

## 2

### Séries Temporais e Modelos Auto-Regressivos

Nesse capítulo, são apresentadas brevemente algumas definições básicas de teoria de probabilidades, para conveniência do leitor. Algumas referências para o assunto são: DeGroot (9), Casella (6), Jonhston e Dinardo (16), Parzen (30), Papoulis (29) e Bruce (4).

#### 2.1

##### Definições básicas

Os modelos utilizados para descrever séries temporais são processos estocásticos, isto é, processos controlados por leis probabilísticas. Assim, é necessário rever alguns conceitos fundamentais da teoria de probabilidade.

##### Probabilidade

Suponha que um experimento seja realizado sob certas condições fixas. Denota-se por  $\Omega$  o conjunto de resultados possíveis. O conjunto  $\Omega$  será chamado *espaço amostral* do experimento, que é um conjunto não-vazio qualquer.

Um evento  $A$  ao qual atribuímos uma probabilidade é chamado de *evento aleatório*. Suponha que a classe dos eventos aleatórios possua certas propriedades intuitivas, que são essenciais para a teoria e para o cálculo das probabilidades. Indica-se com  $\mathbb{A}$  a classe dos eventos aleatórios e seja  $A^c$  o complemento do evento  $A$ . Estipula-se as seguintes propriedades para  $\mathbb{A}$ :

- P1.  $\Omega \in \mathbb{A}$  (define-se  $P(\Omega) = 1$ ).
- P2. Se  $A \in \mathbb{A}$ , então  $A^c \in \mathbb{A}$ .
- P3. Se  $A \in \mathbb{A}$  e  $B \in \mathbb{A}$ , então  $A \cup B \in \mathbb{A}$ .

Seja  $\Omega$  um conjunto não-vazio. Uma classe  $\mathbb{A}$  de subconjuntos de  $\Omega$  satisfazendo P1, P2 e P3 é chamada *álgebra* de subconjuntos de  $\Omega$ .

Se  $\mathbb{A}$  é *uma álgebra de subconjuntos de  $\Omega$* , então valem as seguintes propriedades:

- P4.  $\emptyset \in \mathbb{A}$  e

- P5.  $\forall n, \forall A_1, \dots, A_n \in \mathbb{A}$ , tem-se  $\bigcup_{i=1}^n A_i \in \mathbb{A}$  e  $\bigcap_{i=1}^n A_i \in \mathbb{A}$ .

Pelas propriedades acima, conclui-se que uma álgebra é fechada para um número *finito* de aplicações das operações  $\bigcup$ ,  $\bigcap$ , e  $^c$ . Sem perda de generalidade, suponha que a classe dos eventos aleatórios também satisfaça a propriedade

- P6. Se  $A_n \in \mathbb{A}$  para  $n = 1, 2, 3, \dots$ , então  $\bigcup_{n=1}^{\infty} A_n \in \mathbb{A}$ .

Uma classe  $\mathbb{A}$  de subconjuntos de um conjunto não-vazio  $\Omega$  satisfazendo as propriedades P1, P2 e P6 é chamada  $\sigma$ -álgebra de subconjuntos de  $\Omega$ .

Pode-se atribuir probabilidades em uma certa  $\sigma$ -álgebra  $\mathbb{A}$  de eventos, chamados eventos aleatórios; suponha que a todo  $A_n \in \mathbb{A}$  seja associado um número real  $P(A)$ , chamado *probabilidade* de  $A$ , de modo que os axiomas a seguir sejam satisfeitos.

- **Axioma 1.**  $P(A) \geq 0$ .
- **Axioma 2.**  $P(\Omega) = 1$ .
- **Axioma 3.** (Aditividade finita).

Se  $A_1, \dots, A_n \in \mathbb{A}$  são disjuntos (2 a 2), então  $P(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n P(A_k)$ .

Um *espaço de probabilidade* é uma tripla  $(\Omega, \mathbb{A}, P)$ , onde

- i)  $\Omega$  é um conjunto não-vazio,
- ii)  $\mathbb{A}$  é uma  $\sigma$ -álgebra de subconjuntos de  $\Omega$ , e
- iii)  $P$  é uma probabilidade em  $\mathbb{A}$ .

### Teoria Assintótica

Um conceito importante tratado neste texto é o da *teoria assintótica*. Suponha que  $X$  seja uma variável aleatória com alguma função de densidade de probabilidade (*pdf*) desconhecida. Admite-se que a *pdf* tem uma média  $\mu$  e uma variância  $\sigma^2$  finitas. Recolha uma amostra aleatória desta distribuição  $(x_n)$  e designe a média da amostra por  $\bar{x}_n$ , onde o índice  $n$  indica o número de observações nas quais a média se baseia. Denota-se por  $(f(\bar{x}_n))$  a *pdf* da variável aleatória média amostral  $\bar{x}_n$ . A questão fundamental para a teoria assintótica é saber como uma variável, por exemplo  $\bar{x}_n$ , e a sua *pdf*  $(f(\bar{x}_n))$  se comportam quando  $n \rightarrow \infty$ . Há dois aspectos principais deste comportamento, a serem considerados: o primeiro relaciona-se com a *convergência em probabilidade*, e o segundo *convergência em distribuição*.

