

Análise de Séries Temporais

Fernando Lucambio

Departamento de Estatística
Universidade Federal do Paraná

Novembro, 2020

Na previsão, o objetivo é prever valores futuros de uma série temporal X_{n+m} , $m = 1, 2, \dots$ com base nos dados coletados até o presente, $X_{1:n} = \{X_1, X_2, \dots, X_n\}$. Assumiremos que X_t é estacionário e os parâmetros do modelo são conhecidos.

O problema de previsão quando os parâmetros do modelo são desconhecidos será discutido na próxima seção. O preditor de erro quadrático médio mínimo de X_{n+m} é

$$X_{n+m}^n = E(X_{n+m} | X_{1:n}),$$

porque a esperança condicional minimiza o erro quadrático médio

$$E(X_{n+m} - g(X_{1:n}))^2,$$

onde $g(\cdot)$ é uma função das observações $X_{1:n}$.

Proposição III.3. Melhor previsão linear para processos estacionários.

Dados os dados X_1, X_2, \dots, X_n , o melhor preditor linear

$$X_{n+m}^n = \alpha_0 + \sum_{k=1}^n \alpha_k X_k,$$

de X_{n+m} para $m \geq 1$, é encontrado resolvendo

$$E((X_{n+m} - X_{n+m}^n)X_k) = 0, \quad k = 0, 1, \dots, n,$$

onde $X_0 = 1$, para $\alpha_0, \alpha_1, \dots, \alpha_n$.

Se $E(X_t) = \mu$, a primeira equação ($k=0$) na Proposição III.3 implica que

$$E(X_{n+m}^n) = E(X_{n+m}) = \mu.$$

Então, tomando esperança, temos

$$\mu = \alpha_0 + \sum_{k=1}^n \alpha_k \mu \quad \text{ou} \quad \alpha_0 = \mu \left(1 - \sum_{k=1}^n \alpha_k \right).$$

Assim, a forma melhor preditor linear é

$$X_{n+m}^n = \mu + \sum_{k=1}^n \alpha_k (X_k - \mu).$$

Assim, até discutirmos a estimação, não há perda de generalidade ao considerar o caso que $\mu = 0$, em cujo caso, $\alpha_0 = 0$.

Exemplo III.19. Previsão para o modelo $AR(2)$.

Suponhamos o processo $AR(2)$ causal, $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + W_t$ e uma observação X_t . Então, a previsão um passo à frente de X_2 baseada em X_1 é

$$x_2^1 = \frac{\gamma(1)}{\gamma(0)} X_1 = \rho(1) X_1.$$

Devido a que $\phi_1 X_2 + \phi_2 X_1$ satisfaz as equações de predição na Proposição III.3, verificamos que, para $n \geq 2$,

$$X_{n+1}^n = \phi_1 X_n + \phi_2 X_{n-1}.$$

Isto é, $\phi_{n,1} = \phi_1$, $\phi_{n,2} = \phi_2$ e $\phi_{n,j} = 0$, para $j = 3, 4, \dots, n$.

A partir do Exemplo III.19, deve ficar claro que, se a série temporal é um processo causal $AR(p)$, então, para $n \geq p$,

$$X_{n+1}^n = \phi_1 X_n + \phi_2 X_{n-1} + \dots + \phi_p X_{n-p+1}.$$

Para os modelos *ARMA* em geral, as equações de predição não serão tão simples quanto o caso *AR*. Além disso, para n grande, o uso de sistemas de equações é proibitivo porque requer a inversão de uma matriz grande.

Existem, no entanto, soluções iterativas que não exigem nenhuma inversão de matrizes. Em particular, mencionamos a solução recursiva devido a Levinson (1947) e Durbin (1960).

As equações gerais de predição fornecem pouca informação sobre previsão para modelos *ARMA* em geral. Existem várias maneiras diferentes de expressar essas previsões e cada uma delas ajuda a entender a estrutura especial da predição do *ARMA*.

