

Sara Monteiro Morgado Dias Nunes

Subamostragem em Séries Temporais



*Tese submetida à Faculdade de Ciências da Universidade do Porto
para obtenção do grau de Mestre em Matemática Aplicada*

Departamento de Matemática Aplicada
Faculdade de Ciências da Universidade do Porto

Dezembro / 2001

Dissertação orientada por

Doutora Maria Eduarda da Rocha Pinto da Silva

Professora Auxiliar do Departamento de Matemática Aplicada
da Faculdade de Ciências da Universidade do Porto

à minha mãe

Agradecimentos

Agradeço à Professora Doutora Maria Eduarda Silva pelo apoio científico concedido ao desenvolvimento deste trabalho. As suas excelentes qualidades profissionais e humanas e a disponibilidade demonstrada ao longo deste percurso foram muito estimulantes e contribuíram de forma decisiva para a conclusão deste trabalho.

Agradeço à Direcção da Escola Superior de Gestão do Instituto Politécnico de Castelo Branco por todas as facilidades que me concedeu para que esta etapa fosse superada.

Agradeço ao colega João Renato Sebastião pelas suas comentários, sugestões e pela disponibilidade que demonstrou em ajudar-me na manutenção dos meios informáticos necessários à prossecução deste projecto.

Agradeço à minha família e ao Pedro pela compreensão que ao longo deste tempo demonstraram, pelo seu apoio incondicional, dando-me ânimo nos momentos de cansaço e criando, em meu maior, todas as condições para que esta prova pudesse ser vencida.

Também a todos aqueles que, de muitas formas, me incentivaram e apoiaram contribuindo para que este trabalho hoje possa tornar-se uma realidade.

Ao Pedro e

à minha mãe

Muito Obrigada

Agradecimentos

Agradeço à Professora Doutora Maria Eduarda Silva pelo apoio científico concedido na orientação deste trabalho. As suas excelentes qualidades profissionais e humanas e a disponibilidade demonstrada ao longo deste percurso foram muito estimulantes e contribuíram de forma decisiva para a conclusão deste trabalho.

Agradeço à Direcção da Escola Superior de Gestão do Instituto Politécnico de Castelo Branco por todas as facilidades que me concedeu para que esta etapa fosse superada.

Agradeço ao colega João Renato Sebastião pelos seus comentários, sugestões e pela disponibilidade que demonstrou em ajudar-me na manutenção dos meios informáticos necessários à prossecução deste projecto.

Agradeço à minha família e ao Pedro pela compreensão que ao longo deste tempo demonstraram, pelo seu apoio incondicional, dando-me ânimo nos momentos de cansaço e criando, em meu redor, todas as condições para que esta prova pudesse ser vencida.

Também a todos aqueles que, de muitas formas, me incentivaram e apoiaram contribuindo para que este trabalho hoje fosse uma realidade deixo o meu

Muito Obrigada.

Resumo

Neste trabalho considera-se a subamostragem em séries temporais a qual permite aproximar a distribuição amostral de uma estatística e as suas características. Esta técnica proporciona a construção intervalos de confiança assintóticos para um parâmetro associado à distribuição de probabilidade de uma série temporal $\{X_1, \dots, X_n\}$, mostrando-se válida sob condições mais gerais que os métodos de reamostragem.

A metodologia de subamostragem é aplicada a séries temporais geradas por modelos autoregressivos, com o objectivo de construir intervalos de confiança para o parâmetro autoregressivo. Faz-se um estudo com o objectivo de inferir sobre a relação entre o parâmetro autoregressivo e a dimensão das subamostras a utilizar na aplicação da subamostragem. Verifica-se que os intervalos de confiança obtidos por subamostragem apresentam uma cobertura comparável à dos intervalos de confiança clássicos. Conclui-se que subamostras de pequena dimensão conduzem a melhores resultados na construção de intervalos de confiança para o parâmetro autoregressivo. Num segundo estudo cujo objectivo é caracterizar a distribuição dos estimadores de máxima verosimilhança calculados nas subamostras, constata-se que estes apresentam uma distribuição aproximadamente normal.

Palavras Chave: bootstrap, jackknife, subamostragem, série temporal e processo autoregressivo.

INDICE

Abstract

In this thesis we consider subsampling in time series analysis. This method is used to approximate the sampling distribution of a statistic, in order to construct confidence regions for a parameter of interest. To illustrate the methodology, we use subsampling to construct confidence intervals for the parameter of an autoregressive process. Using a Monte Carlo simulation study, we find that the confidence intervals thus obtained present a nominal coverage probability similar to the classic ones. We also study the small sample properties of the maximum likelihood estimates obtained by subsampling and the optimal subsample size.

| | |
|--|----|
| 1. Introdução | 1 |
| 1.1. O Problema | 8 |
| 1.2. Séries Temporais – Conceitos Básicos | 10 |
| 1.3. Bootstrap | 13 |
| 1.3.1. Remostragem com Base no Modelo | 14 |
| 1.3.2. Remostragem em Blocos | 15 |
| 1.4. Jackknife | 18 |
| 2. Subamostragem | 22 |
| 2.1. Subamostragem em Séries Temporais | 23 |
| 2.2. Comparação entre a Subamostragem e o Bootstrap em Blocos Móveis | 27 |
| 2.3. Subamostragem com Constante de Normalização Desconhecida | 29 |
| 2.4. Escolha do Tamanho das Subamostras | 35 |
| 2.4.1. Método de Calibração | 34 |
| 2.4.2. Método da Variância Mínima | 36 |

Key Words: bootstrap, jackknife, subsampling, time series and autoregressive process.

| | |
|--|----|
| Índice | 6 |
| Dedicatória | 2 |
| Agradecimentos | 3 |
| Resumo | 4 |
| Abstract | 5 |
| Índice | 6 |
| 1. Introdução | 8 |
| 1.1. O Problema | 8 |
| 1.2. Séries Temporais – Conceitos Básicos | 10 |
| 1.3. Bootstrap | 13 |
| 1.3.1. Reamostragem com Base no Modelo | 14 |
| 1.3.2. Reamostragem em Blocos | 15 |
| 1.4. Jackknife | 19 |
| 2. Subamostragem | 22 |
| 2.1. Subamostragem em Séries Temporais | 23 |
| 2.2. Comparação entre a Subamostragem e o Bootstrap em Blocos Móveis | 27 |
| 2.3. Subamostragem com Constante de Normalização Desconhecida | 29 |
| 2.4. Escolha do Tamanho das Subamostras | 33 |
| 2.4.1. Método de Calibração | 34 |
| 2.4.2. Método da Variância Mínima | 36 |

| | |
|--|----|
| 3. Aplicação da Subamostragem a Modelos Autoregressivos | 38 |
| 3.1. Distribuição Assintótica dos Estimadores $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p$ | 38 |
| 3.2. Estudo de Simulação | 39 |
| 3.2.1. Intervalos de Confiança para o Parâmetro Autoregressivo | 40 |
| 3.2.2. Distribuição dos Estimadores | 43 |
| Conclusões e Trabalho Futuro | 45 |
| Lista de Referências | 46 |
| Anexo I - Demonstração do Teorema A [Politis e Romano (1994)] | 49 |
| Anexo II – Processos Autoregressivos de Ordem p – AR(p) | 52 |
| Anexo III – Estimação dos Parâmetros em Modelos AR(p) | 55 |
| Anexo IV – Intervalos de Confiança para os Parâmetros | 59 |
| Anexo V – Limites dos Intervalos de Confiança para o Parâmetro Autoregressivo | 62 |
| Anexo VI – Histogramas e PP-Plot dos Estimadores de Máxima Verosimilhança do Parâmetro Autoregressivo | 70 |



1. INTRODUÇÃO

1.1. O Problema

Dadas N observações X_1, X_2, \dots, X_N de uma série temporal, suponha-se que optámos por uma estatística $\hat{\theta}$ para estimar um determinado parâmetro θ . O problema que se coloca é o de medir a precisão de $\hat{\theta}$ enquanto estimador de θ .

Foi numa tentativa de estimar o erro padrão de determinada estatística que surgiram os métodos de reamostragem como o **bootstrap** [Efron (1979)] e de subamostragem como o **jackknife** [Tukey (1958)] cujo princípio básico é a utilização repetida da amostra original de modo a obter várias estimativas de θ , que serão usadas para estimar ou aproximar a distribuição de $\hat{\theta}$ e as suas características.

Como é sabido, os métodos de inferência para dados independentes e identicamente distribuídos (i.i.d.) não são consistentes quando os dados em causa são dependentes. Nos últimos anos tem vindo a registar-se um aumento considerável de literatura relativa à modificação destes métodos de forma a poderem ser aplicados a situações de dados dependentes como é o caso das séries temporais.

Historicamente, a primeira extensão destes métodos a dados dependentes consistiu em aplicar o bootstrap de Efron a um conjunto de dados de um modelo de regressão geral reamostrando os resíduos como dados i.i.d. e não as observações originais [Efron e Tibshirani (1986)]. Contudo, esta aproximação é restrita a situações nas quais é possível considerar o modelo de regressão geral como no caso dos modelos autoregressivos.

Uma aproximação mais geral, consistiu em aplicar a reamostragem à sequência de dados original considerando blocos de dados em vez de dados pontuais como no caso i.i.d.. Desta forma, a estrutura de dependência do modelo subjacente é preservada dentro de cada bloco e conseguem-se inferências assintoticamente correctas. Com base nesta ideia, Carlstein (1986) considerou o problema de estimar a variância de uma estatística a partir dos valores da mesma calculada em sub-séries da sequência de dados original. A sua ideia tinha como base a divisão da sequência original de dados em blocos não sobrepostos de tamanho $b < n$, calculando a estatística de interesse nesses blocos e desenvolvendo assim, sob hipóteses fracas, propriedades consistentes para uma distribuição amostral estimada. Mais tarde, Künsch (1989) e Liu e Singh (1992), independentemente, introduziram o bootstrap em blocos

móveis, através do qual a estimação da variância da estatística em causa pode também ser usada para estimar a sua distribuição amostral construindo assim intervalos ou regiões de confiança para parâmetros desconhecidos. O método é análogo ao bootstrap i.i.d. de Efron sendo a única diferença a seguinte: os blocos de observações de tamanho $b < n$, reamostrados com reposição a partir dos dados, assumem a forma de pseudo-sequência em vez de dados pontuais. Note-se que, em contraste com a aproximação de Carlstein (1986), o bootstrap em blocos móveis utiliza blocos sobrepostos, uma técnica geralmente mais eficiente. Politis e Romano (1992) propuseram variantes do bootstrap em blocos móveis que possuem características interessantes tais como blocos de tamanho aleatório e uma extensão periódica da sequência de dados original.

Em suma, os métodos de **reamostragem** podem essencialmente ser usados de duas formas em dados dependentes estritamente estacionários:

- ajustar um modelo ao conjunto de dados, reamostrando em seguida os resíduos tal como no caso de observações i.i.d.;
- dividir os dados em blocos de observações consecutivas e reamostrar esses blocos em vez de dados individuais ou inovações estimadas. Carlstein (1986) propôs a utilização de blocos não sobrepostos, enquanto Künsch (1989) e Liu e Singh (1992), independentemente, introduziram um método de blocos móveis que utiliza blocos sobrepostos.

Como alternativa aos métodos de reamostragem, Politis e Romano (1992, 1994) propõem a construção de intervalos de confiança para parâmetros com base no método de **subamostragem** cujo princípio básico é o cálculo da estatística de interesse em subconjuntos da amostra original de modo a aproximarmos a distribuição amostral de $\hat{\theta}$ e fazermos inferências sobre θ . No contexto das séries temporais, em vez de reamostrar blocos a partir da série temporal original usando-os para gerar novas pseudo-séries temporais, cada subconjunto de observações é encarado como uma sub-série, ela própria também válida. Cada um destes blocos é gerado pelo modelo probabilístico subjacente e por isso é de esperar que possamos obter informação válida acerca da distribuição amostral de uma estatística calculando-a em todas as sub-séries.

Outra característica atractiva do método de subamostragem é o facto de se mostrar válido sob condições muito fracas. Além das condições de regularidade e dependência, o

único requisito no contexto estacionário é que a distribuição amostral da estatística devidamente normalizada, possua uma distribuição limite não degenerada.

Neste trabalho procedemos a uma revisão de alguns métodos de reamostragem com ênfase na sua aplicação a séries temporais e apresentamos a metodologia da subamostragem. Assim, no capítulo 1 apresentamos alguns conceitos básicos relativos a séries temporais e apresentamos também uma revisão acerca das metodologias bootstrap e jackknife. O capítulo 2 é dedicado à metodologia da subamostragem de Politis e Romano (1994) e no capítulo 3 aplicamos a metodologia apresentada a um modelo autoregressivo de 1ª ordem com o objectivo de fazer inferências sobre o parâmetro autoregressivo, apresentando alguns resultados de um estudo de simulação.

1.2. Séries Temporais – Conceitos Básicos

Dado um processo estocástico $\{X_t : t \in \mathfrak{R}\}$, uma **série temporal** é um conjunto de observações do processo estocástico, $X_{t_1}, X_{t_2}, \dots, X_{t_n}$, feitas nos instantes t_1, t_2, \dots, t_n .

Por uma questão de simplicidade consideraremos $t \in T \equiv \{0, \pm 1, \pm 2, \dots\}$, isto é, consideramos apenas séries temporais **discretas** (uma série temporal dir-se-á **contínua** se T for um intervalo da recta real). Consideraremos também que as observações são feitas em intervalos de tempo regulares, isto é, com a mesma amplitude. Note-se que um processo $\{X_t\}$ admite um conjunto infinito de realizações que se denomina “*ensemble*” sendo uma série temporal apenas uma das realizações do “*ensemble*” associado ao processo.

Encarando uma série temporal como uma das infinitas realizações de um processo estocástico, temos um conjunto de variáveis aleatórias ordenadas no tempo.

Apesar de um processo estocástico poder ser visto, de uma forma mais geral, como uma família de variáveis aleatórias indicadas não só no tempo como no espaço ou em ambos, consideramos aqui que um **processo estocástico** é uma colecção ordenada de variáveis aleatórias $\{X_t : t \in T\}$ onde T é um conjunto de índices representando o tempo.

Um processo estocástico $\{X_t : t \in T\}$ diz-se **estritamente estacionário** se e só se a distribuição conjunta de $(X_{t_1}, \dots, X_{t_n})$ é igual à distribuição conjunta de $(X_{t_1+k}, \dots, X_{t_n+k})$ para qualquer inteiro positivo n e para quaisquer inteiros t_1, t_2, \dots, t_n e k , ou seja

$$F_{(X_{t_1}, \dots, X_{t_n})}(x_1, \dots, x_n) = F_{(X_{t_1+k}, \dots, X_{t_n+k})}(x_1, \dots, x_n)$$

em todos os pontos (x_1, \dots, x_n) em que $F_{(X_{t_1}, \dots, X_{t_n})}$ representa uma função distribuição conjunta de $(X_{t_1}, \dots, X_{t_n})$.

O conceito de “equilíbrio estatístico”, entendido como a invariância de qualquer distribuição de dimensão finita em relação a translações no tempo, leva-nos a concluir que é indiferente observar os processos nos pontos t_1, t_2, \dots, t_n ou nos pontos $t_{1+k}, t_{2+k}, \dots, t_{n+k}$. Contudo, a noção de estrita estacionariedade é difícil de verificar na prática pois exige o conhecimento de todas as distribuições marginais, sendo portanto uma exigência bastante forte. Assim, há processos que, não sendo estritamente estacionários, o são em sentido lato.

Um processo estocástico $\{X_t : t \in T\}$ diz-se **fracamente estacionário** ou **estacionário até à ordem m** se, para qualquer inteiro positivo n , para quaisquer inteiros t_1, t_2, \dots, t_n e k , todos os momentos conjuntos até à ordem m de $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$ existem e são iguais aos correspondentes momentos conjuntos até à ordem m de $\{X_{t_1+k}, X_{t_2+k}, \dots, X_{t_n+k}\}$. Então, $E[X_{t_1}^{m_1} X_{t_2}^{m_2} \dots X_{t_n}^{m_n}] = E[X_{t_1+k}^{m_1} X_{t_2+k}^{m_2} \dots X_{t_n+k}^{m_n}]$ para todos os inteiros positivos, m_1, \dots, m_n tais que $m_1 + \dots + m_n \leq m$ e qualquer k .

Em particular, um processo estocástico $\{X_t : t \in T\}$ diz-se **estacionário de 2ª ordem** ou estacionário para a covariância se, para todo o t , $E(X_t^2) < +\infty$,

$$E(X_t) = \mu \quad \text{e} \quad \text{Var}(X_t) = \sigma^2.$$

No estudo de processos estacionários no domínio do tempo são de extrema importância as funções autocovariância e autocorrelação, as quais passamos a definir.

Chamamos **função autocovariância (FACOV)** do processo à função

$$\gamma_k = \text{Cov}(X_t, X_{t+k}) = E[(X_t - \mu)(X_{t+k} - \mu)]$$

a qual se encontra definida para todos os valores de $k = 0, \pm 1, \pm 2, \dots$ e mede a covariância entre dois valores do processo separados por um intervalo (“lag”) de amplitude k .

Da mesma forma, definimos a **função autocorrelação (FAC)** do processo como sendo

$$\rho_k = \text{Corr}(X_t, X_{t+k}) = \frac{\text{cov}(X_t, X_{t+k})}{\sqrt{\text{var}(X_t) \text{var}(X_{t+k})}}$$

Mas como $\text{var}(X_{t+k}) = \text{var}(X_t) = \gamma_0$

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

Para cada k , a função ρ_k mede a correlação entre pares de valores do processo separados por um intervalo de amplitude k . Intuitivamente, podemos interpretar ρ_k como uma medida da semelhança entre cada realização e a mesma realização deslocada para a esquerda (ou direita) k unidades de tempo. Nos modelos usuais, quando k tende para infinito, γ_k e ρ_k tendem para zero pois, à medida que k aumenta, é de esperar que a capacidade de memória do processo diminua. A representação gráfica da FAC denomina-se por **correlograma**.

As principais **propriedades das funções autocovariância e autocorrelação** são as seguintes:

i) $\gamma_0 = \sigma^2$ e $\rho_0 = 1$

ii) $|\gamma_k| \leq \gamma_0$ e $|\rho_k| \leq 1$

iii) $\gamma_k = \gamma_{-k}$ e $\rho_k = \rho_{-k}$

iv) As funções γ_k e ρ_k são semidefinidas positivas, isto é

$$\sum_{i=1}^n \sum_{j=1}^n \gamma_{|t_i - t_j|} z_i z_j \geq 0 \quad \text{e} \quad \sum_{i=1}^n \sum_{j=1}^n \rho_{|t_i - t_j|} z_i z_j \geq 0$$

para qualquer conjunto de números reais z_1, \dots, z_n e instantes de tempo t_1, \dots, t_n .

Para além da análise da correlação entre X_t e X_{t+k} interessa analisar a correlação entre X_t e X_{t+k} depois de eliminar o efeito que sobre elas exercem as variáveis intermédias $X_{t+1}, X_{t+2}, \dots, X_{t+k-1}$, isto é

$$\gamma(X_t, X_{t+k} \mid X_{t+1}, X_{t+2}, \dots, X_{t+k-1})$$

Na análise temporal, esta correlação condicional designa-se por **autocorrelação parcial**.

Destas considerações decorre uma interessante propriedade relativa aos processos Gaussianos. $\{X_t\}$ diz-se um **processo Gaussiano** se, para qualquer inteiro positivo n e para quaisquer inteiros t_1, t_2, \dots, t_n , a distribuição conjunta de $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$ é normal

multivariada. Ora, uma vez que a estrutura probabilística de tais processos fica completamente especificada pelas funções média e covariância, concluímos que qualquer processo Gaussiano estacionário até à 2ª ordem é estritamente estacionário uma vez que as referidas funções são invariantes sob translações no tempo.

Neste trabalho consideraremos apenas processos estacionários de 2ª ordem.

1.3. Bootstrap

O bootstrap faz parte da grande família dos métodos de reamostragem e foi originalmente proposto por Efron em 1979, tendo revolucionado a Estatística.

Suponhamos que dispomos de uma amostra aleatória $Y = (X_1, \dots, X_n)$ pertencente a uma determinada distribuição de probabilidade desconhecida F e que pretendemos estimar um parâmetro de interesse $\theta = \theta(F)$ com base em Y . Teremos então que calcular o estimador $\hat{\theta} = s(Y)$. Embora as ideias subjacentes ao bootstrap sejam antigas, o interesse da comunidade científica pelo método surgiu essencialmente após os trabalhos de Efron (1979) e Efron e Tibshirani (1986) onde o bootstrap é apresentado como um método de base computacional adequado a vários problemas de estimação. Através deste método podemos, por exemplo, construir de modo simples estimadores para a distribuição de $\hat{\theta}$ e as suas características com base na amostra Y , como refere Brito (2000).

Efron e Tibshirani (1993) apresentam de forma muito clara o funcionamento deste método:

Começamos por construir a distribuição empírica \hat{F} , estimador de F , definida como $\hat{F}(x) = \frac{\#\{X_i \leq x\}}{n}$. A função distribuição empírica obtém-se atribuindo peso $1/n$ a cada um dos valores observados X_i , $i = 1, \dots, n$. Com \hat{F} fixo, definimos uma **amostra bootstrap** $Y^* = (X_1^*, \dots, X_n^*)$ como sendo uma amostra aleatória de tamanho n obtida de \hat{F} , com reposição, a partir da população dos n indivíduos (X_1, \dots, X_n) : $\hat{F} \rightarrow (X_1^*, \dots, X_n^*)$. A notação * indica que Y^* não é o conjunto de dados original Y mas sim uma versão reamostrada de Y .

Seleccionamos então B amostras bootstrap independentes Y^{*1}, \dots, Y^{*B} , cada uma consistindo em n dados obtidos, com reposição, a partir de Y (para estimar o erro padrão, geralmente $50 < B < 200$). Associada a cada conjunto de dados bootstrap Y^* está uma **réplica**



bootstrap de $\hat{\theta}$: $\hat{\theta}^*(b) = s(Y^{*b})$, $b=1, \dots, B$, a qual resulta da aplicação da mesma função a Y^* que foi aplicada a Y .

Finalmente, estimamos o erro padrão de $\hat{\theta}$ - $se_F(\hat{\theta})$ - pelo desvio padrão das B réplicas:

$$\hat{se}_B = \left\{ \frac{\sum_{b=1}^B [\hat{\theta}^*(b) - \hat{\theta}^*(\cdot)]^2}{B-1} \right\}^{1/2} \quad \text{onde} \quad \hat{\theta}^*(\cdot) = \frac{\sum_{b=1}^B \hat{\theta}^*(b)}{B},$$

verificando-se que o desvio padrão empírico aproxima o desvio padrão da população quando o número de réplicas aumenta pois $\lim_{B \rightarrow \infty} \hat{se}_B = se_{\hat{F}} = se_{\hat{F}}(\hat{\theta}^*)$.

Como atrás referimos, no caso de se tratarem de dados dependentes, esta aproximação não é viável pois, ao reamostrarmos observações desta forma eliminamos a estrutura de dependência que caracteriza os dados. Este facto conduz-nos à ideia de construir amostras bootstrap tomando blocos a partir das observações originais e não simplesmente dados pontuais.

Centrando a atenção apenas no contexto das séries temporais, Davison e Hinkley (1997) consideram duas formas distintas de reamostragem: reamostragem **com base no modelo** e reamostragem **por blocos**, as quais passamos a descrever sucintamente.

1.3.1. Reamostragem com base no modelo

Neste caso procedemos a uma abordagem análoga à reamostragem com base num modelo de regressão. A ideia é ajustar um modelo aos dados, construir resíduos a partir do modelo ajustado e depois gerar novas séries incorporando amostras aleatórias dos resíduos dentro do modelo ajustado. Os resíduos são geralmente centrados novamente de forma a apresentarem a mesma média que as inovações do modelo. A situação mais simples consiste em ajustar um modelo autoregressivo de 1ª ordem [AR(1)] - $X_t = aX_{t-1} + \varepsilon_t$ com $t = \dots, -1, 0, 1, \dots$ e $|a| < 1$ - a uma série observada X_1, \dots, X_n , originando o coeficiente autoregressivo estimado \hat{a} e as inovações estimadas $e_t = X_t - \hat{a}X_{t-1}$, $t = 2, \dots, n$; (não é possível obter e_1 porque X_0 é desconhecido). Na reamostragem com base no modelo procedemos por amostragem equiprovável, com reposição, a partir dos resíduos centrados $e_2 - \bar{e}, \dots, e_n - \bar{e}$ para obtermos inovações simuladas $\varepsilon_0^*, \dots, \varepsilon_n^*$ e depois colocar $X_0^* = \varepsilon_0^*$ e $X_t^* = \hat{a}X_{t-1}^* + \varepsilon_t^*$, $t = 1, \dots, n$. Na verdade, as séries assim geradas não são estacionárias. É

conveniente iniciar a série em equilíbrio ou então gerar uma série mais longa de inovações e começar a equação $X_t^* = \hat{\alpha}X_{t-1}^* + \varepsilon_t^*$, $t = 1, \dots, n$ em $j = -k$, sendo o período de $-k$ a zero suficientemente longo para assegurar a estacionariedade das observações X_1^*, \dots, X_n^* . Os valores X_{-k}^*, \dots, X_0^* são colocados de parte.

A reamostragem com base no modelo é, em geral, simples de aplicar e conduz a bons resultados quando o modelo ajustado aos dados é correcto. O maior inconveniente deste procedimento é que, na prática, têm que ser identificados a partir dos dados não apenas os parâmetros do modelo mas também a sua estrutura. Se a estrutura escolhida é incorrecta, as séries reamostradas serão geradas a partir de um modelo errado, e logo não terão as mesmas propriedades estatísticas que os dados originais. Surge então a necessidade de um método em que não seja necessário ajustar previamente aos dados qualquer modelo: a reamostragem em blocos.

1.3.2. Reamostragem em blocos

Neste caso reamostramos blocos de observações consecutivas em vez de inovações. A versão mais simples desta ideia consiste em dividir os dados em l **blocos não sobrepostos** de tamanho b , onde supostamente $n = lb$. Fazemos $Y_1 = (X_1, \dots, X_b)$, $Y_2 = (X_{b+1}, \dots, X_{2b})$, etc., originando blocos Y_1, \dots, Y_l . O procedimento consiste em gerar amostras bootstrap, com probabilidades iguais $1/l$, de $\{Y_1, \dots, Y_l\}$, e depois “colá-las” de uma extremidade à outra para formar uma nova série. Note-se que, nesta perspectiva, o bootstrap de Efron pode ser encarado como um caso particular do bootstrap em blocos móveis, caso em que são utilizados blocos de tamanho $b = l$.

Suponhamos, por exemplo, que a série original é X_1, \dots, X_{12} e tomamos $b = 4$ e $l = 3$. Então os blocos são $Y_1 = (X_1, X_2, X_3, X_4)$, $Y_2 = (X_5, X_6, X_7, X_8)$ e $Y_3 = (X_9, X_{10}, X_{11}, X_{12})$. Se os blocos reamostrados forem $Y_1^* = Y_2$, $Y_2^* = Y_1$ e $Y_3^* = Y_2$, a nova série de tamanho 12 é

$$\{X_t^*\} = Y_1^*, Y_2^*, Y_3^* = X_5, X_6, X_7, X_8, X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8.$$

Em geral, as séries reamostradas encontram-se mais próximas do ruído branco que a série original.

A ideia que está por detrás da reamostragem por blocos é que, se os blocos são suficientemente longos, a dependência original será preservada nas séries reamostradas onde

as estatísticas $\hat{\theta}$, calculadas a partir de $\{X_t^*\}$, têm aproximadamente a mesma distribuição que as estatísticas $\hat{\theta}$ calculadas a partir de réplicas da série original. Obviamente, esta aproximação será tanto melhor quanto mais fraca for a dependência e, por outro lado, quanto mais longos forem os blocos, de forma a reterem melhor a dependência. Por outro lado, os valores de $\hat{\theta}^*$ devem ser tão numerosos quanto possível, de forma a conduzirem a uma boa estimativa da distribuição de $\hat{\theta}$, o que se proporciona quando usamos blocos pequenos. Trabalhos de natureza teórica sugerem que o tamanho dos blocos b deve ser de ordem n^r , $r \in]0,1[$. Neste caso, tanto o tamanho dos blocos b como o número de blocos $l = n/b$ tendem para infinito quando $n \rightarrow \infty$, dependendo ainda o valor de r do tipo de estatística $\hat{\theta}$ em causa.

Podem ser consideradas diversas variantes desta técnica de reamostragem. Uma delas consiste em formar **blocos sobrepostos**. Por exemplo, considerando novamente que a série original é X_1, \dots, X_{12} , e mantendo $b=4$ formamos $n-b+1=9$ blocos: $Y_1 = (X_1, X_2, X_3, X_4)$, $Y_2 = (X_2, X_3, X_4, X_5)$, ..., $Y_9 = (X_9, X_{10}, X_{11}, X_{12})$. Isto provoca alguns efeitos terminais pois as primeiras e últimas $b-1$ observações originais aparecem em menos blocos que as restantes. Porém, tais efeitos podem ser eliminados acrescentando $b-1$ blocos (no exemplo acima, acrescentando os blocos $Y_{10} = (X_{10}, X_{11}, X_{12}, X_1)$, $Y_{11} = (X_{11}, X_{12}, X_1, X_2)$ e $Y_{12} = (X_{12}, X_1, X_2, X_3)$). Desta forma garantimos que cada uma das observações originais possui igual probabilidade de aparecer nas séries simuladas. Além disso, esta “correção terminal” elimina um pequeno problema existente na técnica de blocos não sobrepostos: o último bloco seria menor que os restantes sempre que n/b não fosse um inteiro.

