Avaliação da Performance de algoritmo automático de identificação de modelos ARMA(1,1), MA(1), AR(1) e ruído brando.

INTRODUÇÃO:

Falar de séries temporais (importância, onde encontrar, modelos sugeridos que tentam “explicar” o comportamento destas.)

Restringir o tema do trabalho para o modelo ARMA (quem sugeriu, importância, aplicação, notação)

Um aspecto muito importante é o processo de especificação do modelo apropriado à série temporal de interesse. Restringindo-se à classe de modelos ARMA o processo de especificação deve definir a ordem dos parâmetros das partes auto regressiva e de médias móveis do modelo, ou seja, os valores “p” e “q” devem ser identificados com base no comportamento da série temporal. A identificação da ordem do modelo não é uma tarefa simples e direta, depende da análise crítica de uma pessoa que tenha certa familiaridade com as técnicas de identificação e teoria por trás dessa classe de modelos. Mesmo a identificação da ordem de um modelo ARMA de uma série temporal sendo altamente dependente da análise de um ser humano, existem algoritmos automáticos que tentam realizar esta tarefa com base nos dados.

A função “auto.arima” disponível no pacote estatístico R, dentre outras tarefas, identifica a ordem apropriada de um modelo ARMA com base numa série temporal fornecida como parâmetro da função. A utilização da função “auto.arima” faria com que o processo de estimação de um modelo ARMA fosse muito mais simples e rápido, mas primeiro é preciso avaliar a precisão do algoritmo. Uma maneira de avaliar a precisão seria simular séries segundo modelos ARMA, aplicar nestas a função “auto.arima” e comparar a ordem do modelo sugerido pela função com a ordem real do modelo simulado, a precisão pode ser estimada pela proporção de ordens sugeridas que concordam com as ordens simuladas.

OBJETIVO:

Este trabalho visa avaliar a performance na identificação automática das ordens de modelos ARMA(p,q) utilizando a função “auto.arima” disponível no pacote estatístico R.

METODOLOGIA:

Utilizando o pacote estatístico R versão XXXX e a função “arima.sim” disponível neste, foram simuladas três quantidades de tamanho amostrais (100, 500, 1000) de séries temporais segundo modelos ARMA(1,1), ARMA(1,0), ARMA(0,1) e ARMA(0,0) com diferentes combinações de valores para os parâmetros phi e theta (phi, theta = 0, 0,25, 0,5 e 0,75).

A variação dos parâmetros, ordens do modelo e valores de phi e theta objetivam identificar possíveis diferenças de performance evidente na utilização da função “auto.arima” para diferentes situações. Optou-se por restringir os valores p e q como valores máximos iguais a 1, por se tratar de modelos mais simples, e valores de phi e theta variando entre 0, 0,25 , 0,5 e 0,75 por se tratar de valores que gerariam séries invertíveis, estacionárias e que não proporcionem cancelamento de raízes (veja a notação utilizada pelas funções “arima.sim” e “auto.arima” em (1)).

Aplicou-se a função “auto.arima” em todas as séries geradas, definindo os parâmetros “seasonal” como ‘FALSE’ e “stationary” como ‘TRUE’, de modo a restringir a busca da função por modelos estacionários e não sazonais, mantendo o restante dos parâmetros como o padrão da função. Comparou-se as ordens dos modelos sugeridos pela função com as ordens do modelo simulado, caso estas concordassem computa-se um acerto. Esta situação foi repetida 1000 vezes e então identificou-se a quantidade de acertos para cada série simulada, assim como a proporção de vezes em que ocorreu um acerto.

