Bartlett's Formula

Jared Fisher

Lecture 7a

Announcements

Announcements

- Homework 4 is due tomorrow, Wednesday March 17 by 11:59pm.
- Project checkpoint 3 will be due Wednesday March 31, after Spring Break
- There is no lecture, lab, or office hours next week.

Recap: Theoretical vs Sample ACVF (and ACF)

Where we are at

- We introduced the sample ACVF and ACF equations at the beginning of the semester. (These don't change)
- We discussed how to find the theoretical ACVF last time (which is used to get the theoretical ACF).

Autocovariance of ARMA by “diving polynomials”

- Recall the covariance of MA(∞):

$$\gamma(h) = \sigma_W^2 \sum_{j=0}^{\infty} \theta_j \theta_{j+h}, \qquad \text{for } h \geq 0$$

- Recall that a causal, stationary ARMA process can be explicitly written as MA(∞):

$$X_t = \psi(B) W_t = \psi_0 W_t + \psi_1 W_{t-1} + \psi_2 W_{t-2} + \dots$$

  where $\psi(z) = \theta(z)/\phi(z)$.

- Note that ψ_0 will always equal one.

Autocovariance of ARMA by "dividing polynomials"

- Thus the ACVF of ARMA(p,q):

$$\gamma_X(h) = \text{cov}(X_t, X_{t+h}) = \sigma_W^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+h}, \text{ for } h \geq 0,$$

  where σ_W^2 is the variance of the white noise process W_t.

- Solve for ψ by "dividing polynomials" $\psi(z) = \theta(z)/\phi(z)$, or $\phi(z)\psi(z) = \theta(z)$ which expanded is

$$(1 - \phi_1 z - \cdots - \phi_p z^p)(\psi_0 + \psi_1 z + \dots) = 1 + \theta_1 z + \cdots + \theta_q z^q.$$

Autocovariance of ARMA by "solving difference equations"

- Another way: solve the difference equation $\phi(B)X_t = \theta(B)W_t$ by taking covariance of both sides with respect to X_{t-k}:

$$cov(\phi(B)X_t, X_{t-k}) = cov(\theta(B)W_t, X_{t-k})$$

- $$\gamma_X(k) - \phi_1\gamma_X(k-1) - \cdots - \phi_p\gamma_X(k-p) = c_k$$

  with

$$c_k = \begin{cases} (\psi_0\theta_k + \psi_1\theta_{k+1} + \cdots + \psi_{q-k}\theta_q)\,\sigma_W^2 & \text{for } 0 \le k \le q \\ 0 & \text{for } k > q. \end{cases}$$

Example: ARMA(1,1)

- Recall: in our theory work, we assume ϕ, θ, σ^2,etc. are known such that our results are functions of the parameters
- $\{X_t\}$ satisfies $X_t - \phi X_{t-1} = W_t + \theta W_{t-1}$.
- Note, $p = 1, \phi_0 = 1, \phi_1 = \phi$, and $q = 1, \theta_0 = 1, \theta_1 = \theta$.
- Then

$$\gamma_X(k) - \phi\gamma_X(k-1) = \begin{cases} (\psi_0\theta_k + \psi_1\theta_{k+1})\,\sigma_W^2 & \text{for } 0 \leq k \leq 1 \\ 0 & \text{for } k > 1. \end{cases}$$

Example: ARMA(1,1)

- Then for $k = 0$:

$$\gamma_X(0) - \phi\gamma_X(-1) = \gamma_X(0) - \phi\gamma_X(1) = (1 + \psi_1\theta_1)\sigma_W^2$$

- For $k = 1$:

$$\gamma_X(1) - \phi\gamma_X(0) = \theta\sigma_W^2$$

- For $k \geq 2$:

$$\gamma_X(k) = \phi\gamma_X(k-1)$$

Example: ARMA(1,1)

- The number ψ_1 is the coefficient of z in

$$(1-\phi z)(1+\psi_1 z+\psi^2 z^2+\dots)=1+\theta z$$

  and equals $\theta+\phi$.

- Solving the first two equations, we get

$$\gamma_X(0)=\sigma_W^2\frac{1+\theta^2+2\phi\theta}{1-\phi^2}\quad \gamma_X(1)=\sigma_W^2\frac{(\theta+\phi)(1+\theta\phi)}{1-\phi^2}.$$

- The last equation(s) gives

$$\gamma_X(k)=\phi^{k-1}\sigma_W^2\frac{(\theta+\phi)(1+\theta\phi)}{1-\phi^2}.$$

Example: ARMA(1,1)

- Autocorrelations:

$$\rho_X(k) = \frac{(\theta + \phi)(1 + \theta\phi)}{1 + \theta^2 + 2\phi\theta}\phi^{k-1} \text{ for } k \geq 1.$$

- Note that the autocorrelation at lag one is not equal to ϕ as this is not AR(1). But what happens if $\theta = 0$?
- After lag one, subsequent autocorrelations decay exponentially with factor ϕ.

So what?

The Question on everyone's mind:

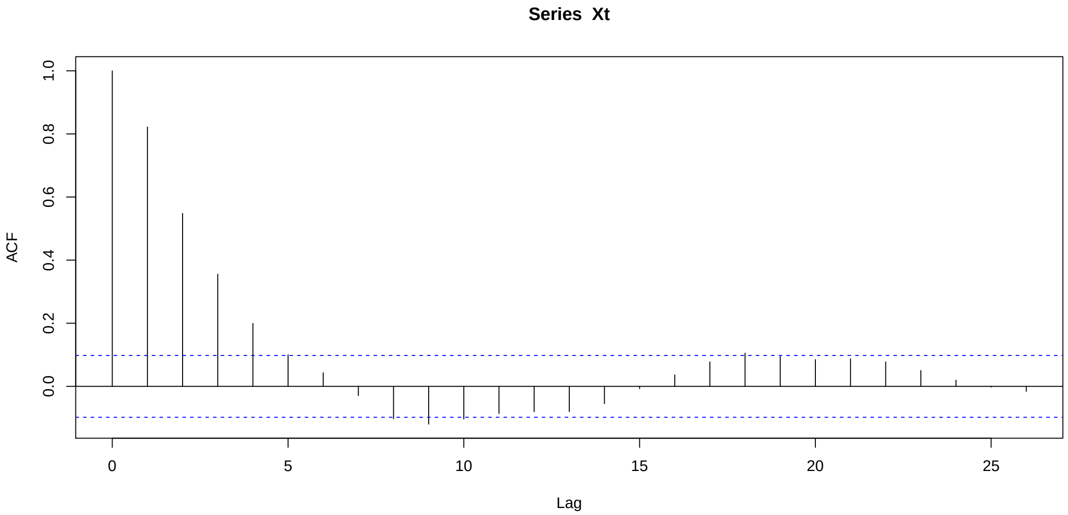
- How can one decide whether ARMA(p,q) for some orders p, q is a good model for an observed time series $\{X_t\}$?
- Let's go back to our ARMA(1,1) example. How can we see if this is a reasonable model of a stationary process? Why did we just work to find

$$\rho_X(k) = \frac{(\theta + \phi)(1 + \theta\phi)}{1 + \theta^2 + 2\phi\theta}\phi^{k-1} \text{ for } k \geq 1?$$

- Recall the big picture: the stationary process has constant mean, constant variance, and covariance that only depends on the lag: $cov(X_t, X_{t+h}) = \gamma_X(h)$.
- But $\gamma_X(h)$ may be different for different values of h, and thus the ACF $\rho_X(h)$ can be too.
- We can check to see if our sample ACF values r_h from the data are reasonably close to the ACF values $\rho_X(h)$ from a proposed model.

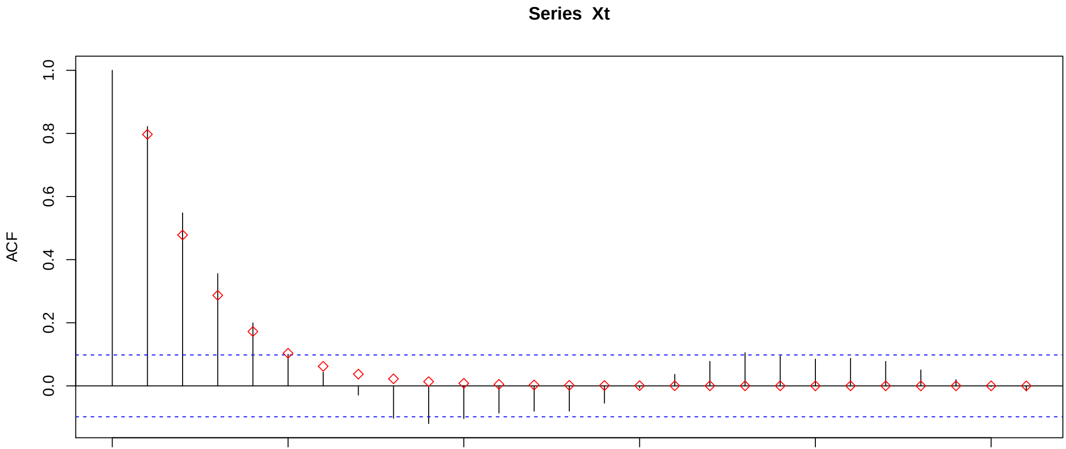
ARMA(1,1) ACF plot (sample values r_h)

```
Xt = arima.sim(model = list(ar = .6,ma=.8),n=400)
acf(Xt)
```



ARMA(1,1) ACF plot (add $\rho_X(h)$ in red)

```
acf(Xt)
k = 1:26; theta = 0.8; phi = 0.6
rho = ((theta+phi)*(1+theta*phi))/(1+theta^2+2*phi*theta)*phi^(k-1)
points(k,rho,col='red',pch=5)
```



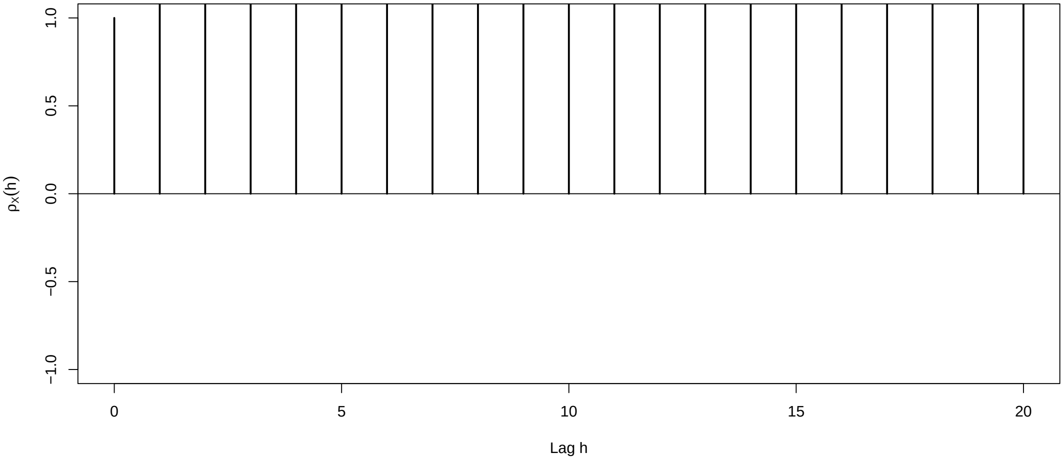
Other ACF Plots (theoretical, so no uncertainty/blue lines)

Make a new function

```
plot.theoretical.acf = function(ar=numeric(),ma=numeric()
                                ,lag.max=20){
  plot(0:lag.max, ARMAacf(ar=ar,ma=ma,lag.max = lag.max)
       ,type='h'
       ,lwd=2
       ,ylim=c(-1,1)
       ,xlab='Lag h'
       ,ylab=expression(rho[X](h))
       )
  abline(h=0)
}
```