O maior inconveniente da reamostragem por blocos prende-se com o facto de se registar uma dependência menor nas séries geradas que nos dados originais, o que poderá conduzir a aproximações erradas. Como atrás referimos, não é apropriado tomar blocos de tamanho $b=1$ na reamostragem de dados dependentes pois nesse caso as séries reamostradas seriam apenas ruído branco, o que pode contudo continuar a verificar-se quando os valores de b são pequenos. Isto sugere uma estratégia intermédia entre os dois tipos de reamostragem aqui apresentados (com base no modelo e por blocos). A ideia é branquear a série ajustando-lhe o modelo que supostamente elimina mais a dependência entre as observações originais. Em seguida, uma série de inovações é gerada reamostrando, por blocos, os resíduos a partir do modelo ajustado. Finalmente, a série de inovações é colorida através da aplicação do modelo estimado às inovações reamostradas. Por exemplo, se utilizarmos um modelo AR(1)

para branquear os dados originais, novas séries são geradas aplicando a equação $X_t^* = \hat{a}X_{t-1}^* + \hat{\varepsilon}_t^*$, $t = 1, \dots, n$, com a série de inovações $\{\hat{\varepsilon}_t^*\}$ amostrada não independentemente mas em blocos tomados a partir da série dos resíduos estimados centrados $e_2 - \bar{e}, \dots, e_n - \bar{e}$.

Uma técnica diferente, que permite eliminar o efeito de ruído branco presente na reamostragem por blocos, consiste em reamostrar **blocos de blocos**.

Suponhamos que o foco de interesse é uma estatística $\hat{\theta}$, estimador de θ , que depende apenas de blocos de m observações sucessivas. Um exemplo é a autocovariância amostral no lag k : $\hat{\gamma}_k = \frac{1}{n-k} \sum_{j=1}^{n-k} (X_j - \bar{X})(X_{j+k} - \bar{X})$ onde $m=k+1$. Excepto quando $b \gg m$, em geral a distribuição de $(\hat{\theta} - \hat{\theta}^*)$, onde $\hat{\theta}^*$ denota os valores da estatística $\hat{\theta}$ calculada nas séries $\{X_t^*\}$, constitui uma aproximação pobre de $(\hat{\theta} - \theta)$ uma vez que um número significativo de pares (X_j^*, X_{j+k}^*) na série reamostrada pertencerá à junção de blocos, estando esta mais próxima do ruído branco que a série original. Para implementar a reamostragem por blocos de blocos definimos um novo processo m -variado $\{X'_j\}$ para o qual $X'_j = (X_j, \dots, X_{j+m-1})$. A reescrita de $\hat{\theta}$ envolve as médias dos X'_j e a reamostragem de blocos dos novos "dados" X'_1, \dots, X'_{n-m+1} , sendo cada observação um bloco de dados originais. Por exemplo para a autovariância num lag 1 colocamos

$$(X'_1, \dots, X'_{n-1}) \equiv \begin{pmatrix} X'_{11} & X'_{12} & \dots & X'_{1,n-1} \\ X'_{21} & X'_{22} & \dots & X'_{2,n-1} \end{pmatrix} = \begin{pmatrix} X_1 & X_2 & \dots & X_{n-1} \\ X_2 & X_3 & \dots & X_n \end{pmatrix}$$

e escrevemos $\hat{\theta}^* = \frac{1}{n-1} \sum (X'_{1j} - \bar{X}'_1)(X'_{2j} - \bar{X}'_2)$. O ponto chave é que $\hat{\theta}^*$ não deve comparar observações adjacentes em cada linha. Com $n=12$ e $b=4$, uma réplica bootstrap poderia ser

$$\{X'_j\} = \begin{pmatrix} X_5 X_6 X_7 X_8 & X_1 X_2 X_3 X_4 & X_7 X_8 X_9 X_{10} \\ X_6 X_7 X_8 X_9 & X_2 X_3 X_4 X_5 & X_8 X_9 X_{10} X_{11} \end{pmatrix}.$$

Uma vez que a versão bootstrap de $\hat{\theta}^*$ baseada nesta série apenas conterá produtos de observações adjacentes (centradas) dos dados originais, o ruído branco será reduzido, apesar de não ser removido inteiramente.

Outro inconveniente da reamostragem em blocos é que as séries geradas não são estacionárias, porque, próximo da ligação entre blocos e no centro dos blocos, a distribuição

conjunta das observações reamostradas difere. Isto pode ser ultrapassado tomando blocos de tamanho aleatório. No **bootstrap estacionário** são tomados blocos cujos tamanhos B são distribuídos geometricamente, com densidade $Pr(B=t)=(1-p)^{t-1}p$, $t=1,2,\dots$. Isto assegura que as séries reamostradas sejam estacionárias com um tamanho médio de bloco $b = \frac{1}{p}$ enquanto as séries bootstrap com tamanhos de bloco fixos não são estacionárias.

Suponhamos agora que pretendemos usar o bootstrap em blocos para estimar uma determinada característica K , com base numa série de n observações. Diferentes tamanhos de blocos - b - conduzem, a diferentes estimadores bootstrap $\hat{K}(n,b)$. Que valor de b devemos usar?

O tamanho ótimo de bloco é definido como sendo aquele que minimiza o erro quadrático médio (EQM) assintótico do estimador bootstrap. Sabe-se que o tamanho ótimo de bloco cresce com n , dependendo da característica que estamos a estimar. Hall et al (1995) demonstraram que, tanto nos blocos sobrepostos como não sobrepostos com tamanho fixo, o tamanho b do bloco ótimo verifica $b \sim n^r$, onde $r = \frac{1}{3}$ para estimar o viés ou a variância, $r = \frac{1}{4}$ para estimar uma função distribuição unilateral (ex: $F(\theta_n \leq \tau)$) e $r = \frac{1}{5}$ para estimar uma função distribuição simétrica (ex: $F(|\theta_n| \leq \tau)$). Davison e Hinkley (1997) referem que, para n e b grandes, o EQM de $\hat{K}(n,b)$ é proporcional a

$$\frac{1}{n^d} \left(\frac{C_1}{b^2} + \frac{C_2 b^c}{n} \right)$$

onde C_1 e C_2 dependem apenas de K e da estrutura de dependência da série. Na expressão acima, $d=2$ e $c=1$ se K for o viés ou a variância, $d=1$ e $c=2$ se K é um quantil unilateral e $d=2$ e $c=3$ se é um quantil bilateral. Desta expressão decorre que, para n grande, o EQM de $\hat{K}(n,b)$ é minimizado tomando $b \propto n^{\frac{1}{d+2}}$. Não conhecemos porém a constante de proporcionalidade, que pode ser estimada da seguinte forma: estabelecemos um valor inicial para b de forma a obter $\hat{K}(n,b)$. Depois tomamos $m < n$ e $k < b$ e calculamos os valores de $\hat{K}(m,k)$ a partir das $m-n+1$ séries X_t, \dots, X_{t+m-1} , $t=1, \dots, n-m+1$. O EQM para $\hat{K}(m,k)$, calculado a partir de séries de m observações e com blocos de tamanho k , vem então

$$EQM(m,k) = \frac{1}{n-m+1} \sum_{j=1}^{n-m+1} \left\{ \hat{K}_j(m,k) - \hat{K}(n,b) \right\}^2.$$

Repetindo este procedimento para diferentes valores de k mas igual valor de m , obtemos o valor de \hat{k} para o qual $EQM(m,k)$ é minimizado.

Hall et al (1995) mostraram também que blocos sobrepostos são mais eficientes que blocos não sobrepostos.

1.4. Jackknife

Tal como o bootstrap, o jackknife é uma técnica de base computacional que permite estimar o enviesamento e o erro padrão de um estimador em condições teoricamente complexas. Esta técnica foi originalmente proposta em 1949 por Maurice Quenouille para estimar o enviesamento de um estimador e consiste no cálculo da estatística em amostras construídas a partir da amostra inicial por eliminação consecutiva de um dado observado. Mais tarde, Tukey (1958) reorganizou o trabalho anteriormente feito, utilizando o jackknife para estimar o erro padrão, dando-lhe este nome. Vejamos, resumidamente, em que consiste esta técnica:

Suponhamos que dispomos de uma amostra $Y = (X_1, X_2, \dots, X_n)$ e de um estimador $\hat{\theta} = s(Y)$. Pretendemos estimar o viés e o erro padrão de $\hat{\theta}$. O ponto fulcral desta técnica consiste em gerar amostras jackknife, e para isso deixamos de fora uma observação de cada vez:

$$Y_{(i)} = (X_1, X_2, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$$

A i -ésima **amostra jackknife** é constituída pelo conjunto de dados ao qual foi retirada a i -ésima observação, sendo $\hat{\theta}_{(i)} = s(Y_{(i)})$ a i -ésima **réplica jackknife** de $\hat{\theta}$.

O estimador jackknife do viés é definido por

$$\hat{bias}_{jack} = (n-1)(\hat{\theta}_{(\cdot)} - \hat{\theta}) \quad \text{onde} \quad \hat{\theta}_{(\cdot)} = \sum_{i=1}^n \frac{\hat{\theta}_{(i)}}{n}.$$

Por sua vez, o estimador jackknife do erro padrão é definido por

$$\hat{se}_{jack} = \left[\frac{n-1}{n} \sum (\hat{\theta}_{(i)} - \hat{\theta}_{(\cdot)})^2 \right]^{1/2}.$$

Na prática, verifica-se que o jackknife não é consistente para estatísticas que não sejam estáveis, como acontece, em geral, quando as observações em causa são dependentes. De forma a contornar esta inconsistência, Wu (1986) e mais tarde Shao e Wu (1989), propuseram o **jackknife delete-d**. Neste caso, em vez de deixarmos uma observação de fora de cada vez, deixamos d observações. De facto, devemos deixar de fora mais do que $d = \sqrt{n}$ observações, de forma a obter consistência para o estimador jackknife do erro padrão.

Seja $\hat{\theta}_{(s)}$ o valor da estatística $\hat{\theta}$ aplicada ao conjunto de dados com o subconjunto s removido. O estimador jackknife delete-d do erro padrão vem

$$\hat{se}_{jack} = \left[\frac{n-d}{d} \frac{1}{n C_d} \sum (\hat{\theta}_{(s)} - \hat{\theta}_{(\cdot)})^2 \right]^{1/2} \quad \text{onde} \quad \hat{\theta}_{(\cdot)} = \sum \frac{\hat{\theta}_{(s)}}{n C_d}$$

O somatório é tomado sobre todos os subconjuntos s de tamanho $n-d$ escolhidos sem reposição de X_1, X_2, \dots, X_n .

Se n for muito grande e $\sqrt{n} < d < n$, o número de amostras jackknife ${}^n C_d$ pode ser muito elevado. Neste caso, em vez de calcularmos $\hat{\theta}$ em todos estes subconjuntos, podemos gerar uma amostra aleatória de subconjuntos, o que torna o jackknife delete-d mais parecido com o bootstrap.

Numa tentativa de estender a metodologia jackknife a dados dependentes, Künsch (1989) sugere que se apaguem sucessivamente blocos de b observações consecutivas. A estatística de interesse é então calculada em cada conjunto de dados obtido desta forma, calculando-se depois a variância amostral dos valores da estatística assim obtidos. Esta metodologia mostra-se eficiente em alguns casos mas levanta alguns problemas quando o conjunto de dados em causa apresenta dependência forte. Assim, ao omitirmos um determinado bloco temos que ter algum cuidado na transição entre as observações que antecedem e precedem o bloco apagado. Alonso et al (2000) propõem uma técnica alternativa na qual encaram estes blocos de observações apagadas como dados que faltam, sendo estes substituídos por estimadores dos valores que faltam conservando assim a estrutura de dependência das observações.

De um modo geral, quando os dados em causa são dependentes, é mais usual recorrer à metodologia bootstrap que aos métodos jackknife por se tratar de um método mais geral e adaptável a uma maior diversidade de situações. De facto, mais de duas décadas se passaram desde que Efron (1979) introduziu o bootstrap, o qual se veio a revelar um método bastante eficaz na estimação da distribuição amostral de uma estatística. Podemos mesmo considerar

que o bootstrap revolucionou o campo da Estatística, alcançando um enorme sucesso. Porém, apesar disso, é sabido que o bootstrap falha quando aplicado a um número considerável de situações, sendo um desses contra-exemplos o caso em que o estimador em causa não apresenta uma distribuição limite contínua. Para ilustrarmos esta situação, consideremos, por exemplo, o modelo autoregressivo de 1ª ordem - AR(1):

$$X_t = \mu + aX_{t-1} + \varepsilon_t, \quad t = 1, \dots, n$$

onde a é uma constante satisfazendo as condições de estacionariedade e $\{\varepsilon_t\}$ é uma sucessão de variáveis aleatórias i.i.d. com distribuição $N(0, \sigma_\varepsilon^2)$.

Basawa et al (1991) atribuem a falha do bootstrap neste modelo à descontinuidade da distribuição limite do estimador de mínimos quadrados de a no círculo unitário. Estes autores referem que a causa principal da falha do estimador bootstrap em $|a|=1$ é o facto de a distribuição limite de \hat{a}_n ser drasticamente diferente para os três casos seguintes: $|a|<1$, $|a|>1$ e $|a|=1$.

Relativamente ao modelo acima referido, Politis et al (1999) referem ainda que a inferência sobre a não é trivial pois a distribuição limite do estimador depende dos parâmetros subjacentes a e μ . Assim, se $|a|<1$, o estimador de mínimos quadrados converge para uma distribuição limite normal à velocidade de $n^{1/2}$. Para $a=1$, e $\mu=0$, converge para uma distribuição não standard à velocidade de n . Finalmente, para $a=1$ e $\mu \neq 0$ o estimador converge para uma distribuição normal à velocidade de $n^{3/2}$. Para construirmos um intervalo de confiança para a , é necessário saber qual é o modelo correcto e logo, a descontinuidade da distribuição limite conduz à falha do bootstrap.

Relativamente aos métodos bootstrap, além do problema da descontinuidade da distribuição limite do estimador, Zaman (1996) refere que, ao aproximarmos a distribuição $F(x)$ de um estimador pela sua distribuição empírica \hat{F} , pode acontecer que de facto \hat{F} não aproxime correctamente $F(x)$ e, nesse caso, o bootstrap falha.

Surge então a necessidade de um método mais geral válido também nas situações que representam um contra-exemplo para o bootstrap. Foi neste contexto que Politis e Romano (1994) propuseram a metodologia da subamostragem que desenvolvemos no capítulo seguinte.

2. SUBAMOSTRAGEM

Suponhamos que dispomos de n observações $\{X_1, \dots, X_n\}$ e que optámos por uma estatística $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ para estimar um determinado parâmetro $\theta = \theta(P) \in \mathfrak{R}$, sendo P a medida de probabilidade subjacente. O objectivo do método de subamostragem introduzido por Politis e Romano (1994) é a construção de intervalos ou regiões de confiança para θ através da aproximação da distribuição amostral de $(\hat{\theta}_n - \theta(P))$ com base nos valores da estatística calculada em subconjuntos dos dados. Por exemplo, no caso em que os dados são n observações i.i.d., são geradas subamostras de tamanho $b < n$ amostrando, sem reposição, b observações a partir da sequência de dados original de tamanho n . Posteriormente, a estatística de interesse é calculada no conjunto inteiro de dados e depois calculada em todos os ${}^n C_b$ subconjuntos de dados de tamanho b . Os valores resultantes do cálculo da estatística em cada subamostra são convenientemente normalizados de forma a aproximarem a verdadeira distribuição amostral em causa. Assim, a subamostragem constitui um método muito geral para a construção de intervalos de confiança de 1ª ordem assintoticamente válidos pois, sob certas hipóteses em b (de forma a que $b \rightarrow \infty$ e $b/n \rightarrow 0$ quando $n \rightarrow \infty$ como veremos adiante), o método é válido sempre que a estatística original, convenientemente normalizada, apresenta uma distribuição limite sob o verdadeiro modelo. Outros métodos como o bootstrap, impõem que a distribuição da estatística seja de alguma forma estável enquanto tal condição não é necessária no método de subamostragem, funcionando este sob condições muito fracas.

Na verdade, o uso da subamostragem para aproximar a variância de uma estatística não constitui uma ideia nova. Os estimadores jackknife do viés e da variância baseados no cálculo da estatística em todas as subamostras de tamanho $n-1$ foram estudados por Quenouille-Tukey e aproximam-se da média e variância da distribuição amostral estimada estudada por Politis e Romano (1992) com $b=n-1$. Também Hartigan (1969) havia introduzido o que Efron (1982) apelidou de método de subamostragem aleatória, o qual se baseava no cálculo da estatística de interesse em todos os $2^n - 1$ subconjuntos de dados, produzindo assim limites de confiança exactos. Mais tarde, Hartigan (1975) adaptou estes resultados a um contexto mais geral de estimadores com distribuição assintótica normal. O jackknife e a subamostragem aleatória são métodos similares no sentido em que ambos usam

subconjuntos de dados para aproximarem o erro padrão de uma estatística ou a sua distribuição amostral.

Apesar de partilhar algumas propriedades com os métodos atrás referidos, o bootstrap de Efron (1979) veio corrigir algumas deficiências no jackknife, permitindo concretizar um objectivo mais ambicioso: aproximar uma distribuição amostral completa. Shao e Wu (1989) mostraram que, baseando o estimador jackknife da variância na estatística calculada em subamostras com d observações apagadas, muitas deficiências do estimador da variância jackknife usual (com $d=1$) podem ser ultrapassadas. Mais tarde, Wu (1990) usou estes resultados para aproximar uma distribuição amostral através do que chamou um histograma jackknife, mas apenas em situações i.i.d. onde a normalidade assintótica está assegurada. Posteriormente, Politis e Romano (1992) mostraram como os valores resultantes da subamostragem podem servir para estimar com exactidão uma distribuição amostral sem qualquer pressuposto de normalidade assintótica, assumindo apenas a existência de uma distribuição limite.

2.1. Subamostragem em Séries Temporais

Uma vez que o nosso foco de interesse são as séries temporais, limitamo-nos a apresentar aqui a metodologia da subamostragem de Politis e Romano (1994) para dados dependentes. Contudo, é de referir que, num contexto de dados independentes, a subamostragem funciona de forma muito similar àquela que a seguir apresentamos. Tal como no caso dos dados independentes, a principal condição a verificar em dados dependentes, é que a estatística de interesse, devidamente normalizada, possua uma distribuição limite não degenerada. A diferença crucial é que apenas blocos de dados consecutivos, de tamanho $b < n$, serão permitidos como subamostras, em vez de qualquer colecção de b dados amostrados sem reposição a partir da série original de forma a preservar a estrutura de dependência das observações.

Sejam $\{X_1, \dots, X_n\}$ n observações de uma série temporal estacionária que toma valores num espaço amostral S e seja P uma medida de probabilidade. O objectivo é construir um intervalo de confiança para um determinado parâmetro $\theta(P) \in \mathfrak{R}$, com base nas observações X_1, \dots, X_n . Assim, seja $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ um estimador do parâmetro de interesse $\theta(P)$.

Pretendemos estimar a verdadeira distribuição amostral de $\hat{\theta}_n$ de modo a fazer inferências sobre θ .

Assumamos que a sequência $\{X_t\}$ satisfaz uma certa condição de dependência fraca. Mais concretamente, dada uma sequência aleatória $\{Z_t\}$, seja \mathcal{F}_n^m a σ -álgebra gerada por $\{Z_t, n \leq t \leq m\}$ e definamos a sequência α -mixing correspondente como

$$\alpha_z(k) = \sup_n \sup_{A, B} |P(A \cap B) - P(A)P(B)|,$$

onde A e B são eventos nas σ -álgebras $\mathcal{F}_{-\infty}^n$ e $\mathcal{F}_{n+k}^{+\infty}$, respectivamente. A sequência $\{Z_t\}$ diz-se α -mixing ou que satisfaz uma condição de mistura forte se $\alpha_z(k) \rightarrow 0$ quando $k \rightarrow \infty$. Sherman (1997) refere que, intuitivamente, a condição $\alpha_z(k) \rightarrow 0$ quando $k \rightarrow \infty$ significa que observações separadas por um lag grande comportam-se aproximadamente como sendo independentes. No caso particular em que $\{X_1, \dots, X_n\}$ são observações i.i.d., $\alpha_z(k) = 0$ para $k > 0$.

O ponto fulcral do método de subamostragem consiste em aproximar a distribuição amostral da estatística em causa calculando-a novamente em subamostras dos dados observados, isto é, pretendemos aproximar a distribuição amostral de $\hat{\theta}_n$ através da distribuição empírica das réplicas $\hat{\theta}_b$ onde o tamanho b da subamostra é um inteiro positivo menor que o tamanho da amostra original n . No caso das séries temporais usamos apenas blocos de observações consecutivas, de tamanho b , como subamostras válidas sendo a primeira $\{X_1, X_2, \dots, X_b\}$ e a última $\{X_{n-b+1}, X_{n-b+2}, \dots, X_n\}$. Note-se que existem $q = n - b + 1$ blocos destes, sendo, obviamente $q = n - b + 1 \ll n C_b$, o número de subamostras a considerar no caso de dados independentes.

Seja $\hat{\theta}_{n,b,t} = \hat{\theta}_b(X_t, \dots, X_{t+b-1})$ o estimador de $\theta(P)$ baseado na subamostra $\{X_t, \dots, X_{t+b-1}\}$ (note-se que, de acordo com esta notação $\hat{\theta}_n = \hat{\theta}_{n,n,1}$). Seja $J_{b,t}(P)$ a distribuição amostral de $\tau_b(\hat{\theta}_{n,b,t} - \theta(P))$, onde τ_b é uma constante de normalização apropriada. Defina-se também a correspondente função distribuição acumulada como $J_{b,t}(x, P) = \text{Prob}_P \{ \tau_b(\hat{\theta}_{n,b,t} - \theta(P)) \leq x \}$.

Por conveniência, notemos por $J_n(P) = J_{n,1}(P)$ a distribuição amostral de $\tau_n(\hat{\theta}_n - \theta(P))$ e notemos a função distribuição acumulada correspondente por $J_n(\cdot, P)$.

Essencialmente, a única hipótese que precisamos colocar para estimar $J_n(P)$ de forma consistente é a seguinte:

Hipótese A:

Existe uma lei limite $J(P)$ tal que $J_n(P)$ converge fracamente para $J(P)$ quando $n \rightarrow \infty$.

Isto significa que o estimador, devidamente centrado e normalizado, possui uma distribuição limite. É difícil desenvolver uma teoria assintótica sem tal exigência.

As únicas condições a impor a b é que $b/n \rightarrow 0$ e $b \rightarrow \infty$ quando $n \rightarrow \infty$.

A aproximação subamostragem a $J_n(x, P)$ é então definida por:

$$L_{n,b}(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} I\{\tau_b(\hat{\theta}_{n,b,t} - \hat{\theta}_n) \leq x\}$$

onde $I\{E\}$ é a função indicatriz do evento $E = \{\tau_b(\hat{\theta}_{n,b,t} - \hat{\theta}_n) \leq x\}$.

Para cada t , $\{X_t, \dots, X_{t+b-1}\}$ é uma subamostra de tamanho b do verdadeiro modelo P . Logo, a distribuição exacta de $\tau_b(\hat{\theta}_{n,b,t} - \theta(P))$ é $J_b(P)$. A estacionariedade implica que a distribuição empírica dos $n-b+1$ valores de $\tau_b(\hat{\theta}_{n,b,t} - \theta(P))$ constitui uma boa aproximação a $J_n(P)$. Assim, sendo $\theta(P)$ desconhecido, podemos substituir $\theta(P)$ por $\hat{\theta}_n$ uma vez que $\tau_b(\hat{\theta}_n - \theta(P))$ é de ordem τ_b/τ_n em probabilidade e assumiremos que $\tau_b/\tau_n \rightarrow 0$.

Politis e Romano (1994) demonstram um resultado que estabelece as condições sob as quais o método de subamostragem conduz a resultados assintoticamente válidos para estatísticas gerais, funcionando sob condições mínimas. Nenhum teorema equivalente é considerado nos métodos de reamostragem como o bootstrap em blocos móveis. Pelo contrário, tais métodos exigem uma análise de cada caso muito mais complexa.

Teorema A [Politis e Romano (1994)]:

Notemos a sequência α -mixing correspondente a $\{X_t\}$ por $\alpha_X(\cdot)$.

Assumamos a Hipótese A e que $\tau_b/\tau_n \rightarrow 0$, $b/n \rightarrow 0$ e $b \rightarrow \infty$ quando $n \rightarrow \infty$.

Assumamos também que $\alpha_X(m) \rightarrow 0$ quando $m \rightarrow \infty$.

- i) Se x é um ponto de continuidade de $J(\cdot, P)$, então $L_{n,b}(x) \rightarrow J(x, P)$ em probabilidade.
- ii) Se $J(\cdot, P)$ é contínua, então $\sup_x |L_{n,b}(x) - J(x, P)| \rightarrow 0$ em probabilidade.
- iii) Para $\alpha \in [0, 1]$, seja $c_{n,b}(1-\alpha) = \inf \{x : L_{n,b}(x) \geq 1-\alpha\}$.

Correspondentemente, definamos $c(1-\alpha, P) = \inf \{x : J(x, P) \geq 1-\alpha\}$

Se dado $J(\cdot, P)$ é contínua em $c(1-\alpha, P)$, então

$$Prob_p \left\{ \tau_n \left[\hat{\theta}_n - \theta(P) \right] \leq c_{n,b}(1-\alpha) \right\} \rightarrow 1-\alpha \text{ quando } n \rightarrow \infty.$$

Então, a probabilidade de cobertura assintótica sob P do intervalo

$$I_1 = \left[\hat{\theta}_n - \tau_n^{-1} c_{n,b}(1-\alpha), \infty \right] \text{ é } 1-\alpha.$$

A demonstração do Teorema A encontra-se no Anexo I.

Note-se que, além da condição de mistura, a principal dificuldade em aplicar o teorema é verificar quando é que o estimador, devidamente normalizado, possui uma distribuição limite não degenerada, cuja forma, contudo, não precisa ser conhecida. Muito mais trabalho é geralmente necessário para demonstrar a validade dos métodos bootstrap. Por outro lado, o facto de não ser necessário que conheçamos $J(x, P)$ para que a subamostragem funcione, faz desta uma metodologia muito geral.

Um intervalo de confiança bilateral de caudas iguais pode ser obtido através da intersecção de dois intervalos unilaterais. Assim, o intervalo bilateral análogo a I_1 é

$$I_{ET} = \left[\hat{\theta}_n - \tau_n^{-1} c_{n,b} \left(1 - \frac{\alpha}{2} \right), \hat{\theta}_n - \tau_n^{-1} c_{n,b} \left(\frac{\alpha}{2} \right) \right].$$

I_{ET} diz-se um intervalo de caudas iguais por ter probabilidade aproximadamente igual em cada cauda:

$$Prob_p \left\{ \theta < \hat{\theta}_n - \tau_n^{-1} c_{n,b} \left(1 - \frac{\alpha}{2} \right) \right\} \approx Prob_p \left\{ \theta > \hat{\theta}_n - \tau_n^{-1} c_{n,b} \left(\frac{\alpha}{2} \right) \right\} \approx \frac{\alpha}{2}$$

Como aproximação alternativa, também podem ser construídos intervalos de confiança simétricos bilaterais. Para isso estimamos a função distribuição bilateral

$$J_{n|1}(x, P) = Prob_p \left\{ \tau_n \left[\hat{\theta}_n - \theta(P) \right] \leq x \right\}.$$



A aproximação por subamostragem a $J_{n,|l|}(x, P)$ é então definida por

$$L_{n,b,|l|}(x) = \frac{1}{n-b+1} \sum_{t=1}^{n-b+1} I\left\{\tau_b \left| \hat{\theta}_{n,b,t} - \hat{\theta}_n \right| \leq x\right\}.$$

Denotando um quantil $(1-\alpha)$ de $L_{n,b,|l|}$ por $c_{n,b,|l|}(1-\alpha)$, o intervalo subamostragem simétrico é então dado por

$$I_{sym} = \left[\hat{\theta}_n - \tau_n^{-1} c_{n,b,|l|}(1-\alpha), \hat{\theta}_n + \tau_n^{-1} c_{n,b,|l|}(1-\alpha) \right].$$

Intervalos de confiança simétricos não constituem necessariamente uma escolha superior pois a assimetria de um intervalo de caudas iguais pode conter informação útil acerca da localização do verdadeiro parâmetro e da assimetria da distribuição amostral do estimador.

Num estudo de simulação que apresentamos no capítulo seguinte verificamos que, ao aplicarmos a metodologia da subamostragem ao parâmetro autoregressivo (a) de um modelo AR(1), intervalos de confiança de caudas iguais (construídos a partir de I_{ET}) apresentam uma cobertura superior à dos intervalos de confiança simétricos (construídos a partir de I_{sym}) no caso em que $a > 0$.

