Homework 4 Solutions

Ruoqi Yu

10/27/2021

1 Theory: Bartlett's Formula

1a

Let X_t be ARMA(0,0), i.e. white noise, and show that

$$\sqrt{n}\begin{pmatrix}\hat{r_1}\\ \vdots \\ \hat{r_k}\end{pmatrix} \xrightarrow{d} N_k(0, I_k)$$

The process $X_t = W_t$, where W_t is white noise, is certainly causal and invertible (using the geometric sum definition, ie. $W_t = \sum \psi_i X_{t-i}$ and $X_t = \sum \phi_i W_{t-i}$). We can also assume that the white noise conditions are met. Therefore, applying Bartlett's formula to $\sqrt{n}\begin{pmatrix}\hat{r_1}\\ \vdots \\ \hat{r_k}\end{pmatrix}$ we can calculate the asymptotic covariance matrix W. We split into cases:

1. $i \neq j$

$$\begin{aligned}
w_{ij} &= \sum_{m\geq 1} (\rho(m+i) + \rho(m-i) - 2\rho(i)\rho(m))(\rho(m+j) + \rho(m-j) - 2\rho(j)\rho(m)) \\
&= \sum_{m\geq 1} (\rho(m+i) + \rho(m-i))(\rho(m+j) + \rho(m-j)) \qquad \text{as white noise} \Rightarrow \rho(h) = 0 \text{ for h} \neq 0 \\
&= \sum_{m\geq 1} \rho(m-i)\rho(m-j) \qquad \text{since } m, i, j \geq 1 \text{ and } \rho(h) \neq 0 \text{ iff } h = 0 \\
&= 0 \qquad \text{since } i \neq j\text{, so if m=i, then } \rho(m-j) = 0
\end{aligned}$$

2. $i = j$. Using the above calculation, we start at:

$$\begin{aligned}
w_{ij} &= \sum_{m\geq 1} \rho(m-i)\rho(m-j) = \sum_{m\geq 1} (\rho(m-i))^2 \\
&= 1 \qquad \text{as } m - i = 0 \text{ for only one } m
\end{aligned}$$

We have shown that W is the $k \times k$ identity matrix for white noise. In addition, all $\rho(h) = 0$, $|h| \geq 1$ for a white noise. So Bartlett's formula gives:

$$\sqrt{n}\begin{pmatrix}\hat{r_1}\\ \vdots \\ \hat{r_k}\end{pmatrix} \xrightarrow{d} N_k(0, I_k)$$

1b

Calculate the asymptotic variance of $\hat{r}_k$ for AR(1) process X_t.

Because of Bartlett's formula, the ACF previously found yields the expected values, and for the variances W_{kk}/n we need to find W_{kk}. Note that we assume causality: $|\phi| < 1$.

$$
\begin{aligned}
W_{kk} &= \sum_{m=1}^{\infty} \left[\rho(m+k) + \rho(m-k) - 2\rho(k)\rho(m)\right]^2 \\
&= \sum_{m=1}^{\infty} \left[\phi^{|m+k|} + \phi^{|m-k|} - 2\phi^k\phi^m\right]^2 \\
&= \sum_{m=1}^{\infty} \left[\phi^{|m-k|} - \phi^{k+m}\right]^2 \\
&= \sum_{m=1}^{\infty} \left[\phi^{2|m-k|} - 2\phi^{|m-k|+k+m} + \phi^{2k+2m}\right] \\
&= \sum_{m=1}^{k} \left[\phi^{2|m-k|} - 2\phi^{|m-k|+k+m}\right] + \sum_{m=k+1}^{\infty} \left[\phi^{2|m-k|} - 2\phi^{|m-k|+k+m}\right] + \sum_{m=1}^{\infty} \left[\phi^{2k+2m}\right] \\
&= \sum_{m=1}^{k} \left[\phi^{2k-2m} - 2\phi^{k-m+k+m}\right] + \sum_{m=k+1}^{\infty} \left[\phi^{2m-2k} - 2\phi^{m-k+k+m}\right] + \sum_{m=1}^{\infty} \left[\phi^{2k+2m}\right] \\
&=\phi^{2k} \sum_{m=1}^{k} \left[\phi^{-2m}\right] - 2k\phi^{2k} + \left(\phi^{-2k} - 2\right) \sum_{m=k+1}^{\infty} \left[\phi^{2m}\right] + \phi^{2k} \sum_{m=1}^{\infty} \left[\phi^{2m}\right] \\
&=\phi^{2k} \left[\frac{1-(\phi^{-2})^{k+1}}{1-\phi^{-2}} - 1\right] - 2k\phi^{2k} + \left(\phi^{-2k} - 2\right) \left[\frac{(\phi^2)^{k+1}}{1-\phi^2}\right] + \phi^{2k} \left[\frac{\phi^2}{1-\phi^2}\right] \\
&=\phi^{2k} \left[\frac{\phi^{-2}-\phi^{-2k-2}}{1-\phi^{-2}}\right] - 2k\phi^{2k} + \left(\phi^{-2k} - 2\right) \left[\frac{\phi^{2k+2}}{1-\phi^2}\right] + \frac{\phi^{2k+2}}{1-\phi^2} \\
&=\frac{\phi^{2k-2}-\phi^{-2}}{1-\phi^{-2}} - 2k\phi^{2k} + \frac{\phi^2-\phi^{2k+2}}{1-\phi^2} \\
&=\frac{1-\phi^{2k}}{1-\phi^2} - 2k\phi^{2k} + \frac{\phi^2-\phi^{2k+2}}{1-\phi^2} \\
&=\frac{(1-\phi^{2k})(1+\phi^2)}{1-\phi^2} - 2k\phi^{2k}
\end{aligned}
$$