AR(1), $\phi = 1.5\ldots$? Be careful!

```
plot.theoretical.acf(ar=1.5)
```

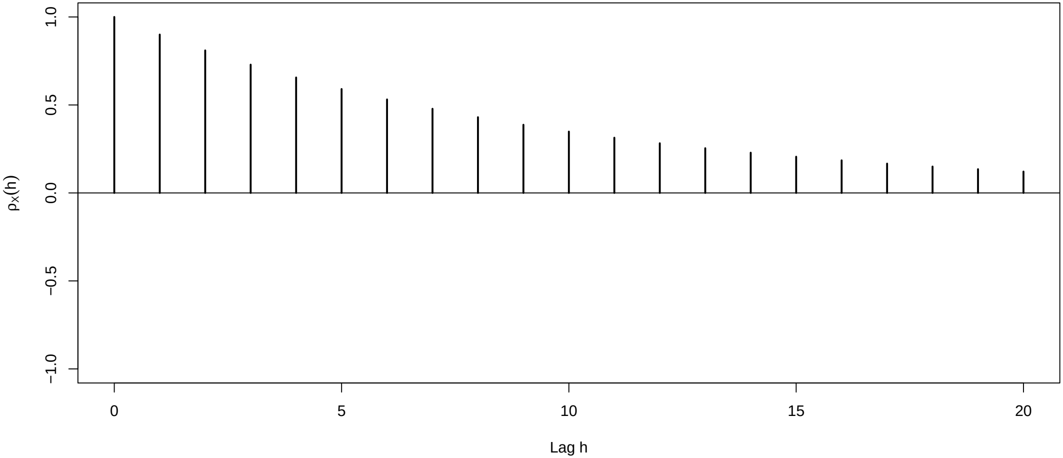


Careful

- The ARMAacf() function doesn't check for causality/invertibility, but arima.sim() will.
- Using *arima.sim(model=list(ar=1.5),n=20)* yields
- *Error in arima.sim(model = list(ar = 1.5), n = 20) : 'ar' part of model is not stationary*

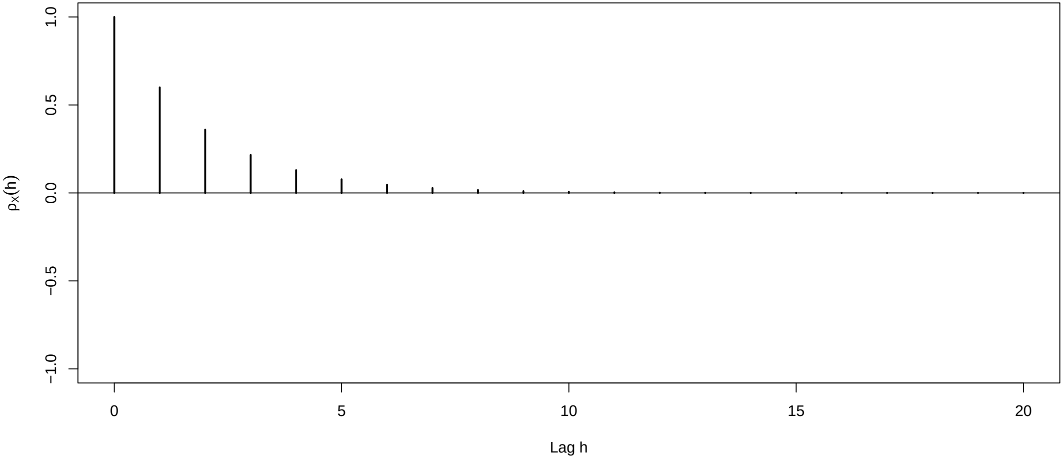
AR(1), $\phi = .9$

```r
plot.theoretical.acf(ar=.9)
```



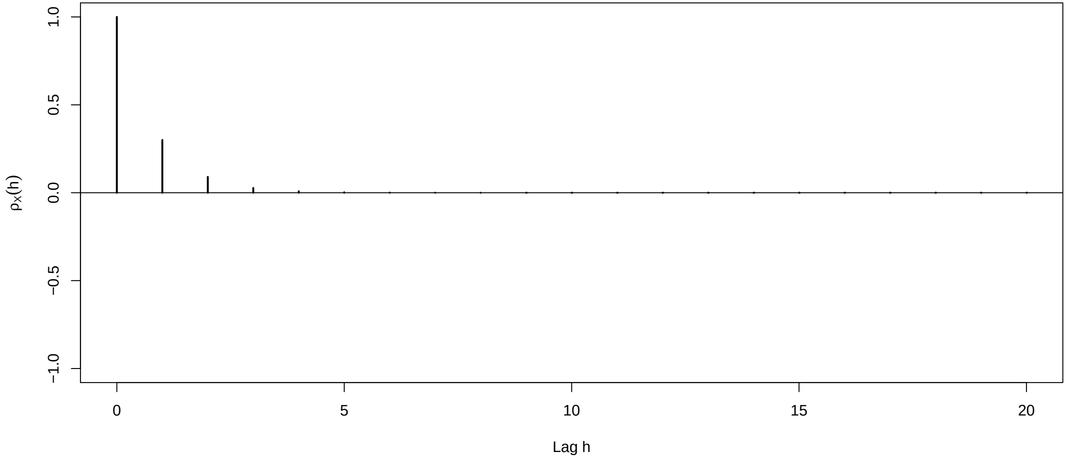
AR(1), $\phi = .6$

```
plot.theoretical.acf(ar=.6)
```