Ao longo do tempo, assumimos que X_t é um processo causal e inversível *ARMA*(p, q), $\phi(B)X_t = \theta(B)W_t$, onde $W_t \sim N(0, \sigma_w^2)$ independentes. No caso de média não zero $E(X_t) = \mu_t$, simplesmente substitua X_t com $X_t - \mu_t$ no modelo. Primeiro, consideramos dois tipos de previsões. Escrevemos X_{n+m}^n para significar o preditor de erro quadrático médio mínimo de X_{n+m} com base nos dados $\{X_n, \dots, X_1\}$, isto é,

$$X_{n+m}^n = E(X_{n+m} | X_n, \dots, X_1).$$

Para modelos *ARMA*, é mais fácil calcular o preditor de X_{n+m} assumindo que temos a história completa do processo

$$\{X_n, X_{n-1}, \dots, X_1, X_0, X_{-1}, \dots\}.$$

Vamos denotar o preditor de X_{n+m} com base no passado infinito como

$$\tilde{X}_{n+m}^n = E(X_{n+m} \mid X_n, X_{n-1}, \dots, X_1, X_0, X_{-1}, \dots).$$

Em geral, X_{n+m}^n e \tilde{X}_{n+m}^n não são iguais, mas a ideia aqui é que, para grandes amostras, \tilde{X}_{n+m}^n proporcionará uma boa aproximação para X_{n+m}^n .

Exemplo III.23. Previsões de longo alcance.

Considere prever o processo *ARMA* com a média μ_x . Substituindo X_{n+m} por $X_{n+m} - \mu_x$ e considerando a esperança condicional, deduzimos que a previsão m -passos à frente pode ser escrita como

$$\tilde{X}_{n+m} = \mu_x + \sum_{j=m}^{\infty} \psi_j W_{n+m-j}.$$

Deve ficar claro que as previsões do *ARMA* se ajustam rapidamente à média com um erro de previsão constante à medida que o horizonte de previsão, m , cresce. Esse efeito pode ser visto na figura do Exemplo III.25, onde a série Recrutamento está prevista para 24 meses.

Proposição III.7. Previsão truncada para modelos ARMA.

Para modelos $ARMA(p, q)$, os preditores truncados são

$$\tilde{X}_{t+1}^n = \phi_1 \tilde{X}_{n+m-1}^n + \cdots + \phi_p \tilde{X}_{n+m-p}^n + \theta_1 \tilde{W}_{n+m-1}^n + \cdots + \theta_q \tilde{W}_{n+m-q}^n,$$

para $m = 1, 2, \dots$, onde $\tilde{X}_t^n = X_t$, para $1 \leq t \leq n$ e $\tilde{X}_t^n = 0$ para $t \leq 0$. Os erros de previsão truncados são dados por $\tilde{W}_t^n = 0$ para $t \leq 0$ ou $t > n$ e

$$\tilde{W}_t^n = \phi(B)\tilde{X}_t^n - \theta_1 \tilde{W}_{t-1}^n - \cdots - \theta_q \tilde{W}_{t-q}^n,$$

para $1 \leq t \leq n$.

Exemplo III.24. Previsões de uma série ARMA(1, 1).

Dados dados X_1, \dots, X_n , para fins de previsão, escrevamos o modelo como

$$X_{n+1} = \phi X_n + W_{n+1} + \theta W_n.$$

Então, a previsão truncada de um passo à frente é

$$\tilde{X}_{n+1}^n = \phi X_n + 0 + \theta \tilde{W}_n^n.$$

Para $m \geq 2$, temos

$$\tilde{X}_{n+m}^n = \phi \tilde{X}_{n+m-1}^n,$$

que pode ser calculado recursivamente, $m = 2, 3, \dots$.