**Convergência em probabilidade** - considere por hipótese que as variáveis  $x$  são independentes e identicamente distribuídas (i.i.d.)  $(\mu, \sigma^2)$ , assim,  $E(\bar{x}_n) = \mu$  e  $\text{var}(\bar{x}_n) = \frac{\sigma^2}{n}$ . Um estimador não tendencioso de  $\mu$  é  $\bar{x}_n$  para qualquer dimensão da amostra, e a variância tende para zero quando  $n$  cresce infinitamente, isto é, quando  $n$  tende para infinito. Formalmente, escolhe-se um  $\varepsilon$  e define-se uma vizinhança em torno de  $\mu$  como

$$P\{\mu - \varepsilon < \bar{x}_n < \mu + \varepsilon\} = P\{|\bar{x}_n - \mu| < \varepsilon\},$$

que representa a probabilidade de  $\bar{x}_n$  pertencer ao intervalo especificado. O intervalo pode ser tomado arbitrariamente pequeno pela escolha adequada de  $\varepsilon$ . A variância,  $\text{var}(\bar{x}_n)$ , tende monotonamente para zero quando  $n$  cresce, ou seja, existe um número  $n^*$  natural e um  $\delta$  ( $0 < \delta < 1$ ), tais que, para todo  $n > n^*$ , tem-se

$$P\{|\bar{x}_n - \mu| < \varepsilon\} > 1 - \delta.$$

Desta forma, pode-se dizer que a variável aleatória  $\bar{x}_n$  converge em probabilidade para a constante  $\mu$ , uma expressão equivalente é  $\lim_{n \rightarrow \infty} P\{|\bar{x}_n - \mu| < \varepsilon\} = 1$ .

**Convergência em distribuição** - O *Teorema Central do Limite* (TCL) estabelece que

$$\lim_{n \rightarrow \infty} P\left\{\frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma} \leq y\right\} = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz. \quad (2-1)$$

O primeiro membro desta equação é o valor limite para a probabilidade de que a variável aleatória  $\frac{\sqrt{n}(\bar{x}_n - \mu)}{\sigma}$  seja menor ou igual a algum valor  $y$ . O segundo membro é a área apropriada sob a *distribuição normal padrão*,  $N(0, 1)$ . Qualquer que seja a variável aleatória  $X_i$  e a forma de  $f(X_i)$ , a distribuição limite da estatística relevante é normal padrão. Estatisticamente, fala-se de *convergência em distribuição*, pode-se escrever a equação (2-1) como

$$\sqrt{n}\bar{x}_n \xrightarrow{d} N(\sqrt{n}\mu, \sigma^2).$$

Se  $Y_1, Y_2, \dots, Y_t$  e  $t = 1, 2, \dots, N$  é uma seqüência de variáveis aleatórias (i.i.d.) com média  $\mu$  e variância  $\sigma^2$ , e pode-se considerar a média de  $Y_t$  como  $E(\bar{Y}_t) = \frac{1}{N} \sum_{t=1}^N Y_t$ , então, pelo Teorema Central do Limite obtém-se

$$\sqrt{N}(\bar{Y}_N - \mu) \xrightarrow{d} N(0, \sigma^2).$$

## Série temporal e processo estocástico

Uma *série temporal* é um conjunto de observações de uma variável ao longo do tempo. As observações podem ser medidas de forma contínua ou em intervalos de tempo discretos. No caso discreto, os intervalos conseguem ser igualmente espaçados ou não. Representa-se o conjunto de valores do tempo pelo símbolo  $\mathbb{T}$ .

Um *processo estocástico*  $Y = \{Y_t : \Omega \rightarrow \mathbb{R}^\nu, \nu \in \mathbb{N}, t \in \mathbb{T}\}$  é uma coleção de variáveis aleatórias (vetoriais ou escalares) definidas em um espaço de probabilidades  $(\Omega, \mathbb{A}, \mathbb{P})$  fixo. Dependendo da natureza de  $\mathbb{T}$ , trata-se de processo estocástico *discreto* ou *contínuo*. O valor da variável  $Y_t$  é o *estado* do processo no instante  $t$ . O conjunto de todos os valores possíveis da variável aleatória  $Y_t$  é o *espaço de estados* de um processo estocástico, que também pode ser discreto ou contínuo.

