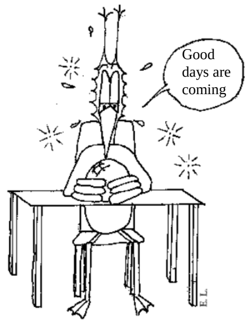
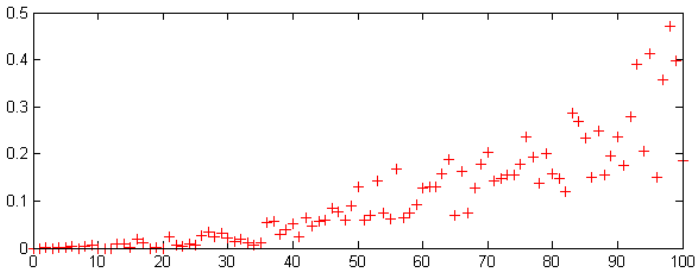
PERFORMANCE EVALUATION
EXERCISES

FORECASTING

With Solutions Jean-Yves Le Boudec, Spring 2021



1. The following data shows the amount of memory claimed by a server process, in percent of the total physical memory, as a function of times in seconds since last reboot. The server should be rebooted 10 seconds before the used memory reaches the threshold $\theta = 90\%$ (of the physical memory). Explain a method for deciding when to reboot.



Solution.

Here there are many possible answers, but given that the used memory seems to increase over time (since last reboot) faster than linearly, a simple model could consist in fitting the used memory M_t to a square function. Note that the variability also seems to increase with M_t. A simple model could thus be

$$M_{t_i} = at_i^2 + t_i^2 \varepsilon_i$$

with for example $\varepsilon_t \sim$ iid $N(0, \sigma^2)$. This is a weighted least square problem, and we can estimate a by minimizing the score

$$\text{score} \quad = \quad \sum_i \left(\frac{M_{t_i} - at_i^2}{t_i^2} \right)^2 = \sum_i \left(\frac{M_{t_i}}{t_i^2} - a \right)^2$$

This is a linear estimation model, the optimal value $\hat{a}$ of a is obtained by cancelling the derivative:

$$-2 \sum_i \left(\frac{M_{t_i}}{t_i^2} - a \right)^2 = 0$$

which gives

$$\hat{a} = \frac{1}{N} \sum_i \frac{M_{t_i}}{t_i^2}$$

where N is the number of observed data points.Then σ is estimated by setting $\hat{\sigma}^2 = 1/N\times$ the optimal score:

$$\hat{\sigma}^2 = \frac{1}{N}\sum_i \left(\frac{M_{t_i}}{t_i^2} - \hat{a}\right)^2$$

This can be done online or offline. If online, the estimation is redone at every time step.

(To be safe, we need to check if the residuals appear reasonably normal.)

The 10-second ahead prediction done at time t with confidence 95% gives an upper bound

$$\hat{U}_t = \hat{a}(t+10)^2 + 1.96\hat{\sigma}$$

Whenever $\hat{U}_t \geq \theta$ we reboot.

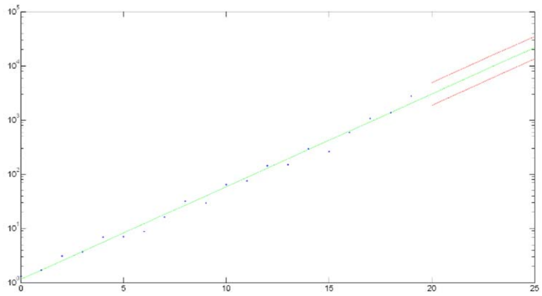
$\frac{\sum_{i=1}^N M_{t_i}}{\left(\sum_{i=1}^N t_i^2\right)^2}$

2. We fit the log of virus expansion data using Laplace noise.

The model is

$$L_i = \ell + \alpha t_i + \varepsilon_i \text{ with } \varepsilon_i \sim \text{ iid Laplace}(\lambda)$$

where L_i is the logarithm of the ith value and t_i the time of measurement.



   (a) Write a linear program that you can use for estimating ℓ, α and λ.
   (b) When X is Laplace noise with parameter λ, for which value of η do we have $\mathbb{P}(|X| > \eta) = 0.05$?
   (c) We want to use the estimated model to predict the virus expansion at a time T. Give the formula for a 95%-prediction interval, assuming we can neglect the estimation uncertainty.

Solution.