Note this can likely be simplified further, but I'm only expecting simplification up to and including the geometric series identities halfway through my derivation above.

So, r_k is approximately normally distributed (as $n \to \infty$) with mean ϕ^k and variance $\frac{W_{kk}}{n}$, where W_{kk} is defined above.

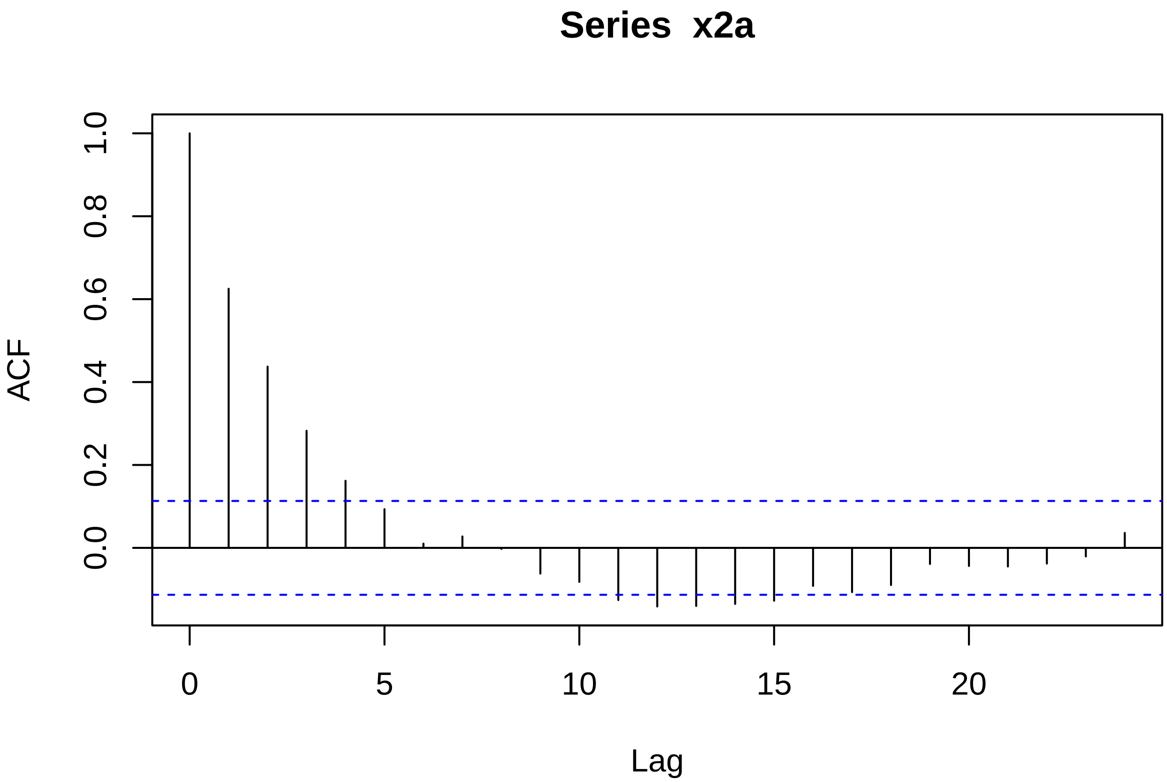
2 Computation: Bartlett's Formula

Assume the AR(1) model with $\phi = 0.7$.

2a

Simulate n=300 observations using arima.sim(). Plot the ACF.