Informalmente, um processo estocástico é um mecanismo gerador de dados (MGD), cujo comportamento não pode ser descrito de forma determinística. A estrutura probabilística de um processo estocástico possibilita completamente a definição por meio da especificação da sua distribuição de probabilidade conjunta  $\mathbb{F}(Y_1, \dots, Y_t) = P(y_1 \leq Y_1, \dots, y_t \leq Y_t)$ . A partir da distribuição conjunta obtém-se qualquer informação probabilística sobre o processo.

Define-se uma série temporal  $Y_t$  como uma realização de um processo estocástico desconhecido. Denota-se por  $y_t$  o valor observado da variável aleatória  $Y_t$ . Uma classe de processos estocásticos muito importante é a classe de *processos estacionários*. Num processo estocástico estacionário, algumas características do processo permanecem as mesmas ao longo do tempo.

Um processo estocástico  $Y = \{Y_t, t \in \mathbb{T}\}$  é *estritamente estacionário*, quando a distribuição conjunta do processo permanece invariante sob uma translação temporal, ou seja,  $\mathbb{F}(Y_1, Y_2, \dots, Y_t) = \mathbb{F}(Y_{1+k}, Y_{2+k}, \dots, Y_{t+k}), \forall k$ .

Um processo estocástico  $Y = \{Y_t, t \in \mathbb{T}\}$  é *estacionário de segunda ordem* quando a média e a variância da variável  $Y_t$  são constantes no tempo, e a correlação entre as variáveis  $Y_t$  depende apenas da distância entre os tempos e diminui com esta distância:

$$E[Y_t] = \mu, \quad \text{Var}[Y_t] = \sigma^2, \quad E[(Y_t - \mu)(Y_{t-k} - \mu)] = \gamma_k, \quad \gamma \text{ decrescente.}$$

Aqui cabe uma observação notacional, que será utilizada ao longo desse texto. Para tempos discretos, a variável  $t$  não designa o valor do tempo, e, sim, um índice. Assim, um processo estocástico discreto pode ter sido avaliado

nos tempos 1.5, 3.1, 5.7, 10.13. Na notação adotada, os quatro números correspondem a escolhas de  $t$  indo de 1 a 4.

Suponha que se tenha uma *série histórica*, isto é, uma série temporal obtida empiricamente. O objetivo de um *modelo*, no sentido desse texto, é poder realizar outras séries temporais que sejam de algum modo semelhantes à série histórica dada. Mais precisamente, cada modelo é determinado por certo conjunto de parâmetros, que são ajustados a partir da série histórica. O modelo assim determinado é o ponto de partida para obter séries temporais que emulam a série histórica.

## 2.2

### Os Modelos Auto-Regressivos

#### 2.2.1

##### Introdução

Em 1970, os estatísticos G. E. P. Box e G. M. Jenkins (2) apresentaram uma metodologia para o desenvolvimento de modelos de previsão em séries temporais e controle. Esse trabalho foi embasado em pesquisas anteriores: Yule, em 1926, introduziu os modelos auto-regressivos (AR), enquanto os modelos de médias móveis (MA) foram introduzidos por Slutsky em 1937. Em 1938, Wold mostrou que qualquer processo estocástico estacionário discreto pode ser representado por modelos auto-regressivos e de médias móveis (ARMA). Box e Jenkins, por sua vez, desenvolveram toda uma técnica de identificação, estimação de parâmetros e verificação de validade do modelo. Essa técnica sofisticada fornece previsões com grande embasamento probabilístico e menor erro de estimativa. Em 1999, foi apresentada uma atualização recente do texto original por Box, Jenkins e Reinsel (3).

A estratégia para a descrição nesse texto do modelo de Box e Jenkins será baseada em um ciclo iterativo, no qual a escolha da estrutura do modelo é baseada nos próprios dados. Os estágios do ciclo iterativo são:

1. Uma classe geral de modelos é considerada para análise (*especificação*);
2. Há uma *identificação* de um modelo, com base na análise de autocorrelações, autocorrelações parciais e outros critérios;
3. Uma fase de *estimação* dos parâmetros do modelo identificado;
4. Finalmente, há a *verificação* ou *diagnóstico* do modelo ajustado, através de uma análise de resíduos, para saber se este é adequado para os fins em vista (previsão, por exemplo).

### 2.2.2

#### Um resumo da metodologia de Box e Jenkins

Neste trabalho, considera-se apenas o *modelo auto-regressivo*  $AR(p)$ . Em linhas gerais, a especificação e a identificação consistem na determinação da ordem  $p$ , número natural. Em geral, os modelos também fazem uso de *parâmetros*, a valores reais, em sua definição. Nesse texto, todos os modelos considerados são *paramétricos*, isto é, são definidos a partir de um número finito de parâmetros.