2.2. Comparação entre a Subamostragem e o Bootstrap em Blocos Móveis

Seja Y_t o bloco de tamanho b dos dados $\{X_1, \dots, X_{t+b-1}\}$ e seja P_n^* a distribuição empírica dos blocos $Y_1, Y_2, \dots, Y_{n-b+1}$. Seja X_1^*, \dots, X_n^* a pseudo-série que é gerada reamostrando l blocos Y_1^*, \dots, Y_l^* i.i.d. de P_n^* e ligando-os. Sendo $\hat{\theta}_n^* = \hat{\theta}_n(X_1^*, \dots, X_n^*)$ o estimador de $\theta(P)$ calculado nas pseudo-séries, o bootstrap em blocos móveis aproxima a distribuição amostral de $\tau_n(\hat{\theta}_n - \theta(P))$ através da distribuição de $\tau_n(\hat{\theta}_n^* - \hat{\theta}_n)$ sob P_n^* , ou seja, por $K_{n,b}(x) = Prob_{P_n^*} \left\{ \tau_n(\hat{\theta}_n^* - \hat{\theta}_n) \leq x \right\}$. Note-se que, o bootstrap em blocos móveis é válido apenas em situações regulares onde a natureza da constante de normalização τ_n é bem conhecida. (Desde que as constantes de normalização sejam iguais na verdadeira distribuição amostral e na aproximação bootstrap, não é necessário incluir τ_n . Contudo, fazê-mo-lo de modo a facilitar a comparação com o método de subamostragem).

Apesar de a distribuição de $\tau_n(\hat{\theta}_n^* - \hat{\theta}_n)$ sob P_n^* , acima notada por $K_{n,b}(\cdot)$, estar bem definida, em geral, não pode ser obtida analiticamente e por isso temos que recorrer às aproximações de Monte Carlo. Assim, um elevado número B de pseudo-séries X_1^*, \dots, X_n^* é gerado por reamostragem a partir dos blocos da série original com reposição, sendo a estatística de interesse $\hat{\theta}_n$ calculada em cada série. Este processo origina B pseudo-estatísticas $\hat{\theta}_{n,1}^*, \dots, \hat{\theta}_{n,B}^*$. A aproximação Monte Carlo da distribuição $K_{n,b}(x)$ é então dada por

$$\hat{K}_{n,b}(x) = \frac{1}{B} \sum_{h=1}^B I\{\tau_n(\hat{\theta}_{n,h}^* - \hat{\theta}_n) \leq x\}.$$

Dada a aproximação $\hat{K}_{n,b}$ da distribuição amostral de $\tau_n(\hat{\theta}_n - \theta(P))$, a inferência sobre $\theta(P)$ pode ser obtida de modo análogo pelo método de subamostragem. Por exemplo, os intervalos de confiança para parâmetros univariados $\theta(P)$ baseiam-se em quantis apropriados de $\hat{K}_{n,b}$.

Os métodos de subamostragem e bootstrap em blocos móveis têm em comum o facto de utilizarem blocos de dados Y_i de tamanho b . A diferença é a seguinte: enquanto na subamostragem encaramos estes blocos como sub-séries ou “mini séries temporais”, no bootstrap em blocos temos que encará-los como peças que permitem construir novas pseudo-séries. Note-se que, na subamostragem, cada subconjunto de tamanho b (tomado sem reposição a partir do conjunto de dados original) é, na verdade, uma amostra de tamanho b do verdadeiro modelo. Logo, é intuitivo que podemos aproximar a distribuição amostral da estatística normalizada com base em tais amostras. Por sua vez, o bootstrap baseia-se no cálculo da estatística de interesse em amostras de tamanho n a partir do modelo estimado o qual apenas supomos estar próximo do verdadeiro modelo. Isto leva a que o método de subamostragem funcione sob hipóteses fracas pois, para além das condições de mistura, a única condição a impor é que o estimador $\hat{\theta}_n$ possua uma distribuição limite não degenerada. Pelo contrário, em geral, o bootstrap em blocos móveis só é válido para estatísticas assintoticamente lineares e parâmetros de uma distribuição marginal m -dimensional, com m fixo, como a média ou a autocorrelação. Assim, para que o bootstrap em blocos móveis seja válido, é, geralmente necessário impor condições fortes à série subjacente, de forma a garantir a existência de uma distribuição limite.

Assim, Sherman (1997) refere que os métodos baseados no uso de subséries permanecem válidos em muitas situações em que o bootstrap falha, como é o caso das séries que satisfazem condições de mistura fortes.

2.3. Subamostragem com constante de normalização desconhecida

Como vimos na secção anterior, para construir um intervalo de confiança para um determinado parâmetro desconhecido $\hat{\theta}(P)$, precisamos de uma aproximação da distribuição amostral de uma estatística normalizada $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$, consistente com $\theta(P)$ para alguma constante de normalização conhecida τ_n . Note-se que, em geral, τ_n depende de P , ainda que tal dependência não seja explicitamente referida. Apesar de ser condição necessária a existência de uma distribuição assintótica para podermos aproximar a distribuição de uma estatística $\hat{\theta}_n$, a constante de normalização τ_n pode ser desconhecida porque desconhecemos P e pode não ser possível assumir que P pertence a uma classe particular de medidas que partilham uma determinada constante de normalização τ_n . Neste caso, a subamostragem pode dar-nos alguma informação acerca da forma de $J(x, P)$ - enquanto função de x - o que pode constituir uma ferramenta de diagnóstico útil. Assim, é possível utilizar a metodologia de subamostragem para encontrar um estimador consistente para τ_n , usando depois a constante de normalização estimada para construir intervalos de confiança assintoticamente correctos. A ideia subjacente assenta no facto de ser possível construir a distribuição subamostragem de $\hat{\theta}_n$ (em vez de $\tau_n \hat{\theta}_n$). Construindo diversas distribuições subamostragem para diversas escolhas de b , obtemos informação acerca da forma de τ_n como função de n .

Assim, Bertail et al (1999), demonstram que a subamostragem pode conduzir a resultados consistentes sob condições mínimas ainda que a distribuição amostral da estatística em causa seja desconhecida, isto é, não é necessário conhecer *a priori* a forma da constante de normalização τ_n para aplicar a subamostragem. Na verdade, a metodologia da subamostragem pode constituir a única forma de construir regiões de confiança exactas.

A título de exemplo, Politis et al (1999), consideram o problema da construção de um intervalo de confiança para o parâmetro autoregressivo (a) num modelo AR(1)

$$X_t = \mu + aX_{t-1} + \varepsilon_t$$

onde $\{\varepsilon_t\}$ é uma sequência de inovações de ruído branco estritamente estacionária e $a \in]-1, 1[$. Como vimos no capítulo 1, o estimador de mínimos quadrados converge segundo a constante $n^{1/2}$ se $|a| < 1$, para $a=1$ e $\mu=0$ a velocidade de convergência é n e para $a=1$ e $\mu \neq 0$ é $n^{3/2}$. Assim, a velocidade de convergência depende de parâmetros desconhecidos.

Bertail et al (1999) apresentam um procedimento simples para estimar a velocidade de convergência, o qual passamos a apresentar de forma sucinta:

Suponhamos, por simplicidade, que $\hat{\theta}_n$ e $\theta(P)$ são valores reais e que $J(x,P)$ é contínua em x em toda a recta real. Seja Y_a a subsequência $\{X_a, X_{a+1}, \dots, X_{a+b-1}\}$ para $a=1, \dots, q$ e $q=n-b+1$. Note-se que Y_a consiste em b observações consecutivas da série original $\{X_1, \dots, X_n\}$ e a ordem das observações é preservada. Seja $\hat{\theta}_{n,b,a}$ o valor da estatística $\hat{\theta}_b$ aplicada à amostra Y_a . A aproximação subamostragem da distribuição amostral de $\tau_n(\hat{\theta}_n - \theta(P))$, baseada em blocos de tamanho b , é definida por

$$L_{n,b}(x|\tau_*) \equiv \frac{1}{q} \sum_{a=1}^q I\{\tau_b(\hat{\theta}_{n,b,a} - \hat{\theta}_n) \leq x\}$$

Note-se que está explícita a dependência da aproximação subamostragem da constante τ_* .

Por outro lado, Politis e Romano (1992, 1994) demonstram que as condições $J_n(x,P) \equiv Prob_p\{\tau_n(\hat{\theta}_n - \theta(P)) \leq x\} \rightarrow J(x,P)$, $b \rightarrow \infty$ e $b/n \rightarrow 0$ quando $n \rightarrow \infty$ asseguram que $\sup_x |L_{n,b}(x|\tau_*) - J(x,P)| = o_p(1)$ quando $n \rightarrow \infty$.

Notemos por $L_{n,b}(x|I)$ a aproximação subamostragem da distribuição amostral de $(\hat{\theta}_n - \theta(P))$.

Dada uma distribuição G na recta real e um número $\alpha \in]0,1[$, denotaremos por $G^{-1}(\alpha)$ a transformação quantil, isto é,

$$G^{-1}(\alpha) = \inf\{x : G(x) \geq \alpha\}$$

o que se reduz à inversa da função G se G for invertível. Note-se que temos $L_{n,b}(x\tau_b^{-1}|I) = L_{n,b}(x|\tau_*)$ e logo é fácil ver que $L_{n,b}^{-1}(\alpha|\tau_*) = \tau_b L_{n,b}^{-1}(\alpha|I)$.

O lema seguinte será necessário:

Lema A [Bertail, Politis e Romano (1999)]:

Seja $k_0 = \sup\{x : J(x,P) = 0\}$ e $k_1 = \sup\{x : J(x,P) = 1\}$. Assumamos que $J(x,P)$ é contínua e estritamente crescente em $]k_0, k_1[$ como função de x .

Se $\sup_x |L_{n,b}(x|\tau_*) - J(x,P)| = o_p(1)$ quando $n \rightarrow \infty$ então $L_{n,b}^{-1}(\alpha|\tau_*) = J^{-1}(\alpha, P) + o_p(1)$ para qualquer $\alpha \in]0,1[$.

Como consequência do lema temos

$$L_{n,b}^{-1}(\alpha | I) = \tau_b^{-1} J^{-1}(\alpha, P) + o_p(\tau_b^{-1}).$$

Uma vez que $L_{n,b}^{-1}(t | I)$ é aproximadamente proporcional a τ_b^{-1} , se construirmos diversas distribuições subamostragem para diferentes dimensões b , conseguimos estimar a constante de normalização.

Mais precisamente, para cada ponto $\alpha > J(0, P)$, temos

$$\log(L_{n,b}^{-1}(\alpha | I)) = \log(J^{-1}(\alpha, P)) - \log(\tau_b) + o_p(1).$$

Consequentemente, se escolhermos $b_1 \neq b_2$, temos

$$\log(L_{n,b_2}^{-1}(\alpha | I)) - \log(L_{n,b_1}^{-1}(\alpha | I)) = \log\left(\frac{\tau_{b_1}}{\tau_{b_2}}\right) + o_p(1).$$

Daqui decorre que, com duas distribuições subamostragem, é possível estimar a forma da função τ_\bullet . Se τ_n é da forma $\tau_n = n^\beta$ onde β é uma constante desconhecida, então

$$\log\left(\frac{b_1}{b_2}\right)^{-1} \left(\log(L_{n,b_2}^{-1}(\alpha | I)) - \log(L_{n,b_1}^{-1}(\alpha | I)) \right) = \beta + o_p\left(\log\left(\frac{b_1}{b_2}\right)^{-1}\right).$$

A média sobre vários pontos $\alpha_j, j=1, \dots, J$ do membro esquerdo da última equação constitui um estimador consistente de β , tendo em conta que escolhemos b_1 e b_2 tal que $\log\left(\frac{b_1}{b_2}\right) \rightarrow \infty$ quando $n \rightarrow \infty$ (tomamos, por exemplo, $b_i = n^{\beta_i}$, onde $1 > \beta_1 > \beta_2 > 0$ são constantes). De forma mais geral, podemos construir I distribuições subamostragem com base em subamostras de tamanho $b_i, i=1, \dots, I$; então, calculando a equação $\log(L_{n,b}^{-1}(\alpha | I)) = \log(J^{-1}(\alpha, P)) - \log(\tau_b) + o_p(1)$ em alguns pontos $\alpha_j > J(0, P), j=1, \dots, J$, asseguramos que

$$y_{i,j} \equiv \log(L_{n,b_i}^{-1}(\alpha_j | I)) = a_j - \beta \log(b_i) + u_{i,j}$$

onde $a_j \equiv \log(J^{-1}(\alpha_j, P))$ e $u_{i,j} = o_p(1)$, $i=1, \dots, I$ e $j=1, \dots, J$.

A equação acima pode ser interpretada como uma análise da covariância e assim, o próprio estimador de β surge naturalmente como:

$$\hat{\beta}_{1,J} \equiv - \frac{\sum_{i=1}^I (y_{i.} - \bar{y})(\log b_i - \bar{\log})}{\sum_{i=1}^I (\log b_i - \bar{\log})^2} \quad \text{onde}$$

$$y_{i.} = J^{-1} \sum_{j=1}^J y_{i,j} = J^{-1} \sum_{j=1}^J \log(L_{n,b_i}^{-1}(\alpha_j | I)) \quad , \quad \bar{y} = (IJ)^{-1} \sum_{i=1}^I \sum_{j=1}^J y_{i,j} \quad \text{e} \quad \bar{\log} = I^{-1} \sum_{i=1}^I \log b_i .$$

Por exemplo, se $I=2$ e $J=1$, o estimador sugerido será

$$\beta_{2,1} = \log \left(\frac{b_1}{b_2} \right)^{-1} \left(\log(L_{n,b_2}^{-1}(\alpha | I)) - \log(L_{n,b_1}^{-1}(\alpha | I)) \right).$$

Bertail et al (1999) demonstram a consistência e validade assintótica de $\hat{\beta}_{1,J}$, enquanto estimador de β .

Vimos que é possível estimar de forma consistente a constante de normalização τ_* em situações gerais. Numa situação em que tenha sido estimada a constante τ_* , falta depois construir um intervalo de confiança para o parâmetro desconhecido em causa - $\theta(P)$. Neste contexto, os autores enunciam e demonstram um teorema que estabelece a validade assintótica da subamostragem, para valores reais de $\hat{\theta}_n$ e θ , com a constante de normalização estimada:

Teorema B [Bertail, Politis e Romano (1999)]:

Assumamos que $J_n(x,P) \equiv \text{Prob}_P \left\{ \hat{\tau}_n (\hat{\theta}_n - \theta(P)) \leq x \right\} \rightarrow J(x,P)$ quando $n \rightarrow \infty$ para $\tau_n = n^\beta$, para um determinado $\beta > 0$ e uma determinada $J(x,P)$ contínua em x . Assumamos também que $b \rightarrow \infty$ e $b/n \rightarrow 0$ quando $n \rightarrow \infty$.

Seja $\hat{\beta} = \beta + o_p((\log n)^{-1})$ e $\hat{\tau}_n = n^{\hat{\beta}}$. Então, $\sup_x |L_{n,b}(x | \hat{\tau}_n) - J(x,P)| = o_p(1)$.

Seja $\alpha \in]0,1[$ e seja $c_n(\alpha) = L_{n,b}^{-1}(\alpha | \hat{\tau}_n)$ o α -ésimo quantil da distribuição subamostragem $L_{n,b}(x | \hat{\tau}_n)$. Logo, $\text{Prob}_P \left\{ \hat{\tau}_n (\hat{\theta}_n - \theta(P)) \leq c_n(\alpha) \right\} \rightarrow \alpha$ quando $n \rightarrow \infty$.

Então, a probabilidade de cobertura assintótica do intervalo $[\hat{\theta}_n - \hat{\tau}_n^{-1} c_n(\alpha), \infty[$ é α .

2.4. Escolha do tamanho das subamostras

No contexto das séries temporais, o maior problema prático que surge na aplicação do método de subamostragem a uma estatística $\theta(P)$ reside na escolha do tamanho b dos blocos de observações consecutivas. Este problema não é exclusivo do método de subamostragem pois existe também no bootstrap em blocos móveis e outros métodos de blocos. Como vimos, as condições assintóticas necessárias são usualmente $b \rightarrow \infty$ e $b/n \rightarrow 0$ quando $n \rightarrow \infty$. Apesar de qualquer b que satisfaça estas condições assegurar a consistência pretendida para o método de subamostragem, tais condições dão-nos muito pouca orientação acerca da escolha de b numa amostra finita. De qualquer forma, é importante notar que a escolha de um “tamanho óptimo” para b depende do objectivo com que estamos a usar a subamostragem.

Suponhamos que o nosso objectivo é construir um intervalo de confiança para um parâmetro desconhecido $\theta(P)$, sendo $\hat{\theta}_n$ o estimador de $\theta(P)$ calculado a partir de n observações $\{X_1, \dots, X_n\}$ de uma série temporal. Neste caso, a distribuição amostral do estimador devidamente normalizado, tem que ser estimada.

Hall et al (1995) mostraram que, para o bootstrap em blocos móveis, no contexto de observações estacionárias, a constante de normalização para b depende da aplicação em causa. Segundo estes autores, a constante de normalização, é $n^{1/3}$ para a estimação do viés ou variância, $n^{1/4}$ para a estimação de uma função distribuição unilateral e $n^{1/5}$ para a estimação de uma função distribuição bilateral. Por esta razão, a condição imposta por Lahiri (1992) ao tamanho dos blocos para o bootstrap em blocos móveis - $b = o(n^{1/4})$ - parece demasiado restritiva.

Contudo, mesmo sendo conhecidas as constantes de proporcionalidade atrás referidas, surgem diversos problemas práticos. Neste contexto, Hall et al (1995) apresentaram uma regra prática para seleccionar o tamanho dos blocos. O primeiro passo é encontrar um tamanho de bloco óptimo para sub-séries de tamanho m menores que a série original, isto é $m < n$. Para tal, é necessário aplicar o método dos blocos móveis a todas as $n-m+1$ sub-séries de tamanho m bem como à série completa de tamanho n , seleccionando depois o tamanho de bloco para o qual o EQM é minimizado, conforme vimos na subsecção 1.3.2.. Assim, notamos por \hat{b}_m o tamanho de bloco óptimo para uma série de tamanho m , sendo depois o tamanho óptimo para a série original estimado por $\hat{b}_n = \binom{n}{m}^k \hat{b}_m$, onde $k=3, 4$ ou 5 , dependendo do contexto em causa.

Como demonstram Politis et al (1997) esta regra não conduz a resultados satisfatórios quando aplicada a diversos estudos de simulação e, como tal, necessitamos de um método que nos permita ultrapassar esta dificuldade. Assim, estes autores desenvolveram dois procedimentos que permitem escolher o tamanho dos blocos nos métodos de subamostragem: o **método de calibração** e o **método da variância mínima**, os quais passamos a descrever de forma sucinta.

2.4.1. Método de Calibração

Podemos pensar na exactidão de uma aproximação ou procedimento de confiança assintótico (tal como os métodos de reamostragem e subamostragem) em termos da sua calibração, ideia sugerida por Loh (1987).

Suponhamos que usamos este procedimento para construir um intervalo de confiança com nível de significância $1-\alpha$. Podemos notar tal nível de significância por $1-\lambda$ pois α é conhecido enquanto λ geralmente não é. Os métodos assintóticos apenas asseguram que $1-\lambda$ tende para $1-\alpha$ quando o tamanho da amostra tende para infinito. Para um tamanho finito da amostra, estes dois níveis são geralmente diferentes. Se conhecermos a função de calibração $h: 1-\alpha \rightarrow 1-\lambda$, podemos construir uma região de confiança exactamente com a cobertura desejada. Primeiro obrigamos a que $1-\lambda$ seja o nível pretendido e depois construímos um intervalo com nível de significância $1-\alpha$ tal que $h(1-\alpha) = 1-\lambda$. Por exemplo, se $h(0.98) = 0.95$ então um intervalo de confiança com nível de significância 98% seria um intervalo de confiança actual de 95%.

Felizmente, a função calibração $h(\cdot)$ pode ser estimada. Uma forma de fazer isso seria assumir um modelo paramétrico aproximado com um parâmetro conhecido θ_0 . Assim, usando uma aproximação Monte Carlo, seria fácil encontrar um estimador $\hat{h}(\cdot)$: iríamos gerar várias pseudo-sequências, depois calcular um intervalo $1-\alpha$ para cada sequência, e registar a proporção de intervalos que continham θ_0 . Mas, em vez disso, é mais fácil optar por uma aproximação não paramétrica usando a distribuição bootstrap P_n^* como mecanismo gerador de pseudo-dados. Para observações independentes, podemos usar o bootstrap de Efron. Num contexto de dados dependentes, é necessário recorrer, por exemplo, ao bootstrap em blocos móveis.

A aplicação do método de calibração à subamostragem pode condicionar o tamanho de bloco razoável. Isto significa que o tamanho dos blocos b fixado vai calibrar os intervalos subamostragem ao usarmos esse tamanho de bloco particular. Então, o problema de encontrar

o tamanho de bloco óptimo é eliminado. O algoritmo seguinte descreve, de modo mais formal, o método de calibração:

- 1) Geramos K pseudo-sequências $X_1^{*k}, \dots, X_n^{*k}$ de acordo com uma distribuição bootstrap P_n^* . Para cada sequência, $k=1, \dots, K$ calculamos um intervalo de confiança com um nível de significância $1-\alpha$ - $CI_{1-\alpha}^k$ - para alguns valores de α na vizinhança de λ .

- 2) Para cada α calculamos $\hat{h}(1-\alpha) = \frac{\#\{\hat{\theta}_n \in CI_{1-\alpha}^k\}}{K}$.

- 3) Interpolamos $\hat{h}(\cdot)$ entre os valores de α .

- 4) Encontramos o valor de α que satisfaz $\hat{h}(1-\alpha) = 1-\lambda$.

- 5) Construimos um intervalo de confiança com nível de significância $1-\alpha$.

No passo 1 deste algoritmo podemos recorrer ao bootstrap em blocos móveis usando um determinado bloco b_{BM} . Desta forma, limitamos os valores de α a $\alpha = \lambda$. Repetindo este algoritmo para diversos valores de b_{BM} , podemos seleccionar o valor de b_{BM} que assegura uma cobertura estimada próxima de $1-\lambda$.

Em situações nas quais não temos uma ideia prévia acerca do tamanho de bloco a usar, podemos fixar o nível de significância no nível desejado, isto é, escolher $\alpha = \lambda$ e depois estimar a função calibração de bloco $g : b \rightarrow 1-\lambda$. Neste caso, a segunda parte do passo 1 é substituída pelo cálculo de um intervalo de confiança com nível de significância $1-\lambda$ - CI_b^k para seleccionarmos os tamanhos de bloco b . Por sua vez, no passo 2 do algoritmo, para cada

b calculamos $\hat{g}(b) = \frac{\#\{\hat{\theta}_n \in CI_b^k\}}{K}$. Um tamanho de bloco razoável deve então satisfazer $\hat{g}(b) \approx 1-\lambda$.

Obviamente que, se já temos à partida um tamanho de bloco b que satisfaz $\hat{g}(b) \approx 1-\lambda$, não é necessária qualquer calibração. Assim, usando esse b particular, construimos depois um intervalo de confiança com nível de significância $1-\alpha = 1-\lambda$.

Uma vez que o método de calibração é baseado na função calibração estimada $\hat{h}(\cdot)$, os intervalos calibrados não são exactos. O grande inconveniente do método de calibração reside no facto de a estimação de $h(\cdot)$ se basear na distribuição bootstrap de P_n^* . Assim, serão obtidos bons resultados se P_n^* for uma boa aproximação do mecanismo de probabilidade subjacente, isto é, se o bootstrap conduzir a intervalos de confiança assintoticamente correctos. Torna-se assim necessário um método alternativo à técnica de calibração o qual possa ser usado nos casos em que o bootstrap de Efron ou em blocos móveis sejam inconsistentes, pelo que, Politis et al (1999) sugerem o método da variância mínima, um método mais geral que passamos a apresentar sucintamente.

2.4.2. Método da Variância Mínima

Esta técnica baseia-se no facto de ser necessário, para assegurar a consistência do método de subamostragem, que a dimensão dos blocos b tenda para infinito com o tamanho n da amostra satisfazendo $b/n \rightarrow 0$. Na realidade, para valores de b próximos de n , as estatísticas $\hat{\theta}_{n,b,t}$ serão quase iguais a $\hat{\theta}_n$, sendo, portanto, difícil obter a distribuição subamostragem. Por outro lado, se b é muito pequeno, os intervalos podem conduzir a uma subestimação ou sobrestimação dependendo da estatística em causa. Assim, surge a ideia de calcular intervalos subamostragem para um número largo de tamanhos de bloco b e depois procurar uma região onde os intervalos não mudem muito. Dentro dessa região, escolhe-se então um intervalo, de acordo com um determinado critério arbitrário.

Embora este método possa assentar numa “inspecção visual” é necessário dispor de um procedimento de selecção automático. O procedimento proposto por Politis et al (1999) baseia-se na minimização de um *desvio padrão móvel*. Assumamos que calculamos intervalos subamostragem para um tamanho de bloco b que varia entre b_{small} e b_{big} . Os pontos extremos dos intervalos de confiança devem mudar de forma diferenciável, quando b muda. Isto pode, em parte, ser violado se utilizarmos uma aproximação estocástica, para amostras de tamanho moderado ou grande. Nesse caso convém imprimir alguma diferenciabilidade aplicando uma *média móvel* aos pontos extremos dos intervalos. Finalmente, um *desvio padrão móvel* aplicado aos pontos extremos determina a variância em torno de um valor de b específico, sendo escolhido o valor de b associado com menor variância. Apresentamos a seguir uma descrição mais formal do algoritmo:

3. APLICAÇÃO DA SUBAMOSTRAGEM A MODELOS AUTOREGRESSIVOS

- 1) Para $b = b_{small}$ até $b = b_{big}$ calculamos um intervalo de confiança para $\theta(P)$ com o nível de confiança pretendido, resultando daí os pontos extremos $I_{b,low}$ e $I_{b,up}$.
- 2) Se tivermos usado uma aproximação estocástica no passo 1, diferenciamos os pontos extremos "lower" e "upper" separadamente, usando uma *média móvel* de distância m . Isto significa substituir $I_{b,low}$ pela média dos valores $\{I_{b-m,low}, I_{b-m+1,low}, \dots, I_{b+1,low}\}$ e o mesmo para $I_{b,up}$.
- 3) Para cada b , calculamos o índice de variância VI_b como o desvio padrão do intervalo entre os pontos extremos numa vizinhança de b . Mais especificamente, para um inteiro pequeno k , seja VI_b igual ao desvio padrão dos pontos extremos $\{I_{b-k,low}, \dots, I_{b+k,low}\}$ mais o desvio padrão dos pontos extremos $\{I_{b-k,up}, \dots, I_{b+k,up}\}$.
- 4) Seleccionamos o valor b^* correspondente ao menor índice de variância e apresentamos $[I_{b^*,low}; I_{b^*,up}]$ como o intervalo de confiança final.

Este algoritmo contém dois parâmetros: m e k . Politis et al (1999) referem que estudos de simulação evidenciam que o algoritmo é pouco sensível à escolha dos referidos parâmetros e por isso empregam-se geralmente $m=2$ ou $m=3$ e $k=2$ ou $k=3$.

Na estimativa de uma função distribuição tanto podemos usar o Método de Calibração como o Método da Variância Mínima. Note-se que a ideia do Método de Calibração é começar com um tamanho de bloco razoável e depois ajustar o nível de significância estimando a função calibração. Esta estimativa é feita gerando pseudo-dados através de um mecanismo apropriado, tal como o bootstrap. O inconveniente desta técnica é que nem sempre é possível encontrar tal mecanismo e por isso é por vezes necessário recorrer ao Método da Variância Mínima, o qual, em geral, se mostra mais robusto.

onde $R_p =$

$$R_p = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^p$$

Sob a hipótese de ruído branco Gaussiano $N(0, \sigma^2)$, Quenneville e Davis (1987) provam o seguinte resultado assintótico:

$$\sqrt{n}(R_p - R_p(\theta)) \rightarrow N(0, V_p(\theta)) \text{ onde } R_p = (a_1, a_2, \dots, a_p)' \text{ e } V_p(\theta) = \sigma^2 R_p'$$

3. APLICAÇÃO DA SUBAMOSTRAGEM A MODELOS AUTOREGRESSIVOS

Este capítulo resultou da aplicação da teoria apresentada a séries geradas por modelos autoregressivos de 1ª ordem - AR(1). Para tal, foram levados a efeito alguns estudos de simulação em MatLab, cujos resultados aqui apresentamos. Apresentamos também uma breve revisão acerca da distribuição assintótica dos estimadores de mínimos quadrados / máxima verosimilhança dos parâmetros autoregressivos - $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p$ - de um modelo AR(p).

Uma abordagem mais detalhada acerca dos modelos AR(p) e da estimação dos parâmetros nestes modelos encontra-se nos Anexos II e III, respectivamente.

3.1. Distribuição assintótica dos estimadores $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p$

Seja $\{X_t\}$ um **processo autoregressivo de ordem p - AR(p)** - satisfazendo a seguinte equação:

$$X_t - a_1 X_{t-1} - \dots - a_p X_{t-p} = \varepsilon_t$$

onde a_1, a_2, \dots, a_p são constantes satisfazendo as condições de estacionariedade e $\{\varepsilon_t\}$ é uma sucessão de variáveis aleatórias i.i.d. com distribuição $N(0, \sigma_\varepsilon^2)$.