(a) By Theorem 3.2, (ℓ, α) solves

$$\min_{\ell,\alpha} \sum_{i=1}^{I} |L_i - \ell - \alpha t_i|$$

which is not a linear program, but is equivalent to

$$\begin{aligned} &\min_{\ell,\alpha,u} \sum_{i=1}^{I} u_i \\ \text{s.c.} \quad & u_i \geq |L_i - \ell - \alpha t_i| \text{ for } i = 1...I \end{aligned}$$

which in turn is equivalent to the following linear program

$$\begin{aligned} &\min_{\ell,\alpha,u} \sum_{i=1}^{I} u_i \\ \text{s.c.} \quad & u_i \geq L_i - \ell - \alpha t_i \text{ for } i = 1...I \\ & u_i \geq -\left(L_i - \ell - \alpha t_i\right) \text{ for } i = 1...I \end{aligned}$$

Once optimal values of ℓ, α are found by solving this linear program, the value of λ is given by Theorem 3.2 and is

$$\hat{\lambda} = \left(\frac{1}{I} \sum_{i=1}^{I} \left| L_i - \hat{\ell} - \hat{\alpha} t_i \right| \right)^{-1}$$

(b) Observe that $|X| \sim \text{Exp}(\lambda)$ and the CDF of an exponential distribution is $F(c) = 1 - e^{-\lambda c}$ hence

$$\mathbb{P}(|X| > \eta) = e^{-\lambda\eta}$$

hence

$$\eta = -\frac{1}{\lambda} \log 0.05 = \frac{1}{\lambda} \cdot 2.996 \approx \frac{3.00}{\lambda}$$

(c) At time T, for the true value we have, according to the model and if we ignore the fact that the estimation is not perfect:

$$L_T = \hat{\ell} + \hat{\alpha} T + \varepsilon \text{ with } \varepsilon \sim \text{Exp}(\hat{\lambda})$$

We know from the previous question that,

$$\mathbb{P}\left(-\frac{3.00}{\hat{\lambda}} \leq \varepsilon \leq \frac{3.00}{\hat{\lambda}} \right) = 0.95$$

Hence

$$\mathbb{P}\left(\hat{\ell} + \hat{\alpha} T - \frac{3.00}{\hat{\lambda}} \leq L_T \leq \hat{\ell} + \hat{\alpha} T + \frac{3.00}{\hat{\lambda}} \right) = 0.95 \tag{1}$$

i. e. a 95% prediction interval for L_T is

$$\hat{\ell} + \hat{\alpha} T \pm \frac{3.00}{\hat{\lambda}}$$

Taking the exponential gives

$$\mathbb{P}\left(e^{\hat{\ell} + \hat{\alpha} T - \frac{3.00}{\hat{\lambda}}} \leq Y_T \leq e^{\hat{\ell} + \hat{\alpha} T + \frac{3.00}{\hat{\lambda}}} \right) = 0.95$$

i.e. a 95% prediction interval for $Y_T = e^{L_T}$ is

$$\left[\hat{a} e^{\hat{\alpha} T} e^{-\frac{3.00}{\hat{\lambda}}};\ \hat{a} e^{\hat{\alpha} T} e^{\frac{3.00}{\hat{\lambda}}} \right]$$

with $\hat{a} = e^{\hat{\ell}}$. Note that the prediction interval is not symmetric around the point prediction $\hat{a} e^{\hat{\alpha} T}$.

3. Δ_k is the differencing filter at lag k.

   (a) Is Δ_{16} stable ? Is it invertible ? If so, is the inverse stable ?

   (b) Say which is true

      i. ☒ Δ_{16} is a FIR filter
      ii. ☐ Δ_{16} is an AR filter
      iii. ☐ Both
      iv. ☐ None

   (c) Compute the $MA(\infty)$ and $AR(\infty)$ representations of Δ_{16}.

   (d) Let $F = \Delta_1 \Delta_{16}$ and $G = \Delta_{16} \Delta_1$. Give the operator- and the input-output-representations of F and G.

(e) Is F stable ? Is it invertible ? If so, is the inverse stable ?

Solution.

(a) The first element of the impulse response is $h_0 = 1 \neq 0$ therefore Δ_{16} is invertible.
Δ_{16} is an ARMA filter with denominator polynomial $= 1$ hence it is stable.
Its inverse is

$$\Delta_{16}^{-1} = \frac{1}{1 - B^{16}}$$

The denominator polynomial is

$$Q(\xi) = \xi^{16} - 1$$

The roots are $e^{\frac{2j\pi}{16}}$, $j = 0...15$. They have module 1 so the inverse filter is *not* stable.

