Previsão

Ricardo Ehlers
ehlers@icmc.usp.br

Departamento de Matemática Aplicada e Estatística
Universidade de São Paulo

Forecasting: principles and practice (available at `https://www.otexts.org/fpp`). A book by Rob J Hyndman (`http://robjhyndman.com`) and George Athanasopoulos

Previsão pontual k passos à frente,

$$\begin{aligned}\hat{x}_t(k) &= E(X_{t+k}|x_t, x_{t-1}, \dots) \\ &= E(X_{t+k}|x_t, x_{t-1}, \dots, x_1)\end{aligned}$$

Erro de previsão k passos à frente,

$$e_{t+k} = \hat{x}_t(k) - x_{t+k}.$$

Exemplo. Holt-Winters com sazonalidade multiplicativa aplicado à série `AirPassengers` (totais mensais de passageiros em linhas aéreas internacionais nos EUA entre 1949 e 1960). Função `HoltWinters()`.

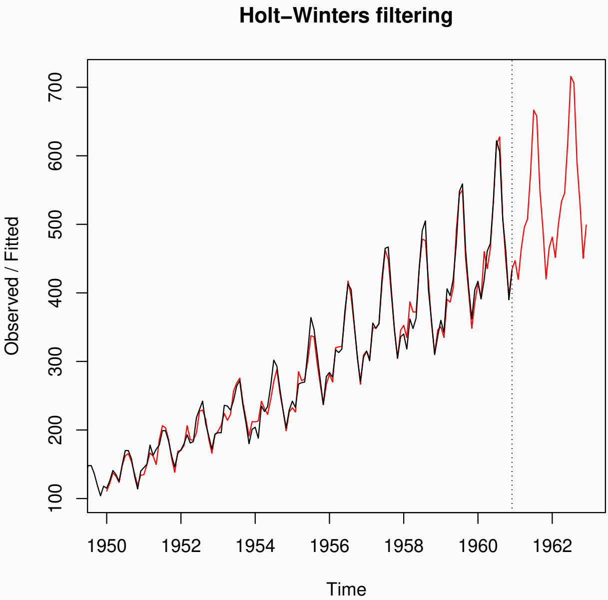
Lembrando a equação de previsão k períodos à frente,
$\hat{x}_t(k) = (L_t + kT_t)\, I_{t-12+k}, \quad k = 1, 2, \ldots.$

```
> data(AirPassengers)
> m = HoltWinters(AirPassengers, seasonal = "mult")
```

Previsões para os anos de 1961 e 1962.

	Jan	Feb	Mar	Apr	May	Jun
1961	447.0559	419.7123	464.8671	496.0839	507.5326	575.4509
1962	481.3732	451.7258	500.1008	533.4477	545.5202	618.2550

	Jul	Aug	Sep	Oct	Nov	Dec
1961	666.5923	657.9137	550.3088	492.9853	420.2073	465.6345
1962	715.8704	706.2524	590.4954	528.7681	450.5242	499.0281


Holt−Winters filtering
Observed / Fitted
Time
100
200
300
400
500
600
700
1950
1952
1954
1956
1958
1960
1962

Intervalos de Predição

Para construir intervalos para as predições pontuais há 2 possiveis abordagens,

1. assumir normalidade (função `HoltWinters()`),

$$x_t = L_{t-1} + T_{t-1} + I_{t-12} + \epsilon_t,\ \epsilon_t \sim N(0, \sigma^2),$$

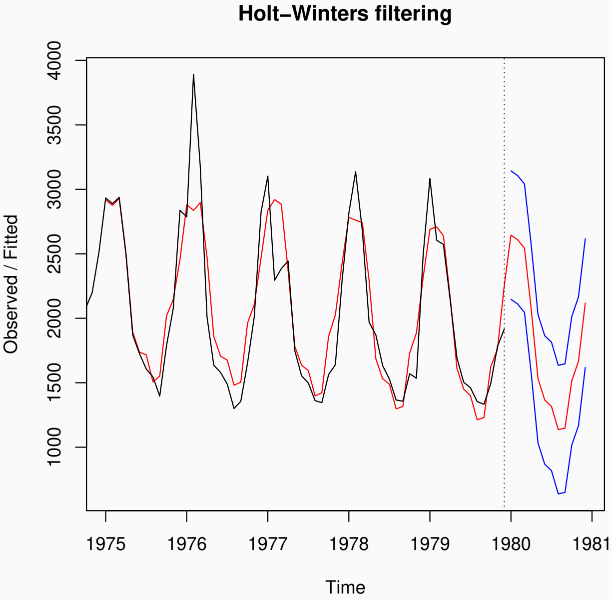
2. usar bootstrap (pacote `forecast`)