A especificação de ordens e parâmetros se realiza através da análise da série histórica, que precisa ter comprimento maior ou igual a 50 para que os resultados sejam satisfatórios. Como ferramenta de identificação, Box e Jenkins sugeriram a comparação das funções de autocorrelação (FAC) e da função de autocorrelação parcial (FACP) (definidas abaixo) da série histórica com os valores teóricos dessas funções para esses modelos.

Suponha que uma série histórica, obtida empiricamente, na verdade seja a saída de um processo estocástico discreto estacionário. Sejam  $t$  e  $t - k$  dois tempos desse processo. O *lag* (*variação do índice ou defasagem*) entre esses dois tempos é  $j - i$ . Nesse caso, a *autocovariância* entre as observações em  $t$  e  $t - k$ , que em geral é dada por

$$\gamma_k = \text{cov}(Y_t, Y_{t-k}) = E[(Y_t - E[Y_t])(Y_{t-k} - E[Y_{t-k}])],$$

que depende apenas do *lag*, e vale

$$\gamma_k = E[(Y_t - \mu)(Y_{t-k} - \mu)],$$

já que, pela estacionariedade do processo,  $E[Y_t] = E[Y_{t-k}] = \mu$ .

A *função de autocovariância* revela a dependência entre os diferentes termos de uma mesma série. Para comparar séries diferentes, é mais conveniente utilizar a *função de autocorrelação FAC* ( $\rho_{k,t}$ ), que é dada pela razão da covariância pela variância. A *FAC* representa uma medida padrão da dependência entre os intervalos de tempo (atrasos), ou seja, é uma medida de associação linear entre o processo e o seu passado, uma medida de “memória”:

$$\rho_{k,t} = \frac{E[(Y_t - E[Y_t])(Y_{t-k} - E[Y_{t-k}])]}{\sqrt{E[(Y_t - E[Y_t])^2]} \cdot \sqrt{E[(Y_{t-k} - E[Y_{t-k}])^2]}}$$

Para processos estacionários de segunda ordem, a função de autocorrelação é definida pela seguinte equação

$$\rho_{k,t} \equiv \rho_k = \frac{E[(Y_t - \mu)(Y_{t-k} - \mu)]}{\sigma^2} = \frac{\gamma_k}{\sigma^2},$$

da própria definição de variância,  $\rho_0 = \frac{\gamma_0}{\sigma^2} = 1$ . Pela estacionariedade,  $\rho_k = \rho_{-k}$ . Vale que, pela desigualdade de Cauchy-Schwartz,  $|\rho_k| \leq 1$  e que a igualdade ocorre exatamente quando há uma relação linear perfeita entre  $Y_t$  e  $Y_{t-k}$ .

À representação gráfica da *FAC* de um processo estacionário dá-se o nome de *correlograma*. Correlogramas não distinguem aspectos importantes de séries temporais: um critério mais sutil faz uso da função de autocorrelação parcial.

A *função de autocorrelação parcial (FACP)* é um instrumento utilizado para facilitar o procedimento de identificação do modelo. O coeficiente de autocorrelação parcial de ordem ( $k$ ), usualmente representado por  $\phi_{kk}$ , mede a correlação entre  $y_t$  e  $y_{t-k}$  depois que a influência de  $y_{t-1}, y_{t-2}, \dots, y_{t-k+1}$ , os termos intermediários sobre  $y_t$ , foram eliminados. Formalmente, a FACP é definida por

$$\phi_{kk} = \text{Corr}[(Y_t - E[Y_t|Y_{t-1}, \dots, Y_{t-k+1}])(Y_{t-k} - E[Y_{t-k}|Y_{t-1}, \dots, Y_{t-k+1}])].$$

### 2.2.3

#### O Ruído Branco Gaussiano

**Definição 1** Diz-se que  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  é um ruído branco, se for composto por uma seqüência de variáveis aleatórias independentes de média  $E(\varepsilon_t) = 0$  e variância  $E(\varepsilon_t^2) = \sigma^2$  e descorrelatadas temporalmente,  $E(\varepsilon_t \varepsilon_\tau) = 0$ , para  $t \neq \tau$ . Se cada variável  $\{\varepsilon_t\}_{t \in \mathbb{Z}}$  for normalmente distribuída  $N(0, \sigma^2)$ , o ruído branco é dito *gaussiano*.