Sob a hipótese de que os $\{\varepsilon_t\}$ são i.i.d. (mas não necessariamente normais), com momentos até à 4ª ordem finitos, Mann e Wald (1943) demonstraram que os estimadores de mínimos quadrados $\{\sqrt{N}(\hat{a}_1 - a_1), \sqrt{N}(\hat{a}_2 - a_2), \dots, \sqrt{N}(\hat{a}_p - a_p)\}$ possuem uma distribuição normal multivariada assintótica com média zero e matriz de variâncias-covariâncias $\sigma_\varepsilon^2 \mathbf{R}_p^{-1}$

$$\text{onde } \mathbf{R}_p = \begin{bmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \dots & \gamma_{p-2} \\ \dots & \dots & \dots & \dots \\ \gamma_{p-1} & \gamma_{p-2} & \dots & \gamma_0 \end{bmatrix}.$$

Sob a hipótese de ruído branco Gaussiano $N(0, \sigma_\varepsilon^2)$, também Brockwell e Davis (1987) provam o seguinte resultado assintótico:

$$\sqrt{N}(\hat{\theta} - \theta) \sim N_p(0, V(\theta)) \text{ onde } \theta = (a_1, a_2, \dots, a_p) \text{ e } V(\theta) = \sigma_\varepsilon^2 \mathbf{R}_p^{-1}.$$

Como aproximação podemos substituir σ_ε^2 por $\hat{\sigma}_\varepsilon^2$ e os elementos de R_p pelas correspondentes autocovariâncias amostrais.

No caso particular de termos $p=1$, o resultado apresentado vem $N \text{var}(\hat{a}) \sim \sigma_\varepsilon^2 \{V_0\}^{-1} = \frac{\sigma_\varepsilon^2}{\sigma_x^2}$. Mas, como num modelo AR(1) temos $\sigma_x^2 = \frac{\sigma_\varepsilon^2}{1-a^2}$, então $\text{var}(\hat{a}) \sim \frac{1-a^2}{N}$. Portanto, para um processo AR(1) temos $\hat{a} \sim N\left(a, \frac{1}{N}(1-a^2)\right)$.

Brockwell e Davis (1987) referem que os estimadores de mínimos quadrados possuem as mesmas propriedades assintóticas que os de máxima verosimilhança, sendo consequentemente válidos para aqueles estimadores os resultados anteriormente estabelecidos.

Segundo Murteira et al (1993), tendo em conta a distribuição assintótica dos estimadores $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p$ num modelo AR(p) e designando por $v_{ii}^2(\theta)$ o i -ésimo elemento da diagonal principal da matriz $V(\hat{\theta})$, pode concluir-se que

$$\frac{\sqrt{N}(\hat{\theta}_i - \theta_i)}{v_{ii}(\hat{\theta})} \sim t_{N-p}$$

onde t_{N-p} denota a distribuição de Student com $N-p$ graus de liberdade.

Priestley (1981) descreve um método geral, desenvolvido por Box e Jenkins, que permite a obtenção de intervalos de confiança para os parâmetros de um modelo AR(p) o qual apresentamos no Anexo IV.

3.2. Estudo de Simulação

O objectivo desta secção é apresentar os resultados de dois estudos de simulação levados a efeito em MatLab. No primeiro estudo aplicámos a metodologia da subamostragem a amostras geradas por modelos AR(1) com o objectivo de construir intervalos de confiança por subamostragem e intervalos de confiança clássicos para o parâmetro autoregressivo, comparando as respectivas probabilidades de cobertura estimadas. Por sua vez, o objectivo do segundo estudo é caracterizar a distribuição dos estimadores de máxima verosimilhança calculados nas subamostras.

3.2.1. Intervalos de Confiança para o Parâmetro Autoregressivo

O objectivo deste estudo é analisar o comportamento da metodologia da subamostragem em pequenas amostras.

Começámos por gerar amostras de 256 observações de um processo AR(1) onde a sequência de inovações $\{\varepsilon_t\}$ tem distribuição $N(0,1)$, e o parâmetro autoregressivo (a) toma sucessivamente os valores $0.2, 0.5, 0.8, 0.95, -0.2, -0.5, -0.8$ e -0.95 . Aplicámos a subamostragem a cada uma das oito amostras geradas usando diversos tamanhos de bloco ($b=4, b=8, b=16, b=32$ e $b=64$). Para cada situação construímos, para o parâmetro autoregressivo, um intervalo de confiança a 95% de caudas iguais (I_{ET}) e um intervalo de confiança a 95% simétrico (I_{sym}). Repetimos este procedimento 500 vezes, registando o número de intervalos de confiança que, em cada simulação, continham o verdadeiro parâmetro e estimamos a probabilidade de cobertura para cada caso. Além dos intervalos de confiança por subamostragem, construímos, a partir de cada uma das oito amostras inicialmente geradas, um intervalo de confiança clássico (I_{class}). Este intervalo de confiança foi construído com base no resultado assintótico de Murteira et al (1993) apresentado na secção 3.1 o qual, para o caso particular de um processo AR(1), vem

$$\frac{\sqrt{N}(\hat{a}-a)}{v_{11}(\hat{a})} \sim t_{N-1} \quad \text{onde } N \text{ é o número de observações de que dispomos e}$$

$$v_{11}(\hat{a}) = \sqrt{\hat{\sigma}_\varepsilon^2 \hat{\gamma}_0^{-1}} = \sqrt{\hat{\sigma}_\varepsilon^2 \cdot \frac{1-\hat{\rho}_1}{\hat{\sigma}_\varepsilon^2}} = \sqrt{1-\hat{\rho}_1}$$

Com base neste resultado, um intervalo de confiança bilateral com nível de significância $1-\alpha$ para a será dado por

$$I_{class} = \left[\hat{a} - t_{N-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{\frac{1-\hat{\rho}_1}{N}}, \hat{a} + t_{N-1} \left(1 - \frac{\alpha}{2} \right) \sqrt{\frac{1-\hat{\rho}_1}{N}} \right]$$

Na tabela 3.1 apresentamos a probabilidade de cobertura estimada para cada um dos intervalos de confiança construídos. Da análise da referida tabela verificamos que, para $a > 0$, intervalos de confiança de caudas iguais apresentam uma probabilidade de cobertura superior à dos intervalos de confiança simétricos. Por sua vez, para $a < 0$, são os intervalos de confiança simétricos que apresentam, uma probabilidade de cobertura superior. Da análise dos valores apresentados concluí-se ainda que, tamanhos de bloco pequenos conduzem a intervalos de confiança para o parâmetro autoregressivo com melhor cobertura.

| | | $b=4$ | $b=8$ | $b=16$ | $b=32$ | $b=64$ | I_{class} |
|-----------|-----------|-------|-------|--------|--------|--------|-------------|
| $a=0.2$ | I_{ET} | 0.98 | 0.91 | 0.88 | 0.86 | 0.82 | 0.97 |
| | I_{sym} | 0.94 | 0.85 | 0.83 | 0.78 | 0.77 | |
| $a=0.5$ | I_{ET} | 0.99 | 0.91 | 0.88 | 0.86 | 0.80 | 0.94 |
| | I_{sym} | 0.93 | 0.78 | 0.77 | 0.76 | 0.69 | |
| $a=0.8$ | I_{ET} | 0.99 | 0.93 | 0.88 | 0.85 | 0.79 | 0.97 |
| | I_{sym} | 0.94 | 0.74 | 0.67 | 0.67 | 0.63 | |
| $a=0.95$ | I_{ET} | 1.00 | 0.99 | 0.95 | 0.90 | 0.83 | 0.92 |
| | I_{sym} | 0.98 | 0.85 | 0.72 | 0.68 | 0.56 | |
| $a=-0.2$ | I_{ET} | 0.98 | 0.91 | 0.92 | 0.85 | 0.79 | 0.95 |
| | I_{sym} | 0.96 | 0.91 | 0.91 | 0.88 | 0.78 | |
| $a=-0.5$ | I_{ET} | 0.97 | 0.92 | 0.90 | 0.86 | 0.79 | 0.94 |
| | I_{sym} | 0.98 | 0.98 | 0.95 | 0.91 | 0.81 | |
| $a=-0.8$ | I_{ET} | 0.98 | 0.95 | 0.88 | 0.86 | 0.81 | 0.95 |
| | I_{sym} | 1.00 | 0.99 | 0.98 | 0.93 | 0.87 | |
| $a=-0.95$ | I_{ET} | 1.00 | 1.00 | 0.94 | 0.95 | 0.81 | 0.94 |
| | I_{sym} | 1.00 | 1.00 | 0.99 | 0.98 | 0.87 | |

Tabela 3.1. Probabilidade de cobertura estimada para intervalos de confiança a 95% por subamostragem - de caudas iguais (I_{ET}) e simétricos (I_{sym}) - e clássicos (I_{class}), para o parâmetro autoregressivo (a). Todos os intervalos foram construídos com base em amostras de 256 observações de processos AR(1). Para os intervalos de confiança por subamostragem, foram usados diversos tamanhos de bloco (b). As estimativas apresentadas são baseadas em 500 réplicas de cada cenário.

Para uma análise mais pormenorizada destes resultados foram elaborados os gráficos relativos aos limites inferiores e superiores dos 500 intervalos de confiança construídos para cada situação. Nos gráficos 3.1, 3.2 e 3.3 apresentamos, para $a=0.2$, os limites dos intervalos de confiança a 95% por subamostragem com $b=4$ (de caudas iguais e simétricos) e clássicos. Nos gráficos 3.4, 3.5, e 3.6 apresentamos as situações análogas para $a=-0.95$. Os gráficos relativos à restantes situações encontram-se no Anexo V. Da análise da tabela 3.1 e dos gráficos do Anexo V podemos retirar as seguintes conclusões:

- Os intervalos de confiança clássicos apresentam uma probabilidade de cobertura comparável à dos intervalos de confiança por subamostragem. Note-se porém que, à medida que $|a|$ aumenta, os intervalos de confiança construídos por subamostragem têm tendência a apresentar amplitudes menores mas probabilidades de cobertura superiores enquanto, para a mesma situação, intervalos de confiança clássicos apresentam amplitudes menores e probabilidades de cobertura inferiores.
- Para $a>0$ e para um determinado valor de b , intervalos de confiança de caudas iguais apresentam, em geral, maiores amplitudes que intervalos de confiança simétricos enquanto que, para $a<0$, esta situação tende a inverter-se.
- Nos intervalos de confiança por subamostragem, fixando o valor de a e aumentando b , a amplitude dos intervalos diminui, sendo esta situação mais evidente quando $a>0$.

Note-se também que, neste caso, a probabilidade de cobertura dos intervalos tende a diminuir.

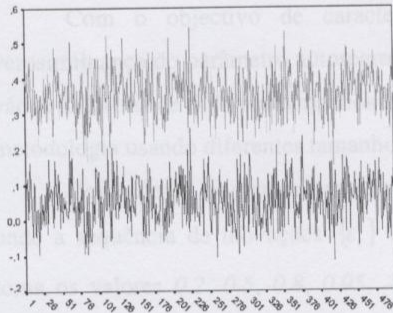


Gráfico 3.1. Limites inferiores e superiores dos 500 intervalos de confiança a 95% de caudas iguais (I_{ET}) para $a=0.2$ e $b=4$.

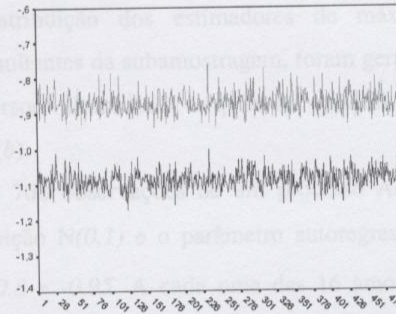


Gráfico 3.4. Limites inferiores e superiores dos 500 intervalos de confiança a 95% de caudas iguais (I_{ET}) para $a=-0.95$ e $b=4$.

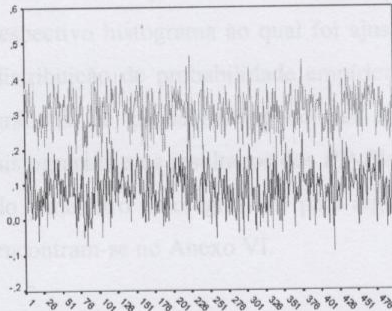


Gráfico 3.2. Limites inferiores e superiores dos 500 intervalos de confiança a 95% simétricos (I_{sym}) para $a=0.2$ e $b=4$.

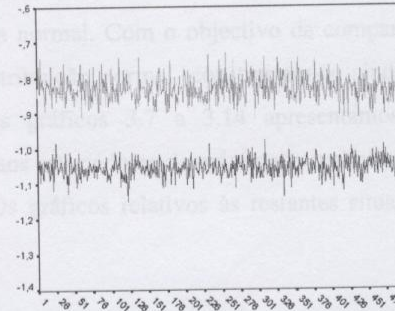


Gráfico 3.5. Limites inferiores e superiores dos 500 intervalos de confiança a 95% simétricos (I_{sym}) para $a=-0.95$ e $b=4$.

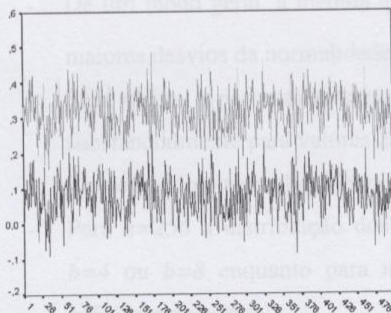


Gráfico 3.3. Limites inferiores e superiores dos 500 intervalos de confiança a 95% clássicos (I_{class}) para $a=0.2$.

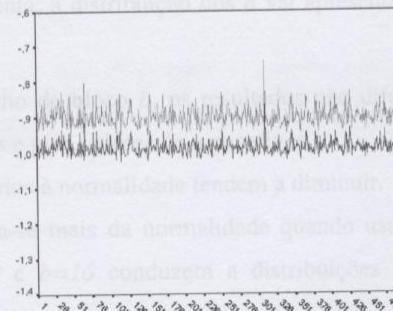


Gráfico 3.6. Limites inferiores e superiores dos 500 intervalos de confiança a 95% clássicos (I_{class}) para $a=-0.95$.

3.2.2. Distribuição dos Estimadores

Como vimos na secção 3.1, a distribuição dos estimadores $\hat{a}_1, \hat{a}_2, \dots, \hat{a}_p$ num processo AR(p) é assintoticamente normal.

Com o objectivo de caracterizar a distribuição dos estimadores de máxima verosimilhança do parâmetro autoregressivo (a) resultantes da subamostragem, foram geradas várias amostras de diversos tamanhos (n) para diversos valores de a , às quais foi aplicada esta metodologia usando diferentes tamanhos de bloco (b).

Começámos por gerar amostras de 256 e 768 observações de um processo AR(1) onde a sequência de inovações $\{\varepsilon_i\}$ tem distribuição $N(0,1)$ e o parâmetro autoregressivo toma os valores 0.2, 0.5, 0.8, 0.95, -0.2, -0.5, -0.8 e -0.95. A cada uma das 16 amostras geradas foi aplicada a subamostragem usando blocos de tamanho 4, 8, 16, 32 e 64. O estimador de máxima verosimilhança do parâmetro autoregressivo (\hat{a}) foi calculado em cada subamostra. A partir dos valores resultantes do cálculo de \hat{a} nas subamostras construiu-se o respectivo histograma ao qual foi ajustada a curva normal. Com o objectivo de comparar a distribuição de probabilidade empírica com a distribuição normal, construíram-se ainda os gráficos de probabilidade normal *PP-Plot*. Nos gráficos 3.7 a 3.14 apresentamos os histogramas e os resultados dos *PP-Plot* relativos aos estimadores de máxima verosimilhança do parâmetro autoregressivo para alguns casos. Os gráficos relativos às restantes situações encontram-se no Anexo VI.

Da análise dos gráficos do Anexo VI podemos retirar as seguintes conclusões:

- De um modo geral, à medida que $|a|$ aumenta, a distribuição dos \hat{a} vai apresentando maiores desvios da normalidade.
- Para n fixo e para um determinado tamanho de bloco b , os resultados não diferem substancialmente para valores de a positivos e para valores de a negativos.
- Aumentando o tamanho da amostra, os desvios à normalidade tendem a diminuir.
- Para $n=256$ a distribuição dos \hat{a} aproxima-se mais da normalidade quando usamos $b=4$ ou $b=8$ enquanto para $n=768$, $b=8$ e $b=16$ conduzem a distribuições mais próximas da normal.
- Fixando um determinado valor de a , à medida que b aumenta, o desvio padrão dos estimadores diminui.

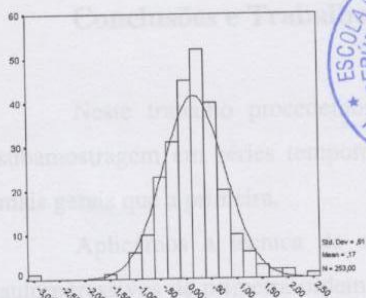


Gráfico 3.7. Histograma dos 253 \hat{a} para $n=256$, $a=0.2$ e $b=4$

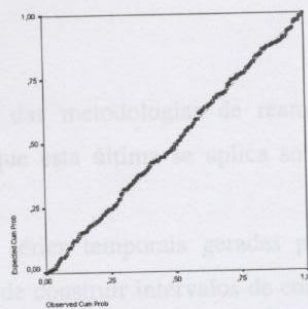


Gráfico 3.8. PP-Plot dos 253 \hat{a} para $n=256$, $a=0.2$ e $b=4$.

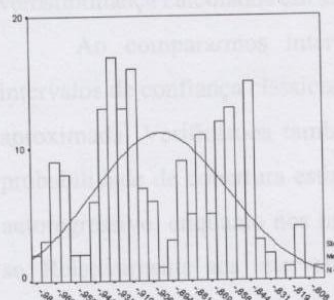


Gráfico 3.9. Histograma dos 193 \hat{a} para $n=256$, $a=-0.95$ e $b=64$

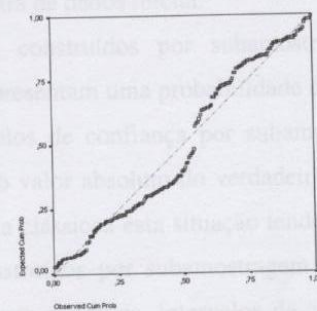


Gráfico 3.10. PP-Plot dos 193 \hat{a} para $n=256$, $a=-0.95$ e $b=64$

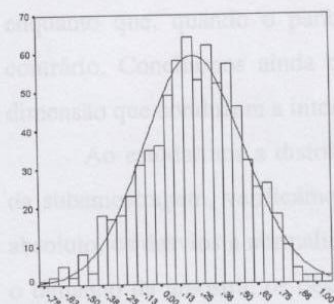


Gráfico 3.11. Histograma dos 761 \hat{a} para $n=768$, $a=0.2$ e $b=8$

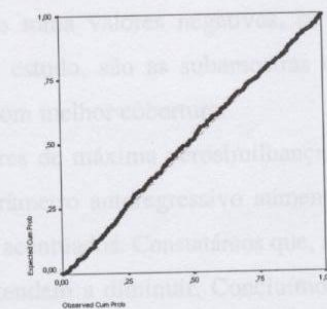


Gráfico 3.12. PP-Plot dos 761 \hat{a} para $n=768$, $a=0.2$ e $b=8$

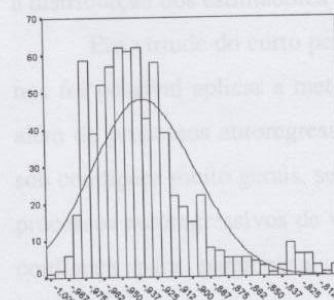


Gráfico 3.13. Histograma dos 705 \hat{a} para $n=768$, $a=-0.95$ e $b=64$

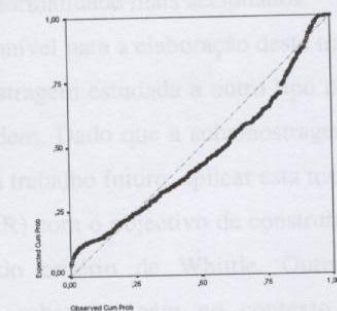


Gráfico 3.14. PP-Plot dos 705 \hat{a} para $n=768$, $a=-0.95$ e $b=64$

Conclusões e Trabalho Futuro

Neste trabalho procedemos a uma revisão das metodologias de reamostragem e subamostragem em séries temporais, constatando que esta última se aplica sob condições mais gerais que a primeira.

Aplicámos a técnica de subamostragem a séries temporais geradas por modelos autoregressivos de primeira ordem com o objectivo de construir intervalos de confiança para o parâmetro autoregressivo e também de estudar a distribuição dos estimadores de máxima verosimilhança calculados em subconjuntos da amostra de dados inicial.

Ao compararmos intervalos de confiança construídos por subamostragem com intervalos de confiança clássicos, verificámos que apresentam uma probabilidade de cobertura aproximada. Verificámos também que, nos intervalos de confiança por subamostragem, a probabilidade de cobertura estimada aumenta com o valor absoluto do verdadeiro parâmetro autoregressivo, enquanto nos intervalos de confiança clássicos esta situação tende a inverter-se. Relativamente aos intervalos de confiança construídos por subamostragem concluímos que, quando o parâmetro autoregressivo toma valores positivos, intervalos de confiança de caudas iguais apresentam uma cobertura superior à dos intervalos de confiança simétricos enquanto que, quando o parâmetro autoregressivo toma valores negativos, se constata o contrário. Concluímos ainda que, na situação em estudo, são as subamostras de pequena dimensão que conduzem a intervalos de confiança com melhor cobertura.

Ao estudarmos a distribuição dos estimadores de máxima verosimilhança resultantes da subamostragem, verificámos que, quando o parâmetro autoregressivo aumenta em valor absoluto, os desvios à normalidade tornam-se mais acentuados. Constatámos que, aumentando o tamanho da amostra, os desvios à normalidade tendem a diminuir. Concluímos ainda que fixando o valor do parâmetro autoregressivo e fazendo aumentar o tamanho das subamostras, a distribuição dos estimadores apresenta desvios à normalidade mais acentuados.

Em virtude do curto período de tempo disponível para a elaboração deste trabalho, não nos foi possível aplicar a metodologia de subamostragem estudada a outro tipo de situações além de processos autoregressivos de primeira ordem. Dado que a subamostragem funciona sob condições muito gerais, seria interessante, num trabalho futuro, aplicar esta metodologia a processos autoregressivos de valores inteiros (INAR) com o objectivo de construir regiões de confiança para parâmetros estimados através do critério de Whittle. Outra aplicação interessante seria estudar o comportamento da subamostragem no contexto das séries financeiras.

Lista de Referências

- ALONSO, A. M., PEÑA, D. e ROMO, J. (2000). Resampling time series by missing values techniques. *Statistics and Economics Series* **19**.
- BASAWA, I. V., MALLIK, A. K., McCORMICK, W. P., REEVES, J. H. e TAYLOR, R. L. (1991). Bootstrapping unstable first-order autoregressive processes. *Annals of Statistics* **19**, 1098-1101.
- BERAN, R. (1984). Bootstrap methods in statistics. *Jahresberichte des Deutschen Mathematischen Vereins* **86**, 14-30.
- BERTAIL, P., POLITIS, D. N. e ROMANO, J. P. (1999). On subsampling estimators with unknown rate of convergence. *Journal of the American Statistical Association* **94**, 569-579.
- BRITO, M. (2000). A metodologia bootstrap. *Actas do VIII Congresso Anual da Sociedade Portuguesa de Estatística*, 15-37.
- BROCKWELL, P. J. e DAVIS, R. A. (1987). *Time Series: Theory and Methods*. Springer-Verlag, Nova Iorque.
- CARLSTEIN, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary time series. *Annals of Statistics* **14**, 1171-1179.
- DAVISON, A. C. e HINKLEY, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge.
- EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics* **7**, 1-26.
- EFRON, B. (1982). *The jackknife, the bootstrap and other resampling plans*. CBMS-NSF Regional Conference Series Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia.
- EFRON, B. e TIBSHIRANI, R. J. (1986). Bootstrap methods for standart errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* **1**, 54-77.
- EFRON, B. e TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, Nova Iorque.
- HALL, P., HOROWITZ, J. L. e JING, B. -Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82**, 561-574.
- HARTIGAN, J. (1969). Using subsample values as typical values. *Journal of the American Statistical Association* **64**, 1303-1317.
- HARTIGAN, J. (1975). Necessary and sufficient conditions for asymptotic joint normality of a statistic and its subsample values. *Annals of Statistics* **3**, 573-580.
- KÜNSCH, H. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* **17**, 1217-1241.
- LAHIRI, S. N. (1992). Edgeworth correction by "moving block" bootstrap for stationary and nonstationary data. In *Exploring the Limits of Bootstrap*, 183-214, R. LePage and L. Billard (editors). John Wiley, New York.

- LIU, R. Y. e SINGH, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In *Exploring the Limits of Bootstrap*, 225-248, R. LePage and L. Billard (editors). John Wiley, New York.
- LOH, W. -Y. (1987). Calibrating confidence coefficients. *Journal of the American Statistical Association* **82**, 155-162.
- MANN, H. B. e WALD, A. (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica* **11**, 173-220.
- MURTEIRA, B. J. F., MÜLLER, D. A. e TURKMAN, K. F. (1993). *Análise de Sucessões Cronológicas*. Ed. McGraw Hill de Portugal.
- POLITIS, D. N. e ROMANO, J. P. (1992). A circular block-resampling procedure for stationary data. In *Exploring the Limits of Bootstrap*, 263-270, R. LePage and L. Billard (editors). John Wiley, New York.
- POLITIS, D. N. e ROMANO, J. P. (1992). *A general theory for large sample confidence regions based on subsamples under minimal assumptions*. Technical Report 399, Department of Statistics, Stanford University.
- POLITIS, D. N. e ROMANO, J. P. (1994). Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics* **22**, 2031-2050.
- POLITIS, D. N., ROMANO, J. P. e WOLF, M. (1997). Subsampling for heteroskedastic time series. *Journal of Econometrics* **81**, 281-317.
- POLITIS, D. N., ROMANO, J. P. e WOLF, M. (1999). *Subsampling*. Springer-Verlag, Nova Iorque.
- PRIESTLEY, M. B. (1981). *Spectral Analysis and Time Series*. Academic Press, Nova Iorque.
- SHAO, J. e WU, C. F. (1989). A general theory for jackknife variance estimation. *Annals of Statistics* **17**, 1176-1197.
- SHERMAN, M. (1997). Subseries methods in regression. *Journal of the American Statistical Association* **92**, 1041-1048.
- TUKEY, J. W. (1958). Bias and confidence in not quite large samples (abstract). *Annals of Mathematical Statistics* **29**, p.614.
- WU, C. F. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics* **14**, 1261-1343.
- WU, C. F. (1990). On the asymptotic properties of the jackknife histogram. *Annals of Statistics* **18**, 1438-1452.
- ZAMAN, A. (1996). *Statistical Foundations for Econometric Techniques*, Academic Press. San Diego.

ANEXO I – Demonstração do Teorema A [Pollis e Romano (1994)]

Teorema A (Pollis e Romano (1994)):

Notemos a sequência α -míng correspondente a $\{X_n\}$ por $\alpha_n(\cdot)$.

Assumamos a Hipótese A e que $\frac{1}{x_n} \rightarrow 0$, $\frac{1}{n} \rightarrow 0$ e $b \rightarrow \infty$.

ANEXOS

Assumamos também que $\alpha_n(\pi) \rightarrow 0$ quando $n \rightarrow \infty$.

(i) Se x é um ponto de continuidade de $J(x, P)$, então $L_n(x) \rightarrow J(x, P)$ em probabilidade.

(ii) Se $J(x, P)$ é contínua, então $\sup_x |L_n(x) - J(x, P)| \rightarrow 0$ em probabilidade.

(iii) Para $\alpha \in [0, 1]$, seja $c_{\alpha}(1-\alpha) = \inf \{x : L_n(x) \geq 1-\alpha\}$.

Correspondentemente, definamos $c(1-\alpha, P) = \inf \{x : J(x, P) \geq 1-\alpha\}$.

Se $J(x, P)$ é contínua em $c(1-\alpha, P)$, então

$\text{Prob}_n \{x_n \left[\hat{\theta}_n - \theta(P) \right] \leq c_{\alpha}(1-\alpha)\} \rightarrow 1-\alpha$ quando $n \rightarrow \infty$.

Então, a probabilidade de cobertura assintótica sob P do intervalo

$$I_n = \left[\hat{\theta}_n - x_n \left[c_{\alpha}(1-\alpha) \right], \hat{\theta}_n \right] \text{ é } 1-\alpha.$$

Sem perda de generalidade, podemos encerrar b como função de n . Neste caso, podemos omitir na notação utilizada os índices n . Por exemplo, $L_n(\cdot) = L(\cdot)$, $c_n(\alpha) = c(\alpha)$, etc. Para simplificar a notação, seja ainda $q = q_n = n - b + 1$. Seja

$$U_n(x) = \frac{1}{q} \sum_{i=1}^q I \{x_n \left[\hat{\theta}_{n+i} - \theta(P) \right] \leq x\}.$$

Para provar (i) é suficiente mostrar que $U_n(x)$ converge em probabilidade para $J(x, P)$ para todo o ponto de continuidade x de $J(x, P)$. Para isso basta notar que

$$L_n(x) = \frac{1}{q} \sum_{i=1}^q I \{x_n \left[\hat{\theta}_{n+i} - \theta(P) \right] + x_n \left[\hat{\theta}(P) - \hat{\theta}_n \right] \leq x\}.$$

Logo, para todo $\epsilon > 0$,

$$U_n(x - \epsilon) \leq L_n(x) \leq U_n(x + \epsilon).$$

ANEXO I – Demonstração do Teorema A [Politis e Romano (1994)]

Teorema A [Politis e Romano (1994)]:

Notemos a sequência α -mixing correspondente a $\{X_t\}$ por $\alpha_x(\cdot)$.