Exemplo. Números mensais de mortes por doenças do pulmão (bronquite, efisema e asma) no Reino Unido entre janeiro de 1974 e dezembro de 1979 (variável `UKLungDeaths`).

```
> data(UKLungDeaths)
> m = HoltWinters(ldeaths, seasonal="addit")
> p = predict(m, n.ahead=12, prediction.interval = T)
```

```
              fit      upr       lwr
Jan 1980 2645.163 3143.119 2147.2079
Feb 1980 2608.152 3106.124 2110.1814
Mar 1980 2543.142 3041.141 2045.1437
Apr 1980 2070.752 2568.793 1572.7115
May 1980 1531.724 2029.826 1033.6215
Jun 1980 1366.184 1864.369  867.9990
Jul 1980 1316.317 1814.610  818.0230
Aug 1980 1136.551 1634.982  638.1205
Sep 1980 1148.062 1646.662  649.4627
Oct 1980 1514.505 2013.309 1015.7009
Nov 1980 1666.974 2166.021 1167.9274
Dec 1980 2117.777 2617.109 1618.4455
```

Previsões para 1980/1981. Intervalos de 95% supondo normalidade.

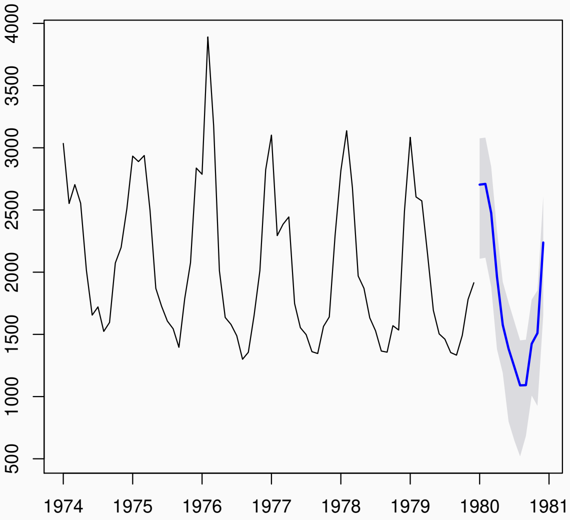


Exemplo. No exemplo anterior, obter intervalos de 95% via bootstrap.

```
> data(UKLungDeaths)
> library(forecast)
> fcast=hw(ldeaths,h=12,level=95,seasonal="additive",bootstrap=T
+          simulate=T)
```

	Point Forecast	Lo 95	Hi 95
Jan 1980	2703.659	2108.1143	3076.183
Feb 1980	2709.518	2116.0733	3081.600
Mar 1980	2475.210	1882.3329	2847.640
Apr 1980	1962.715	1381.5810	2333.469
May 1980	1576.509	1190.2172	1922.497
Jun 1980	1385.025	800.8241	1756.660
Jul 1980	1238.803	647.5784	1604.118
Aug 1980	1090.090	519.2287	1451.997
Sep 1980	1091.898	683.2737	1457.108
Oct 1980	1424.133	1008.6343	1779.182
Nov 1980	1510.164	925.0558	1853.390
Dec 1980	2238.539	1652.6972	2610.691

Previsões para 1980/1981. Intervalos de 95% via bootstrap.



Previsão em Modelos ARIMA

- As previsões podem ser obtidas usando-se diretamente a equação do modelo.
- Substituir $\epsilon_{n+1}, \epsilon_{n+2}, \ldots$ pela sua esperança condicional,
- Substituir $X_{n+1}, X_{n+2}, \ldots$ pela sua esperança condicional,
- Substituir valores passados de X e de ϵ pelos seus valores observados.

Exemplo. Modelo SARIMA$(1,0,0)\times(0,1,1)_{12}$. A equação do modelo é dada por,

$$(1-\alpha B)(1-B^{12})X_t = (1+\theta B^{12})\epsilon_t$$

ou equivalentemente,

$$X_t = X_{t-12} + \alpha(X_{t-1} - X_{t-13}) + \epsilon_t + \theta\epsilon_{t-12}.$$

As previsões 1 e 2 passos à frente ficam,

$$\begin{aligned}
\hat{x}_n(1) &= x_{n-11} + \alpha(x_n - x_{n-12}) + \theta\epsilon_{n-11} \\
\hat{x}_n(2) &= x_{n-10} + \alpha(\hat{x}_n(1) - x_{n-11}) + \theta\epsilon_{n-10}.
\end{aligned}$$

- Note que o valor futuro X_{n+1} foi substituído na segunda equação pela sua esperança condicional $\hat{x}_n(1)$ (previsão feita em $t = n$ para $t = n + 1$).

