(2,1) residuos quadrados
ACF
0.15
0.05
−0.05
0
5
10
15
20
25
30
Lag


ARMA(1,1) residuos
Density
0
10
20
30
−0.10
−0.05
0.00
0.05


ARMA(1,2) residuos
Density
0
10
20
30
−0.10
−0.05
0.00
0.05


ARMA(2,1) residuos
Density
0
10
20
30
−0.10
−0.05
0.00
0.05

- Se o modelo tiver um bom ajuste espera-se que os resíduos se distribuam aleatoriamente em torno de zero com variância aproximadamente constante e sejam não correlacionados.

- Se a variância dos resíduos for crescente uma transformação logarítmica nos dados pode ser apropriada.

- O fenômeno de "não constância" na variância é denominado de volatilidade na literatura de séries temporais e pode ser tratado através de transformações nos dados (e.g. transformações de Box-Cox).

- Uma tendência mais recente consiste em tentar modelar simultaneamente a média e a variância ao invés de usar transformações.

- em modelos de séries temporais os resíduos estão ordenados no tempo e é portanto natural tratá-los também como uma série temporal.

- É importante que os resíduos de um modelo estimado sejam serialmente (i.e. ao longo do tempo) não correlacionados.

- Evidência de correlação serial nos resíduos é uma indicação de que uma ou mais características da série não foi adequadamente descrita pelo modelo.

Testes sobre os resíduos

Ao invés de olhar para as autocorrelações residuais individualmente pode-se testar se um grupo de autocorrelações é significativamente diferente de zero.

Para modelos ARMA sugere-se o uso do teste de Box-Pierce para as hipóteses,

$$H_0 : \rho(1) = \cdots = \rho(m) = 0$$
$$H_1 : \rho(k) \neq 0, \text{para algum } k \in \{1, \ldots, m\}.$$

sendo a estatística de teste dada por

$$Q = n \sum_{k=1}^{m} r_k^2.$$

- Na prática o número m de autocorrelações amostrais é tipicamente escolhido entre 15 e 30.
- Se o modelo ajustado for apropriado então Q terá distribuição aproximadamente qui-quadrado com $m - p - q$ graus de liberdade.
- valores grandes de Q fornecem indicação contra a hipótese de que as autocorrelações são todas nulas, em favor da hipótese de que ao menos uma delas é diferente de zero.

O teste de Box-Pierce não tem bom desempenho em amostras pequenas ou moderadas (a distribuição se afasta da qui-quadrado).

Testes alternativos foram sugeridos e o mais conhecido é o teste de Ljung-Box, cuja estatística de teste é,

$$Q = n(n+2)\sum_{k=1}^{m} \frac{r_k^2}{n-k}.$$

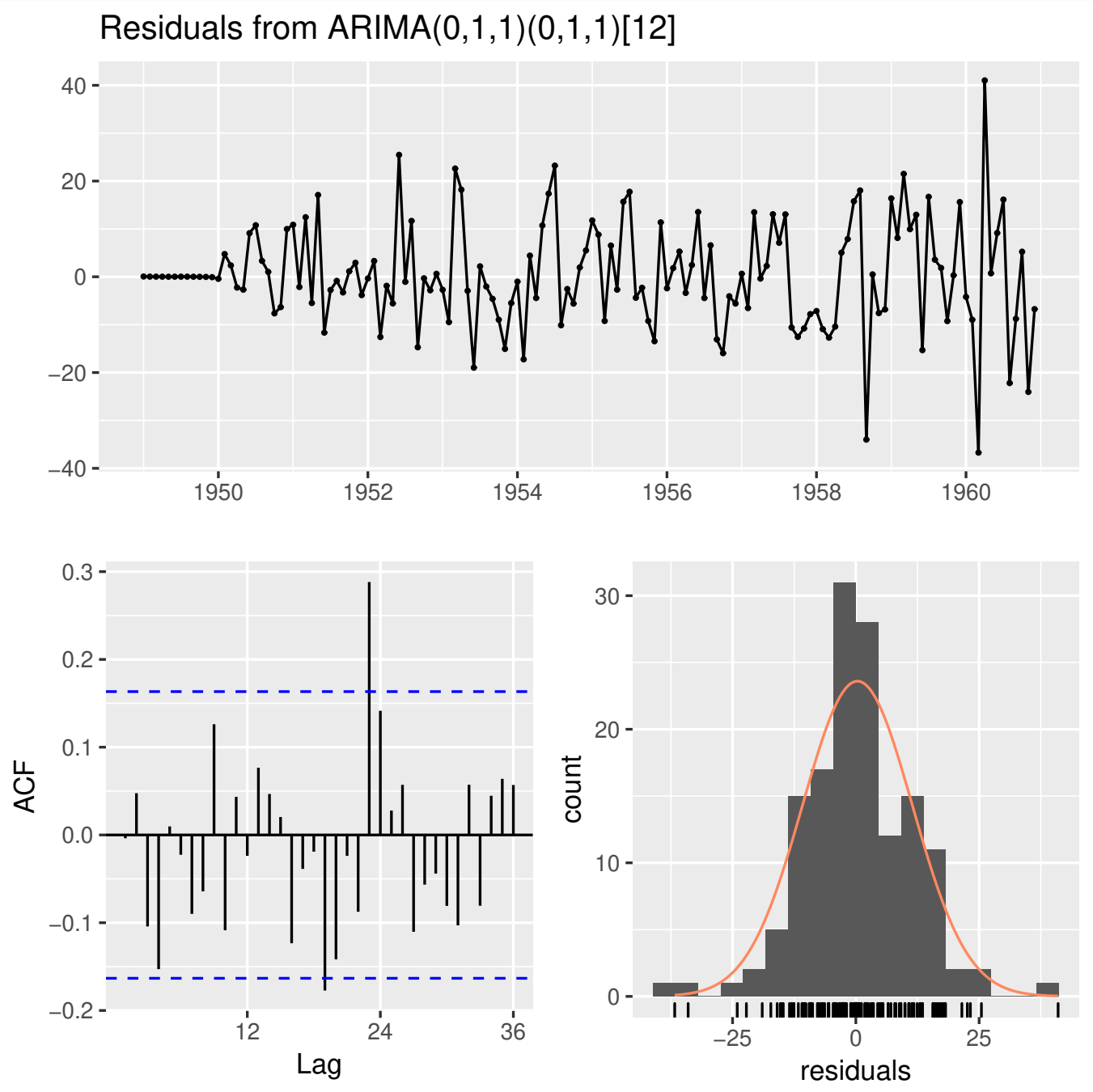
Sua distribuição amostral também é aproximadamente qui-quadrado com $m - p - q$ graus de liberdade.