(b) The impulse response of Δ_{16} is given by

$$\begin{aligned} h_0 &= 1 \\ h_{16} &= -1 \\ h_i &= 0 \text{ else} \end{aligned}$$

hence it non zero only for a finite number of indices.
Δ_{16} is not AR because its inverse is not FIR as we see in the next question.

(c) The MA(∞) representation is the input-output representation of Δ_{16}, and is given by

$$Y_n = X_n - X_{n-16}$$

where X is the input, Y the output and $X_k = 0$ when $k \leq 0$.
The AR(∞) representation requires computing the impulse response of Δ_{16}^{-1}:

$$\Delta_{16}^{-1} = \frac{1}{1 - B^{16}} = 1 + B^{16} + B^{32} + ... + B^{16k} + ...$$

Hence its impulse response is such that $h_{16k} = 1$ for $k = 0, 1, 2, ...$ and 0 otherwise. Let $Y = \Delta_{16}X$, so that $X = (\Delta_{16})^{-1}Y$. The input-output representation of Δ_{16}^{-1} gives:

$$X_n = Y_n + Y_{n-16} + Y_{n-32} + ... + Y_{n-16k} + ...$$

Therefore

$$Y_n = X_n - Y_{n-16} - Y_{n-32} - ... - Y_{n-16k} + ...$$

This is the AR(∞) representation of Δ_{16}. It has an infinite number of non zero coefficients so Δ_{16} is not an AR filter.

(d) $F = G$ because filters commute. The operator representation is:

$$\begin{aligned} F &= (1 - B)(1 - B^{16}) \\ &= 1 - B - B^{16} + B^{17} \end{aligned}$$

Therefore the input-output relation $X \overset{F}{\mapsto} Y$ is

$$Y_n = X_n - X_{n-1} - X_{n-16} + X_{n-17}$$

where $X_k = 0$ if $k \leq 0$

(e) F is an ARMA filter with denominator polynomial $= 1$ hence it is stable.
F is invertible because the first element h_0 of its impulse response is non zero, and

$$F^{-1} = \frac{1}{1 - B - B^{16} + B^{17}}$$

The denominator polynomial of F^{-1} is

$$Q(\xi) = \xi^{17} - \xi^{16} - \xi + 1 = (\xi - 1)(\xi^{16} - 1)$$

Its roots are $e^{\frac{2j\pi}{16}}$, $j = 0...15$. They have module 1 so the inverse filter is *not* stable.

4. We want to forecast the temperature $T_1, T_2, ..$ where there is one measurement every hour. We want to use a differencing filter at lag 24. Let X_n be the differenced time series. Give the formulas to compute T from X and vice versa. We find that X_n looks almost iid with mean μ. We want to use this fact to give a point prediction for T_{n+5}, assuming we are at time n (where n is large). Give the formulas for this point prediction.

Solution. By definition of Δ_{24}:

$$\begin{aligned} X_n &= T_n - T_{n-24} \text{ for } n \geq 25 \\ X_n &= T_n \text{ for } n = 1...24 \end{aligned}$$

which can be used to compute X from T. It can also be used recursively to compute T from X, using $T_n = X_n$ for $n = 1...24$ and

$$T_n = X_n + T_{n-24} \text{ for } n \geq 25 \tag{2}$$

At time n we know $T_{1:n}$ and $X_{1:n}$. From the previous equation, and since we can assume $n > 24$:

$$T_{n+5} = X_{n+5} + T_{n-19}$$

The forecast $\hat{T}_n(5)$ is the conditional expectation given we observed everything up to and including time n.

$$\begin{aligned} \hat{T}_n(5) = \mathbb{E}\left(T_{n+5} | T_{1:n}\right) &= \mathbb{E}\left(X_{n+5} | T_{1:n}\right) + \mathbb{E}\left(T_{n-19} | T_{1:n}\right) \\ &= \mu + T_{n-19} \end{aligned}$$

5. We have a times series Y_t. We computed the differenced time series $X_t = Y_t - Y_{t-1}$ and found that X_t can be modelled as an AR process:

$$X_t = 0.5X_{t-1} + \varepsilon_t \text{ with } \varepsilon_t \sim \text{ iid } N_{0,\sigma^2}$$

for some value of σ that we have estimated.