Para avaliar a precisão das previsões, os intervalos de previsão são normalmente calculados junto com as previsões. Em geral, os intervalos de previsão com probabilidade de cobertura $1 - \alpha$, são da forma

$$X_{n+m}^n \pm c_{\alpha/2} \sqrt{P_{n+m}^n},$$

onde $c_{\alpha/2}$ é escolhido para obter o grau de confiança desejado.

Por exemplo, se o processo for gaussiano, escolher $c_{\alpha/2} = 2$ produz um intervalo de previsão de aproximadamente 95% para X_{n+m} . Se o interesse é envolver os intervalos de previsão para o longo de um período de tempo, então ele deve ser convenientemente ajustado, por exemplo, usando a desigualdade de Bonferroni (ver Johnson e Wichern, 1992, Capítulo 5).

Exemplo III.25. Previsões da série de Recrutamento.

Usando as estimativas dos parâmetros como os valores reais dos parâmetros, a figura abaixo mostra o resultado da previsão da série de Recrutamentos fornecida no Exemplo III.18 em um horizonte de 24 meses, $m = 1, 2, \dots, 24$.

As previsões reais são calculadas como

$$X_{n+m}^n = 6.74 + 1.35X_{n+m-1}^n - 0.46X_{n+m-2}^n,$$

para $n = 453$ e $m = 1, 2, \dots, 12$. Recordemos que $X_t^s = X_t$ quando $t \leq s$.

Os erros de previsão P_{n+m}^n são calculados usando

$$P_{n+m}^n = \sigma_w^2 \sum_{j=0}^{m-1} \psi_j^2.$$

Exemplo III.25. Previsões da série de Recrutamento.

Encontramos que $\hat{\sigma}_w^2 = 89.72$ e do Exemplo III.12, temos que

$$\psi_j = 1.35\psi_{j-1} - 0.46\psi_{j-2}$$

para $j \geq 2$, onde $j \geq 2$ sendo $\psi_0 = 1$ e $\psi_1 = 1.35$. Portanto, para $n = 453$,

$$\begin{aligned} P_{n+1}^n &= 89.72, \\ P_{n+2}^n &= 89.72(1 + 1.35^2), \\ P_{n+3}^n &= 89.72(1 + 1.35^2 + (1.35^2 - 0.46^2)), \end{aligned}$$

e assim por diante.

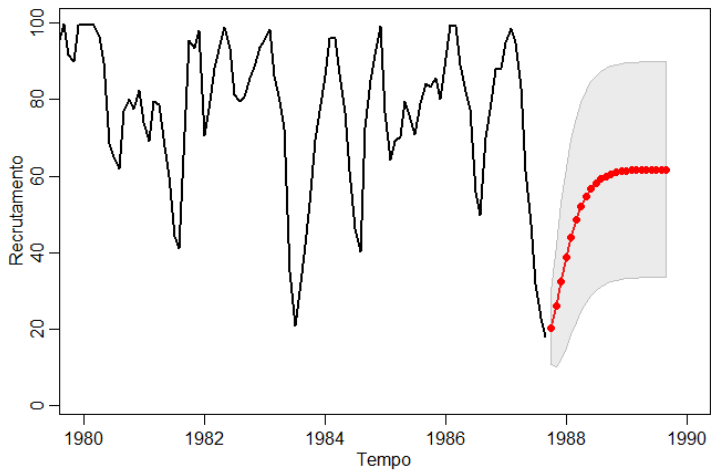
Observe como a previsão se estabiliza rapidamente e os intervalos de previsão são amplos, embora neste caso os limites de previsão sejam baseados apenas em um erro padrão; isto é,

$$X_{n+m}^n \pm \sqrt{P_{n+m}^n}.$$

Exemplo III.25. Previsões da série de Recrutamento.