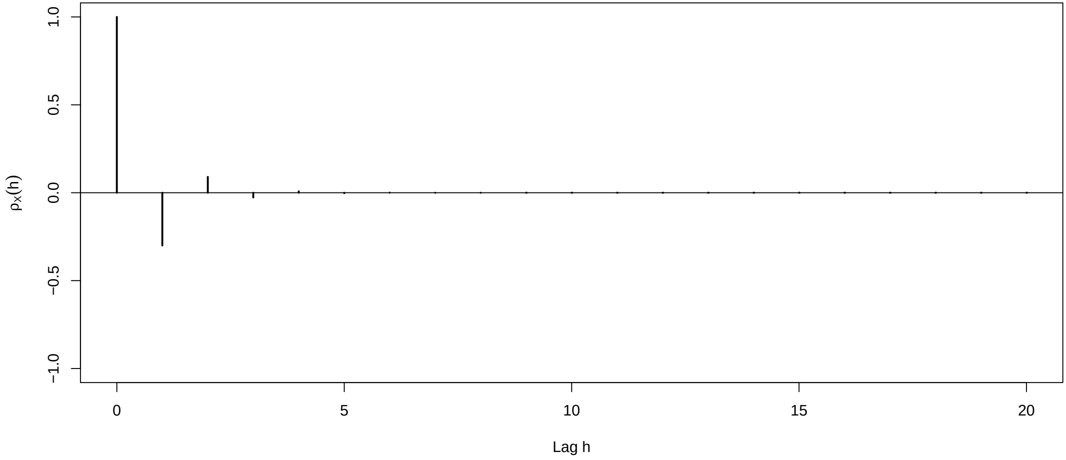
AR(1), $\phi = .3$

```
plot.theoretical.acf(ar=.3)
```



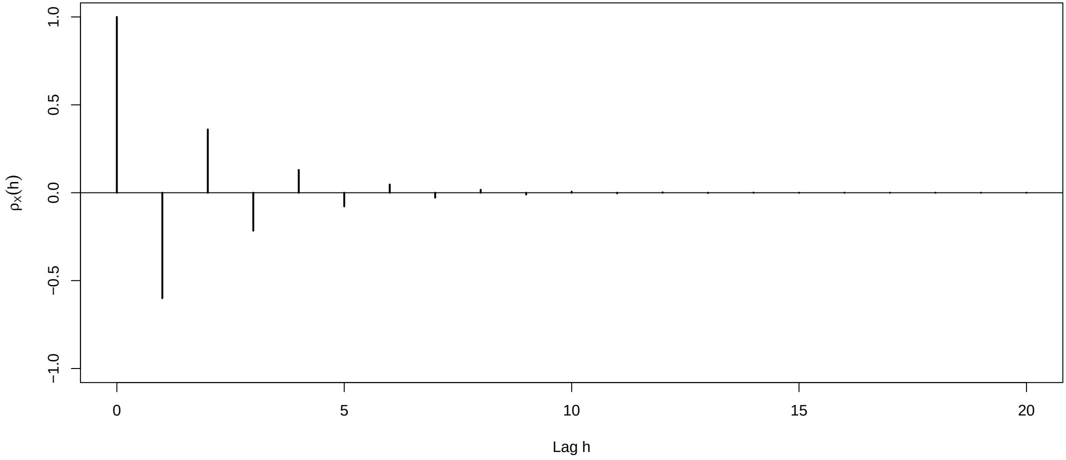
AR(1), $\phi = -0.3$

```
plot.theoretical.acf(ar=-.3)
```



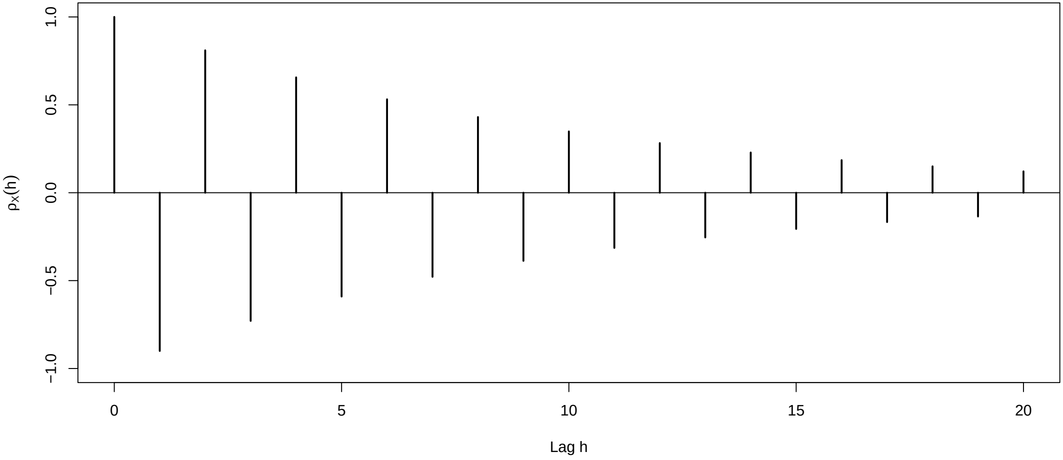
AR(1), $\phi = -0.6$

```r
plot.theoretical.acf(ar=-.6)
```



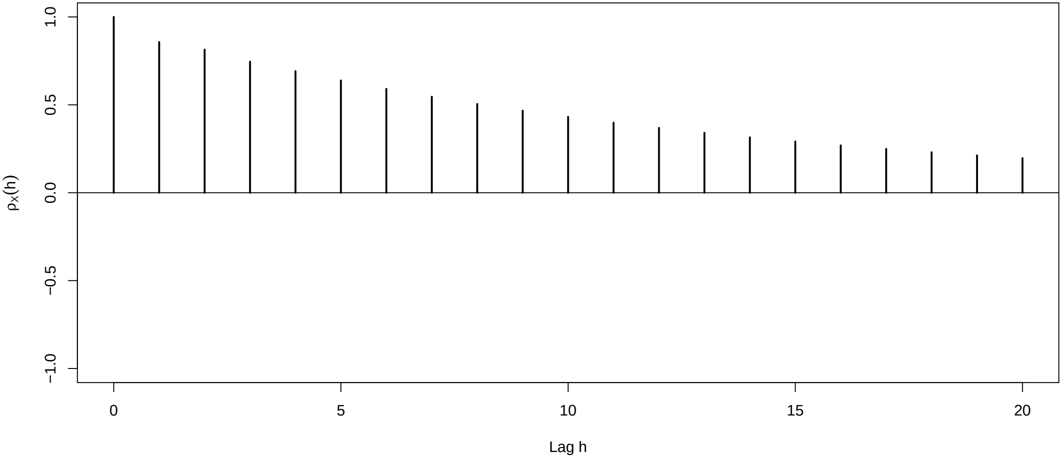
AR(1), $\phi = -0.9$

```r
plot.theoretical.acf(ar=-.9)
```