A FAC de um ruído branco gaussiano é

$$\rho_k = \begin{cases} 1, & \text{se } k = 0; \\ 0, & \text{se } k \neq 0. \end{cases} \quad (2-2)$$

A função de autocorrelação parcial, de um ruído branco é

$$\phi_{kk} = \begin{cases} 1, & \text{se } k = 0; \\ 0, & \text{se } k \neq 0. \end{cases} \quad (2-3)$$

### 2.2.4

#### Movimento Browniano Padrão

**Definição 2** Chama-se de *Movimento Browniano Padrão* ao processo contínuo  $\{W(t), 0 \leq t \leq 1\}$  tal que:

- (i)  $W(0) = 0$ ;
- (ii)  $W(s) - W(t) \stackrel{d}{\sim} N(0, s - t)$ ;
- (iii) para quaisquer instantes  $0 \leq t_1 \leq \dots \leq t_k \leq 1$ , as variáveis aleatórias

$W(t_2) - W(t_1), W(t_3) - W(t_2), \dots, W(t_k) - W(t_{k-1})$  são independentes;  
 (iv) as trajetórias de  $W(t)$  são contínuas com probabilidade um.

A média de  $W(t)$  é  $E(W(t)) = 0$  e a variância é  $E(W^2(t)) = t$  para cada tempo  $t \geq 0$ .

### 2.2.5 Operador LAG

O operador de *lag*, representado por  $L$ , dá o valor anterior na sucessão cronológica. Tem-se, assim,

$$\begin{aligned} L(y_t) &= y_{t-1} \\ L^2(y_t) &= L[L(y_t)] = L(y_{t-1}) = y_{t-2} \\ L^k y_t &= y_{t-k} \\ (1 - L)y_t &= y_t - y_{t-1} = \Delta y_t \\ L(1 - L)y_t &= y_{t-1} - y_{t-2} = \Delta y_{t-1}. \end{aligned} \quad (2-4)$$

Em muitas manipulações algébricas, o operador *lag* pode ser tratado como um escalar. Uma das operações mais importantes é o cálculo da inversa de um expressão em  $L$ . Por exemplo, seja  $A(L) = 1 - \alpha L$  um polinômio de primeira ordem em  $L$ . Considere a multiplicação

$$(1 - \alpha L)(1 + \alpha L + \alpha^2 L^2 + \alpha^3 L^3 + \dots + \alpha^p L^p) = 1 - \alpha^{p+1} L^{p+1}.$$

Quando  $p \rightarrow \infty$ ,  $\alpha^{p+1} L^{p+1} \rightarrow 0$ , desde que  $|\alpha| < 1$ . Pode-se, então, escrever o recíproco ou inverso de  $A(L)$ , da seguinte forma

$$A^{-1}(L) = \frac{1}{(1 - \alpha L)} = 1 + \alpha L + \alpha^2 L^2 + \alpha^3 L^3 + \dots \quad (2-5)$$

### 2.2.6 O modelo auto-regressivo de ordem $p$ $AR(p)$

Sejam  $\varepsilon_t$  um ruído branco gaussiano,  $(\phi_1, \phi_2, \dots, \phi_p)$  coeficientes reais e  $c$  é uma constante real. O processo

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (2-6)$$

é chamado de *modelo auto-regressivo de ordem  $p$* , denotado por  $AR(p)$ . Neste modelo, o valor presente do processo  $Y_t$  depende de um conjunto finito de valores passados até uma defasagem ( $p$ ), e de uma componente de erro  $\varepsilon_t$ .

Condicionalmente aos valores passados, a dependência linear compõe a parte determinística do modelo (localização da distribuição), enquanto que o erro determina o formato e a dispersão da distribuição condicional. Seu nome provém do fato de que modelos em que uma variável depende linearmente de outras e de uma componente de erro são chamados de modelos de regressão.

Seja  $L$  o *lag* que retarda de uma unidade temporal, ou seja,  $LY_t = Y_{t-1}$ . O polinômio  $\Phi_p(L)$  do modelo auto-regressivo de ordem  $p$  é definido a partir de  $L$  pela expressão

$$\Phi_p(L) = (1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p), \quad (2-7)$$

denominado de *operador auto-regressivo*.

Note que os expoentes designam os retardos na variável temporal. Com essa notação, pode-se representar a equação (2-6) através da forma reduzida  $\Phi_p(L)Y_t = c + \varepsilon_t$ . O modelo auto-regressivo de ordem  $p$  contém  $p + 2$  parâmetros desconhecidos ( $c; \phi_1, \phi_2, \dots, \phi_p$  e  $\sigma^2$ ), que devem ser estimados a partir do dados.