Para reproduzir a análise e a figura, use os seguintes comandos:

```
> regr = ar.ols(rec, order=2, demean=FALSE, intercept=TRUE)
> fore = predict(regr, n.ahead=24)
> par(mfrow = c(1,1),mar=c(4,3,1,1),mgp=c(1.6,.6,0), pch=19)
> ts.plot(rec, fore$pred, col=1:2, xlim=c(1980,1990),
          lwd=2, ylab="Recrutamento", xlab="Tempo")
> U = fore$pred+fore$se; L = fore$pred-fore$se
> xx = c(time(U), rev(time(U))); yy = c(L, rev(U))
> polygon(xx, yy, border = 8, col = gray(.6, alpha = .2))
> lines(fore$pred, type="p", col=2)
```



Ao longo desta seção, assumimos que temos n observações, X_1, \dots, X_n , a partir de um processo $ARMA(p, q)$ gaussiano, no qual, inicialmente, os parâmetros de ordem p e q são conhecidos.

Nosso objetivo é estimar os parâmetros $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ e σ_w^2 . Discutiremos o problema de determinar p e q mais adiante nesta seção.

Começamos com o método dos estimadores de momentos. A ideia por trás desses estimadores é a de igualar os momentos da população aos momentos da amostra e depois resolver os parâmetros em termos dos momentos da amostra. Vemos imediatamente que, se $E(X_t) = \mu$, então o método de estimador de momentos é a média amostral \bar{X} . Assim, enquanto se discute o método dos momentos, assumimos $\mu = 0$.

Primeiro, consideramos o caso em que o método leva a estimadores ótimos (eficientes), isto é, modelos $AR(p)$,

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + W_t,$$

onde as primeiras $p + 1$ equações homogêneas gerais levam à seguinte definição.

Definição III.10. Equações de Yule-Walker.

As equações de Yule-Walker são dadas por

$$\begin{aligned} \gamma(h) &= \phi_1 \gamma(h-1) + \cdots + \phi_p \gamma(h-p), & h = 1, 2, \dots, p, \\ \sigma_w^2 &= \gamma(0) - \phi_1 \gamma(1) - \cdots - \phi_p \gamma(p). \end{aligned}$$

Embora o método dos momentos possa produzir bons estimadores, eles podem levar a estimadores sub-ótimos.

Para os modelos $AR(p)$, se o tamanho da amostra for grande, os estimadores de Yule-Walker são aproximadamente normalmente distribuídos e $\hat{\sigma}_w^2$ está próximo ao valor real de σ_w^2 .

Proposição III.8. Resultados em amostras grandes para os estimadores de Yule-Walker.

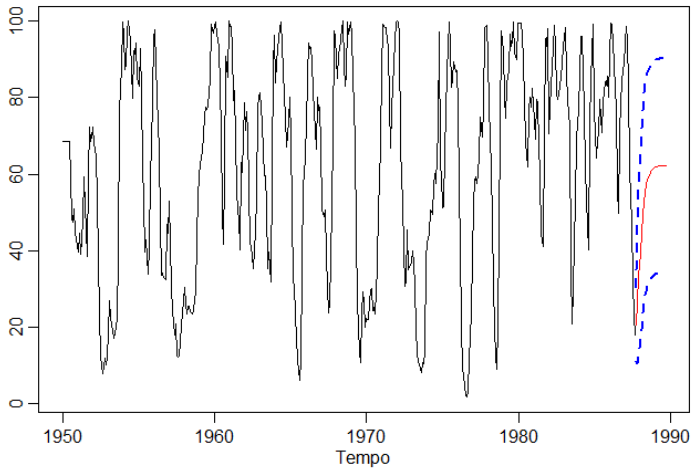
O comportamento assintótico, ou seja, para $n \rightarrow \infty$, dos estimadores de Yule-Walker no caso de processos $AR(p)$ é o seguinte:

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{D} N(0, \sigma_w^2 \Gamma_p^{-1}), \quad \tilde{\sigma}_w^2 \xrightarrow{P} \sigma_w^2.$$

Exemplo III.28. Estimadores de Yule-Walker para a série de Recrutamento.