- Previsões para horizontes maiores podem ser obtidas recursivamente. Por exemplo,

$$\hat{x}_n(3) = x_{n-9} + \alpha(\hat{x}_n(2) - x_{n-10}) + \theta\epsilon_{n-9}.$$

Modelos autoregressivos

Modelo AR(p) em $t+k$,

$$X_{t+k} \;=\; \alpha_1 X_{t+k-1} + \cdots + \alpha_p X_{t+k-p} + \epsilon_{t+k}$$

Previsões,

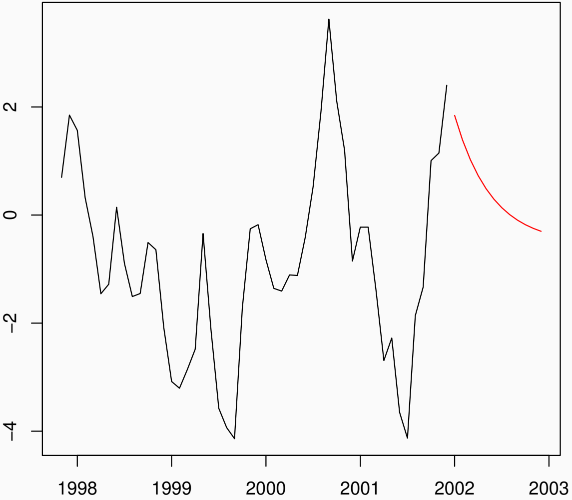
$$\begin{aligned}
\hat{x}_t(1) &= \alpha_1 x_t + \cdots + \alpha_p x_{t-p+1} \\
\hat{x}_t(2) &= \alpha_1 \hat{x}_t(1) + \cdots + \alpha_p x_{t-p+2} \\
&\vdots \\
\hat{x}_t(p+1) &= \alpha_1 \hat{x}_t(p) + \cdots + \alpha_p \hat{x}_t(1)
\end{aligned}$$

de modo que as previsões para horizontes $k > p$ usam apenas as previsões anteriores.

Exemplo. No modelo AR(1) segue que,

$$
\begin{aligned}
\hat{x}_t(k) &= \alpha \hat{x}_t(k-1) \\
&= \alpha^2 \hat{x}_t(k-2) \\
&\vdots \\
&= \alpha^k x_t
\end{aligned}
$$

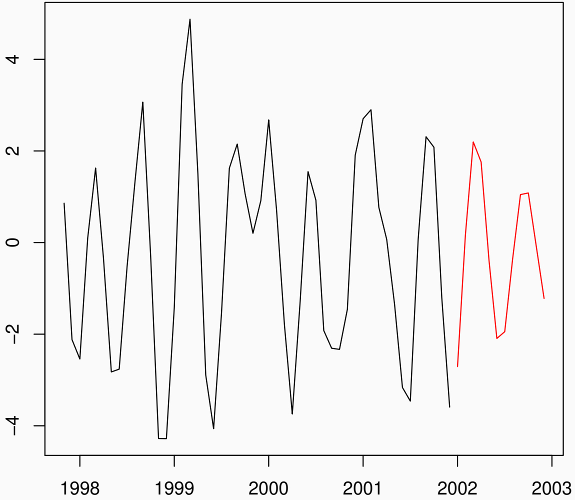
Modelo AR(1) mensal simulado com $n = 50$ e parâmetros $\alpha = 0.8$ e $\sigma^2_\epsilon = 1$. Previsões até 12 passos a frente.



Exemplo. No modelo AR(2) segue que,

$$\begin{aligned}
\hat{x}_t(1) &= \alpha_1 x_t + \alpha_2 x_{t-1} \\
\hat{x}_t(2) &= \alpha_1 \hat{x}_t(1) + \alpha_2 x_t \\
\\
\hat{x}_t(k) &= \alpha_1 \hat{x}_t(k-1) + \alpha_2 \hat{x}_t(k-2), \;\; k = 3, 4, \ldots
\end{aligned}$$

Modelo AR(2) mensal simulado com $n = 50$ e parâmetros $\alpha_1 = 0.8$, $\alpha_2 = -0.7$ e $\sigma_\epsilon^2 = 1$. Previsões até 12 passos a frente.