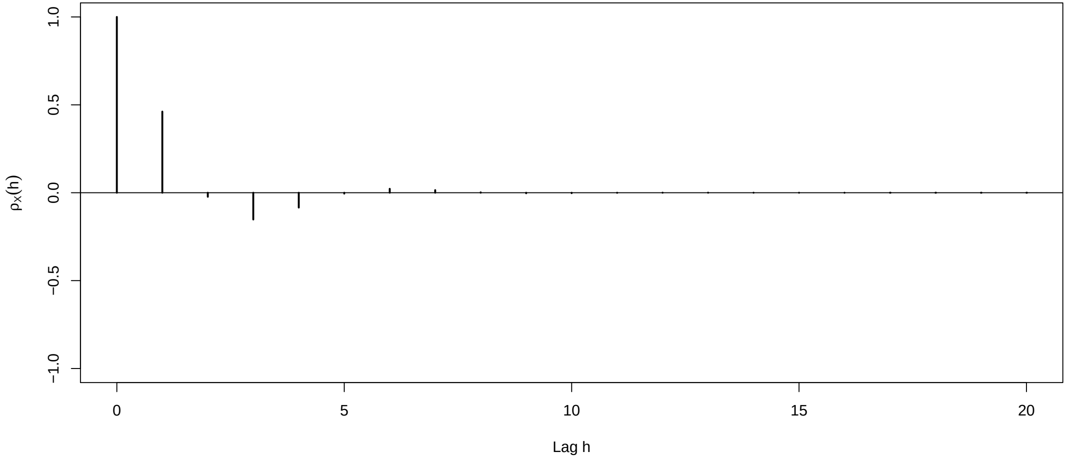
AR(2), $\phi_1 = 0.6$, $\phi_2 = 0.3$

```r
plot.theoretical.acf(ar=c(.6,.3))
```



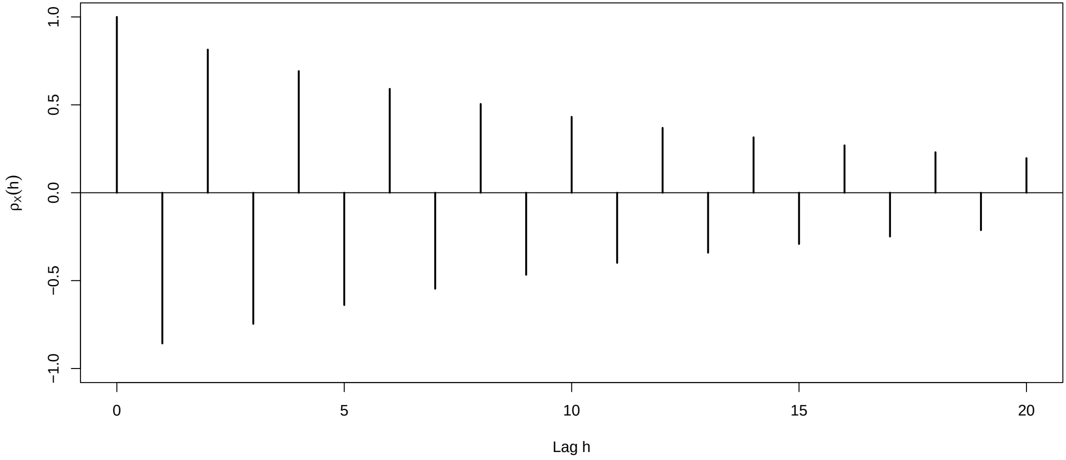
AR(2), $\phi_1 = 0.6$, $\phi_2 = -0.3$

```
plot.theoretical.acf(ar=c(.6,-.3))
```



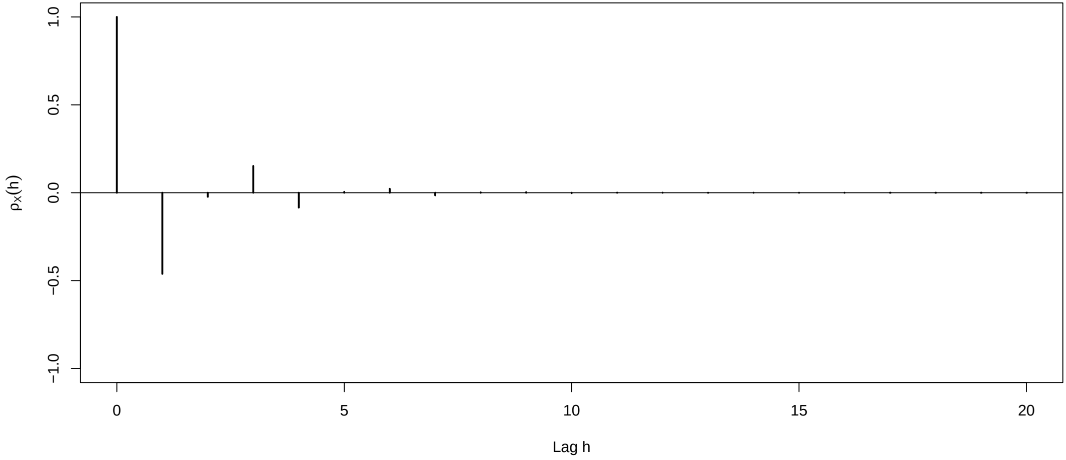
AR(2), $\phi_1 = -0.6$, $\phi_2 = 0.3$

```r
plot.theoretical.acf(ar=c(-.6,.3))
```