Assumindo que a condição de estacionariedade do processo  $AR(p)$  seja satisfeita, a **média** de  $Y_t$  é igual a

$$\mu = c + \phi_1 \mu + \phi_2 \mu + \dots + \phi_p \mu = \frac{c}{1 - (\phi_1 + \phi_2 + \dots + \phi_p)}.$$

Com a finalidade de se calcular as autocovariâncias, escreve-se a equação a seguir, usando a relação entre a média e a constante  $c$  calculada acima, resultando na expressão

$$Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) \dots + \phi_p(Y_{t-p} - \mu) + \varepsilon_t. \quad (2-8)$$

As **autocovariâncias** são obtidas multiplicando-se ambos os lados da equação (2-8) por  $(Y_{t-j} - \mu)$  e tomando-se a esperança:

$$\gamma_j = \begin{cases} \theta_1 \gamma_{j-1} + \theta_2 \gamma_{j-2} + \dots + \theta_p \gamma_{j-p}, & \text{para } j = 1, 2, \dots, \\ \theta_1 \gamma_1 + \theta_2 \gamma_2 + \dots + \theta_p \gamma_p + \sigma^2, & \text{para } j = 0. \end{cases} \quad (2-9)$$

Assim, a **variância** do modelo  $AR(p)$  é dada pela equação  $\gamma_0 = \sum_{t=1}^p \phi_t \gamma_t + \sigma^2$  e a **autocovariância** é  $\gamma_k = \sum_{t=1}^p \phi_t \gamma_{k-i}$ .

A seguir, estão listadas as principais propriedades estatísticas do processo  $AR(p)$ . Uma descrição mais detalhada pode ser encontrada em Box, Jenkins e Reinsel (3) e Hamilton (12).

- Estacionariedade: Um processo  $AR(p)$  é estacionário se as raízes da equação  $\Phi_p(L) = 0$  estiverem todas fora do círculo unitário.
- Invertibilidade: Os processos auto-regressivos são sempre inversíveis.

- Função de autocorrelação (FAC): A função de autocorrelação de um processo  $AR(p)$  tem a forma de uma exponencial amortecida ou de uma senóide amortecida.
- Função de autocorrelação parcial (FACP) de um modelo  $AR(p)$  tem uma quebra brusca após o defasamento  $p$ .

### Modelo auto-regressivo de primeira ordem, AR(1)

Considera-se o caso especial  $AR(1)$ , em que só se emprega uma defasagem do modelo anterior,

$$Y_t = c + \phi Y_{t-1} + \varepsilon_t \quad (2-10)$$

onde  $c$ ,  $\phi$  são constantes e assumindo que  $|\phi| < 1$ , o processo  $Y_t$  é estacionário. Agora, a média é  $E(Y_t) = \frac{c}{1-\phi}$  e a variância é  $E(Y_t - \mu)^2 = \gamma_0 = \frac{\sigma^2}{1-\phi^2}$ . A função de autocovariância de  $Y_t$  é  $\gamma_1 = \left[ \frac{\phi}{1-\phi^2} \right] \cdot \sigma^2$  para  $j = 1$ . Finalmente, a função de autocorrelação é  $\rho_1 = \frac{\gamma_1}{\gamma_0} = \phi$ .

#### 2.2.7

#### Modelos não-lineares

Os modelos lineares descritos anteriormente não são adequados para descrever o comportamento de séries financeiras, que apresentam a heteroscedasticidade condicional, ou seja, a variância condicional varia com o tempo.

Essas séries necessitaram de modelos específicos para descrever a evolução da volatilidade no tempo. Os modelos adequados para esse fim são a classe de modelos ARCH (*Autoregressive Conditional Heteroskedasticity*) e os modelos GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*) que são utilizados nesse estudo.

#### Modelos ARCH( $p$ )

Os modelos ARCH( $p$ ), ou modelos auto-regressivos com heteroscedasticidade condicional, foram introduzidos por Engle (11), com objetivo de estimar a variância da inflação. A idéia principal é que o retorno  $Y_t$  não seja correlacionado serialmente, mas a volatilidade (variância condicional) depende dos retornos passados por meio de uma função quadrática.

O processo

$$\begin{aligned} Y_t &= h_t^{\frac{1}{2}} \varepsilon_t, \\ h_t &= \alpha_0 + \alpha_1 Y_{t-1}^2 + \dots + \alpha_p Y_{t-p}^2 \end{aligned} \quad (2-11)$$

é chamado de *modelo auto-regressivo com heteroscedasticidade condicional de ordem p*, denotado por ARCH(p). Neste modelo,  $\varepsilon_t$  é um ruído branco com média zero e variância um, ou seja,  $\varepsilon_t \sim N(0, 1)$ .

O modelo ARCH(p) será estacionário se  $\alpha_0 > 0$  e  $\alpha_1 \dots \alpha_p \geq 0$ . Os parâmetros da equação (2-11) são estimados pelo método da máxima verossimilhança. Uma descrição mais detalhada sobre este método de estimação pode ser encontrado em Hamilton (12), Casella (6) e Johnston (15), (16).