Previsão em Modelos MA

Modelo MA(q) com $E(\epsilon_t) = 0$,

$$
\begin{aligned}
X_{t+1} &= \epsilon_{t+1} + \beta_1 \epsilon_t + \cdots + \beta_q \epsilon_{t-q+1} \\
X_{t+2} &= \epsilon_{t+2} + \beta_1 \epsilon_{t+1} + \cdots + \beta_q \epsilon_{t-q+2}
\end{aligned}
$$

$$
\begin{aligned}
\hat{x}_t(1) &= \beta_1 \epsilon_t + \cdots + \beta_q \epsilon_{t-q+1} \\
\hat{x}_t(2) &= \beta_2 \epsilon_t + \cdots + \beta_q \epsilon_{t-q+2} \\
&\vdots \\
\hat{x}_t(q) &= \beta_q \epsilon_t \\
\hat{x}_t(q+j) &= 0, \quad j = 1, 2, \dots
\end{aligned}
$$

ou seja,

$$
\hat{x}_t(k) = \begin{cases} \sum_{i=k}^{q} \beta_i \epsilon_{t+k-i}, & k = 1, \dots, q \\ 0, & k > q \end{cases}
$$

Atualizando as Previsões

Para o modelo SARIMA$(1,0,0)\times(0,1,1)_{12}$ a previsão fica

$$\begin{aligned}\hat{x}_{n+1}(1) &= E(X_{n+2}|x_{n+1},\dots,x_1)\\ &= x_{n-10}+\alpha(x_{n+1}-x_{n-11})+\theta\epsilon_{n-10}. \end{aligned}\tag{1}$$

Somando e subtraindo $\alpha\hat{x}_n(1)$ no lado direito obtemos que

$$\begin{aligned}\hat{x}_{n+1}(1) &= x_{n-10}+\alpha(\hat{x}_n(1)-x_{n-11})+\alpha(x_{n+1}-\hat{x}_n(1))+\theta\epsilon_{n-10}\\ &= \hat{x}_n(2)+\alpha(x_{n+1}-\hat{x}_n(1))\end{aligned}$$

ou seja, a previsão atualizada é a previsão anterior mais uma proporção do erro de previsão 1 passo à frente em $t=n+1$.

Previsões usando a forma MA

Processo MA(∞),

$$\begin{aligned} X_t &= \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + \dots \\ X_{n+k} &= \epsilon_{n+k} + \psi_1 \epsilon_{n+k-1} + \dots + \psi_k \epsilon_n + \psi_{k+1} \epsilon_{n-1} + \dots \end{aligned}$$

Previsão k passos à frente,

$$\hat{x}_n(k) = \psi_k \epsilon_n + \psi_{k+1} \epsilon_{n-1} + \dots$$

Erro de previsão k passos à frente,

$$\begin{aligned} e_{n+k} &= x_{n+k} - \hat{x}_n(k) \\ &= \epsilon_{n+k} + \psi_1 \epsilon_{n+k-1} + \cdots + \psi_{k-1}\epsilon_{n+1} + \psi_k \epsilon_n + \dots \\ &\quad -(\psi_k \epsilon_n + \psi_{k+1}\epsilon_{n-1} + \dots) \\ &= \epsilon_{n+k} + \psi_1 \epsilon_{n+k-1} + \cdots + \psi_{k-1}\epsilon_{n+1} \end{aligned}$$

Variância do erro de previsão,

$$Var(e_{n+k}) = (1 + \psi_1^2 + \cdots + \psi_{k-1}^2)\sigma_\epsilon^2.$$

Assumindo que $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ o intervalo de previsão $100(1-\alpha)\%$ fica,

$$\hat{x}_t(k) \pm z_{\alpha/2}\ \sigma_\epsilon \sqrt{1 + \psi_1^2 + \cdots + \psi_{k-1}^2}.$$

- Para σ_ϵ^2 fixo, a variância do erro de previsão aumenta com o horizonte de previsão.
- Mais confiança em previsões de curto prazo.

- Na prática os parâmetros são desconhecidos e serão usadas suas estimativas.
- Valores passados dos ϵ_t serão estimados como erros de previsão.
- Consequentemente os intervalos de previsão serão aproximados.
- A hipótese de normalidade deve ser checada.

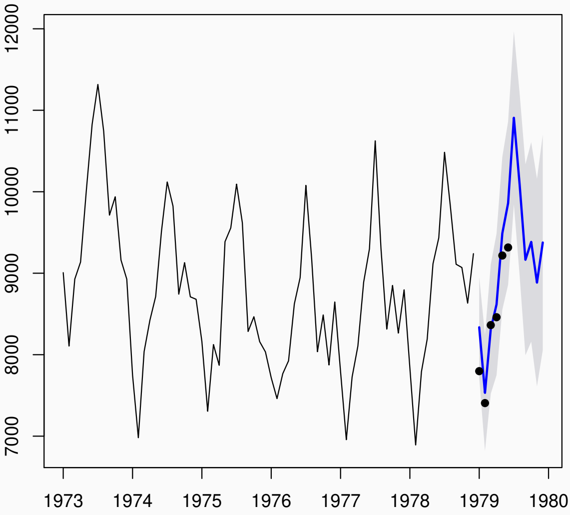
Exemplo. Totais mensais de mortes por acidente nos Estados Unidos entre janeiro de 1973 e dezembro de 1978 (`USAccDeaths`).