(a) Is this a valid ARIMA model ?
(b) Compute a point forecast $\hat{X}_t(2)$
(c) Compute a point forecast $\hat{Y}_t(2)$
(d) Compute the first 3 terms of the impulse response of the filter $\varepsilon \mapsto Y$
(e) Compute a prediction interval for Y_{t+2} done at time t.
(f) Which of the following algorithms is a correct implementation of computing a prediction interval for Y_{t+2} done at time t using the bootstrap from residuals ?

Algorithm A

compute the time series $\varepsilon_s = X_s - 0.5X_{s-1}$ for $s = 3 : t$

for $r = 1 : 999$ do

draw e_s^r, $s = 3 : (t+2)$ with replacement from ε_s, $s = 3 : t$

compute $X_{1:t}^r, Y_{1:t}^r$ and $\hat{Y}_{1:t}^r(2)$ using $X_s^r = 0.5X_{s-1}^r + e_s^r$, $Y_s^r = X_s^r + Y_{s-1}^r$ and the formula you have found for $\hat{Y}_{1:t}^r(2)$

$Y_{t+2}^r = e_{t+2}^r + 1.5e_{t+1}^r + \hat{Y}_t^r(2)$

end do

prediction interval is $[Y_{t+2}^{(25)};\ Y_{t+2}^{(975)}]$

Algorithm B

compute the time series $\varepsilon_s = X_s - 0.5X_{s-1}$ for $s = 3 : t$

for $r = 1 : 999$ do

draw e_1^r, e_2^r with replacement from ε_s, $s = 3 : t$

$Y_{t+2}^r = e_1^r + 1.5e_2^r + \hat{Y}_t(2)$

end do

prediction interval is $[Y_{t+2}^{(25)};\ Y_{t+2}^{(975)}]$

i. ☐ A
ii. ☒ B
iii. ☐ A and B
iv. ☐ None

Solution.

(a) The model is of the form $X = LY$ and $X = F\varepsilon$ where F is an ARMA filter and L is a differencing filter. The only thing to verify is whether the ARMA filter is stable and has a stable inverse.

The filter is

$$F = \frac{1}{1 - 0.5B}$$

The numerator polynomial has no zero so the inverse is stable. The denominator polynomial is $\xi - 0,5$ there is one zero, equal to 0.5, which has module less than 1 so the filter is stable. Thus we have a valid ARIMA model.

(b) We use as point forecast the conditional expectation of X_{t+2} given we have observed X up to time t. Note that L and F are invertible, therefore observing $Y_{1:t}$ is the same as observing $X_{1:t}$ or $X_{1:t}$. Furthermore:

$$\begin{aligned} X_{t+2} &= 0.5X_{t+1} + \varepsilon_{t+2} \\ X_{t+1} &= 0.5X_t + \varepsilon_{t+1} \end{aligned}$$

Take the conditional expectation given the observation up to time t of the above equations and obtain

$$\begin{aligned} \mathbb{E}\left(X_{t+2} | Y_{1:t}\right) &= 0.5\mathbb{E}\left(X_{t+1} | Y_{1:t}\right) \\ \mathbb{E}\left(X_{t+1} | Y_{1:t}\right) &= 0.5X_t \end{aligned}$$

because $\mathbb{E}\left(\varepsilon_{t+2}|\, Y_{1:t}\right) = \mathbb{E}\left(\varepsilon_{t+2}|\, \varepsilon_{1:t}\right) = \mathbb{E}(\varepsilon_{t+2}) = 0$ because ε is iid with 0 mean. Similarly $\mathbb{E}\left(\varepsilon_{t+1}|\, Y_{1:t}\right) = 0$. Thus we find the recursive formulation:

$$\begin{aligned} \hat{X}_t(2) &= 0.5\hat{X}_t(1) \\ \hat{X}_t(1) &= 0.5X_t \end{aligned}$$

and thus

$$\hat{X}_t(2) = 0.25X_t$$

(c) We have

$$\begin{aligned} Y_{t+2} &= X_{t+2} + Y_{t+1} \\ Y_{t+1} &= X_{t+1} + Y_t \end{aligned}$$

Take the conditional expectation given the observation up to time t of the above equations and obtain

$$\begin{aligned} \hat{Y}_t(2) &= \hat{X}_t(2) + \hat{Y}_t(1) \\ \hat{Y}_t(1) &= \hat{X}_t(1) + Y_t \end{aligned}$$

which is sufficient to compute the point forecast. We can also simplify a bit and obtain

$$\begin{aligned} \hat{Y}_t(2) &= \hat{X}_t(2) + \hat{X}_t(1) + Y_t = 0.25X_t + 0.5X_t + Y_t = 0.75X_t + Y_t = 0.75(Y_t - Y_{t-1}) + Y_t \\ &= 1.75Y_t - 0.75Y_{t-1} \end{aligned}$$