No R podemos usar a função Box.test.

```
> args(Box.test)

function (x, lag = 1, type = c("Box-Pierce", "Ljung-Box"), fitdf = 0)
NULL
```

Exemplo. Ajustando um modelo SARIMA(0,1,1)$\times$(0,1,1) à série `AirPassengers`. P-valores dos testes Box-Pierce e Ljung-Box nos residuos.

m-p-q	Box-Pierce	Ljung-Box
1	0.1688	0.1617
2	0.0721	0.0649
3	0.1528	0.1396
4	0.2535	0.2344
5	0.2596	0.2359
6	0.3113	0.2822
7	0.2255	0.1934
8	0.1965	0.1620
9	0.2516	0.2098
10	0.3239	0.2750


Residuals from ARIMA(0,1,1)(0,1,1)[12]
40
20
0
−20
−40
1950
1952
1954
1956
1958
1960
0.3
0.2
0.1
0.0
−0.1
−0.2
ACF
12
24
36
Lag
30
20
10
0
count
−25
0
25
residuals

Testando a Normalidade dos Resíduos

Para uma variável aleatória X tal que $E(X) = \mu$ e $Var(X) = \sigma^2$ define-se os coeficientes de assimetria e curtose como,

$$A(X) = E\left(\frac{(X-\mu)^3}{\sigma^3}\right) \quad \text{e} \quad K(X) = E\left(\frac{(X-\mu)^4}{\sigma^4}\right)$$

A distribuição normal tem assimetria 0 e curtose igual a 3.

Substituindo momentos teóricos pelos seus equivalente amostrais,

$$m_j = \frac{1}{n}\sum_{t=1}^{n}(X_t - \overline{X})^j$$

os estimadores da assimetria e curtose são dados por

$$\hat{A} = \frac{m_3}{\sqrt{m_2^3}} \quad \text{e} \quad \hat{K} = \frac{m_4}{m_2^2}$$

Sob a hipótese de normalidade as variáveis aleatórias $\sqrt{n/6}\hat{A}$ e $\sqrt{n/24}(\hat{K}-3)$ são independentes e têm distribuição assintótica $N(0,1)$.

A estatística

$$\frac{n\hat{A}^2}{6} + \frac{n(\hat{K}-3)^2}{24}$$

tem distribuição assintótica χ^2 com 2 graus de liberdade e pode ser usada para testar a normalidade de X.

- O *Teste de Jarque-Bera* usa esta estatistica para testar normalidade.
- A hipótese nula é que assimetria é igual a zero e curtose é igual a 3. Ou equivalentemente, que o excesso de curtose é igual a zero.
- No `R` podemos usar a função `jarque.bera.test` do pacote `tseries`.
- Existem outros testes de normalidade como teste de Shapiro-Wilk, teste de Kolmogorov–Smirnov, etc.