Foi ajustado um modelo SARIMA$(0, 1, 1) \times (0, 1, 1)_{12}$ com erros normais e feitas previsões para o ano de 1979.

```
> fit = arima(USAccDeaths,order = c(0,1,1),
+              seasonal = list(order=c(0,1,1)))
> library(forecast)
> forecast(fit, h=12,level=95)

         Point Forecast    Lo 95     Hi 95
Jan 1979       8336.061 7717.794  8954.328
Feb 1979       7531.829 6820.351  8243.307
Mar 1979       8314.644 7520.825  9108.462
Apr 1979       8616.869 7748.483  9485.255
May 1979       9488.913 8551.874 10425.951
Jun 1979       9859.757 8858.764 10860.751
Jul 1979      10907.470 9846.369 11968.571
Aug 1979      10086.508 8968.527 11204.489
Sep 1979       9164.959 7992.855 10337.064
Oct 1979       9384.259 8160.423 10608.096
Nov 1979       8884.974 7611.505 10158.443
Dec 1979       9376.574 8055.336 10697.811
```

Forecasts from ARIMA(0,1,1)(0,1,1)[12]



Previsão Bayesiana

- Nas previsões ARIMA substitui-se os parâmetros por suas estimativas pontuais.
- Não conhecermos os valores dos parâmetros é mais uma fonte de incerteza em relação as previsões.
- Esta incerteza pode ser muito grande para ser ignorada.

No contexto Bayesiano a distribuição preditiva de X_{n+k} é dada por,

$$p(x_{n+k}|\mathbf{x}) = \int p(x_{n+k}|\mathbf{x}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}.$$

Neste caso, todos os possíveis valores de $\boldsymbol{\theta}$ estão sendo levados em conta e não apenas a sua estimativa pontual.

Exemplo. No modelo AR(1),

$$x_t = \alpha x_{t-1} + \epsilon_t,\ \epsilon_t \sim N(0, \sigma_\epsilon^2)$$

temos que $\boldsymbol{\theta} = (\alpha, \sigma_\epsilon^2)$ e a distribuição preditiva 1 passo à frente é,

$$p(x_{n+1}|\mathbf{x}) = \int p(x_{n+1}|x_n, \alpha, \sigma_\epsilon^2)\ p(\alpha, \sigma_\epsilon^2|\mathbf{x}) d\alpha d\sigma_\epsilon^2.$$

Suponha que temos uma amostra de N valores simulados

$$(\alpha^{(1)}, \sigma^{2(1)}), \dots, (\alpha^{(N)}, \sigma^{2(N)}),$$

da distribuição a posteriori.

1. Simule valores

$$x_{n+1}^{(j)}|x_n, \alpha^{(j)}, \sigma^{2(j)} \sim N(\alpha^{(j)}x_n, \sigma^{2(j)}),\ j = 1, \ldots, N.$$

2. Para $k = 2, 3, \ldots$, simule valores

$$\begin{aligned} x_{n+2}^{(j)}|x_{n+1}^{(j)}, \alpha^{(j)}, \sigma^{2(j)} &\sim N(\alpha^{(j)}x_{n+1}^{(j)}, \sigma^{2(j)}),\ j = 1, \ldots, N \\ x_{n+3}^{(j)}|x_{n+2}^{(j)}, \alpha^{(j)}, \sigma^{2(j)} &\sim N(\alpha^{(j)}x_{n+2}^{(j)}, \sigma^{2(j)}),\ j = 1, \ldots, N \\ &\vdots \end{aligned}$$

3. Tomando a previsão pontual como a média da distribuição preditiva,

$$E(X_{n+k}|\mathbf{x}) \approx \frac{1}{N}\sum_{j=1}^{N} x_{n+k}^{(j)},\ k = 1, 2, \ldots$$

Exemplo. Foram gerados 200 valores de um processo AR(1) com parâmetros $\alpha = 0.8$ e $\sigma^2_\epsilon = 1$. Os parâmetros foram estimados por máxima verossimilhança e abordagem Bayesiana sendo,