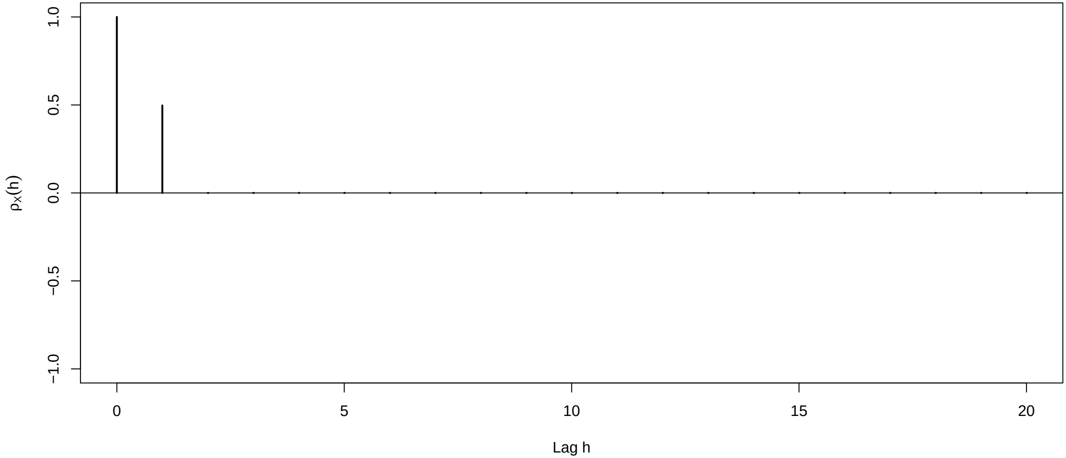
AR(2), $\phi_1 = -0.6$, $\phi_2 = -0.3$

```r
plot.theoretical.acf(ar=c(-.6,-.3))
```



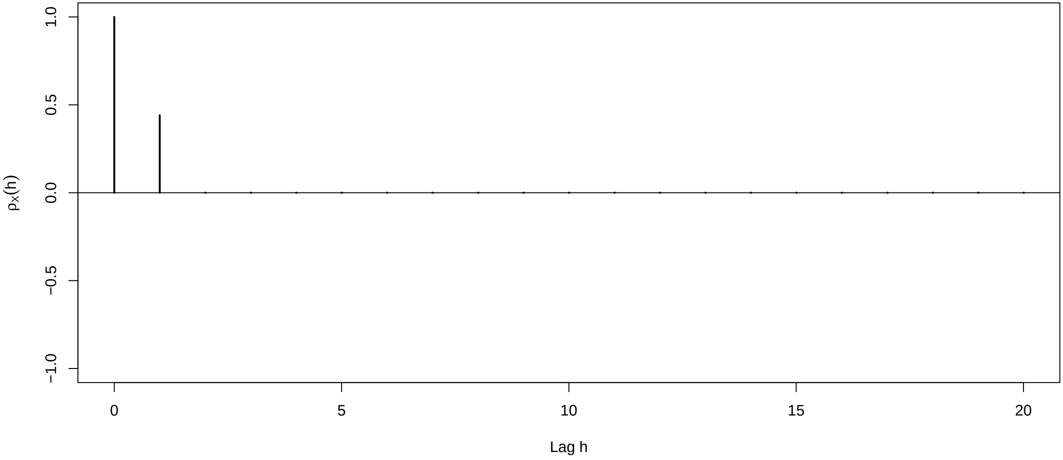
MA(1), $\theta = 0.9$

```r
plot.theoretical.acf(ma=0.9)
```



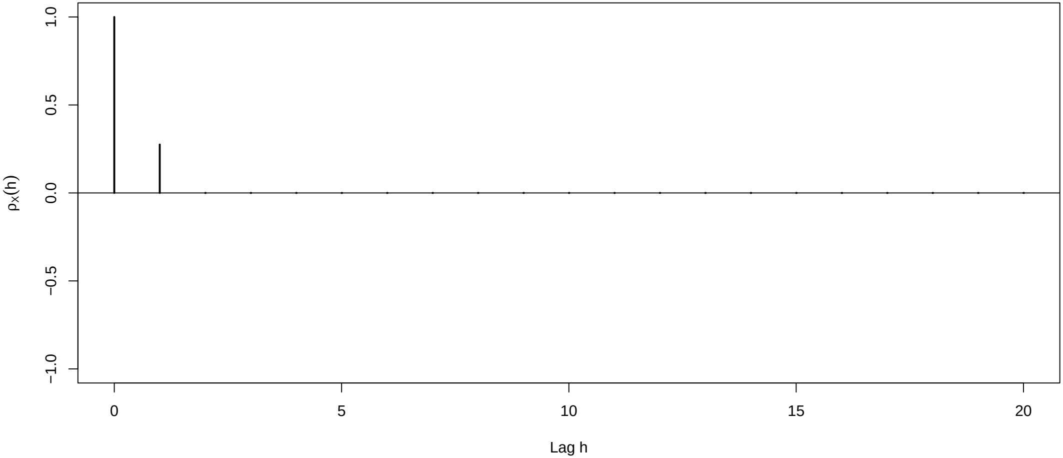
MA(1), $\theta = 0.6$

```
plot.theoretical.acf(ma=0.6)
```