Assumamos a Hipótese A e que $\tau_b/\tau_n \rightarrow 0$, $b/n \rightarrow 0$ e $b \rightarrow \infty$ quando $n \rightarrow \infty$.

Assumamos também que $\alpha_x(m) \rightarrow 0$ quando $m \rightarrow \infty$.

i) Se x é um ponto de continuidade de $J(\cdot, P)$, então $L_{n,b}(x) \rightarrow J(x, P)$ em probabilidade.

ii) Se $J(\cdot, P)$ é contínua, então $\sup_x |L_{n,b}(x) - J(x, P)| \rightarrow 0$ em probabilidade.

iii) Para $\alpha \in [0, 1]$, seja $c_{n,b}(1-\alpha) = \inf\{x : L_{n,b}(x) \geq 1-\alpha\}$.

Correspondentemente, definamos $c(1-\alpha, P) = \inf\{x : J(x, P) \geq 1-\alpha\}$.

Se $J(\cdot, P)$ é contínua em $c(1-\alpha, P)$, então

$\text{Prob}_P\{\tau_n[\hat{\theta}_n - \theta(P)] \leq c_{n,b}(1-\alpha)\} \rightarrow 1-\alpha$ quando $n \rightarrow \infty$.

Então, a probabilidade de cobertura assintótica sob P do intervalo

$I_1 = [\hat{\theta}_n - \tau_n^{-1}c_{n,b}(1-\alpha), \infty]$ é $1-\alpha$.

Sem perda de generalidade, podemos encarar b como função de n . Neste caso, podemos omitir na notação utilizada os índices b . Por exemplo, $L_n(\cdot) \equiv L_{n,b}(\cdot)$, $c_n(\alpha) \equiv c_{n,b}(\alpha)$, etc. Para simplificar a notação, seja ainda $q = q_n = n - b + 1$. Seja

$$U_n(x) = \frac{1}{q} \sum_{i=1}^q I\{\tau_b [\hat{\theta}_{n,b,i} - \theta(P)] \leq x\}.$$

Para provar i), é suficiente mostrar que $U_n(x)$ converge em probabilidade para $J(x, P)$ para todo o ponto de continuidade x de $J(\cdot, P)$. Para isso basta notar que

$$L_n(x) = \frac{1}{q} \sum_{i=1}^q I\{\tau_b [\hat{\theta}_{n,b,i} - \theta(P)] + \tau_b [\theta(P) - \hat{\theta}_n] \leq x\}.$$

Logo, para todo $\varepsilon > 0$,

$$U_n(x-\varepsilon)I\{E_n\} \leq L_n(x)I\{E_n\} \leq U_n(x+\varepsilon),$$

onde $I\{E_n\}$ é a função indicatriz do evento $E_n = \{\tau_b|\theta(P) - \hat{\theta}_n| \leq \varepsilon\}$. Mas, a probabilidade do evento E_n tende para um. Então, com probabilidade a tender para um

$$U_n(x-\varepsilon) \leq L_n(x) \leq U_n(x+\varepsilon).$$

Se $x+\varepsilon$ e $x-\varepsilon$ são pontos de continuidade de $J(\cdot, P)$, então $U_n(x \pm \varepsilon)$ converge em probabilidade para $J(x \pm \varepsilon, P)$, implicando

$$J(x-\varepsilon, P) - \varepsilon \leq L_n(x) \leq J(x+\varepsilon, P) + \varepsilon$$

com probabilidade tendendo para um. Seja agora $\varepsilon \rightarrow 0$ tal que $x \pm \varepsilon$ são pontos de continuidade de $J(\cdot, P)$ pelo que $L_n(x) \rightarrow J(x, P)$ em probabilidade. Então, podemos restringir a nossa atenção a $U_n(x)$.

Como $E(U_n(x)) = J_b(x, P)$, a demonstração de *i*) reduz-se, pela Hipótese A, a mostrar que $\text{Var}(U_n(x)) \rightarrow 0$ quando $n \rightarrow \infty$. Definamos

$$I_{b,t} = I\{\tau_b|\hat{\theta}_{n,b,t} - \theta(P)| \leq x\}, \quad t=1, \dots, q,$$

$$s_{q,h} = \frac{1}{q} \sum_{t=1}^{q-h} \text{Cov}(I_{b,t}, I_{b,t+h}).$$

Então

$$\text{Var}(U_n(x)) = \frac{1}{q} \left(s_{q,0} + 2 \sum_{h=1}^{q-1} s_{q,h} \right) = \frac{1}{q} \left(s_{q,0} + 2 \sum_{h=1}^{b-1} s_{q,h} + 2 \sum_{h=b}^{q-1} s_{q,h} \right) = A^* + A$$

$$\text{onde } A^* = \frac{1}{q} \left(s_{q,0} + 2 \sum_{h=1}^{b-1} s_{q,h} \right) \text{ e } A = \frac{2}{q} \sum_{h=b}^{q-1} s_{q,h}$$

É fácil ver que $|A^*| = O\left(\frac{b}{q}\right)$.

O lema seguinte será necessário:

Lema B:

Seja $\{X_t\}$ uma sequência aleatória com sequência mixing correspondente $\alpha_X(\cdot)$.

Sejam as variáveis A e B $\mathcal{F}_{-\infty}^n$ e \mathcal{F}_{n+k}^∞ mensuráveis, respectivamente, com $|A| \leq C_1$ e $|B| \leq C_2$.

Então $|\text{Cov}(A, B)| \leq 4C_1C_2\alpha_X(k)$.

Aplicando o Lema B,
para $h \geq b$,

$$|Cov(I_{b,t}, I_{b,t+h})| \leq 4\alpha_X(h-b+1)$$

e logo,

$$|A| \leq \frac{8}{q} \sum_{h=1}^{q-b} \alpha_X(h).$$

Pela monotonia dos coeficientes *mixing*, $\alpha_X(m) \rightarrow 0$ quando $m \rightarrow \infty$ e logo $M^{-1} \sum_{m=1}^M \alpha_X(m) \rightarrow 0$ quando $M \rightarrow \infty$, o que implica que A converge para zero. Então, tanto A como A^* convergem para zero, o que completa a demonstração de *i*).

Para demonstrar *ii*), dada uma subsequência $\{n_k\}$, podemos extrair uma subsequência $\{n_{k_j}\}$ tal que $L_{n_{k_j}}(x) \rightarrow J(x, P)$ para todo o x em algum conjunto denso numerável da recta real. Daqui decorre que, num conjunto de probabilidade um, $L_{n_{k_j}}(x)$ tende fracamente para $J(x, P)$. Pela continuidade de $J(\cdot, P)$, esta convergência é uniforme pelo Teorema de Polya.

A demonstração de *iii*) é muito similar à demonstração do Teorema 1 de Beran (1984).

ANEXO II - Processos Autoregressivos de Ordem p - AR(p)

Definição: $\{X_t\}$ diz-se um **processo autoregressivo de ordem p - AR(p)** - se satisfaz a seguinte equação:

$$X_t - a_1 X_{t-1} - \dots - a_p X_{t-p} = \varepsilon_t$$

onde a_1, a_2, \dots, a_p são constantes e $\{\varepsilon_t\}$ é um processo puramente aleatório.

Escrevendo a equação anterior na forma $X_t = a_1 X_{t-1} + \dots + a_p X_{t-p} + \varepsilon_t$, vemos que X_t depende, em parte de X_{t-1}, \dots, X_{t-p} e, em parte, do “distúrbio aleatório”. De facto, X_t tem regressão linear em X_{t-1}, \dots, X_{t-p} com ε_t fazendo o papel de “erro”. O facto de X_t ter regressão linear no seu próprio passado dá origem à terminologia “**processo autoregressivo**”.

Um processo AR(p) pode também ser escrito na forma $\alpha(B)X_t = \varepsilon_t$ onde B é o operador “backward” definido por $B^j X_t = X_{t-j}$ e $\alpha(B) = 1 - a_1 B - a_2 B^2 - \dots - a_p B^p$. Se R_1, R_2, \dots, R_p forem as inversas das raízes do polinómio $\alpha(x) = 1 - a_1 x - a_2 x^2 - \dots - a_p x^p$, ou seja, as raízes do polinómio $\psi(x) = x^p - a_1 x^{p-1} - a_2 x^{p-2} - \dots - a_p$ (que admitimos serem distintas por uma questão de simplicidade), podemos escrever:

$$\alpha(B) = \prod_{i=1}^p (1 - R_i B)$$

A solução desta equação vem $X_t = f(t) + \alpha^{-1}(B)\varepsilon_t$ onde $f(t)$ é a solução da equação homogénea associada $\alpha(B)X_t = 0$ e é da forma $f(t) = A_1 R_1^t + A_2 R_2^t + \dots + A_p R_p^t$. Para que $X_t = f(t) + \alpha^{-1}(B)\varepsilon_t$ represente um processo **assimptoticamente estacionário** (isto é, para t suficientemente grande, o processo assume um comportamento estacionário), $f(t)$ tem que tender para zero quando t tende para infinito. A solução geral da equação toma a forma:

$$X_t = A_1 R_1^t + A_2 R_2^t + \dots + A_p R_p^t + \sum_{j=0}^{+\infty} (R_1^j A_1 + R_2^j A_2 + \dots + R_p^j A_p) \varepsilon_{t-j}$$

Como é necessário que $E(X_t)$ seja constante, temos $A_1 = A_2 = \dots = A_p = 0$. Por outro lado, para que a soma da série seja convergente em média quadrática, é necessário que $|R_j| < 1, \forall j$, isto é, todas as raízes de $\psi(x)$ têm que pertencer ao interior do círculo unitário, ou seja, as raízes de $\alpha(x)$ terão que se encontrar no exterior do círculo unitário pois só assim

$\lim_{t \rightarrow +\infty} f(t) = 0$. Portanto, o estado de equilíbrio de $\{X_t\}$ é descrito unicamente pelo 2º termo do membro direito da última equação, sendo esta expressão chamada a solução estacionária de $X_t - a_1 X_{t-1} - \dots - a_p X_{t-p} = \varepsilon_t$ (a solução complementar também se chama parte transitória por causa do seu decaimento para zero).

Para determinar a **FAC** do processo AR(p) multiplicamos ambos os membros da equação que definem o processo por X_{t-k} , $k \geq 0$ e tomamos valores esperados. Obtemos então:

$$E(X_{t-k} X_t) = a_1 E(X_{t-k} X_{t-1}) + \dots + a_p E(X_{t-k} X_{t-p}) + E(\varepsilon_t X_{t-k})$$

Assim,

| | |
|------------|---|
| para $k=0$ | $\sigma_x^2 (1 - a_1 \rho_1 - \dots - a_p \rho_p) = \sigma_\varepsilon^2$ |
| para $k=1$ | $\rho_1 - a_1 \rho_0 - \dots - a_p \rho_{p-1} = 0$ |
| para $k=2$ | $\rho_2 - a_1 \rho_1 - \dots - a_p \rho_{p-2} = 0$ |
| para $k=p$ | $\rho_p - a_1 \rho_{p-1} - \dots - a_p \rho_0 = 0$ |

Em geral, para $k \geq p$: $\rho_k - a_1 \rho_{k-1} - \dots - a_p \rho_{k-p} = 0$

As equações obtidas fazendo $k=1, \dots, p$ estabelecem pois a correspondência entre $(\rho_1, \rho_2, \dots, \rho_p)$ e (a_1, a_2, \dots, a_p) e são conhecidas por **equações de Yule-Walker**.

Continuando a supor que as p raízes de $\psi(x)$ são distintas, a solução da última equação vem

$$\rho_k = B_1 R_1^k + B_2 R_2^k + \dots + B_p R_p^k$$

podendo os coeficientes B_1, B_2, \dots, B_p ser determinados a partir das equações para $k=0, 1, \dots, p-1$.

Vemos que a FAC tem um decaimento exponencial (se as raízes forem reais) ou sinusoidal para zero (se existir um par de raízes complexas conjugadas) à medida que k aumenta.

Por sua vez, a **FACP** vem:

$$\varphi_{11} = \rho_1 = \frac{a_1}{1 - a_2}$$

$$\varphi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2} = a_2$$

ANEXO III - Estimação dos Parâmetros em Modelos AR(p)

$$\varphi_{pp} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{p-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{p-3} & \rho_2 \\ \dots & \dots & \dots & & \dots & \dots \\ \rho_{p-1} & \rho_{p-2} & \rho_{p-3} & \dots & \rho_1 & \rho_p \end{bmatrix}$$

onde ε_t é um processo de ruído branco com parâmetros (μ, σ^2) e (a_1, \dots, a_p) satisfizerem as condições de estacionariedade. Assim, dadas N observações (X_1, \dots, X_N) pretendemos estimar $\mu, (a_1, a_2, \dots, a_p)$ e σ^2 . Observando a equação acima, reparamos na sua semelhança com o modelo de regressão linear múltipla clássico uma vez que se expressar X_t como função linear de $(X_{t-1} - \mu), \dots, (X_{t-p} - \mu)$ com $\mu, a_1, a_2, \dots, a_p$ actuando como coeficientes de regressão e ε_t como resíduo embora, nestes casos, X_{t-1}, \dots, X_{t-p} não são variáveis independentes e dependentes. Assim, visto que X_1, X_2, \dots, X_p são valores observados, podemos utilizar o Método dos Mínimos Quadrados, isto é, vamos procurar estimadores de $\mu, a_1, a_2, \dots, a_p$ que minimizem

$$Q(\mu, a_1, \dots, a_p) = \sum_{t=p+1}^N \varepsilon_t^2 = \sum_{t=p+1}^N \{X_t - \mu - a_1(X_{t-1} - \mu) - \dots - a_p(X_{t-p} - \mu)\}^2$$

Note-se que não podemos incluir os termos $\varepsilon_1, \dots, \varepsilon_p$ uma vez que estas dependem de variáveis aleatórias que não foram observadas. Porém, para N grande comparado com p (o que sucede quase sempre na prática) o efeito de ignorar os p primeiros ε_t 's será pequeno e portanto a expressão acima constitui uma boa aproximação à soma dos quadrados $\sum_{t=1}^N \varepsilon_t^2$. (De facto, esta aproximação é equivalente a ter $\varepsilon_1 = \varepsilon_2 = \dots = \varepsilon_p = 0$.)

Se assumirmos que os resíduos ε_t 's são Gaussianos, sendo não correlacionados são também independentes. Nestas circunstâncias, os estimadores obtidos pelo Método dos Mínimos Quadrados são também estimadores de máxima verossimilhança pois a densidade conjunta de X_1, \dots, X_N será dada por

ANEXO III - Estimação dos Parâmetros em Modelos AR(p)

Consideremos o modelo AR(p) com média μ não nula:

$$X_t - \mu = a_1(X_{t-1} - \mu) + \dots + a_p(X_{t-p} - \mu) + \varepsilon_t$$

onde ε_t é um processo de ruído branco e os coeficientes (a_1, a_2, \dots, a_p) satisfazem as condições de estacionariedade. Assim, dadas N observações X_1, X_2, \dots, X_N , pretendemos estimar μ , (a_1, a_2, \dots, a_p) e σ_ε^2 . Observando a equação acima, reparamos na sua semelhança com o modelo de regressão linear múltipla clássico uma vez que se expressa X_t como função linear de $(X_{t-1} - \mu), \dots, (X_{t-p} - \mu)$ com $\mu, a_1, a_2, \dots, a_p$ actuando como coeficientes de regressão e ε_t como resíduo embora, neste caso, X_t desempenhe simultaneamente o papel de variável dependente e independente. Assim, visto que $X_{t-1}, X_{t-2}, \dots, X_{t-p}$ são valores observados, podemos utilizar o **Método dos Mínimos Quadrados**, isto é, vamos procurar estimadores de $\mu, a_1, a_2, \dots, a_p$ que minimizem

$$Q(\mu, a_1, \dots, a_p) = \sum_{t=p+1}^N \varepsilon_t^2 = \sum_{t=p+1}^N \{X_t - \mu - a_1(X_{t-1} - \mu) - \dots - a_p(X_{t-p} - \mu)\}^2$$

Note-se que não podemos incluir os termos $\varepsilon_1, \dots, \varepsilon_p$ uma vez que estes dependem de variáveis aleatórias que não foram observadas. Porém, para N grande comparado com p (o que sucede quase sempre na prática) o efeito de ignorar os p primeiros ε_t 's será pequeno e portanto a expressão acima constitui uma boa aproximação à soma dos quadrados $\sum_{t=1}^N \varepsilon_t^2$. (De facto, esta aproximação é equivalente a ter $\varepsilon_1 = \varepsilon_2 = \dots = \varepsilon_p = 0$.)

Se assumirmos que os resíduos ε_t 's são Gaussianos, sendo não correlacionados são também independentes. Nestas circunstâncias, os estimadores obtidos pelo Método dos Mínimos Quadrados são também estimadores de **máxima verosimilhança** pois a densidade conjunta de $\varepsilon_{p+1}, \dots, \varepsilon_N$ será dada por

$$L(\varepsilon_{p+1}, \dots, \varepsilon_N) = \left(\frac{1}{2\pi\sigma_\varepsilon^2} \right)^{\frac{N-p}{2}} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} \left(\sum_{t=p+1}^N \varepsilon_t^2 \right) \right].$$

Fazendo a transformação de $(\varepsilon_{p+1}, \dots, \varepsilon_N)$ para (X_{p+1}, \dots, X_N) :

$$(X_t - \mu) - a_1(X_{t-1} - \mu) - \dots - a_p(X_{t-p} - \mu) = \varepsilon_t, \quad p+1 \leq t \leq N$$

vem que o jacobiano desta transformação é igual à unidade e

$$L(X_{p+1}, \dots, X_N) = \left(\frac{1}{2\pi\sigma_\varepsilon^2} \right)^{\frac{N-p}{2}} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=p+1}^N \{X_t - \mu - a_1(X_{t-1} - \mu) - \dots - a_p(X_{t-p} - \mu)\}^2 \right].$$

De facto, em rigor, esta densidade não representa a densidade conjunta de X_{p+1}, \dots, X_N mas sim a densidade condicional de X_{p+1}, \dots, X_N dado que as observações iniciais X_1, \dots, X_p estão fixas nos seus valores observados x_1, \dots, x_p . Contudo, como N é grande em comparação com p , esta modificação provoca apenas uma pequena alteração na função verosimilhança que podemos considerar bem aproximada pela expressão acima indicada.

Assim, o logaritmo da verosimilhança vem dado por

$$L(\mu, a_1, \dots, a_p) = -(N-p) \ln \sqrt{2\pi\sigma_\varepsilon^2} - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=p+1}^N \{X_t - \mu - a_1(X_{t-1} - \mu) - \dots - a_p(X_{t-p} - \mu)\}^2.$$

Se conhecermos σ_ε^2 , os estimadores de μ, a_1, \dots, a_p são obtidos minimizando a soma dos quadrados

$$Q(\mu, a_1, \dots, a_p) = \sum_{t=p+1}^N \{X_t - \mu - a_1(X_{t-1} - \mu) - \dots - a_p(X_{t-p} - \mu)\}^2$$

o que corresponde exactamente a obter os estimadores de **mínimos quadrados**. Para encontrar o mínimo de Q derivamos respectivamente em ordem a μ e a_1, \dots, a_p , obtendo

$$\sum_{t=p+1}^N \{X_t - \hat{\mu} - \hat{a}_1(X_{t-1} - \hat{\mu}) - \dots - \hat{a}_p(X_{t-p} - \hat{\mu})\} = 0 \quad \text{e}$$

$$\sum_{t=p+1}^N \{X_t - \hat{\mu} - \hat{a}_1(X_{t-1} - \hat{\mu}) - \dots - \hat{a}_p(X_{t-p} - \hat{\mu})\} (X_{t-j} - \hat{\mu}) = 0, \quad j=1, \dots, p.$$

Então, da primeira igualdade obtemos $\hat{\mu} = \frac{\bar{X}_0 - \hat{a}_1 \bar{X}_1 - \dots - \hat{a}_p \bar{X}_p}{1 - \hat{a}_1 - \dots - \hat{a}_p}$ onde

$$\bar{X}_j = \frac{1}{N-p} \sum_{t=p+1-j}^{N-j} X_t, \quad j=0, \dots, p-1.$$

Considerando ainda p pequeno em comparação com N , podemos aproximar $\bar{X}_0, \bar{X}_1, \dots, \bar{X}_p$ (que representam as médias dos diferentes conjuntos de $N-p$ observações) por $\bar{X} = \frac{1}{N} \sum_{t=1}^N X_t$, obtendo a expressão acima simplificada: $\hat{\mu} = \bar{X}$.

Substituindo $\hat{\mu}$ por \bar{X} na segunda igualdade temos

$$\sum_{t=p+1}^N \{X_t - \bar{X} - \hat{a}_1(X_{t-1} - \bar{X}) - \dots - \hat{a}_p(X_{t-p} - \bar{X})\}(X_{t-j} - \bar{X}) = 0, \quad j=1, \dots, p.$$

Se fizermos agora a aproximação $\sum_{t=p+1}^N (X_{t-i} - \bar{X})(X_{t-j} - \bar{X}) \approx N\hat{\gamma}(j-i)$, com $0 \leq i \leq p$

e $1 \leq p \leq j$, em que $\hat{\gamma}_k$ é a função autocovariância amostral, as equações acima transformam-se em $\hat{\gamma}_j = \hat{a}_1 \hat{\gamma}_{j-1} + \hat{a}_2 \hat{\gamma}_{j-2} + \dots + \hat{a}_p \hat{\gamma}_{j-p}$, $j=1, \dots, p$.

Em linguagem matricial, podemos escrever $\hat{r} = \hat{R}_p \cdot \hat{a}$, onde

$$\hat{a} = (\hat{a}_1 \hat{a}_2 \dots \hat{a}_p)', \quad \hat{r} = (\hat{\gamma}_1 \hat{\gamma}_2 \dots \hat{\gamma}_p)' \quad \text{e} \quad \hat{R}_p = \begin{bmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \dots & \hat{\gamma}_{p-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \dots & \hat{\gamma}_{p-2} \\ \dots & \dots & \dots & \dots \\ \hat{\gamma}_{p-1} & \hat{\gamma}_{p-2} & \dots & \hat{\gamma}_0 \end{bmatrix}.$$

Assim, podemos obter boas aproximações para os estimadores de máxima verosimilhança (ou de mínimos quadrados) $\hat{a}_1, \dots, \hat{a}_p$ calculando as $p+1$ primeiras autocovariâncias amostrais $\hat{\gamma}_0, \hat{\gamma}_1, \dots, \hat{\gamma}_p$ e resolvendo o sistema de equações lineares $\hat{\gamma}_j = \hat{a}_1 \hat{\gamma}_{j-1} + \hat{a}_2 \hat{\gamma}_{j-2} + \dots + \hat{a}_p \hat{\gamma}_{j-p}$, $j=1, \dots, p$. Note-se que este sistema de equações é idêntico às equações de Yule-Walker, as quais expressam os valores teóricos dos parâmetros a_1, a_2, \dots, a_p em termos da função autocorrelação teórica ρ_k ou da função autocovariância γ_k . Assim, estes estimadores calculam-se pelo **Método dos Momentos**, isto é, substituindo nas equações de

Yule-Walker as autocorrelações teóricas ρ_k pelas estimadas $\hat{\rho}_k$ e resolvendo o sistema de equações resultante

$$\hat{\rho}_1 = a_1 + a_2 \hat{\rho}_1 + \dots + a_p \hat{\rho}_{p-1}$$

$$\hat{\rho}_2 = a_1 \hat{\rho}_1 + a_2 + \dots + a_p \hat{\rho}_{p-2}$$

...

$$\hat{\rho}_p = a_1 \hat{\rho}_{p-1} + a_2 \hat{\rho}_{p-2} + \dots + a_p$$

em ordem aos parâmetros a_1, a_2, \dots, a_p . Os estimadores assim obtidos são chamados **estimadores de Yule-Walker**. Prova-se que, para N suficientemente grande, estes estimadores não diferem substancialmente dos obtidos pelo método dos mínimos quadrados. Porém, se algum dos parâmetros estiver próximo de um, os estimadores de Yule-Walker podem ser bastante pobres.

Quanto ao estimador para σ_ε^2 , podemos utilizar o resultado da Teoria da Regressão que nos dá o **estimador centrado**

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N-2p-1} Q(\hat{\mu}, \hat{a}_1, \dots, \hat{a}_p) \approx \frac{N-p}{N-2p-1} (\hat{y}_0 - \hat{a}_1 \hat{y}_1 - \dots - \hat{a}_p \hat{y}_p)$$

onde o denominador é obtido a partir da expressão geral “número de observações utilizadas ($N-p$) menos o número de parâmetros estimados”.

O estimador de máxima verossimilhança viria

$$\tilde{\sigma}_\varepsilon^2 = \frac{1}{N-p} Q(\hat{\mu}, \hat{a}_1, \dots, \hat{a}_p).$$

ANEXO IV - Intervalos de Confiança para os Parâmetros

Priestley (1981) descreve um método geral desenvolvido por Box e Jenkins (1970) que permite a obtenção de intervalos de confiança para os parâmetros dos modelos AR, MA e ARMA baseando-se quer na Teoria das Hipóteses Lineares (usando o Método dos Mínimos Quadrados), quer na Teoria da Máxima Verosimilhança. Passamos a apresentar sucintamente o segundo.

Para todos os modelos gaussianos de séries temporais (isto é, admitindo que os $\{\varepsilon_t\}$ formam uma sucessão de variáveis aleatórias i.i.d. $N(0, \sigma_\varepsilon^2)$), podemos exprimir a aproximação ao logaritmo da função verosimilhança na forma

$$L(\theta) = \text{const} - \frac{1}{2\sigma_\varepsilon^2} Q(\theta).$$

onde θ denota o conjunto de parâmetros desconhecidos (por exemplo, num modelo AR(p) $\theta = (\mu, a_1, \dots, a_p)$) e $Q(\theta) = \sum_t \varepsilon_t^2$ onde os limites do somatório sobre t dependem dos valores iniciais apropriados ao modelo.

Seja $\hat{\theta}$ o estimador de máxima verosimilhança (ou de mínimos quadrados) de θ . Então, podemos escrever

$$L(\theta) = L\{\hat{\theta} + (\theta - \hat{\theta})\} = L(\hat{\theta}) + \sum_{i=1}^q (\theta_i - \hat{\theta}_i) \frac{\partial L(\hat{\theta})}{\partial \theta_i} + \frac{1}{2} \sum_{i=1}^q \sum_{j=1}^q (\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j) \frac{\partial^2 L(\hat{\theta})}{\partial \theta_i \partial \theta_j} + \dots$$

Assumindo que $L(\theta)$ é localmente quadrática (logo, podemos ignorar as derivadas de $L(\theta)$ de ordem superior à segunda) e notando que $\frac{\partial L(\hat{\theta})}{\partial \theta_i} = 0$, para todo o i (sendo $\hat{\theta}$ o estimador de máxima verosimilhança de θ) e escrevendo $L_{ij} = \frac{\partial^2 L(\hat{\theta})}{\partial \theta_i \partial \theta_j}$, temos

$$L(\theta) \approx L(\hat{\theta}) = \frac{1}{2} \sum_{i=1}^q \sum_{j=1}^q (\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j) L_{ij}.$$

Demonstra-se que para N grande podemos substituir L_{ij} por $E \left[\frac{\partial^2 L(\theta)}{\partial \theta_i \partial \theta_j} \right]$ sendo, neste caso, $-\{L_{ij}\}^{-1}$ uma aproximação da matriz de variâncias-covariâncias de $\{\hat{\theta}_i\}$ (a notação $\{L_{ij}\}^{-1}$ significa o inverso da matriz cujo elemento (i,j) é L_{ij}). Com esta aproximação, a forma quadrática $\left\{ -\sum_i \sum_j L_{ij} (\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j) \right\}$ possui uma distribuição χ_q^2 .

Usando as igualdades anteriores, temos

$$\{Q(\theta) - Q(\hat{\theta})\} \sim 2\sigma_\varepsilon^2 \cdot \frac{1}{2} \chi_q^2 = \sigma_\varepsilon^2 \chi_q^2.$$

Quando σ_ε^2 não é conhecido, estimamo-lo previamente por $\sigma_\varepsilon^2 = \frac{1}{N-q} \cdot Q(\hat{\theta})$.

Então, se o modelo é linear temos $(N-q)\sigma_\varepsilon^2 = \sigma_\varepsilon^2 \chi_{N-q}^2$, sendo independente de $\{Q(\theta) - Q(\hat{\theta})\}$. Assim, neste caso,

$$\frac{Q(\theta) - Q(\hat{\theta})}{(N-q)\sigma_\varepsilon^2} = \frac{q}{N-q} \cdot F_{q, N-q}$$

onde $F_{q, N-q}$ denota uma distribuição F com $q, N-q$ graus de liberdade.

Consequentemente pode ser obtida uma região de confiança a $(100-\alpha)\%$ para θ através de

$$\frac{Q(\theta) - Q(\hat{\theta})}{Q(\hat{\theta})} \leq \frac{q}{N-q} \cdot F_{q, N-q}(\alpha) \quad \text{ou} \quad Q(\theta) \leq Q(\hat{\theta}) \left\{ 1 + \frac{q}{N-q} \cdot F_{q, N-q}(\alpha) \right\}.$$

Box e Jenkins (1970) sugerem ainda que, para N grande, podemos encarar σ_ε^2 como uma constante e aproximar a região anterior por

$$Q(\theta) \leq Q(\hat{\theta}) \left\{ 1 + \frac{\chi_q^2(\alpha)}{N} \right\}.$$

Esta equação pode ser aplicada directamente não só aos modelos AR como também MA e ARMA, desde que, em cada caso, a função “soma dos quadrados” $Q(\theta)$ seja calculada directamente dos resíduos ε_t e dos valores do parâmetro θ .