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

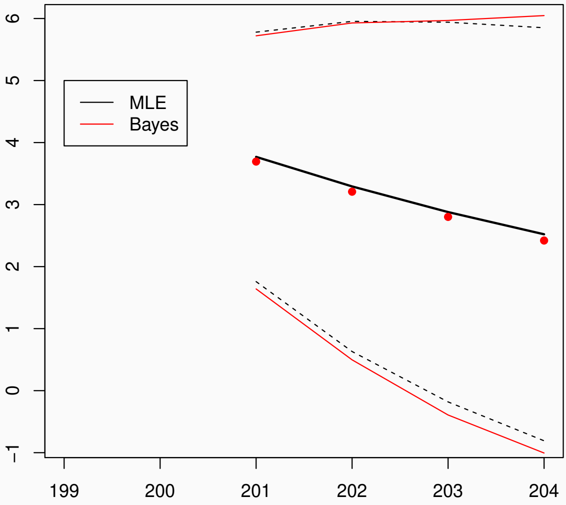
Previsões via máxima verossimilhança,

```
    Point Forecast      Lo 95    Hi 95
201       3.769036  1.7580699 5.780002
202       3.292556  0.6310009 5.954112
203       2.879443 -0.1814368 5.940324
204       2.521270 -0.8084027 5.850943
```

Previsões Bayesianas,

	Mean	Lo.95	Hi.95
201	3.6933	1.6401	5.7210
202	3.2076	0.4968	5.9283
203	2.8017	-0.3923	5.9685
204	2.4205	-1.0056	6.0472

Forecasts from ARIMA(1,0,0) with non−zero mean



Exemplo. Foram gerados 100 valoes de um processo MA(1),

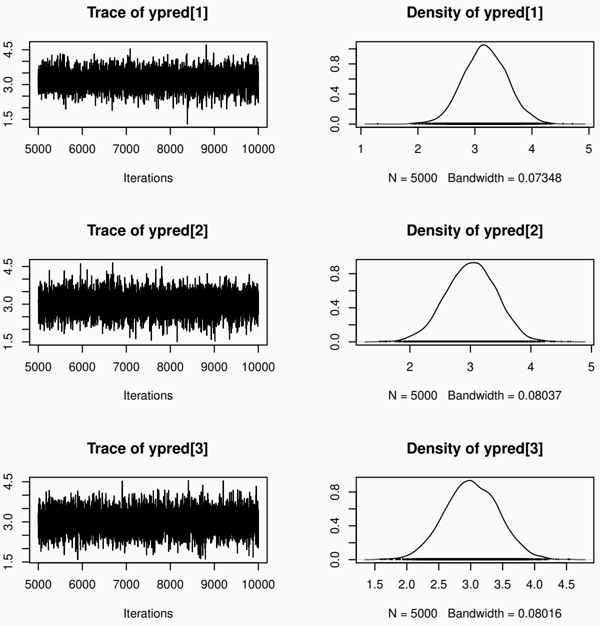
$$X_t = \mu + \epsilon_t + \beta\epsilon_{t-1}, \ \epsilon_t \sim N(0, \sigma_\epsilon^2),$$

com $\mu = 3$, $\beta = -0.5$ e $\sigma_\epsilon = 0.4$. Foi usado o pacote `rjags` para estimação e previsão.

```
1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:
            Mean      SD  Naive SE Time-series SE
beta     -0.4927 0.12023 0.0017003      0.0028440
mu        3.0222 0.01979 0.0002799      0.0002822
sigma     0.3670 0.02699 0.0003817      0.0003947
ypred[1]  3.1814 0.37494 0.0053025      0.0053025
ypred[2]  3.0246 0.40052 0.0056641      0.0060874
ypred[3]  3.0184 0.41100 0.0058124      0.0058124
ypred[4]  3.0254 0.41217 0.0058290      0.0058290

2. Quantiles for each variable:
           2.5%     25%     50%     75%   97.5%
beta     -0.718 -0.5787 -0.4971 -0.4113 -0.2490
mu        2.982  3.0100  3.0222  3.0346  3.0621
sigma     0.319  0.3482  0.3653  0.3837  0.4254
ypred[1]  2.414  2.9297  3.1842  3.4368  3.8937
ypred[2]  2.231  2.7564  3.0183  3.2953  3.7903
ypred[3]  2.205  2.7430  3.0232  3.2978  3.8226
ypred[4]  2.214  2.7511  3.0218  3.3012  3.8252
```

- $x_{n+k}^{(1)}, \ldots, x_{n+k}^{(N)}$ é uma amostra de tamanho N da distribuição preditiva k passos a frente.
- A média aritmética deste valores é uma aproximação para $E(X_{n+k}|\mathbf{x})$.
- Pode-se aproximar toda a distribuição preditiva por exemplo através de um histograma.


Trace of ypred[1]
Density of ypred[1]
Iterations
N = 5000 Bandwidth = 0.07348
Trace of ypred[2]
Density of ypred[2]
Iterations
N = 5000 Bandwidth = 0.08037
Trace of ypred[3]
Density of ypred[3]
Iterations
N = 5000 Bandwidth = 0.08016

Transformações e Previsões

Suponha que a série a ser modelada é uma transformação da série original.