Seja  $Y_t$  o valor de uma série temporal e  $I_{t-1}$  a informação disponível no instante  $t - 1$ , tem-se que a **média** de  $Y_t$ , o processo ARCH(p), é igual a

$$E(Y_t | I_t) = \mu_t$$

e a **variância** é igual a

$$Var(Y_t | I_{t-1}) = E[(Y_t - \mu_t)^2 | I_{t-1}] = h_t,$$

onde  $\mu_t$  e  $h_t$  são funções das componentes de  $I_{t-1}$ . Se  $h_t$  não for constante, então, a série  $Y_t$  é heteroscedástica.

Considera-se o caso especial o ARCH(1), em que só se emprega uma defasagem do modelo anterior:

$$\begin{aligned} Y_t &= h_t^{\frac{1}{2}} \varepsilon_t, \\ h_t &= \alpha_0 + \alpha_1 Y_{t-1}^2 \end{aligned} \tag{2-12}$$

com  $\alpha_0 > 0, \alpha_1 \geq 0$ .

Note que as médias condicionadas do modelo podem obter a equação

$$E_{t-1}(Y_t) = [\alpha_0 + \alpha_1 Y_{t-1}^2]^{\frac{1}{2}} E_{t-1}(\varepsilon_t) = 0,$$

resultando em  $E_{t-2}[E_{t-1}(Y_t)] = E_{t-2}(0) = 0$  e, de forma sucessiva, para todos os períodos anteriores. Portanto, a média não condicionada de  $Y_t$  é  $E(Y_t) = 0$ .

Para calcular a variância, tem-se  $Y_t^2 = \varepsilon_t^2[\alpha_0 + \alpha_1 Y_{t-1}^2]$  e, portanto,  $E_{t-1}(Y_t^2) = \sigma_\varepsilon^2[\alpha_0 + \alpha_1 Y_{t-1}^2] = \alpha_0 + \alpha_1 Y_{t-1}^2$ .

Analogamente,

$$\begin{aligned} E_{t-2}E_{t-1}(y_t^2) &= E_{t-2}[\alpha_0 + \alpha_1 Y_{t-1}^2] \\ &= \alpha_0 + \alpha_1 E_{t-2}(Y_{t-1}^2) \\ &= \alpha_0 + \alpha_1(\alpha_0 + \alpha_1 Y_{t-2}^2) \\ &= \alpha_0 + \alpha_0 \alpha_1 + \alpha_1^2 Y_{t-2}^2. \end{aligned} \tag{2-13}$$

Procedendo do mesmo modo, verifica-se que  $E_0 \dots E_{t-2} E_{t-1}(Y_t^2) = \alpha_0(1 + \alpha_1 + \alpha_1^2 + \dots + \alpha_1^{t-1}) + \alpha_1^t Y_0^2$  e desde que  $|\alpha_1| < 1$ , pode-se calcular o limite quando  $t \rightarrow \infty$ , a variância não condicionada é igual a

$$\text{Var}(Y_t) = \sigma^2 = \frac{\alpha_0}{1 - \alpha_1}.$$

A variância será positiva se  $\alpha_0 > 0$ . O processo é, assim, homoscedástico. A autocovariância de primeira ordem condicionada é

$$E_{t-1}(Y_t Y_{t-1}) = Y_{t-1} E_{t-1}(Y_t) = 0.$$

Assim, todas as autocovariâncias de ordem superior são também iguais a zero. Uma descrição mais detalhada destes modelos pode ser encontrada em Hamilton (12) e Jonhston (16).

### Modelos GARCH( $p, q$ )

A generalização dos modelos ARCH( $p$ ), onde a variância condicional é uma média móvel dos resíduos ao quadrado e dos valores passados da variância condicional, foi proposta por Bollersslev (1), denominada modelo GARCH ( $p, q$ ) ou generalização dos modelos auto-regressivos com heteroscedasticidade condicional.

O processo

$$\begin{aligned} Y_t &= h_t^{\frac{1}{2}} \varepsilon_t, \\ h_t &= \alpha_0 + \sum_{i=1}^p \alpha_i Y_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j} \end{aligned} \quad (2-14)$$

é chamado de *modelo auto-regressivo com heteroscedasticidade condicional de ordem ( $p, q$ )*, denotado por GARCH( $p, q$ ). Neste modelo,  $\varepsilon_t$  é um ruído branco com média zero e variância um, ou seja,  $\varepsilon_t \sim N(0, 1)$ ,  $p \geq 0$ ,  $q > 0$ ,  $\alpha_0 > 0$ ,  $\alpha_i \geq 0, i = 1, \dots, p$ ,  $\beta_j \geq 0, j = 1, \dots, q$ .