No Exemplo III.18, ajustamos um modelo $AR(2)$ à série de Recrutamento usando mínimos quadrados ordinários (OLS). Para os modelos AR , os estimadores obtidos via OLS e Yule-Walker são quase idênticos; veremos isso quando discutirmos a estimação de soma condicional de quadrados.

```
> rec.yw = ar.yw(rec, order=2)
> rec.yw$x.mean # = 62.26 (média estimada)
[1] 62.26278
> rec.yw$ar # = 1.33, -.44 (estimativas de coeficiente)
[1] 1.3315874 -0.4445447
> sqrt(diag(rec.yw$asy.var.coef)) # = .04, .04 (erros padrão)
[1] 0.04222637 0.04222637
> rec.yw$var.pred # = 94.80 (estimativa da variância do erro)
[1] 94.79912
```



No caso dos modelos $AR(p)$, os estimadores Yule-Walker são ótimos no sentido de que a distribuição assintótica é a melhor distribuição normal assintótica.

Isso porque, dadas as condições iniciais, os modelos $AR(p)$ são modelos lineares e os estimadores Yule-Walker são essencialmente estimadores de mínimos quadrados.

Se usarmos o método dos momentos para os modelos MA ou $ARMA$, não obteremos estimadores ótimos porque tais processos não são lineares nos parâmetros.

No caso do modelo $AR(1)$, podemos escrever a verossimilhança como

$$L(\mu, \phi, \sigma_w^2) = f(X_1)f(X_2|X_1) \cdots f(X_n|X_{n-1}),$$

onde deixamos de escrever os parâmetros nas densidades, $f(\cdot)$, para facilitar a notação.

Dado que $X_t|X_{t-1} \sim N(\mu + \phi(X_{t-1} - \mu), \sigma_w^2)$, temos

$$f(X_t|X_{t-1}) = f_W((X_t - \mu) - \phi(X_{t-1} - \mu)),$$

onde $f_W(\cdot)$ é a função de densidade de W_t , isto é, a densidade normal com média zero e variância σ_w^2 .

Podemos então escrever a verossimilhança como

$$L(\mu, \phi, \sigma_w^2) = f(X_1) \prod_{t=2}^n f_W((X_t - \mu) - \phi(X_{t-1} - \mu)).$$

Ou estimadores de Yule-Walker e os estimadores de mínimos quadrados condicionais são aproximadamente os mesmos.

Para modelos *ARMA* gerais, é difícil escrever a verossimilhança como uma função explícita dos parâmetros. Em vez disso, é vantajoso escrever a verossimilhança em termos das inovações ou erros de previsão em um passo à frente $X_t - X_t^{t-1}$.

Para um modelo *ARMA*(p, q) normal, seja

$$\beta = (\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)^\top$$

o vetor de dimensão $p + q + 1$ dos parâmetros do modelo. A verossimilhança pode ser escrita como

$$L(\beta, \sigma_w^2) = \prod_{t=1}^n f(X_t | X_{t-1}, \dots, X_1).$$

A distribuição condicional de X_t dado X_{t-1}, \dots, X_1 é gaussiana com média X_t^{t-1} e variância $P_t^{t-1} = \gamma(0) \prod_{j=1}^{t-1} (1 - \phi_{jj}^2)$.

Exemplo III.30. Os algoritmos de Newton–Raphson e de escore.

Duas rotinas comuns de otimização numérica para realizar a estimação de máxima verossimilhança são Newton–Raphson e escore. Vamos dar um breve relato das ideias matemáticas aqui. A implementação real desses algoritmos é muito mais complicada do que nossa discussão pode implicar. Para detalhes, o leitor é encaminhado para, por exemplo, Press et al. (1993).