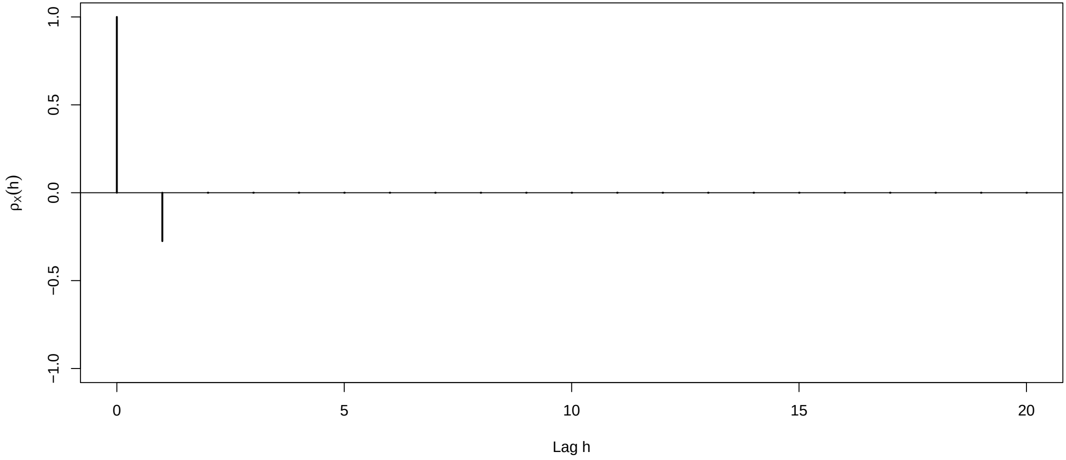
MA(1), $\theta = 0.3$

```
plot.theoretical.acf(ma=0.3)
```



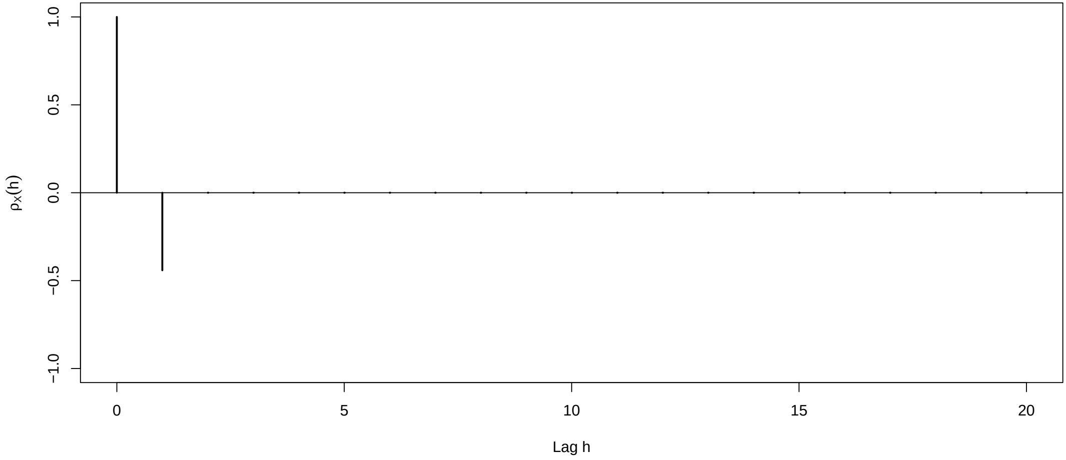
MA(1), $\theta = -0.3$

```r
plot.theoretical.acf(ma=-0.3)
```



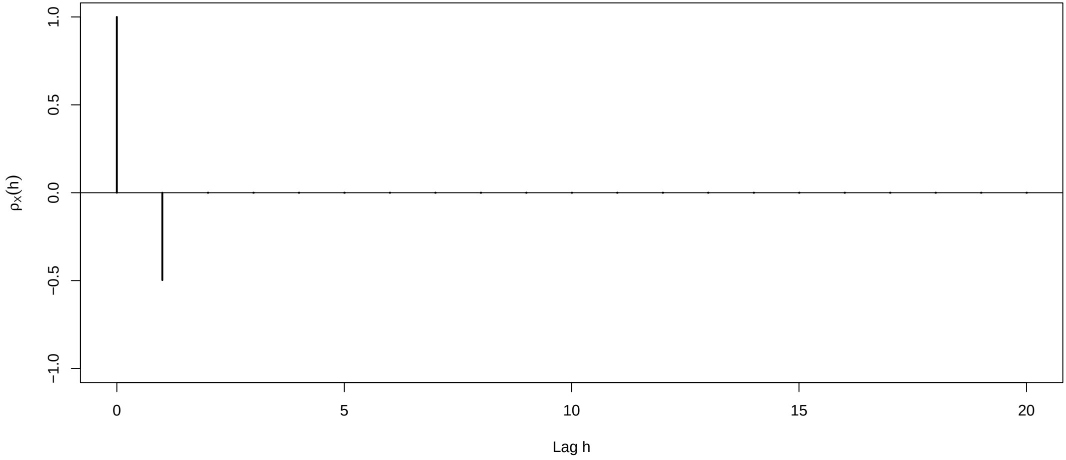
MA(1), $\theta = -0.6$

```
plot.theoretical.acf(ma=-0.6)
```



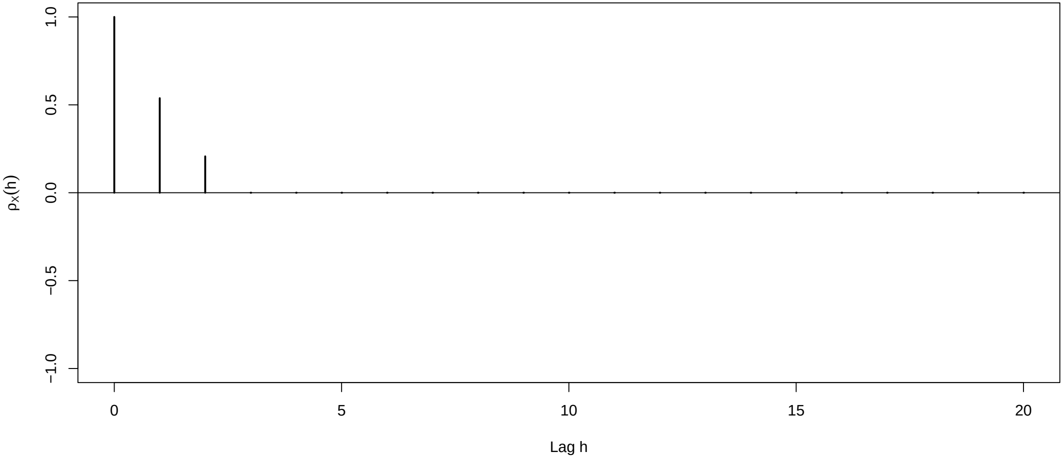
MA(1), $\theta = -0.9$

```
plot.theoretical.acf(ma=-0.9)
```