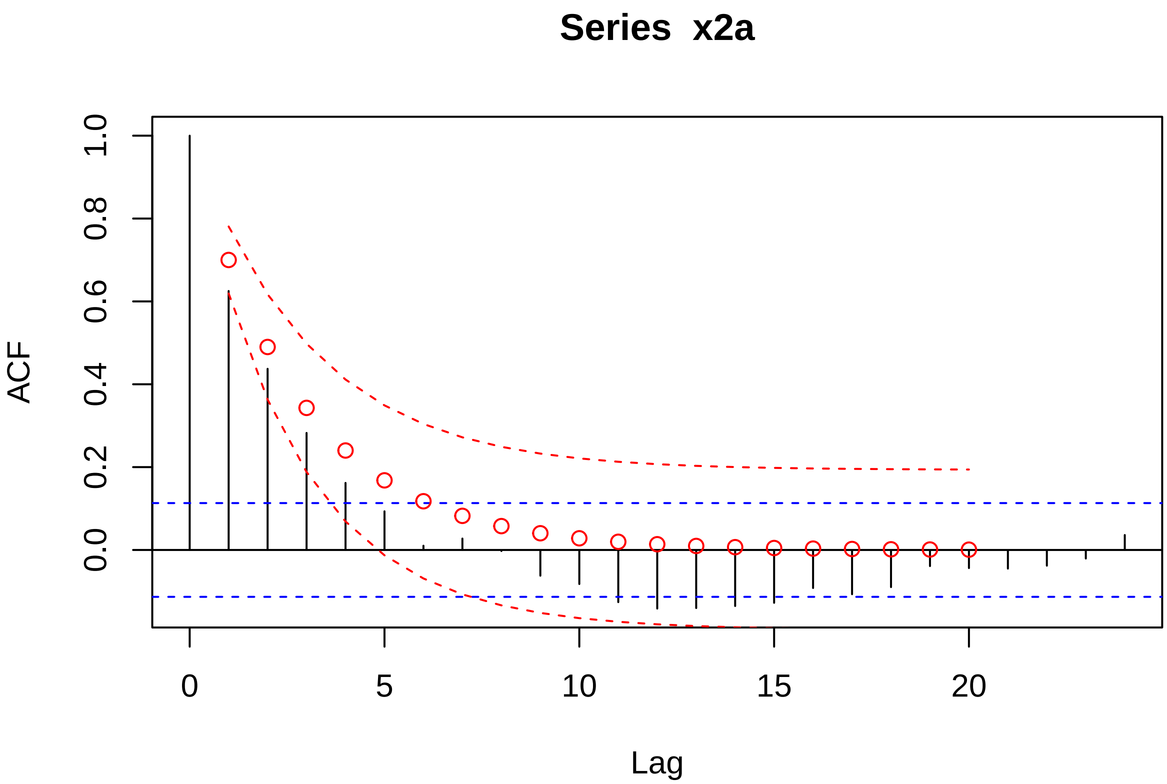
```r
n=300
x2a = arima.sim(model=list(ar=0.7),n=n)
acf(x2a)
```



2b

Add the mean and 95% interval for the distribution of AR(1)'s sample autocorrelations $r_1, ..., r_{20}$ in red. Comment.

```r
phi = 0.7
k = 1:20
Wkk = ( 1-phi^(2*k) )*(1+phi^2)/(1-phi^2) - 2*k*phi^(2*k)
acf(x2a)
points(k,phi^k,col='red')
lines(k,phi^k-1.96*sqrt(Wkk/n),col='red',lty=2)
lines(k,phi^k+1.96*sqrt(Wkk/n),col='red',lty=2)
```



As expected, the generated AR(1) process's sample autocorrelations fit nicely inside the red bands, but not inside the blue bands. Clearly, AR(1) with $\phi = 0.7$ is a better fit than AR(0), aka white noise.

3

Consider an invertible MA(1) model $X_t = W_t + \theta W_{t-1}$ for some i.i.d. white noise process $\{W_t\}$ with variance σ^2.

3a

Derive the minimum mean-square error one-step prediction $\tilde{X}_{n+1} = E(X_{n+1} | X_n, X_{n-1}, X_{n-2}, \ldots)$.

We are forecasting $X_{n+1} = W_{n+1} + \theta W_n$.

As invertible, we can express W_n as

$$W_n = \sum_{j=0}^{\infty} \pi_j X_{n-j}$$

where $\pi_j = (-\theta)^j$

$$\begin{aligned}
\tilde{X}_{n+1} &= E(X_{n+1}|X_n, X_{n-1}, X_{n-2}, \dots) \\
&= E(W_{n+1} + \theta W_n|X_n, X_{n-1}, X_{n-2}, \dots) \\
&= E(W_{n+1}|X_n, X_{n-1}, X_{n-2}, \dots) + \theta E(W_n|X_n, X_{n-1}, X_{n-2}, \dots) \\
&= 0 + \theta E\left(\sum_{j=0}^{\infty}(-\theta)^j X_{n-j} \middle| X_n, X_{n-1}, X_{n-2}, \dots\right) \\
&= \theta \sum_{j=0}^{\infty}(-\theta)^j X_{n-j} \\
&\textit{or} \\
&= -\sum_{j=1}^{\infty}(-\theta)^j X_{n-j+1}
\end{aligned}$$

3b

Derive the mean squared error $E[(\tilde{X}_{n+1} - X_{n+1})]^2$.