Se $L(\theta)$ é localmente quadrática, $Q(\theta)$ também o é. Assim, podemos escrever

$$Q(\theta) \approx Q(\hat{\theta}) = \frac{1}{2} \sum_{i=1}^q \sum_{j=1}^q (\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j) Q_{ij} \quad \text{onde} \quad Q_{ij} = \frac{\partial^2 Q(\hat{\theta})}{\partial \theta_i \partial \theta_j} = -2\sigma_\varepsilon^2 L_{ij}.$$

Então, a expressão obtida acima para a região de confiança, pode ser escrita na seguinte forma:

$$\sum_{i=1}^q \sum_{j=1}^q (\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j) Q_{ij} = q\sigma_\varepsilon^2 F_{q, N-q}(\alpha).$$

Consequentemente, pode ser obtida uma região de confiança a $(100-\alpha)\%$ para θ através de

$$\sum_{i=1}^q \sum_{j=1}^q (\theta_i - \hat{\theta}_i)(\theta_j - \hat{\theta}_j) Q_{ij} \leq q\sigma_\varepsilon^2 F_{q, N-q}(\alpha).$$

Na prática podemos aproximar Q_{ij} usando as segundas diferenças de $Q(\hat{\theta})$, isto é, podemos estimar Q_{ij} numericamente. De forma alternativa, em alguns casos, podemos calcular Q_{ij} analiticamente a partir da expressão da função verosimilhança.

ANEXO V – Limites dos Intervalos de Confiança para o Parâmetro Autoregressivo

Foram geradas amostras de 256 observações de processos AR(1) onde o verdadeiro parâmetro autoregressivo (a) toma sucessivamente os valores $0.2, 0.5, 0.8, 0.95, -0.2, -0.5, -0.8$ e -0.95 . Em cada gráfico apresentamos os limites superiores e inferiores de 500 intervalos de confiança a 95% para o parâmetro autoregressivo. Para cada uma das oito amostras geradas construímos 500 intervalos de confiança clássicos (I_{class}) e intervalos de confiança por subamostragem - 500 intervalos de caudas iguais (I_{ET}) e 500 intervalos simétricos (I_{sym}) - onde foram tomadas subamostras de tamanho $b=4, b=8, b=16, b=32$ e $b=64$.

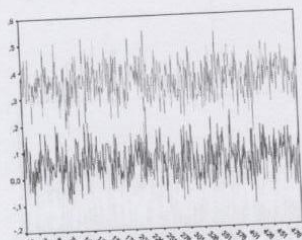


Gráfico 1. Limites dos I_{ET} para $a=0.2$ e $b=4$

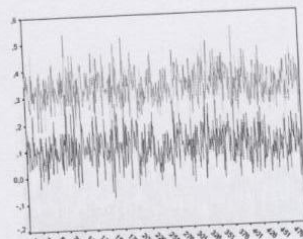


Gráfico 2. Limites dos I_{ET} para $a=0.2$ e $b=8$

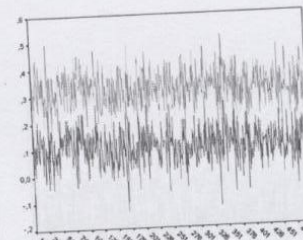


Gráfico 3. Limites dos I_{ET} para $a=0.2$ e $b=16$

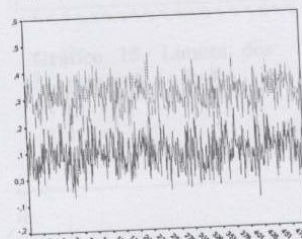


Gráfico 4. Limites dos I_{sym} para $a=0.2$ e $b=4$

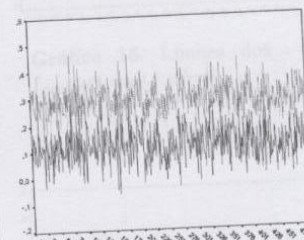


Gráfico 5. Limites dos I_{sym} para $a=0.2$ e $b=8$

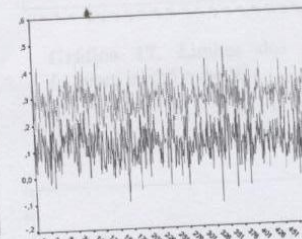


Gráfico 6. Limites dos I_{sym} para $a=0.2$ e $b=16$

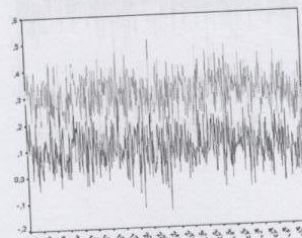


Gráfico 7. Limites dos I_{ET} para $a=0.2$ e $b=32$

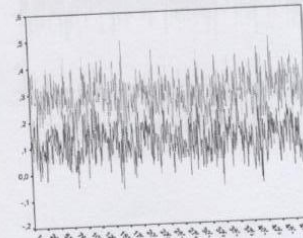


Gráfico 8. Limites dos I_{ET} para $a=0.2$ e $b=64$

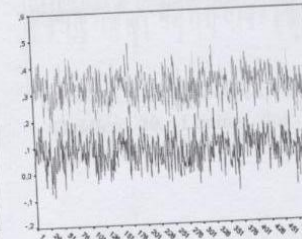


Gráfico 9. Limites dos I_{class} para $a=0.2$

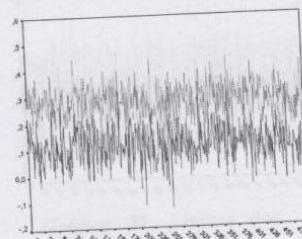


Gráfico 10. Limites dos I_{sym} para $a=0.2$ e $b=32$

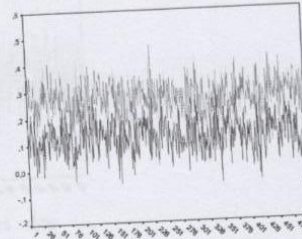


Gráfico 11. Limites dos I_{sym} para $a=0.2$ e $b=64$

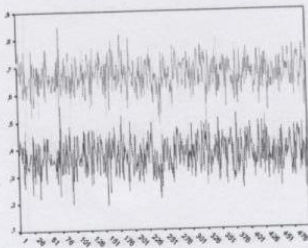


Gráfico 12. Limites dos I_{ET} para $a=0.5$ e $b=4$

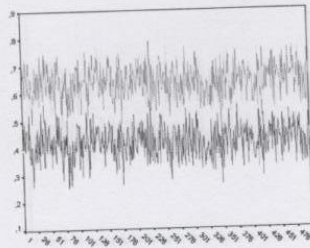


Gráfico 13. Limites dos I_{ET} para $a=0.5$ e $b=8$

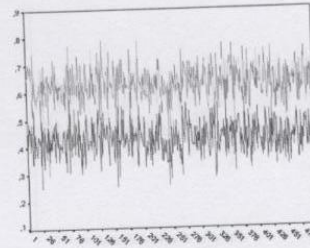


Gráfico 14. Limites dos I_{ET} para $a=0.5$ e $b=16$

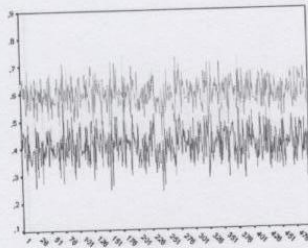


Gráfico 15. Limites dos I_{sym} para $a=0.5$ e $b=4$

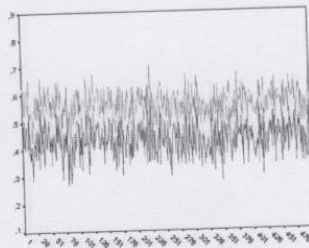


Gráfico 16. Limites dos I_{sym} para $a=0.5$ e $b=8$

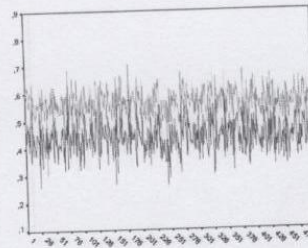


Gráfico 17. Limites dos I_{sym} para $a=0.5$ e $b=16$

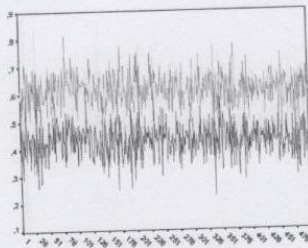


Gráfico 18. Limites dos I_{ET} para $a=0.5$ e $b=32$

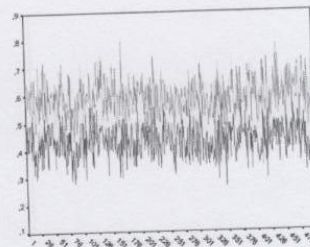


Gráfico 19. Limites dos I_{ET} para $a=0.5$ e $b=64$

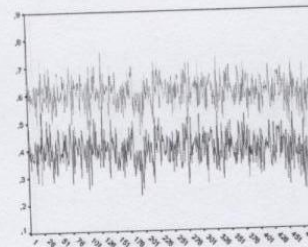


Gráfico 20. Limites dos I_{class} para $a=0.5$

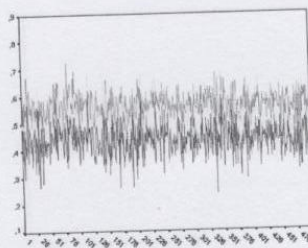


Gráfico 21. Limites dos I_{sym} para $a=0.5$ e $b=32$

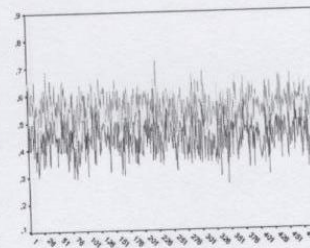


Gráfico 22. Limites dos I_{sym} para $a=0.5$ e $b=32$

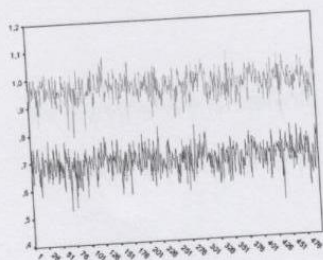


Gráfico 23. Limites dos I_{ET} para $a=0.8$ e $b=4$

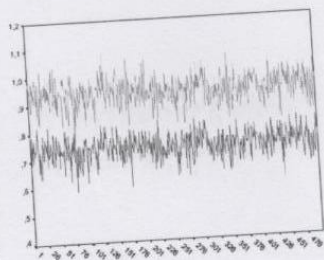


Gráfico 24. Limites dos I_{ET} para $a=0.8$ e $b=8$

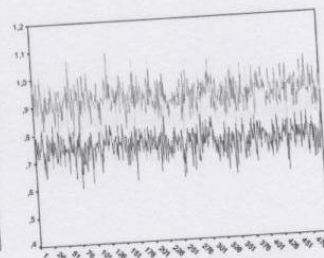


Gráfico 25. Limites dos I_{ET} para $a=0.8$ e $b=16$

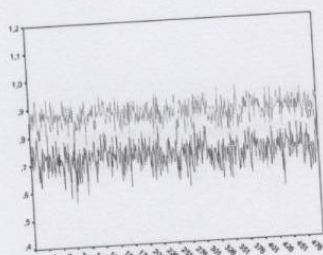


Gráfico 26. Limites dos I_{sym} para $a=0.8$ e $b=4$

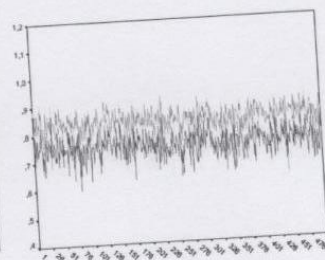


Gráfico 27. Limites dos I_{sym} para $a=0.8$ e $b=8$

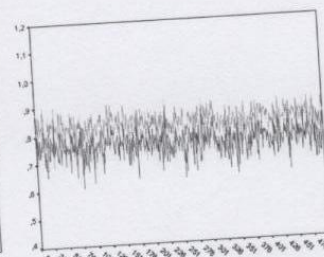


Gráfico 28. Limites dos I_{sym} para $a=0.8$ e $b=16$

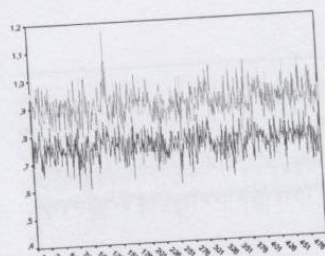


Gráfico 29. Limites dos I_{ET} para $a=0.8$ e $b=32$

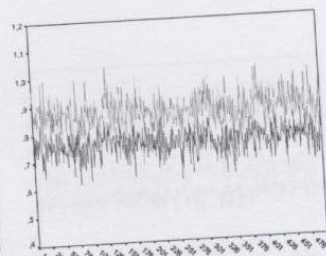


Gráfico 30. Limites dos I_{ET} para $a=0.8$ e $b=64$

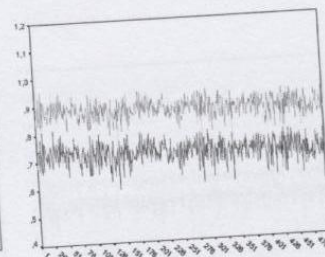


Gráfico 31. Limites dos I_{class} para $a=0.8$

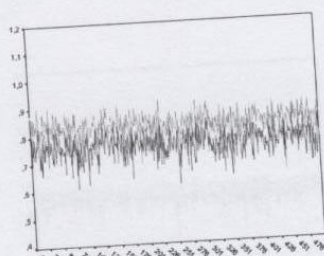


Gráfico 32. Limites dos I_{sym} para $a=0.8$ e $b=32$

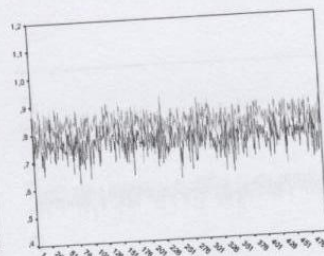


Gráfico 33. Limites dos I_{sym} para $a=0.8$ e $b=32$

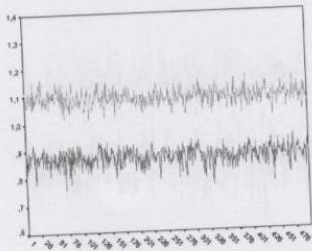


Gráfico 34. Limites dos I_{ET} para $a=0.95$ e $b=4$

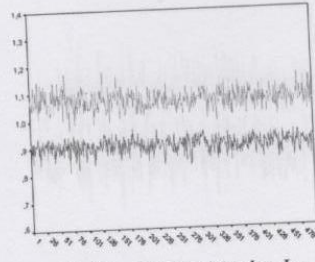


Gráfico 35. Limites dos I_{ET} para $a=0.95$ e $b=8$

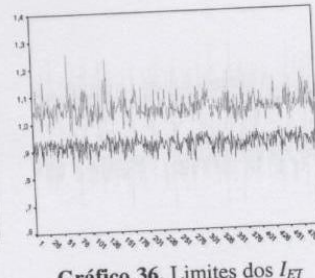


Gráfico 36. Limites dos I_{ET} para $a=0.95$ e $b=16$

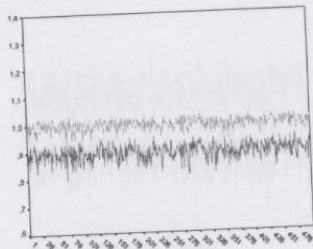


Gráfico 37. Limites dos I_{sym} para $a=0.95$ e $b=4$

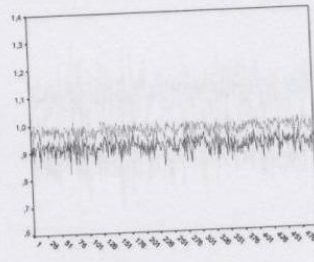


Gráfico 38. Limites dos I_{sym} para $a=0.95$ e $b=8$

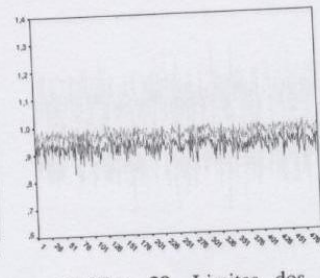


Gráfico 39. Limites dos I_{sym} para $a=0.95$ e $b=16$

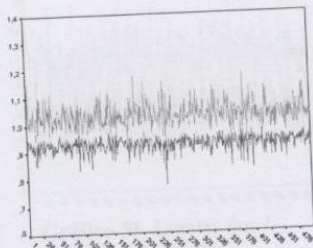


Gráfico 40. Limites dos I_{ET} para $a=0.95$ e $b=32$

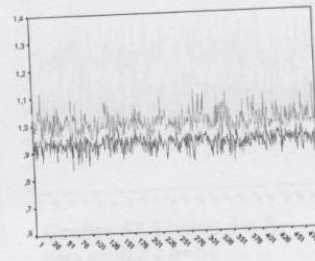


Gráfico 41. Limites dos I_{ET} para $a=0.95$ e $b=64$

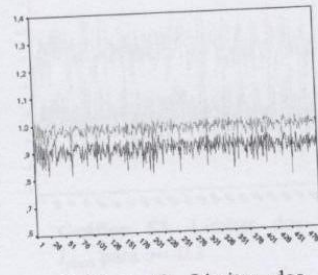


Gráfico 42. Limites dos I_{class} para $a=0.95$

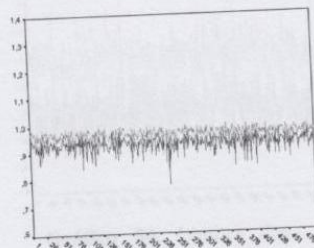


Gráfico 43. Limites dos I_{sym} para $a=0.95$ e $b=32$

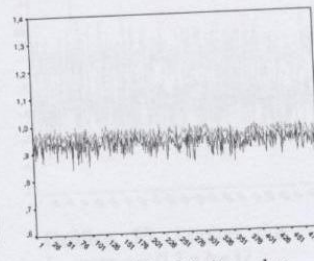


Gráfico 44. Limites dos I_{sym} para $a=0.95$ e $b=32$

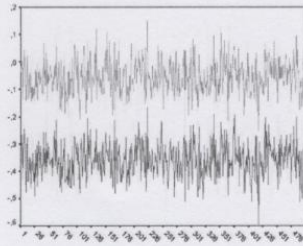


Gráfico 45. Limites dos I_{ET} para $a=-0.2$ e $b=4$

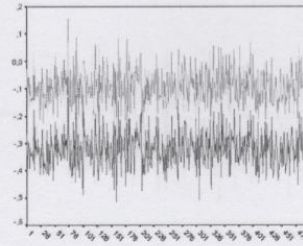


Gráfico 46. Limites dos I_{ET} para $a=-0.2$ e $b=8$

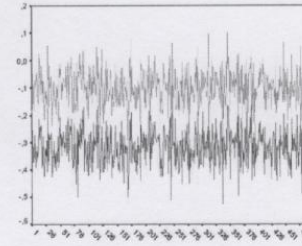


Gráfico 47. Limites dos I_{ET} para $a=-0.2$ e $b=16$

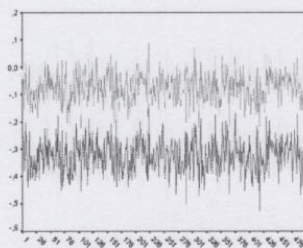


Gráfico 48. Limites dos I_{sym} para $a=-0.2$ e $b=4$

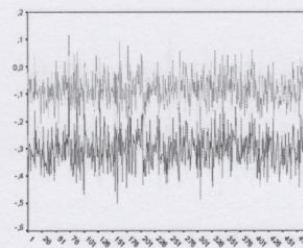


Gráfico 49. Limites dos I_{sym} para $a=-0.2$ e $b=8$

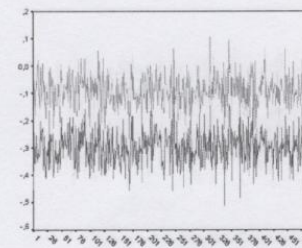


Gráfico 50. Limites dos I_{sym} para $a=-0.2$ e $b=16$

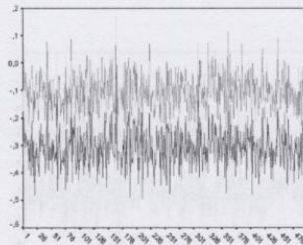


Gráfico 51. Limites dos I_{ET} para $a=-0.2$ e $b=32$

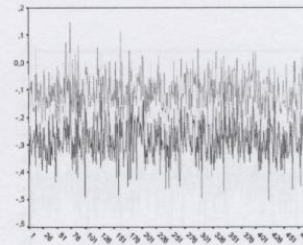


Gráfico 52. Limites dos I_{ET} para $a=-0.2$ e $b=64$

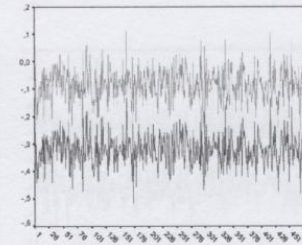


Gráfico 53. Limites dos I_{class} para $a=-0.2$

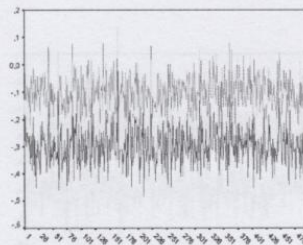


Gráfico 54. Limites dos I_{sym} para $a=-0.2$ e $b=32$

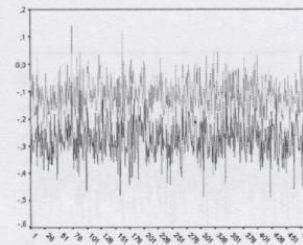


Gráfico 55. Limites dos I_{sym} para $a=-0.2$ e $b=32$

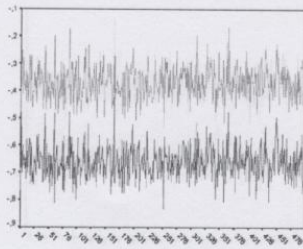


Gráfico 56. Limites dos I_{ET}
para $a=-0.5$ e $b=4$

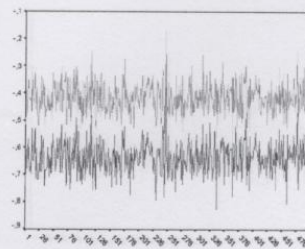


Gráfico 57. Limites dos I_{ET}
para $a=-0.5$ e $b=8$

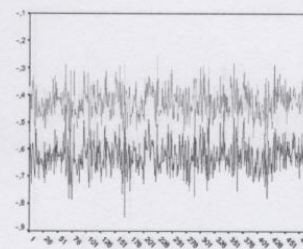


Gráfico 58. Limites dos I_{ET}
para $a=-0.5$ e $b=16$

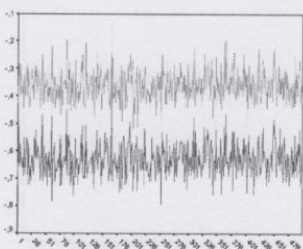


Gráfico 59. Limites dos
 I_{sym} para $a=-0.5$ e $b=4$

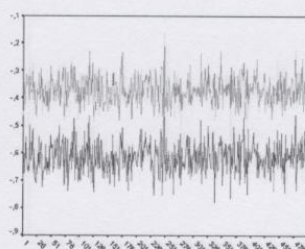


Gráfico 60. Limites dos
 I_{sym} para $a=-0.5$ e $b=8$

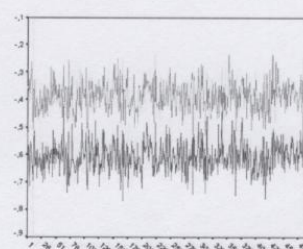


Gráfico 61. Limites dos
 I_{sym} para $a=-0.5$ e $b=16$

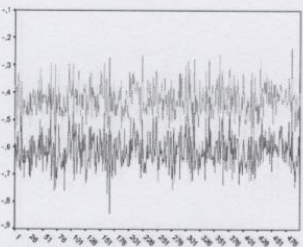


Gráfico 62. Limites dos I_{ET}
para $a=-0.5$ e $b=32$

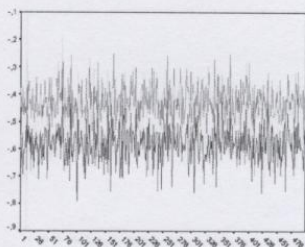


Gráfico 63. Limites dos I_{ET}
para $a=-0.5$ e $b=64$

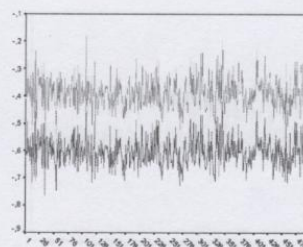


Gráfico 64. Limites dos
 I_{class} para $a=-0.5$

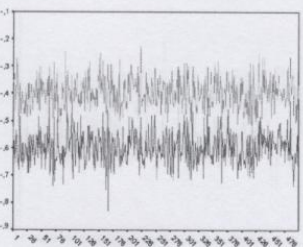


Gráfico 65. Limites dos
 I_{sym} para $a=-0.5$ e $b=32$

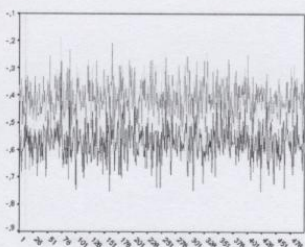


Gráfico 66. Limites dos
 I_{sym} para $a=-0.5$ e $b=32$



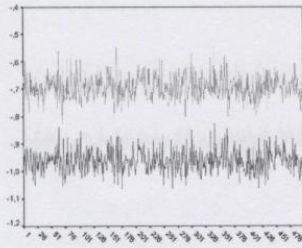


Gráfico 67. Limites dos I_{ET} para $a=-0.8$ e $b=4$

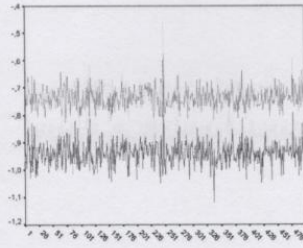


Gráfico 68. Limites dos I_{ET} para $a=-0.8$ e $b=8$

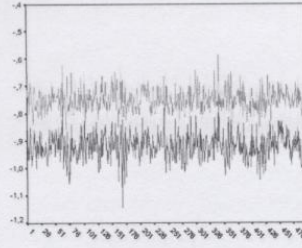


Gráfico 69. Limites dos I_{ET} para $a=-0.8$ e $b=16$

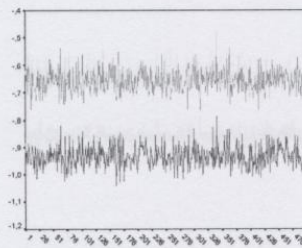


Gráfico 70. Limites dos I_{sym} para $a=-0.8$ e $b=4$

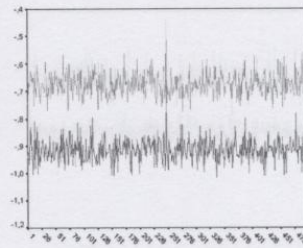


Gráfico 71. Limites dos I_{sym} para $a=-0.8$ e $b=8$

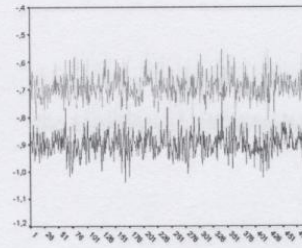


Gráfico 72. Limites dos I_{sym} para $a=-0.8$ e $b=16$

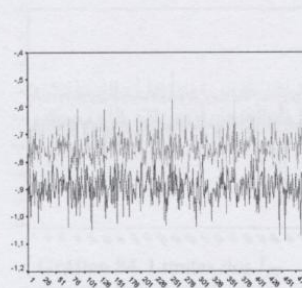


Gráfico 73. Limites dos I_{ET} para $a=-0.8$ e $b=32$

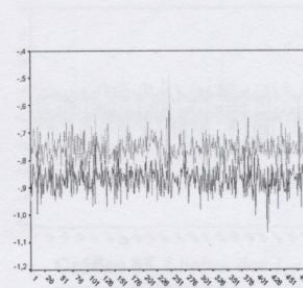


Gráfico 74. Limites dos I_{ET} para $a=-0.8$ e $b=64$

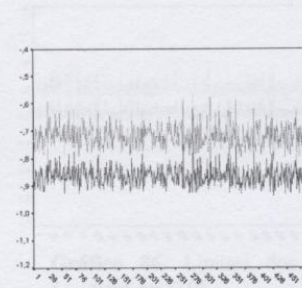


Gráfico 75. Limites dos I_{class} para $a=-0.8$

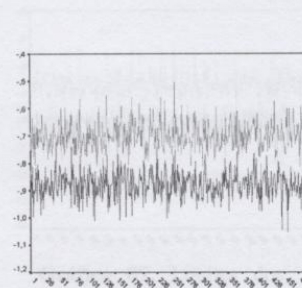


Gráfico 76. Limites dos I_{sym} para $a=-0.8$ e $b=32$

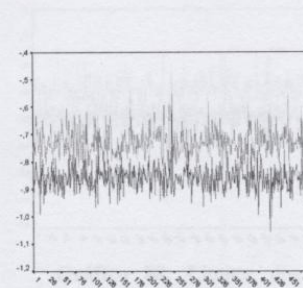


Gráfico 77. Limites dos I_{sym} para $a=-0.8$ e $b=32$

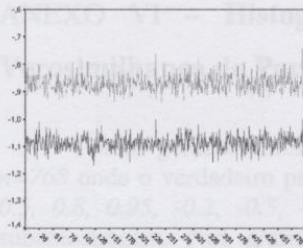


Gráfico 78. Limites dos I_{ET}
para $a=-0.95$ e $b=4$

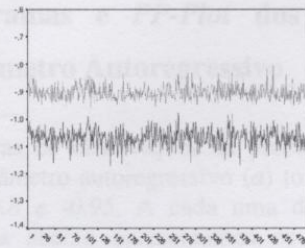


Gráfico 79. Limites dos I_{ET}
para $a=-0.95$ e $b=8$

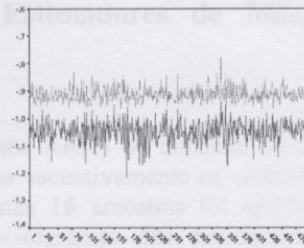


Gráfico 80. Limites dos I_{ET}
para $a=-0.95$ e $b=16$

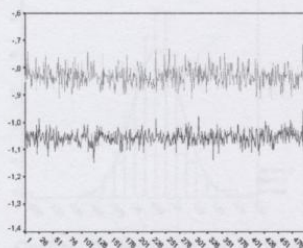


Gráfico 81. Limites dos
 I_{sym} para $a=-0.95$ e $b=4$

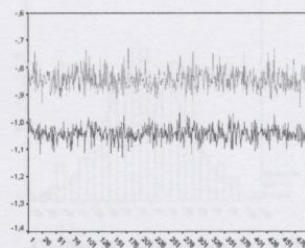


Gráfico 82. Limites dos
 I_{sym} para $a=-0.95$ e $b=8$

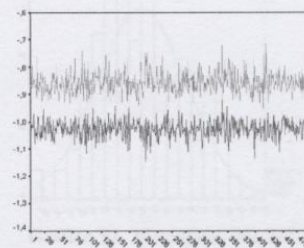


Gráfico 83. Limites dos
 I_{sym} para $a=-0.95$ e $b=16$

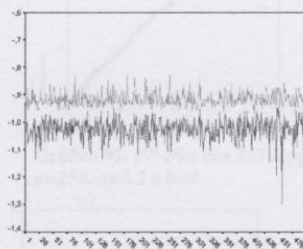


Gráfico 84. Limites dos I_{ET}
para $a=-0.95$ e $b=32$

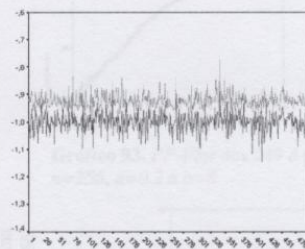


Gráfico 85. Limites dos I_{ET}
para $a=-0.95$ e $b=64$

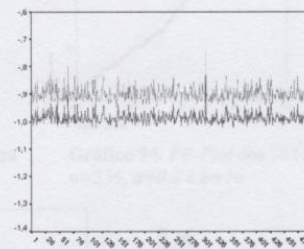


Gráfico 86. Limites dos
 I_{class} para $a=-0.95$

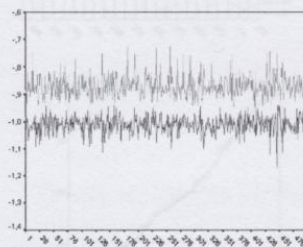


Gráfico 87. Limites dos
 I_{sym} para $a=-0.95$ e $b=32$

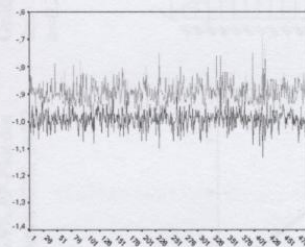


Gráfico 88. Limites dos
 I_{sym} para $a=-0.95$ e $b=32$

ANEXO VI – Histogramas e *PP-Plot* dos Estimadores de Máxima Verosimilhança do Parâmetro Autoregressivo