- Previsões pontuais e intervalos de previsão são obtidos para a série transformada.
- Porém em geral deseja-se as previsões na escala original.

- A abordagem mais simples é aplicar a transformação inversa, se $X_t = g(Y_t)$ então,

$$\hat{y}_n(k) = g^{-1}(\hat{x}_n(k)).$$

- Porém, $E[g^{-1}(X_{n+k})] \neq g^{-1}(E[X_{n+k}])$.
- Aplicando a transformação inversa aos limites do intervalo de previsão este continua sendo de $100(1-\alpha)\%$, porém assimétrico.

Exemplo. Série `AirPassengers` com transformação logaritmica e modelo SARIMA(0,1,1)x(0,1,1). Dados até dezembro de 1960 e previsões de 1 a 12 meses à frente para 1961 nas 2 escalas.

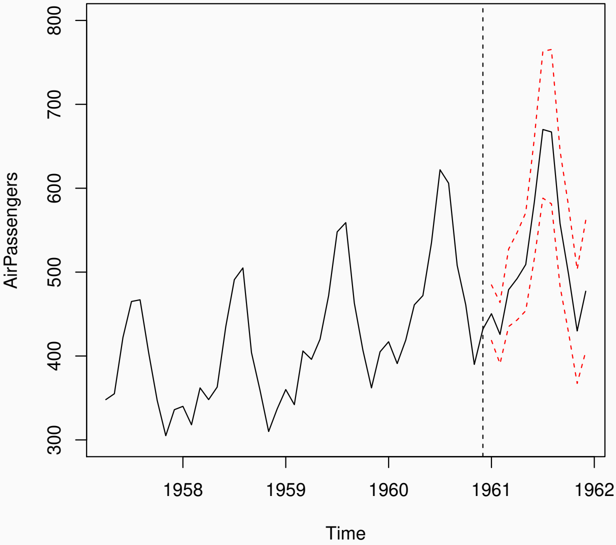
Previsões na escala transformada.

```
         previsão       li       ls
Jan 1961 6.110186 6.036754 6.183617
Feb 1961 6.053775 5.968209 6.139341
Mar 1961 6.171715 6.075533 6.267896
Apr 1961 6.199300 6.093564 6.305037
May 1961 6.232556 6.118059 6.347053
Jun 1961 6.368779 6.246145 6.491412
Jul 1961 6.507294 6.377031 6.637556
Aug 1961 6.502906 6.365438 6.640375
Sep 1961 6.324698 6.180382 6.469014
Oct 1961 6.209008 6.058156 6.359860
Nov 1961 6.063487 5.906370 6.220604
Dec 1961 6.168025 6.004883 6.331166
```

Previsões na escala original.

	prev	li	ls	prev-li	ls-prev
Jan 1961	450.4224	418.5325	484.7421	31.88989	34.31973
Feb 1961	425.7172	390.8053	463.7479	34.91191	38.03070
Mar 1961	479.0068	435.0815	527.3668	43.92531	48.35996
Apr 1961	492.4045	442.9974	547.3219	49.40709	54.91741
May 1961	509.0550	453.9826	570.8081	55.07236	61.75316
Jun 1961	583.3449	516.0199	659.4539	67.32507	76.10897
Jul 1961	670.0108	588.1790	763.2276	81.83178	93.21682
Aug 1961	667.0776	581.3992	765.3821	85.67843	98.30451
Sep 1961	558.1894	483.1767	644.8476	75.01262	86.65824
Oct 1961	497.2078	427.5862	578.1656	69.62164	80.95777
Nov 1961	429.8720	367.3703	503.0072	62.50164	73.13521
Dec 1961	477.2426	405.4037	561.8114	71.83882	84.56888

Previsões na escala original.



Na abordagem Bayesiana, suponha que temos uma amostra de N valores $\boldsymbol{\theta}^{(1)}, \dots, \boldsymbol{\theta}^{(N)}$ da distribuição $p(\boldsymbol{\theta}|\mathbf{x})$ a partir da qual obteve-se uma amostra de previsões k passos à frente,

$$x_{n+k}^{(1)}, \dots, x_{n+k}^{(N)},\ k = 1, 2, \dots$$

Se $X_t = g(Y_t)$ obtem-se uma amostra de previsões k passos à frente na escala original,

$$g^{-1}(x_{n+k}^{(1)}), \dots, g^{-1}(x_{n+k}^{(N)}),\ k = 1, 2, \dots$$

e a previsão pontual na escala original fica,

$$\hat{y}_n(k) \approx \frac{1}{N} \sum_{j=1}^{N} g^{-1}(x_{n+k}^{(j)}).$$

Performance preditiva

Algumas medidas baseadas nos erros de previsão dentro da amostra,

- Erro quadrático médio (MSE):
$$\frac{1}{n-m}\sum_t e_t^2$$
- Erro absoluto médio (MAE):
$$\frac{1}{n-m}\sum_t |e_t|$$
- Erro percentual absoluto médio (MAPE),
$$\frac{1}{n-m}\sum_t \left|\frac{100e_t}{x_t}\right|$$

Estatística U de Theil,

$$U = \sqrt{\frac{\sum_{t=1}^{n-1}(x_{t+1} - \hat{x}_t(1))^2}{\sum_{t=1}^{n-1}(x_{t+1} - x_t)^2}}.$$

compara os erros de previsão do modelo com aqueles de um passeio aleatório.