Seja $\ell(\beta)$ a função de critério dos k parâmetros $\beta = (\beta_1, \dots, \beta_k)$ que queremos minimizar em relação a β . Por exemplo, considere alguma das funções de verossimilhança dadas acima. Suponha que $\ell(\hat{\beta})$ é o extremo que estamos interessados em encontrar e $\hat{\beta}$ é encontrado resolvendo $\partial\ell(\beta)/\partial\beta_j = 0$, para $j = 1, \dots, k$.

Exemplo III.30. Os algoritmos de Newton–Raphson e de escore.

Seja $\ell^{(1)}(\beta)$ o vetor $k \times 1$ de derivadas parciais

$$\ell^{(1)}(\beta) = \left(\frac{\partial \ell}{\partial \beta_1}, \dots, \frac{\partial \ell}{\partial \beta_k} \right)^\top.$$

Observe que, $\ell^{(1)}(\hat{\beta}) = 0$, o vetor $k \times 1$ de zeros. Seja $\ell^{(2)}(\beta)$ a matriz $k \times k$ das derivadas parciais de segundo ordem

$$\ell^{(2)}(\beta) = \left(-\frac{\partial^2 \ell(\beta)}{\partial \beta_i \partial \beta_j} \right)_{i,j=1}^k,$$

não singular. Seja $\beta_{(0)}$ um estimador inicial suficientemente bom de β . Então, usando a expansão de Taylor, temos:

$$0 = \ell^{(1)}(\hat{\beta}) \approx \ell^{(1)}(\beta_{(0)}) - \ell^{(2)}(\beta_{(0))}(\hat{\beta} - \beta_{(0)}).$$

Exemplo III.30. Os algoritmos de Newton–Raphson e de escore.

Configurando o lado direito igual a zero e resolvendo por $\hat{\beta}$, chame a solução de $\beta_{(1)}$, obtemos

$$\beta_{(1)} = \beta_{(0)} + \left(\ell^{(2)}(\beta_{(0)}) \right)^{-1} \ell^{(1)}(\beta_{(0)}).$$

O algoritmo de Newton–Raphson procede iterando esse resultado, substituindo $\beta_{(0)}$ por $\beta_{(1)}$ para obter $\beta_{(2)}$, e assim por diante, até a convergência. Sob um conjunto de condições apropriadas, a sequência de estimadores, $\beta_{(1)}, \beta_{(2)}, \dots$ converge para $\hat{\beta}$.

No método de escore, substituímos $\ell^{(2)}(\beta)$ por $E(\ell^{(2)}(\beta))$, a matriz de informação. Sob condições apropriadas, o inverso da matriz de informação é a matriz de variâncias e covariâncias assintótica do estimador $\hat{\beta}$.

Exemplo III.31. Estimação por máxima verossimilhança para a série de Recrutamento.

Até agora, ajustamos um modelo $AR(2)$ à série Recrutamento usando mínimos quadrados ordinários (Exemplo III.18) e usando os estimadores de Yule-Walker (Exemplo III.28).

A seguir, uma sessão R usada para ajustar um modelo $AR(2)$ via estimação de máxima verossimilhança à série de Recrutamento; estes resultados podem ser comparados com os resultados do Exemplo 3.18 e do Exemplo III.28.

Exemplo III.31. Estimação por máxima verossimilhança para a série de Recrutamento.

```
> rec.mle = ar.mle(rec, order=2)
> rec.mle$x.mean # 62.26
[1] 62.26153
> rec.mle$ar # 1.35, -0.46
[1] 1.3512809 -0.4612736
> sqrt(diag(rec.mle$asy.var.coef)) # 0.04, 0.04
[1] 0.04099159 0.04099159
> rec.mle$var.pred # 89.34
[1] 89.33597
```

Proposição III.10. Distribuição em amostras grandes dos estimadores do processo *ARMA*.