Foram geradas amostras de observações de processos AR(1) de tamanho $n=256$ e $n=768$ onde o verdadeiro parâmetro autoregressivo (a) toma sucessivamente os valores 0.2 , 0.5 , 0.8 , 0.95 , -0.2 , -0.5 , -0.8 e -0.95 . A cada uma destas 16 amostras foi aplicada a subamostragem. Nos gráficos abaixo apresentamos os histogramas dos estimadores de máxima verosimilhança do parâmetro autoregressivo calculados em subamostras de tamanho $b=4$, $b=8$, $b=16$, $b=32$ e $b=64$, aos quais foi ajustada a curva normal. A cada distribuição obtida foi aplicado o teste de normalidade *PP-Plot*.

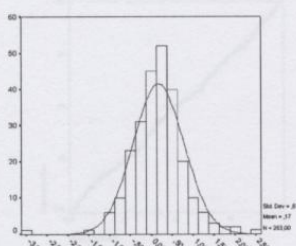


Gráfico 89. Histograma dos 253 \hat{a} para $n=256$, $a=0.2$ e $b=4$

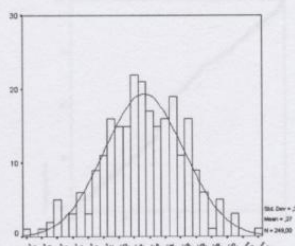


Gráfico 90. Histograma dos 249 \hat{a} para $n=256$, $a=0.2$ e $b=8$

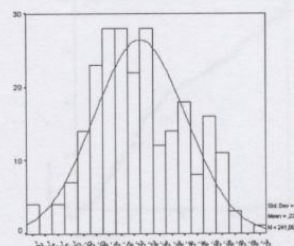


Gráfico 91. Histograma dos 241 \hat{a} para $n=256$, $a=0.2$ e $b=16$

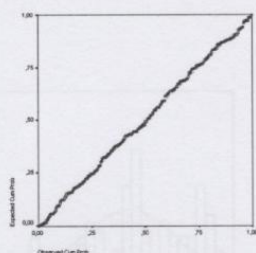


Gráfico 92. *PP-Plot* dos 253 \hat{a} para $n=256$, $a=0.2$ e $b=4$

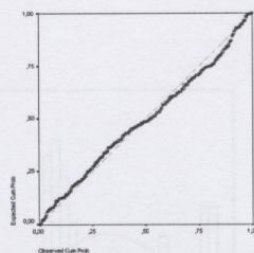


Gráfico 93. *PP-Plot* dos 249 \hat{a} para $n=256$, $a=0.2$ e $b=8$

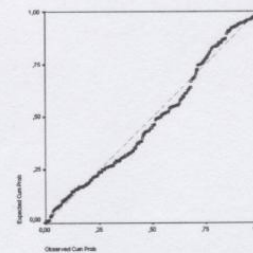


Gráfico 94. *PP-Plot* dos 241 \hat{a} para $n=256$, $a=0.2$ e $b=16$

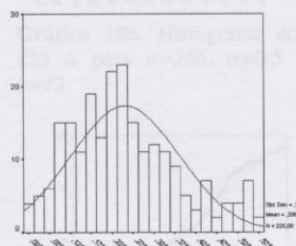


Gráfico 95. Histograma dos 225 \hat{a} para $n=256$, $a=0.2$ e $b=32$

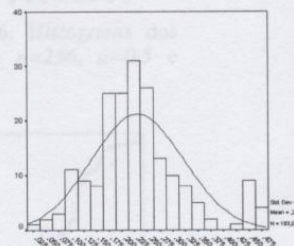


Gráfico 96. Histograma dos 193 \hat{a} para $n=256$, $a=0.2$ e $b=64$

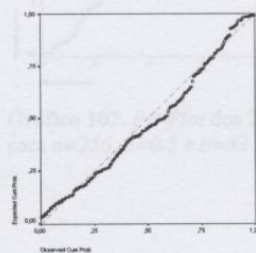


Gráfico 97. *PP-Plot* dos 225 \hat{a} para $n=256$, $a=0.2$ e $b=32$

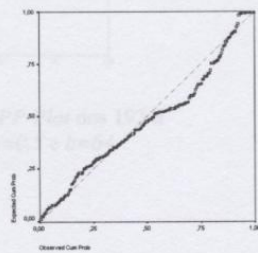


Gráfico 98. *PP-Plot* dos 193 \hat{a} para $n=256$, $a=0.2$ e $b=64$

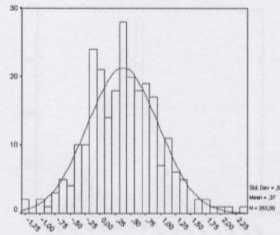


Gráfico 99. Histograma dos 253 \hat{a} para $n=256$, $a=0.5$ e $b=4$

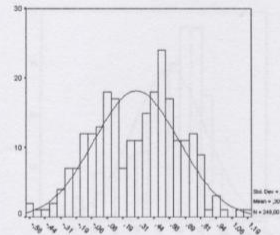


Gráfico 100. Histograma dos 249 \hat{a} para $n=256$, $a=0.5$ e $b=8$

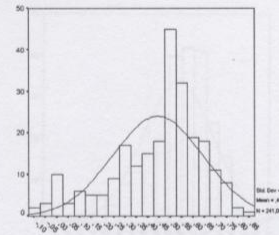


Gráfico 101. Histograma dos 241 \hat{a} para $n=256$, $a=0.5$ e $b=16$

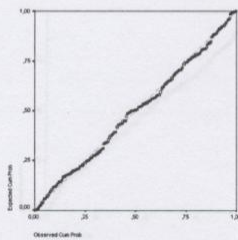


Gráfico 102. PP-Plot dos 253 \hat{a} para $n=256$, $a=0.5$ e $b=4$

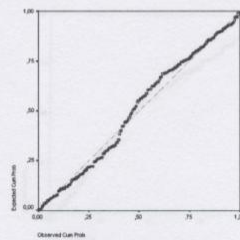


Gráfico 103. PP-Plot dos 249 \hat{a} para $n=256$, $a=0.5$ e $b=8$

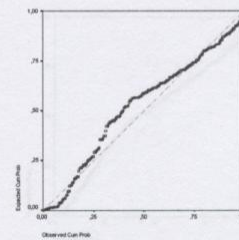


Gráfico 104. PP-Plot dos 241 \hat{a} para $n=256$, $a=0.5$ e $b=16$

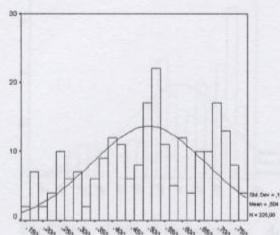


Gráfico 105. Histograma dos 225 \hat{a} para $n=256$, $a=0.5$ e $b=32$

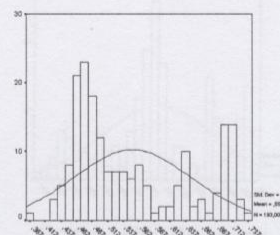


Gráfico 106. Histograma dos 193 \hat{a} para $n=256$, $a=0.5$ e $b=64$

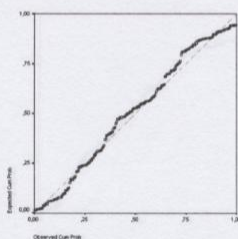


Gráfico 107. PP-Plot dos 225 \hat{a} para $n=256$, $a=0.5$ e $b=32$

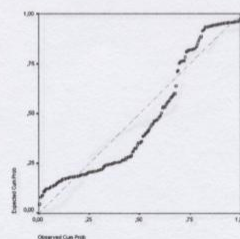


Gráfico 108. PP-Plot dos 193 \hat{a} para $n=256$, $a=0.5$ e $b=64$

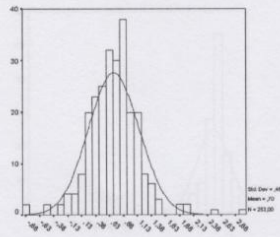


Gráfico 109. Histograma dos 253 \hat{a} para $n=256$, $a=0.8$ e $b=4$

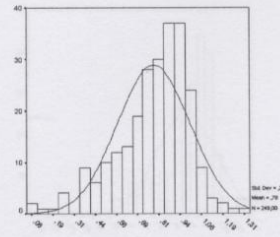


Gráfico 110. Histograma dos 249 \hat{a} para $n=256$, $a=0.8$ e $b=8$

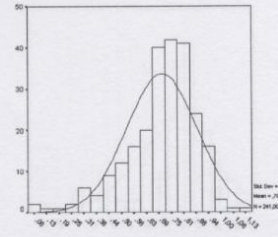


Gráfico 111. Histograma dos 241 \hat{a} para $n=256$, $a=0.8$ e $b=16$

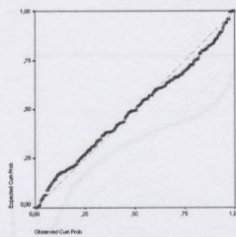


Gráfico 112. PP-Plot dos 253 \hat{a} para $n=256$, $a=0.8$ e $b=4$

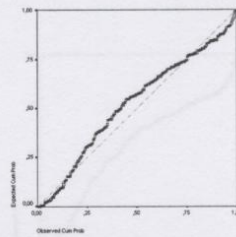


Gráfico 113. PP-Plot dos 249 \hat{a} para $n=256$, $a=0.8$ e $b=8$

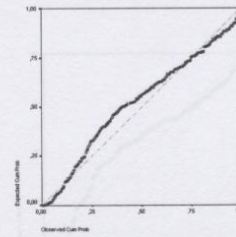


Gráfico 114. PP-Plot dos 241 \hat{a} para $n=256$, $a=0.8$ e $b=16$

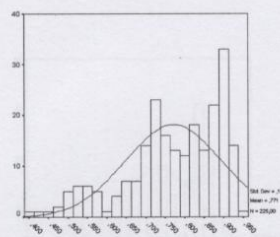


Gráfico 115. Histograma dos 225 \hat{a} para $n=256$, $a=0.8$ e $b=32$

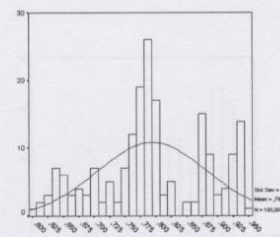


Gráfico 116. Histograma dos 193 \hat{a} para $n=256$, $a=0.8$ e $b=64$

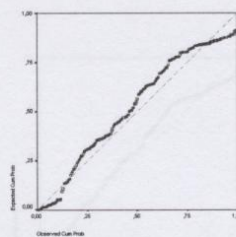


Gráfico 117. PP-Plot dos 225 \hat{a} para $n=256$, $a=0.8$ e $b=32$

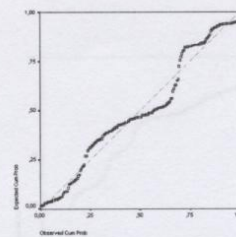


Gráfico 118. PP-Plot dos 193 \hat{a} para $n=256$, $a=0.8$ e $b=64$

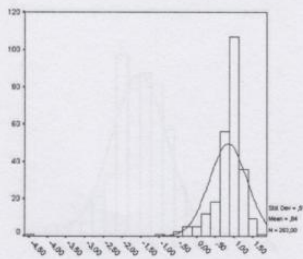


Gráfico 119. Histograma dos 253 \hat{a} para $n=256$, $a=0.95$ e $b=4$

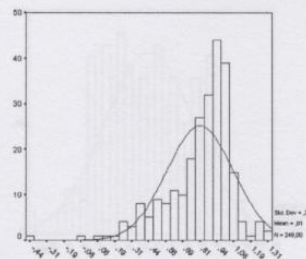


Gráfico 120. Histograma dos 249 \hat{a} para $n=256$, $a=0.95$ e $b=8$

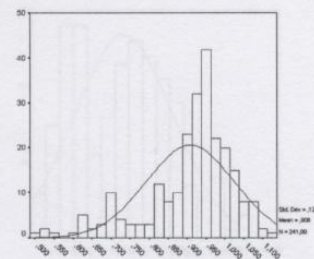


Gráfico 121. Histograma dos 241 \hat{a} para $n=256$, $a=0.95$ e $b=16$

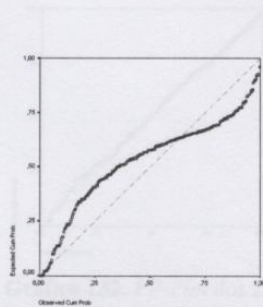


Gráfico 122. PP-Plot dos 253 \hat{a} para $n=256$, $a=0.95$ e $b=4$

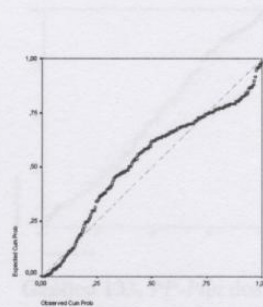


Gráfico 123. PP-Plot dos 249 \hat{a} para $n=256$, $a=0.95$ e $b=8$

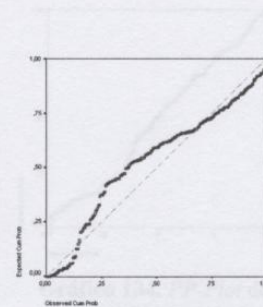


Gráfico 124. PP-Plot dos 241 \hat{a} para $n=256$, $a=0.95$ e $b=16$

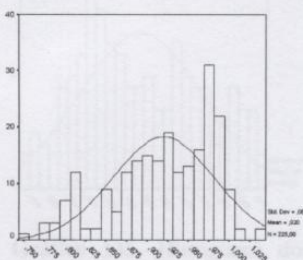


Gráfico 125. Histograma dos 225 \hat{a} para $n=256$, $a=0.95$ e $b=32$

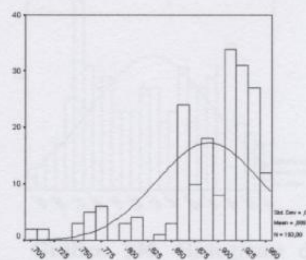


Gráfico 126. Histograma dos 193 \hat{a} para $n=256$, $a=0.95$ e $b=64$

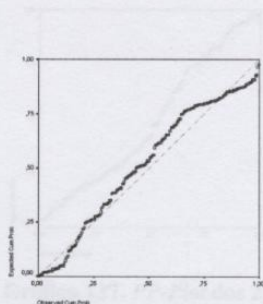


Gráfico 127. PP-Plot dos 225 \hat{a} para $n=256$, $a=0.95$ e $b=32$

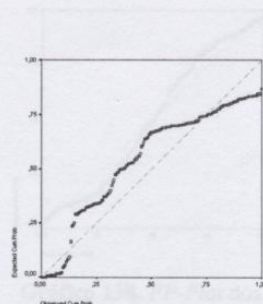


Gráfico 128. PP-Plot dos 193 \hat{a} para $n=256$, $a=0.95$ e $b=64$

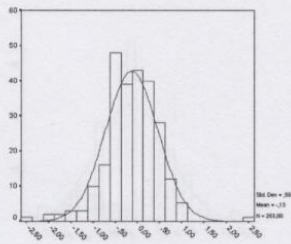


Gráfico 129. Histograma dos 253 \hat{a} para $n=256$, $a=-0.2$ e $b=4$

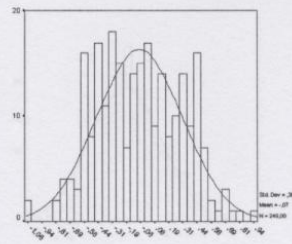


Gráfico 130. Histograma dos 249 \hat{a} para $n=256$, $a=-0.2$ e $b=8$

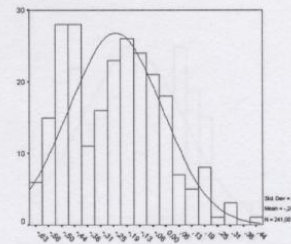


Gráfico 131. Histograma dos 241 \hat{a} para $n=256$, $a=-0.2$ e $b=16$

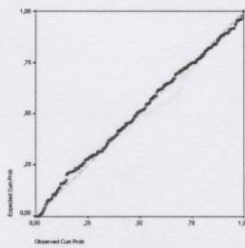


Gráfico 132. PP-Plot dos 253 \hat{a} para $n=256$, $a=-0.2$ e $b=4$

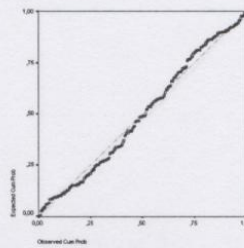


Gráfico 133. PP-Plot dos 249 \hat{a} para $n=256$, $a=-0.2$ e $b=8$

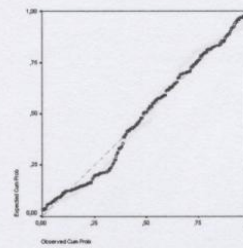


Gráfico 134. PP-Plot dos 241 \hat{a} para $n=256$, $a=-0.2$ e $b=16$

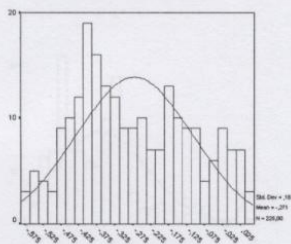


Gráfico 135. Histograma dos 225 \hat{a} para $n=256$, $a=-0.2$ e $b=32$

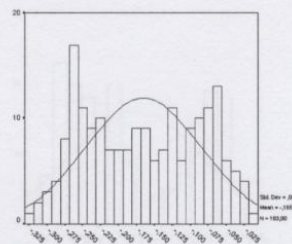


Gráfico 136. Histograma dos 193 \hat{a} para $n=256$, $a=-0.2$ e $b=64$

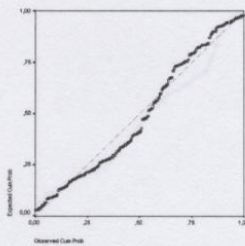


Gráfico 137. PP-Plot dos 225 \hat{a} para $n=256$, $a=-0.2$ e $b=32$

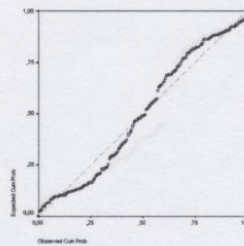


Gráfico 138. PP-Plot dos 193 \hat{a} para $n=256$, $a=-0.2$ e $b=64$

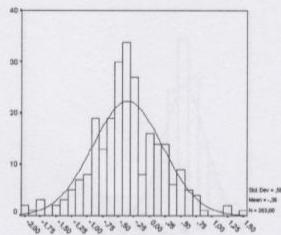


Gráfico 139. Histograma dos 253 \hat{a} para $n=256$, $a=-0.5$ e $b=4$

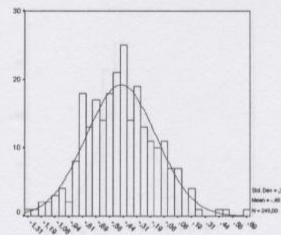


Gráfico 140. Histograma dos 249 \hat{a} para $n=256$, $a=-0.5$ e $b=8$

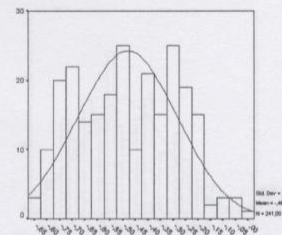


Gráfico 141. Histograma dos 241 \hat{a} para $n=256$, $a=-0.5$ e $b=16$

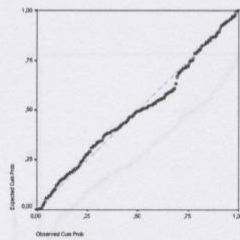


Gráfico 142. PP-Plot dos 253 \hat{a} para $n=256$, $a=-0.5$ e $b=4$

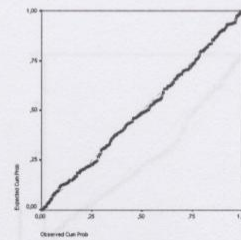


Gráfico 143. PP-Plot dos 249 \hat{a} para $n=256$, $a=-0.5$ e $b=8$

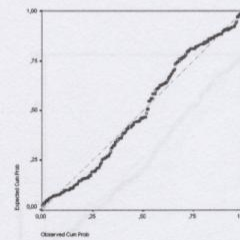


Gráfico 144. PP-Plot dos 241 \hat{a} para $n=256$, $a=-0.5$ e $b=16$

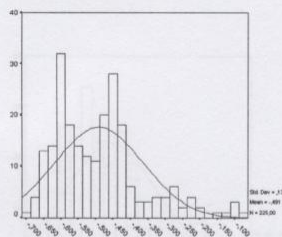


Gráfico 145. Histograma dos 225 \hat{a} para $n=256$, $a=-0.5$ e $b=32$

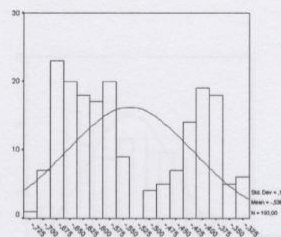


Gráfico 146. Histograma dos 193 \hat{a} para $n=256$, $a=-0.5$ e $b=64$

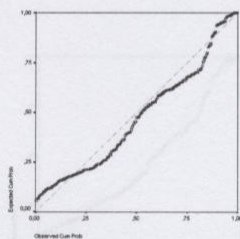


Gráfico 147. PP-Plot dos 225 \hat{a} para $n=256$, $a=-0.5$ e $b=32$

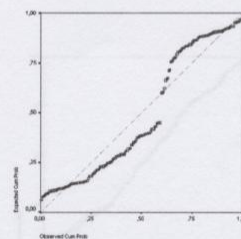


Gráfico 148. PP-Plot dos 193 \hat{a} para $n=256$, $a=-0.5$ e $b=64$

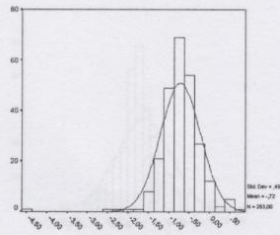


Gráfico 149. Histograma dos 253 \hat{a} para $n=256$, $a=-0.8$ e $b=4$

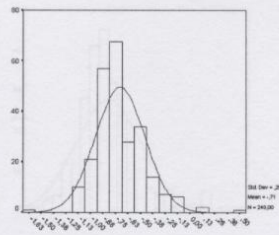


Gráfico 150. Histograma dos 249 \hat{a} para $n=256$, $a=-0.8$ e $b=8$

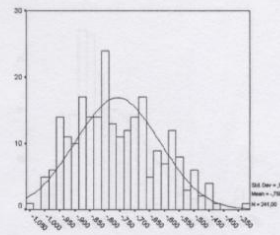


Gráfico 151. Histograma dos 241 \hat{a} para $n=256$, $a=-0.8$ e $b=16$

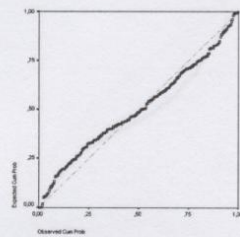


Gráfico 152. PP-Plot dos 253 \hat{a} para $n=256$, $a=-0.8$ e $b=4$

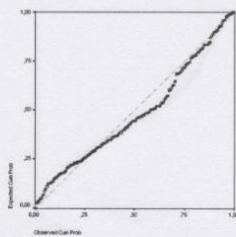


Gráfico 153. PP-Plot dos 249 \hat{a} para $n=256$, $a=-0.8$ e $b=8$

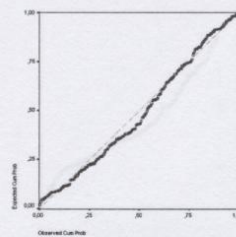


Gráfico 154. PP-Plot dos 241 \hat{a} para $n=256$, $a=-0.8$ e $b=16$

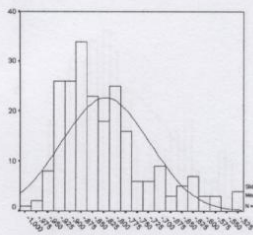


Gráfico 155. Histograma dos 225 \hat{a} para $n=256$, $a=-0.8$ e $b=32$

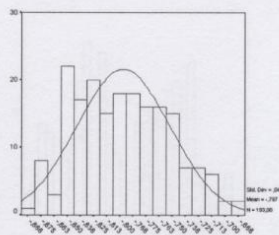


Gráfico 156. Histograma dos 193 \hat{a} para $n=256$, $a=-0.8$ e $b=64$

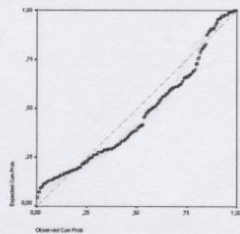


Gráfico 157. PP-Plot dos 225 \hat{a} para $n=256$, $a=-0.8$ e $b=32$

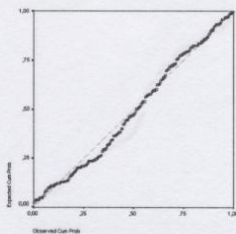


Gráfico 158. PP-Plot dos 193 \hat{a} para $n=256$, $a=-0.8$ e $b=64$

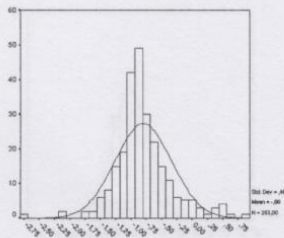


Gráfico 159. Histograma dos 253 \hat{a} para $n=256$, $a=-0.95$ e $b=4$

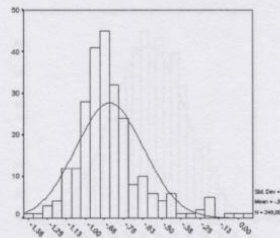


Gráfico 160. Histograma dos 249 \hat{a} para $n=256$, $a=-0.95$ e $b=8$

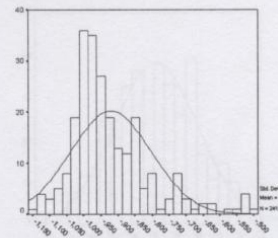


Gráfico 161. Histograma dos 241 \hat{a} para $n=256$, $a=-0.95$ e $b=16$

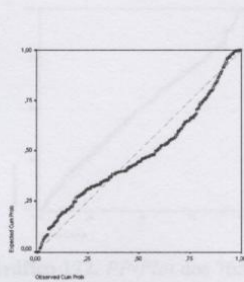


Gráfico 162. PP-Plot dos 253 \hat{a} para $n=256$, $a=-0.95$ e $b=4$

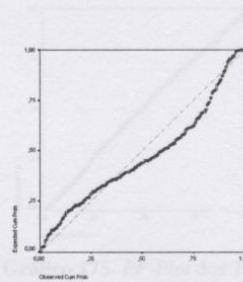


Gráfico 163. PP-Plot dos 249 \hat{a} para $n=256$, $a=-0.95$ e $b=8$

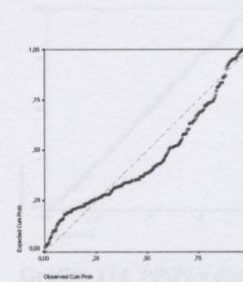


Gráfico 164. PP-Plot dos 241 \hat{a} para $n=256$, $a=-0.95$ e $b=16$

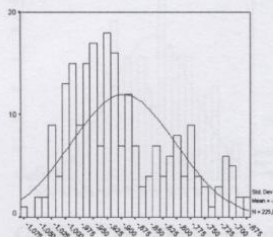


Gráfico 165. Histograma dos 225 \hat{a} para $n=256$, $a=-0.95$ e $b=32$

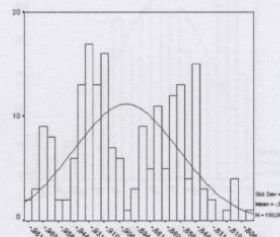


Gráfico 166. Histograma dos 193 \hat{a} para $n=256$, $a=-0.95$ e $b=64$