First, we can rephrase 3a as

$$\tilde{X}_{n+1} = \theta W_n$$

.

$$\begin{aligned}
E[(\tilde{X}_{n+1} - X_{n+1})]^2 &= E[(\theta W_n - X_{n+1})]^2 \\
&= E[(-W_{n+1})]^2 \\
&= var(-W_{n+1}) + [E(-W_{n+1})]^2 \\
&= (-1)^2 var(W_{n+1}) + [-0]^2 \\
&= var(W_{n+1}) \\
&= \sigma^2
\end{aligned}$$

3c

Now consider the truncated estimate $\tilde{X}^n_{n+1}$, which equals $\tilde{X}_{n+1}$ but with unobserved data being set to zero, that is, $0 = X_0 = X_{-1} = \dots$. Show that

$$E((X_{n+1} - \tilde{X}^n_{n+1})^2) = \sigma^2(1 + \theta^{2+2n}).$$

By definition we have

$$\tilde{X}^n_{n+1} = -\sum_{j=1}^{n}(-\theta)^j X_{n+1-j}$$

and thus

$$\begin{aligned}
X_{n+1} - \tilde{X}^n_{n+1} &= W_{n+1} + \theta W_n - \tilde{X}^n_{n+1} \\
&= W_{n+1} - \left[\sum_{j=1}^{\infty}(-\theta)^j X_{n-j+1}\right] + \sum_{j=1}^{n}(-\theta)^j X_{n+1-j} \\
&= W_{n+1} - \sum_{j=n+1}^{\infty}(-\theta)^j X_{n+1-j} \\
&= W_{n+1} - (-\theta)^{n+1}\sum_{j=0}^{\infty}(-\theta)^j X_{-j} \\
&= W_{n+1} - (-\theta)^{n+1} W_0
\end{aligned}$$

and thus

$$\begin{aligned}
E((X_{n+1} - \tilde{X}_{n+1}^n)^2) &= E((W_{n+1} - (-\theta)^{n+1}W_0)^2) \\
&= Var(W_{n+1} - (-\theta)^{n+1}W_0) + (E(W_{n+1} - (-\theta)^{n+1}W_0))^2 \\
&= Var(W_{n+1}) + Var((-\theta)^{n+1}W_0) + 0^2 \\
&= \sigma^2 + (-\theta)^{2(n+1)}\sigma^2 \\
&= \sigma^2(1 + \theta^{2+2n}).
\end{aligned}$$

3d

Comment on the difference between the MSE's of these two forecasts.

The difference MSE of the estimates in 3b and 3c is $\sigma^2\theta^{2+2n}$. Because the process is invertible, we know $|\theta| < 1$. Thus when n is large, this difference is very small, so the truncated estimate works almost as good as the non-truncated one.

4

Consider an invertible MA(q) model $X_t = \theta(B)W_t$ for some white noise $\{W_t\}$ with variance σ^2.

4a

Show that for any $m > q$ the best linear predictor of X_{n+m} based on $X_1, \ldots, X_n$ is always zero.

The best linear predictor is $a_1^\star X_1 + \ldots + a_n^\star X_n$ where $a^\star$ solves a linear system of equations $\Delta a^\star = \zeta$ with $\zeta_i = Cov(X_{n+m}, X_i) = \gamma(n + m - i)$ for $i = 1, \ldots, n$. When $m > q$, then $n + m - i \geq m > q$ and thus $\zeta_i = Cov(X_{n+m}, X_i) = \gamma(n + m - i) = 0$ for all $i = 1, \ldots, n$. Thus, we deduce that $a_1^\star = \ldots = a_n^\star = 0$.

4b

Now assume that the white noise $\{W_t\}$ is also i.i.d.. Show that for any $m > q$ the best predictor (minimum mean-square error forecast) of X_{n+m} based on the full history $X_n, X_{n-1}, X_{n-2}, \ldots$ is also zero.

We have that X_{n+m} only depends on $W_{n+m}, \ldots, W_{n+m-q}$ and $X_n, X_{n-1}, \ldots$ only depends on $W_n, W_{n-1}, \ldots$. Thus, when $m > q$ for i.i.d. noise, we have that X_{n+m} is independent of $X_n, X_{n+1}, \ldots$ and thus $E(X_{n+m}|X_n, \ldots) = E(X_{n+m}) = 0$.