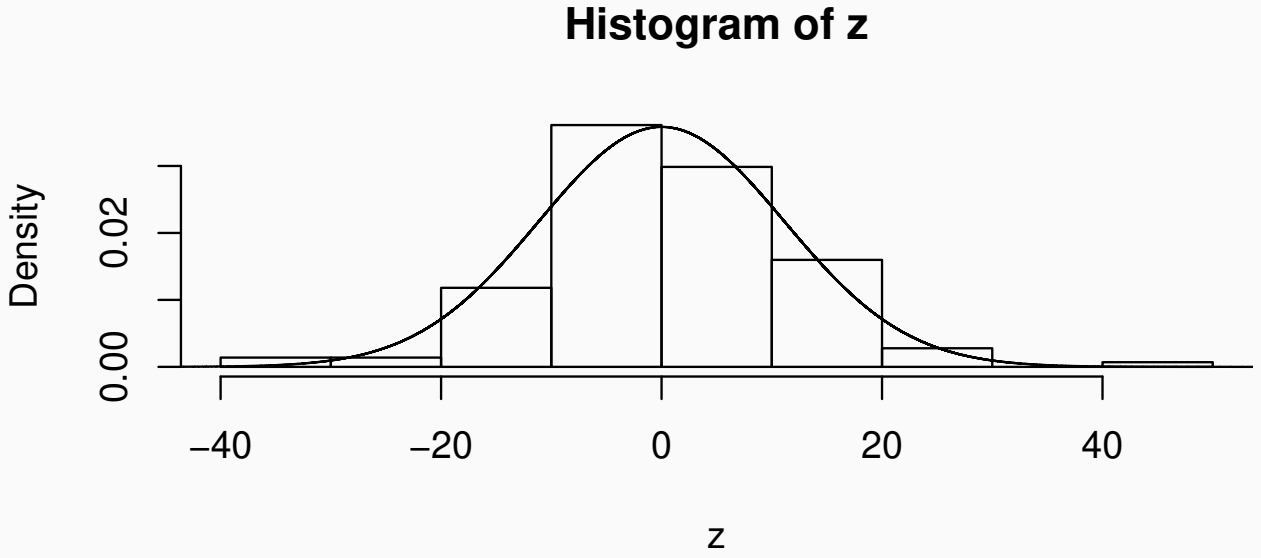
Exemplo. Análise de residuos do modelo SARIMA(0,1,1)×(0,1,1) ajustado à série `AirPassengers`.

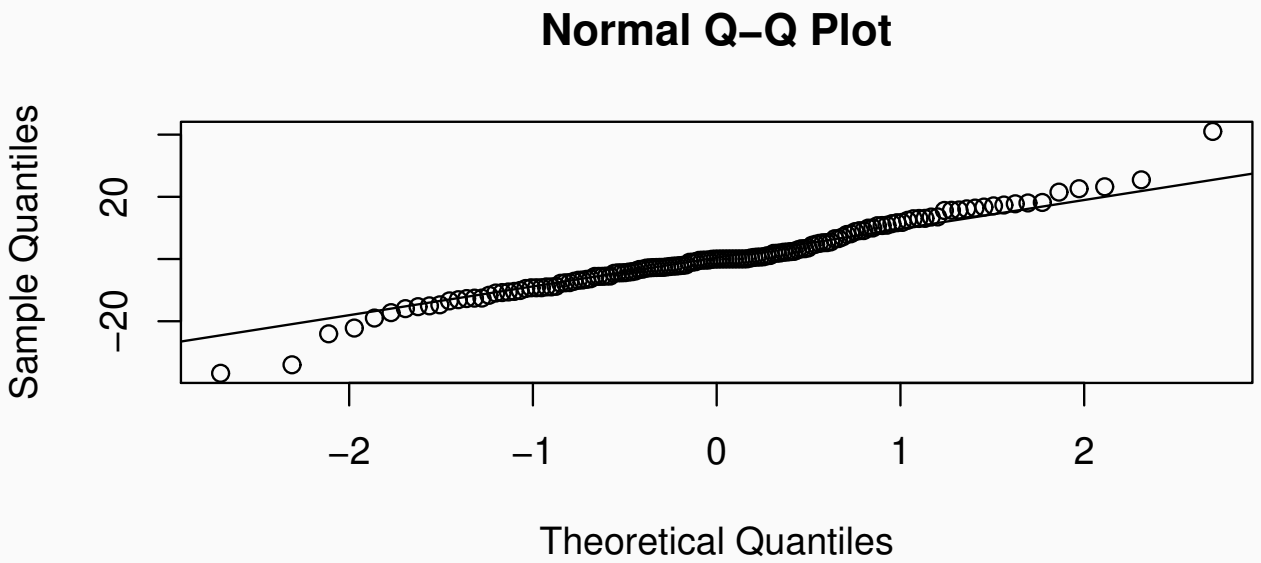
Teste de normalidade de Shapiro-Wilk,

```
        Shapiro-Wilk normality test

data:  Residuals
W = 0.97603, p-value = 0.0125
```

Teste de normalidade de Jarque-Bera,

```
        Jarque Bera Test

data:  Residuals
X-squared = 12.481, df = 2, p-value = 0.001949
```


Histogram of z
Density
0.00
0.02
−40
−20
0
20
40
z


Normal Q−Q Plot
Sample Quantiles
−20
20
−2
−1
0
1
2
Theoretical Quantiles