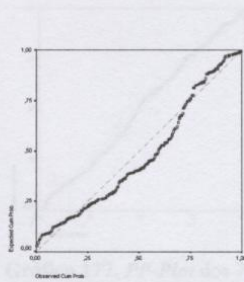


Gráfico 167. PP-Plot dos 225 \hat{a} para $n=256$, $a=-0.95$ e $b=32$

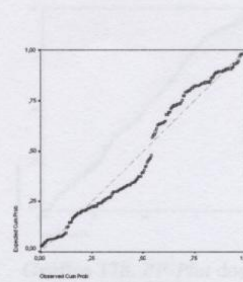


Gráfico 168. PP-Plot dos 193 \hat{a} para $n=256$, $a=-0.95$ e $b=64$

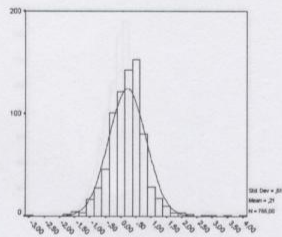


Gráfico 169. Histograma dos 765 \hat{a} para $n=768$, $a=0.2$ e $b=4$

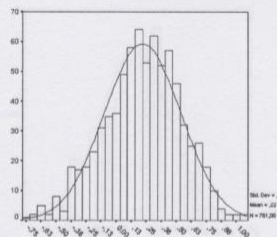


Gráfico 170. Histograma dos 761 \hat{a} para $n=768$, $a=0.2$ e $b=8$

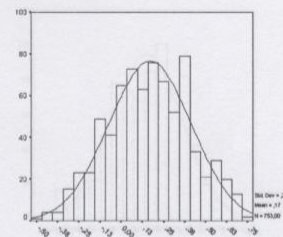


Gráfico 171. Histograma dos 753 \hat{a} para $n=768$, $a=0.2$ e $b=16$

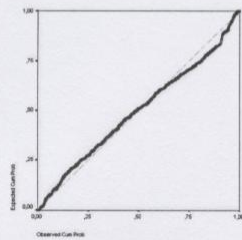


Gráfico 172. PP-Plot dos 765 \hat{a} para $n=768$, $a=0.2$ e $b=4$

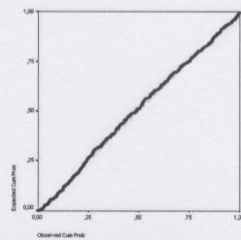


Gráfico 173. PP-Plot dos 761 \hat{a} para $n=768$, $a=0.2$ e $b=8$

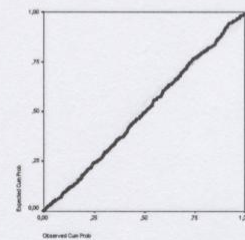


Gráfico 174. PP-Plot dos 763 \hat{a} para $n=768$, $a=0.2$ e $b=16$

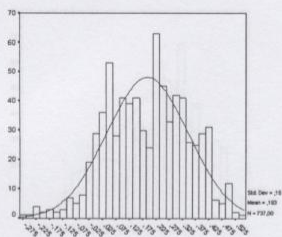


Gráfico 175. Histograma dos 737 \hat{a} para $n=768$, $a=0.2$ e $b=32$

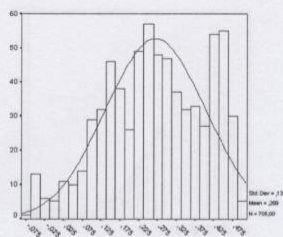


Gráfico 176. Histograma dos 705 \hat{a} para $n=768$, $a=0.2$ e $b=64$

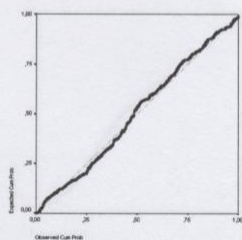


Gráfico 177. PP-Plot dos 737 \hat{a} para $n=768$, $a=0.2$ e $b=32$

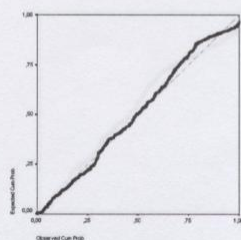


Gráfico 178. PP-Plot dos 705 \hat{a} para $n=768$, $a=0.2$ e $b=64$

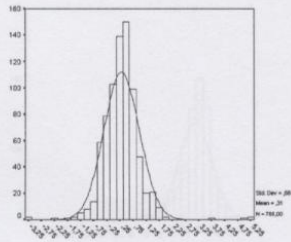


Gráfico 179. Histograma dos 765 \hat{a} para $n=768$, $a=0.5$ e $b=4$

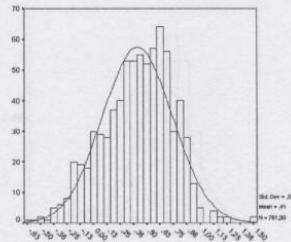


Gráfico 180. Histograma dos 761 \hat{a} para $n=768$, $a=0.5$ e $b=8$

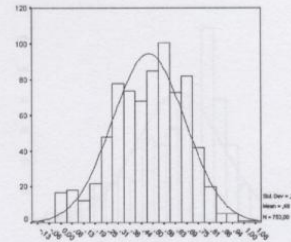


Gráfico 181. Histograma dos 753 \hat{a} para $n=768$, $a=0.5$ e $b=16$

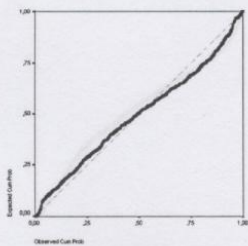


Gráfico 182. PP-Plot dos 765 \hat{a} para $n=768$, $a=0.5$ e $b=4$

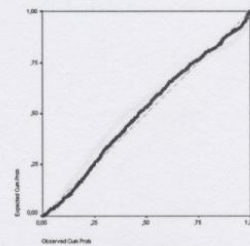


Gráfico 183. PP-Plot dos 761 \hat{a} para $n=768$, $a=0.5$ e $b=8$

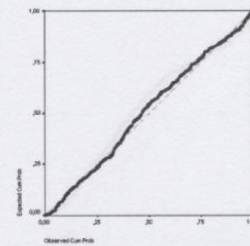


Gráfico 184. PP-Plot dos 763 \hat{a} para $n=768$, $a=0.5$ e $b=16$

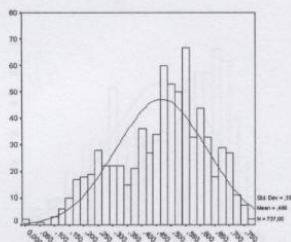


Gráfico 185. Histograma dos 737 \hat{a} para $n=768$, $a=0.5$ e $b=32$

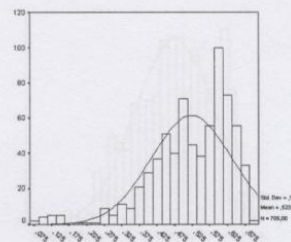


Gráfico 186. Histograma dos 705 \hat{a} para $n=768$, $a=0.5$ e $b=64$

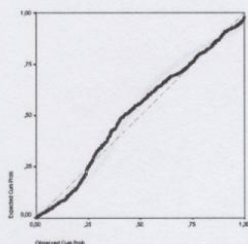


Gráfico 187. PP-Plot dos 737 \hat{a} para $n=768$, $a=0.5$ e $b=32$

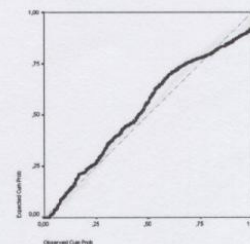


Gráfico 188. PP-Plot dos 705 \hat{a} para $n=768$, $a=0.5$ e $b=64$

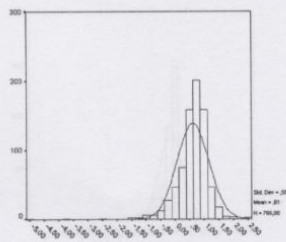


Gráfico 189. Histograma dos 765 \hat{a} para $n=768$, $a=0.8$ e $b=4$

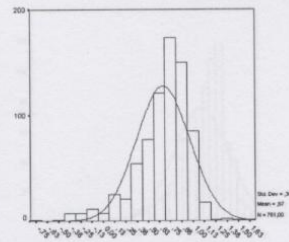


Gráfico 190. Histograma dos 761 \hat{a} para $n=768$, $a=0.8$ e $b=8$

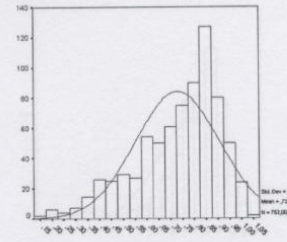


Gráfico 191. Histograma dos 753 \hat{a} para $n=768$, $a=0.8$ e $b=16$

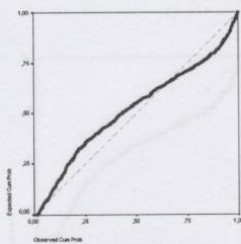


Gráfico 192. PP-Plot dos 765 \hat{a} para $n=768$, $a=0.8$ e $b=4$

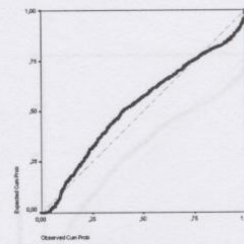


Gráfico 193. PP-Plot dos 761 \hat{a} para $n=768$, $a=0.8$ e $b=8$

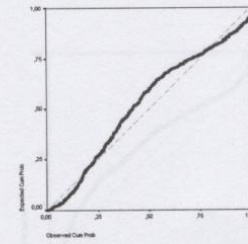


Gráfico 194. PP-Plot dos 763 \hat{a} para $n=768$, $a=0.8$ e $b=16$

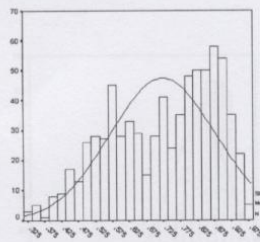


Gráfico 195. Histograma dos 737 \hat{a} para $n=768$, $a=0.8$ e $b=32$

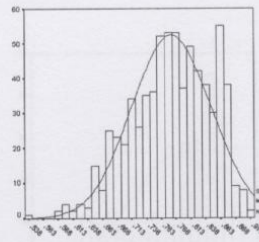


Gráfico 196. Histograma dos 705 \hat{a} para $n=768$, $a=0.8$ e $b=64$

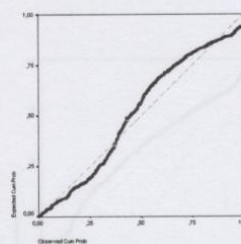


Gráfico 197. PP-Plot dos 737 \hat{a} para $n=768$, $a=0.5$ e $b=32$

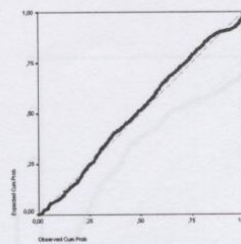


Gráfico 198. PP-Plot dos 705 \hat{a} para $n=768$, $a=0.8$ e $b=64$

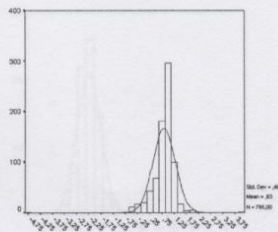


Gráfico 199. Histograma dos 765 \hat{a} para $n=768$, $a=0.95$ e $b=4$

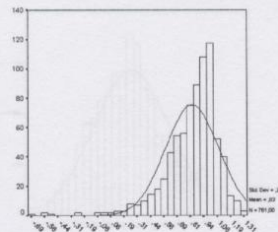


Gráfico 200. Histograma dos 761 \hat{a} para $n=768$, $a=0.95$ e $b=8$

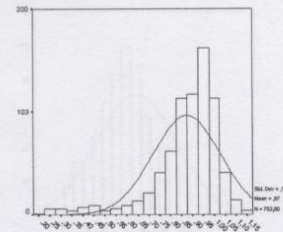


Gráfico 201. Histograma dos 753 \hat{a} para $n=768$, $a=0.95$ e $b=16$

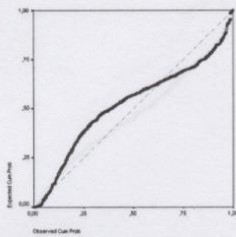


Gráfico 202. PP-Plot dos 765 \hat{a} para $n=768$, $a=0.95$ e $b=4$

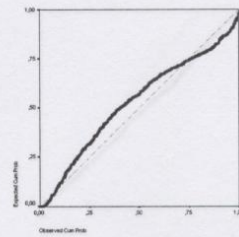


Gráfico 203. PP-Plot dos 761 \hat{a} para $n=768$, $a=0.95$ e $b=8$

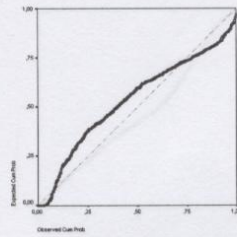


Gráfico 204. PP-Plot dos 763 \hat{a} para $n=768$, $a=0.95$ e $b=16$

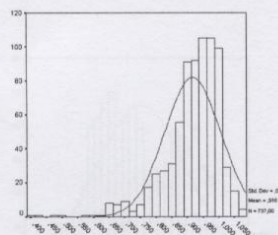


Gráfico 205. Histograma dos 737 \hat{a} para $n=768$, $a=0.95$ e $b=32$

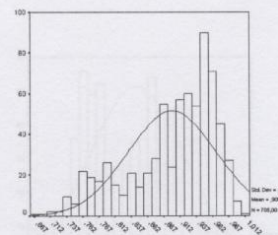


Gráfico 206. Histograma dos 705 \hat{a} para $n=768$, $a=0.95$ e $b=64$

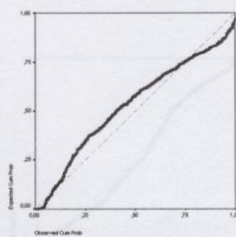


Gráfico 207. PP-Plot dos 737 \hat{a} para $n=768$, $a=0.95$ e $b=32$

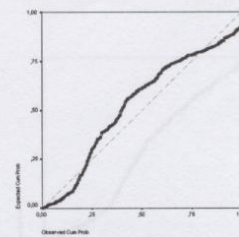


Gráfico 208. PP-Plot dos 705 \hat{a} para $n=768$, $a=0.95$ e $b=64$

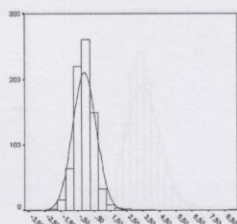


Gráfico 209. Histograma dos 765 \hat{a} para $n=768$, $a=-0.2$ e $b=4$

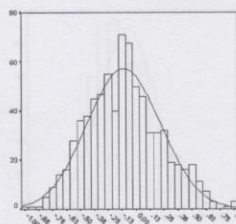


Gráfico 210. Histograma dos 761 \hat{a} para $n=768$, $a=-0.2$ e $b=8$

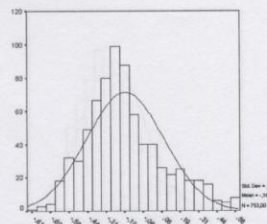


Gráfico 211. Histograma dos 753 \hat{a} para $n=768$, $a=-0.2$ e $b=16$

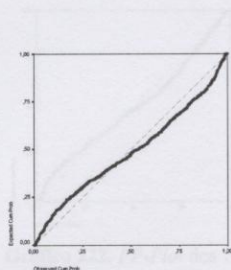


Gráfico 212. PP-Plot dos 765 \hat{a} para $n=768$, $a=-0.2$ e $b=4$

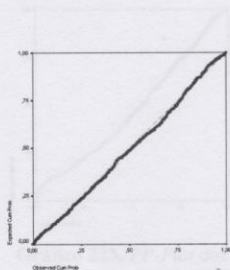


Gráfico 213. PP-Plot dos 761 \hat{a} para $n=768$, $a=-0.2$ e $b=8$

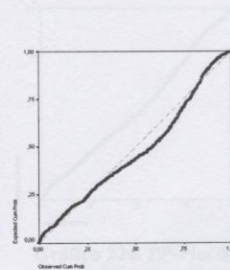


Gráfico 214. PP-Plot dos 763 \hat{a} para $n=768$, $a=-0.2$ e $b=16$

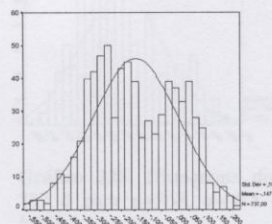


Gráfico 215. Histograma dos 737 \hat{a} para $n=768$, $a=-0.2$ e $b=32$

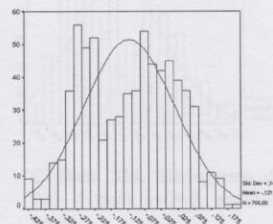


Gráfico 216. Histograma dos 705 \hat{a} para $n=768$, $a=-0.2$ e $b=64$

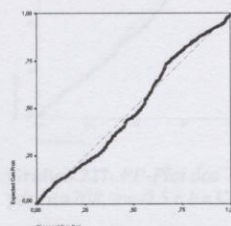


Gráfico 217. PP-Plot dos 737 \hat{a} para $n=768$, $a=-0.2$ e $b=32$

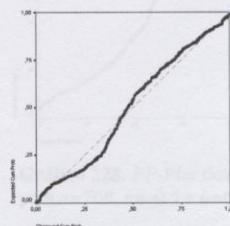


Gráfico 218. PP-Plot dos 705 \hat{a} para $n=768$, $a=-0.2$ e $b=64$

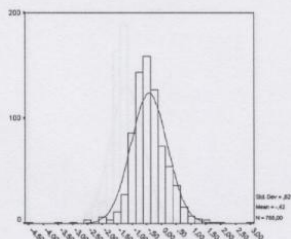


Gráfico 219. Histograma dos 765 \hat{a} para $n=768$, $a=-0.5$ e $b=4$

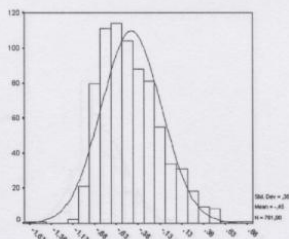


Gráfico 220. Histograma dos 761 \hat{a} para $n=768$, $a=-0.5$ e $b=8$

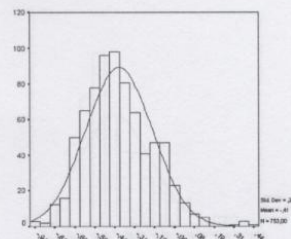


Gráfico 221. Histograma dos 753 \hat{a} para $n=768$, $a=-0.5$ e $b=16$

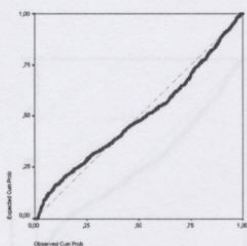


Gráfico 222. PP-Plot dos 765 \hat{a} para $n=768$, $a=-0.5$ e $b=4$

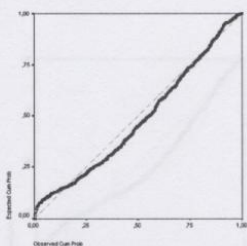


Gráfico 223. PP-Plot dos 761 \hat{a} para $n=768$, $a=-0.5$ e $b=8$

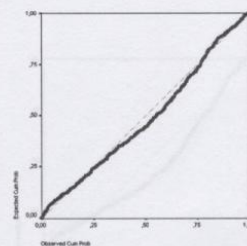


Gráfico 224. PP-Plot dos 763 \hat{a} para $n=768$, $a=-0.5$ e $b=16$

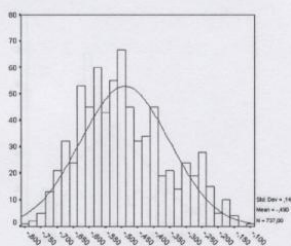


Gráfico 225. Histograma dos 737 \hat{a} para $n=768$, $a=-0.5$ e $b=32$

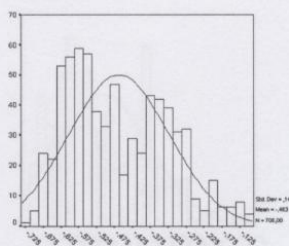


Gráfico 226. Histograma dos 705 \hat{a} para $n=768$, $a=-0.5$ e $b=64$

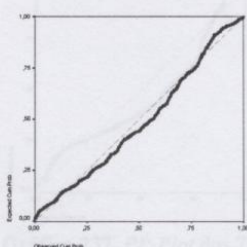


Gráfico 227. PP-Plot dos 737 \hat{a} para $n=768$, $a=-0.5$ e $b=32$

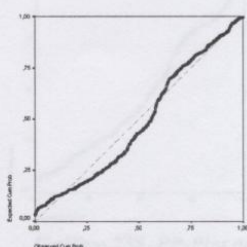


Gráfico 228. PP-Plot dos 705 \hat{a} para $n=768$, $a=-0.5$ e $b=64$

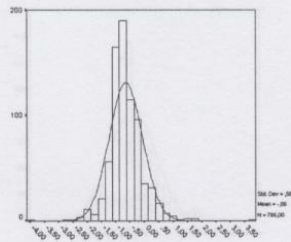


Gráfico 229. Histograma dos 765 \hat{a} para $n=768$, $a=-0.8$ e $b=4$

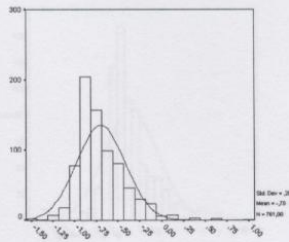


Gráfico 230. Histograma dos 761 \hat{a} para $n=768$, $a=-0.8$ e $b=8$

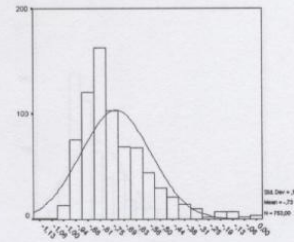


Gráfico 231. Histograma dos 753 \hat{a} para $n=768$, $a=-0.8$ e $b=16$

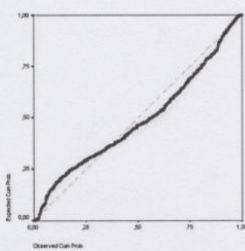


Gráfico 232. PP-Plot dos 765 \hat{a} para $n=768$, $a=-0.8$ e $b=4$

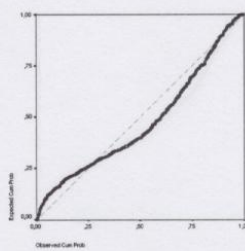


Gráfico 233. PP-Plot dos 761 \hat{a} para $n=768$, $a=-0.8$ e $b=8$

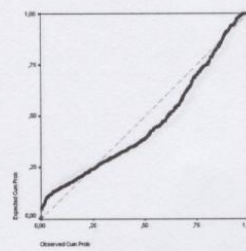


Gráfico 234. PP-Plot dos 763 \hat{a} para $n=768$, $a=-0.8$ e $b=16$

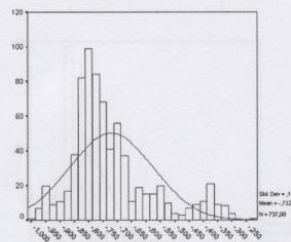


Gráfico 235. Histograma dos 737 \hat{a} para $n=768$, $a=-0.8$ e $b=32$

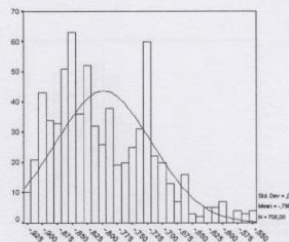


Gráfico 236. Histograma dos 705 \hat{a} para $n=768$, $a=-0.8$ e $b=64$

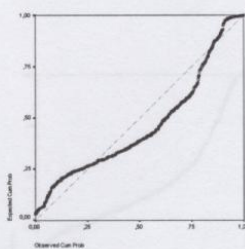


Gráfico 237. PP-Plot dos 737 \hat{a} para $n=768$, $a=-0.8$ e $b=32$

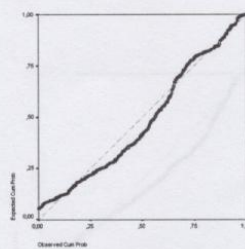


Gráfico 238. PP-Plot dos 705 \hat{a} para $n=768$, $a=-0.8$ e $b=64$

Gráfico 239. PP-Plot dos 737 \hat{a} para $n=768$, $a=-0.8$ e $b=32$

Gráfico 238. PP-Plot dos 705 \hat{a} para $n=768$, $a=-0.8$ e $b=64$

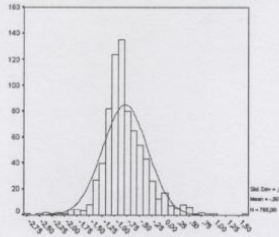


Gráfico 239. Histograma dos 765 \hat{a} para $n=768$, $a=-0.95$ e $b=4$

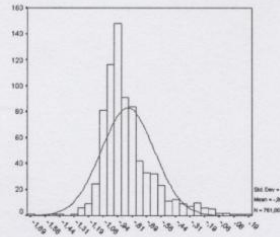


Gráfico 240. Histograma dos 761 \hat{a} para $n=768$, $a=-0.95$ e $b=8$

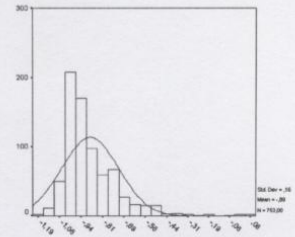


Gráfico 241. Histograma dos 753 \hat{a} para $n=768$, $a=-0.95$ e $b=16$

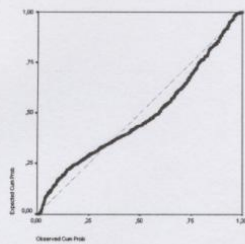


Gráfico 242. PP-Plot dos 765 \hat{a} para $n=768$, $a=-0.95$ e $b=4$

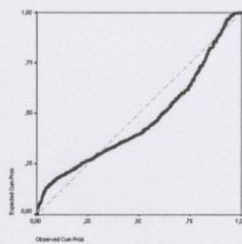


Gráfico 243. PP-Plot dos 761 \hat{a} para $n=768$, $a=-0.95$ e $b=8$

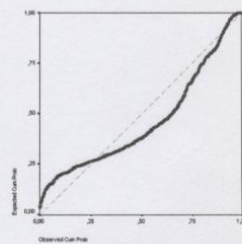


Gráfico 244. PP-Plot dos 763 \hat{a} para $n=768$, $a=-0.95$ e $b=16$

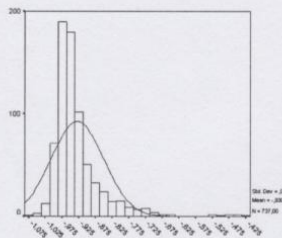


Gráfico 245. Histograma dos 737 \hat{a} para $n=768$, $a=-0.95$ e $b=32$

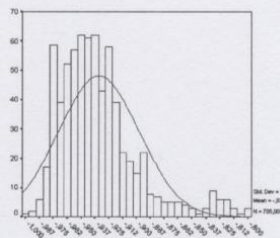


Gráfico 246. Histograma dos 705 \hat{a} para $n=768$, $a=-0.95$ e $b=64$

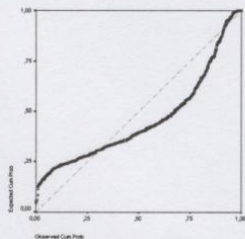


Gráfico 247. PP-Plot dos 737 \hat{a} para $n=768$, $a=-0.95$ e $b=32$

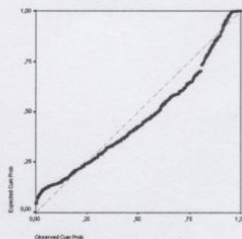


Gráfico 248. PP-Plot dos 705 \hat{a} para $n=768$, $a=-0.95$ e $b=64$