RESULTADOS:

Abaixo são apresentadas as proporções de acerto para todas as combinações de parâmetros phi e theta para séries simuladas de tamanho 100, 500 e 1000 nas tabelas 1, 2 e 3 respectivamente.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| n = 100 | MA | | | |
| AR | 0 | .25 | .5 | .75 |
| 0 | 0,861 | 0,724 | 0,78 | 0,371 |
| .25 | 0,333 | 0,274 | 0,5 | 0,044 |
| .5 | 0,689 | 0,655 | 0,754 | 0,185 |
| .75 | 0,769 | 0,656 | 0,732 | 0,293 |
| Média Geral: | | | | 0,53875 |
|  |  |  |  |  |
| n = 500 | MA | | | |
| AR | 0 | .25 | .5 | .75 |
| 0 | 0,716 | 0,622 | 0,519 | 0,545 |
| .25 | 0,466 | 0,604 | 0,557 | 0,335 |
| .5 | 0,627 | 0,637 | 0,575 | 0,553 |
| .75 | 0,671 | 0,625 | 0,564 | 0,613 |
| Média Geral: | | | | 0,576813 |
|  |  |  |  |  |
| n = 1000 | MA | | | |
| AR | 0 | .25 | .5 | .75 |
| 0 | 0,687 | 0,584 | 0,544 | 0,591 |
| .25 | 0,577 | 0,613 | 0,545 | 0,55 |
| .5 | 0,631 | 0,617 | 0,592 | 0,646 |
| .75 | 0,668 | 0,603 | 0,566 | 0,677 |
| Média Geral: | | | | 0,605688 |

Considerando tudo a média global foi de aproximadamente 57% de acerto.

Discussões:

Observando os resultados obtidos, nota-se maior variabilidade nos valores das proporções de acerto para séries simuladas dos modelos propostos com tamanho 100 do que quando comparadas às de tamanho 500, assim como maior variabilidade destas advindas séries de tamanho 500 em relação às de tamanho 1000. Comparando-se as médias gerais de proporções de acertos com os tamanhos das séries simuladas, nota-se que existe um crescimento nestas proporções quando se têm um tamanho maior da série a ser analisada. Estas informações são um indício de que o algoritmo parece ter performance melhor conforme for maior o tamanho da série analisada.

Por meio da tabela 1, nota-se proporções de acertos bem abaixo da média de proporções para séries de tamanho 100 para séries simuladas a partir de modelos MA(1) com valor de theta igual a 0,75 e modelos ARMA(1,1) com pares de parâmetros phi e theta (0.25, 0,25),(0,25, 0,5), (0,25, 0,75), (0,5, 0,75) e (0,75, 0,75). Em especial para o modelo ARMA(1,1) com pares de parâmetros (0,25, 0,75), observou-se uma acurácia de apenas 0,044.

Já para séries de tamanho 500, destacam-se por ter proporções com desvios grandes em relação à média global de acertos, para este tamanho de séries, os modelos ARMA(1,1) com pares de parâmetros phi e theta iguais a (0.25, 0,25) e (0,25, 0,75).

Para séries de tamanho 1000 nenhum modelo parece ter se destacado dos demais por apresentar menor acurácia na identificação das ordens do modelo por meio da função “auto.arima”.

CONCLUSÃO:

Com exceção dos modelos com parâmetro de médias móveis igual a 0,75 para séries de tamanho 100, onde a função auto.arima apresentou baixa acurácia na identificação das ordens do modelo ARMA em relação às séries simuladas, não se identificou claro destaque da baixa acurácia em algum modelo e/ou situação específica.

Contudo, de um modo geral a identificação das ordens dos modelos apresentados aqui por meio da função “auto.arima” não parece apresentar uma performance razoável, uma vez que mesmo para séries simuladas com 1000 observações, que é quantidade que em algumas situações difícil de se obter em séries reais, a função auto.arima obteve, em termos gerais, proporções de acerto por volta de 60 %. É difícil mensurar a acurácia de uma análise das ordens de um modelo ARMA por um humano, uma vez que é dependente da habilidade deste, porém considerando que a análise fosse feita por um indivíduo treinado para tal fim, acredito que a acurácia na identificação da ordem dos parâmetros seria, apesar de mais demorada, muito mais acurada do que a função auto.arima. Deste modo, o presente trabalho reforça a importância das análises criteriosas de um humano para a identificação do modelo adequado às observações de uma série temporal.