- Nos critérios MSE, MAE e MAPE m é o número de parâmetros a serem estimados.
- Na estatistica U compara-se as previsões do modelo com as de um passeio aleatório.
- MSE e MAE dependem da escala dos dados.
- MAPE não depende da escala mas é sensivel a valores muito pequenos de x_t.

Uma proposta alternativa,

MASE (*Mean absolute scaled error*, Hyndman & Koehler, 2006, *International Journal of Forecasting*),

$$MASE = \frac{1}{n-m} \sum_{t} \left| \frac{e_t}{\frac{1}{n-1} \sum_{t=2}^{n} |x_t - x_{t-1}|} \right|$$

Validação cruzada

Dadas n observações $x_1, \ldots, x_n$ de uma série temporal,

1. Ajuste um modelo a série $x_1, \ldots, x_t$.
2. Calcule a previsão 1 passo a frente $\hat{x}_t(1)$ e o erro de previsão $e^*_{t+1} = x_{t+1} - \hat{x}_t(1)$.
3. Repita os passos 1 e 2 para $t = m, \ldots, n-1$.
4. Calcule MSE, MAE, etc. usando os erros $e^*_{m+1}, \ldots, e^*_n$.

Previsões Usando Todos os Modelos

Sejam as variáveis aleatórias X qualquer e Y discreta. Então,

$$E(X) \;=\; E[E(X|Y=y)] = \sum_y E(X|Y=y)P(Y=y)$$

- Suponha que existem h modelos candidatos $M_1, M_2, \ldots, M_h$ e deseja-se fazer a previsão de X_{n+k}.
- Tratando X_{n+k} e M_i como variáveis aleatórias então pelas regras de esperança condicional,

$$\hat{x}_n(k) = E(X_{n+k}|\mathbf{x}) = \sum_{i=1}^{h} E(X_{n+k}|\mathbf{x}, M_i)P(M_i|\mathbf{x}).$$

- A previsão pontual é uma mistura discreta de previsões pontuais sob cada modelo M_i.

A mesma lógica se aplica a qualquer função de X_{n+k}. Em particular

$$E(X_{n+k}^2|\mathbf{x}) = \sum_{i=1}^{k} E(X_{n+k}^2|\mathbf{x}, M_i)P(M_i|\mathbf{x}).$$

que pode ser usado para quantificar a incerteza sobre X_{n+k},

$$\begin{aligned}
Var(X_{n+k}|\mathbf{x}) &= E(X_{n+k}^2|\mathbf{x}) - [E(X_{n+k}|\mathbf{x})]^2 \\
&= \sum_{i=1}^{h} E(X_{n+k}^2|\mathbf{x}, M_i)P(M_i|\mathbf{x}) - [E(X_{n+k}|\mathbf{x})]^2 \\
&= \sum_{i=1}^{h} [Var(X_{n+k}|\mathbf{x}, M_i) + E^2(X_{n+k}|\mathbf{x}, M_i)]P(M_i|\mathbf{x}) - [\hat{x}_n(k)]^2
\end{aligned}$$

- As probabilidades $P(M_i|\mathbf{x})$ são em geral desconhecidas.
- Possivel abordagem: substituir $P(M_i|\mathbf{x})$ pelos pesos padronizados.
- E.g. modelos AR(p) com defasagem máxima $p_{\max}$,

$$E(X_{n+k}|\mathbf{x}) = \sum_{i=1}^{p_{\max}} E(X_{n+k}|\mathbf{x}, AR(i))\; w_i^*.$$

Exemplo. Totais anuais de linces canadenses capturados em armadilhas entre 1821 e 1934. Previsões 1 passo a frente.

	previsoes
1	0.4968692
2	0.4789037
3	0.4733455
4	0.4547391
5	0.4628502
6	0.4627748
7	0.4737786
8	0.5164067
9	0.5208899
10	0.5076302
11	0.5426978
12	0.5508093
13	0.5411680
14	0.5413534
15	0.5503112
pesos AIC	0.5456198
pesos BIC	0.4765837