Sob condições apropriadas, para processos *ARMA*, os estimadores de máxima verossimilhança, mínimos quadrados incondicionais e mínimos quadrados condicionais, cada um iniciado pelo método de momentos, todos fornecem estimadores ótimos de σ_w^2 e β , no sentido de que $\hat{\sigma}_w^2$ é consistente e a distribuição assintótica de $\hat{\beta}$ é a melhor distribuição normal assintótica. Em particular, quando $n \rightarrow \infty$

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} N(0, \sigma_w^2 \Gamma_{p,q}^{-1}).$$

A matriz de variâncias e covariâncias assintótica do estimador $\hat{\beta}$ é o inverso da matriz de informação.

Exemplo III.34. Algumas distribuições assintóticas específicas.

A seguir estão alguns casos específicos da Proposição III.10.

- ▶ AR(1): $\gamma_x(0) = \frac{\sigma_w^2}{1 - \phi^2}$, assim $\sigma_w^2 \Gamma_{1,0}^{-1} = 1 - \phi^2$. Então,

$$\hat{\phi} \sim N\left(\phi, \frac{1 - \phi^2}{n}\right).$$

- ▶ AR(2): Pode-se verificar que

$$\gamma_x(0) = \left(\frac{1 - \phi_2}{1 + \phi_2}\right) \frac{\sigma_w^2}{(1 - \phi_2)^2 - \phi_1^2},$$

e que $\gamma_x(1) = \phi_1 \gamma_x(0) + \phi_2 \gamma_x(1)$. Em particular, temos

$$\begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \frac{1}{n} \begin{pmatrix} 1 - \phi_2^2 & -\phi_1(1 + \phi_2) \\ -\phi_1(1 + \phi_2) & 1 - \phi_2^2 \end{pmatrix}\right).$$

Exemplo III.34. Algumas distribuições assintóticas específicas.

A seguir estão alguns casos específicos da Proposição III.10.

- ▶ **MA(1):** Neste caso escrevemos $\theta(B)Y_t = W_t$ ou $Y_t + \theta Y_{t-1} = W_t$. Então, análogo ao caso AR(1) temos que, $\gamma_Y(0) = \frac{\sigma_w^2}{1 - \theta^2}$, deste modo $\sigma_w^2 \Gamma_{1,0}^{-1} = 1 - \theta^2$. Portanto,

$$\hat{\theta} \sim N\left(\theta, \frac{1 - \theta^2}{n}\right).$$

- ▶ **MA(2):** Escrevendo $Y_t + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} = W_t$, então, análogo ao caso AR(2), temos

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \frac{1}{n} \begin{pmatrix} 1 - \theta_2^2 & -\theta_1(1 + \theta_2) \\ -\theta_1(1 + \theta_2) & 1 - \theta_2^2 \end{pmatrix}\right).$$

Exemplo III.34. Algumas distribuições assintóticas específicas.

A seguir estão alguns casos específicos da Proposição III.10.

- ▶ **ARMA(1, 1):** Para calcular $\Gamma_{\phi, \theta}$, devemos encontrar $\gamma_{XY}(0)$, onde $X_t - \phi X_{t-1} = W_t$ e $Y_t + \theta Y_{t-1} = W_t$. Temos então

$$\gamma_{XY}(0) = \text{Cov}(X_t, Y_t) = -\phi\theta\gamma_{XY}(0) + \sigma_w^2.$$

Resolvendo, encontramos $\gamma_{XY}(0) = \frac{\sigma_w^2}{1 + \phi\theta}$. Então

$$\begin{pmatrix} \hat{\phi} \\ \hat{\theta} \end{pmatrix} \sim N \left(\begin{pmatrix} \phi \\ \theta \end{pmatrix}, \frac{1}{n} \begin{pmatrix} \frac{1}{1 - \phi^2} & \frac{1}{1 + \phi\theta} \\ \frac{1}{1 + \phi\theta} & \frac{1}{1 - \theta^2} \end{pmatrix}^{-1} \right).$$