(d) We have

$$\begin{aligned} (1-B)Y &= X \\ (1-0.5B)X = \varepsilon \end{aligned}$$

thus

$$Y = \frac{1}{(1-0.5B)(1-B)}\varepsilon$$

We can use power series calculus:

$$\begin{aligned} \frac{1}{1-B} &= 1 + B + B^2 + ... + B^n + ... \\ \frac{1}{1-0.5B} &= 1 + 0.5B + 0.25B^2 + ... + (0.5)^n B^n + ... \\ \frac{1}{(1-0.5B)(1-B)} &= 1 + 1.5B + 1.75B^2 + ... \end{aligned}$$

thus the first three terms of the required impulse response are

$$h_0 = 1 \;\; h_1 = 1.5, \;\; h_2 = 1.75$$

Alternatively, we can use Matlab:

```
>>h=filter([1],[1 -1],filter([1],[1 -0.5],[1 0 0 0 0 0 0]))
h =
1.0000 1.5000 1.7500 1.8750 1.9375 1.9688 1.9844
```

(e) From the previous question we have

$$Y_t = \varepsilon_t + 1.5\varepsilon_{t-1} + 1.75\varepsilon_{t-2} + h_3\varepsilon_{t-3} + ... + h_{t-1}\varepsilon_1$$

where h_i is the impulse response of the filter $\varepsilon \mapsto Y$. Thus

$$\begin{aligned} Y_{t+2} &= \varepsilon_{t+2} + 1.5\varepsilon_{t+1} + 1.75\varepsilon_t + h_3\varepsilon_{t-1} + ... + h_{t+1}\varepsilon_1 \\ &= \varepsilon_{t+2} + 1.5\varepsilon_{t+1} + \mathcal{F}_{1:t} \end{aligned} \tag{3}$$

with $\mathcal{F}_{1:t} = 1.75\varepsilon_t + h_3\varepsilon_{t-1} + ... + h_{t+1}\varepsilon_1$. Observe that $\mathcal{F}_{1:t}$ depends only on the random variables up to time t. Now take the conditional expectation of (3) up to time t, i.e. assume that the values of $\varepsilon_1 ... \varepsilon_t$ are given and non-random and do a forecast. We have $\mathbb{E}\left(\mathcal{F}_{1:t} | \varepsilon_{1:t}\right) = \mathcal{F}_{1:t}$ (the forecast of $\mathcal{F}_{1:t}$ is itself because it is already revealed), $\mathbb{E}\left(\varepsilon_{t+1} | \varepsilon_{1:t}\right) = 0$ and $\mathbb{E}\left(\varepsilon_{t+2} | \varepsilon_{1:t}\right) = 0$ thus

$$\mathbb{E}\left(Y_{t+2} | \varepsilon_{1:t}\right) = \mathcal{F}_{1:t}$$

in other words, $\hat{Y}_t(2) = \mathcal{F}_{1:t}$ and we can rewrite (3) as

$$Y_{t+2} = \varepsilon_{t+2} + 1.5\varepsilon_{t+1} + \hat{Y}_t(2) \tag{4}$$

(this is called the innovation equation; recall that $\hat{Y}_t(2)$ was computed earlier and is known at time t). We can now use (4) to find a prediction interval. Conditional to the observations up to time t, ε_{t+2} and ε_{t+1} are iid N_{0,σ^2} and thus the right-handside of (**??**) is $\sim N_{0,v}$ with the variance v given by

$$v = \sigma^2 + (1.5)^2\sigma^2 = 3.25\sigma^2$$

Thus the conditional distribution of Y_{t+2} given the observations up to time t is normal with mean $\hat{Y}_t(2)$ and variance $3.25\sigma^2$. Therefore, a 95% prediction interval is

$$\hat{Y}_t(2) \pm 1.96\sqrt{3.25}\sigma$$

(f) A is simulating the entire time series, therefore it is producing a sample of the unconditional distribution of Y_{t+2}. It is not the prediction, it is what can be said about Y_{t+2} for an observer who knows the statistics of the time series but did not observe $Y_{1:t}$.

B is simulating the time series from $t+1$ to $t+2$ given the data up to time t, therefore it is producing a sample of the conditional distribution of Y_{t+2} given the observed past. It is a correct implementation.