O modelo GARCH( $p, q$ ) será estacionário se, e somente se,  $\sum_{i=1}^q (\alpha_i + \beta_i) < 1$ . Os parâmetros da equação (2-14) são estimados pelo método da máxima verossimilhança. A média do modelo  $Y_t$  é  $E(Y_t) = E(Y_t | I_{t-1}) = 0$  e a variância é  $E(Y_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^q (\alpha_i + \beta_i)}$ .

Considera-se o caso especial do modelo o *GARCH (1,1)* cuja equação é

$$\begin{aligned} Y_t &= h_t^{\frac{1}{2}} \varepsilon_t, \\ h_t &= \alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1 h_{t-1}, \end{aligned} \quad (2-15)$$

com  $\alpha_0 > 0$ ,  $\alpha_1 \leq 0$ ,  $\beta_1 < 1$ ,  $\alpha_1 + \beta_1 < 1$ . Agora, a média de  $Y_t$  é  $E(Y_t) = E(Y_t|I_{t-1}) = 0$  e a variância é  $E(Y_t^2) = \frac{\alpha_0}{1-(\alpha_1+\beta_1)}$ . Uma descrição mais detalhada desses modelos pode ser encontrada em Hamilton (12) e Jonhston (16).

## 2.3

### Regressão Quantílica

Muitas aplicações estatísticas são mostradas e elaboradas através dos modelos lineares de regressão, associados ao método de estimação de mínimos quadrados condicional à média. A *regressão quantílica* desenvolvida por Koenker e Basett (21) em 1978 oferece uma estratégia sistemática para examinar que covariáveis influenciam na localização, escala e na forma do modelo responsável pela distribuição.

Todo valor real da variável aleatória  $X$  pode ser caracterizado pela função de distribuição

$$F(x) = P(X \leq x), \quad (2-16)$$

sendo que para qualquer  $\tau$ , ( $0 < \tau < 1$ ), tem-se que

$$F^{-1}(\tau) = \inf\{x : F(x) \geq \tau\}, \quad (2-17)$$

é chamado de  $\tau$ -ésimo quantil de  $X$ .

A função perda é descrita pela seguinte função linear

$$\rho_\tau(u) = \begin{cases} \tau u, & \text{se } u \geq 0; \\ (\tau - 1)u, & \text{se } u < 0, \end{cases} \quad (2-18)$$

para algum  $\tau \in [0, 1]$ , deve-se achar  $\hat{x}$  que minimiza a perda esperada.

Em 1964, Fox e Rubin estudaram admissibilidade de um estimador quantílico sobre a função de perda. Procuram minimizar

$$E\rho_\tau(X - \hat{x}) = (\tau - 1) \int_{-\infty}^{\hat{x}} (x - \hat{x})dF(x) + \tau \int_{-\hat{x}}^{\infty} (x - \hat{x})dF(x). \quad (2-19)$$

Diferenciando-se em relação  $\hat{x}$ , obtém-se

$$0 = (\tau - 1) \int_{-\infty}^{\hat{x}} dF(x) - \tau \int_{-\hat{x}}^{\infty} dF(x) = F(\hat{x}) - \tau \quad (2-20)$$

onde  $F$  é monótona, e algum elemento de  $\{x : F(x) = \tau\}$  minimiza perda

esperada.

### 2.3.1

#### Estimação

De modo geral, o método de estimação do modelo de função quantil condicional da média é resolver o seguinte problema:

$$\min_{\mu \in \mathbb{R}} \sum_{i=1}^n (y_i - \mu)^2. \quad (2-21)$$

A média condicional de  $y$  dado  $x$  é  $\mu(x) = x^\top \beta$ , onde  $\beta$  pode ser estimado através da solução de

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n (y_i - x_i^\top \beta)^2 \quad (2-22)$$

Similarmente, visto que a  $\tau$ -ésima função quantil condicional,  $\hat{\alpha}(\tau)$ , é resolvida da seguinte forma:

$$\min_{\alpha \in \mathbb{R}} \sum_{i=1}^n \rho_\tau(y_i - \alpha), \quad (2-23)$$

estabelecendo um  $\tau$ -ésimo específico a função quantil condicional é dada por:

$$Q_y(\tau|x) = x_t^\top \beta(\tau), \quad (2-24)$$

e considerando a solução de  $\hat{\beta}(\tau)$

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau(y_i - x_t^\top \beta). \quad (2-25)$$

Essa é a origem da idéia elaborada por Koenker e Bassett(1978).

Uma descrição mais detalhada sobre o modelo de regressão quantílica pode ser encontrada em Koenker (19), (20), Koenker e Basset (21) e Machado (22).