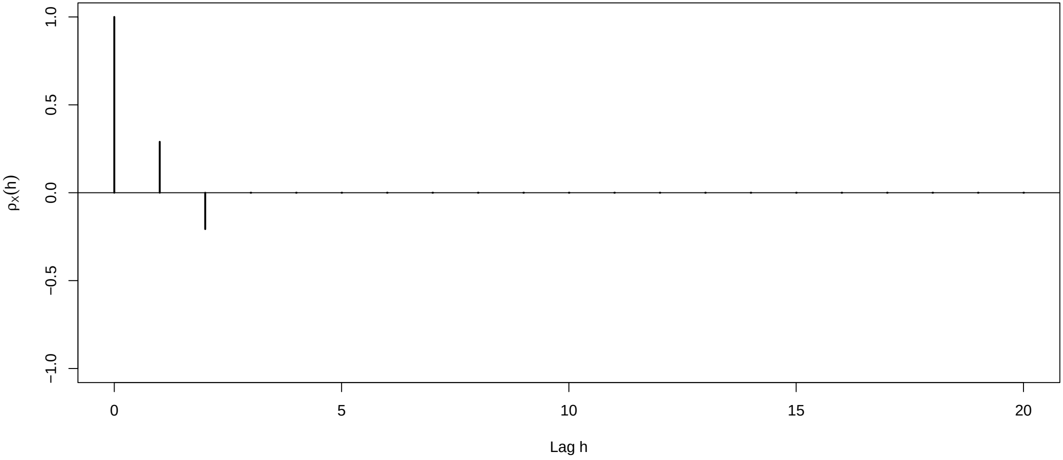
MA(2), $\theta_1 = 0.6$, $\theta_2 = 0.3$

```r
plot.theoretical.acf(ma=c(0.6,0.3))
```



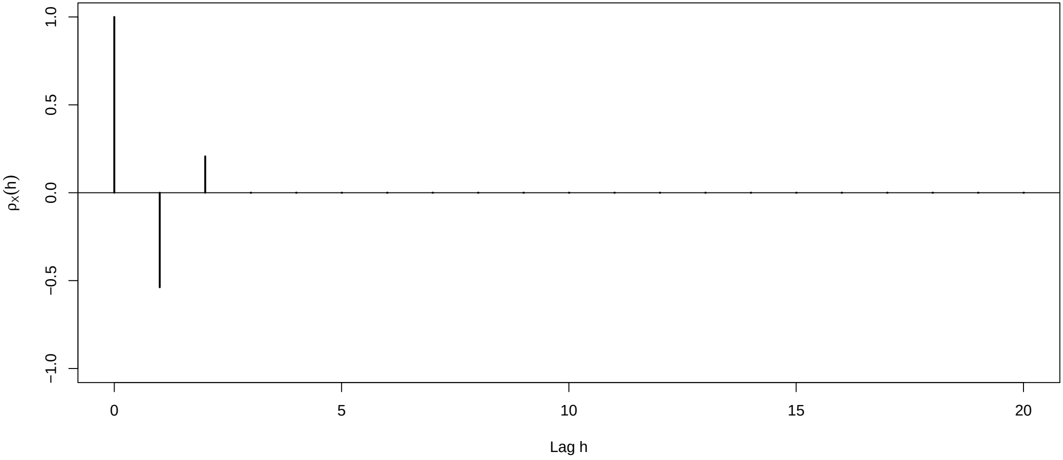
MA(2), $\theta_1 = 0.6$, $\theta_2 = -0.3$

```
plot.theoretical.acf(ma=c(0.6,-0.3))
```



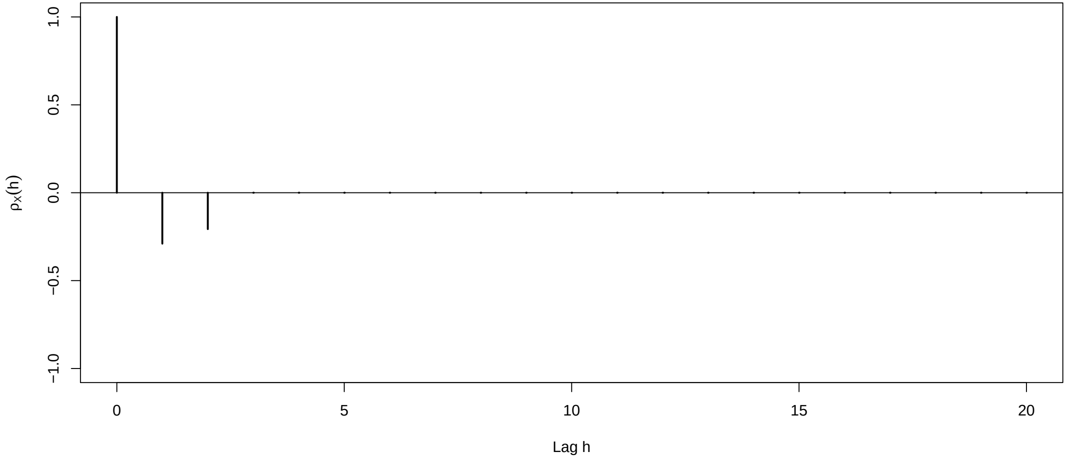
MA(2), $\theta_1 = -0.6$, $\theta_2 = 0.3$

```
plot.theoretical.acf(ma=c(-0.6,0.3))
```



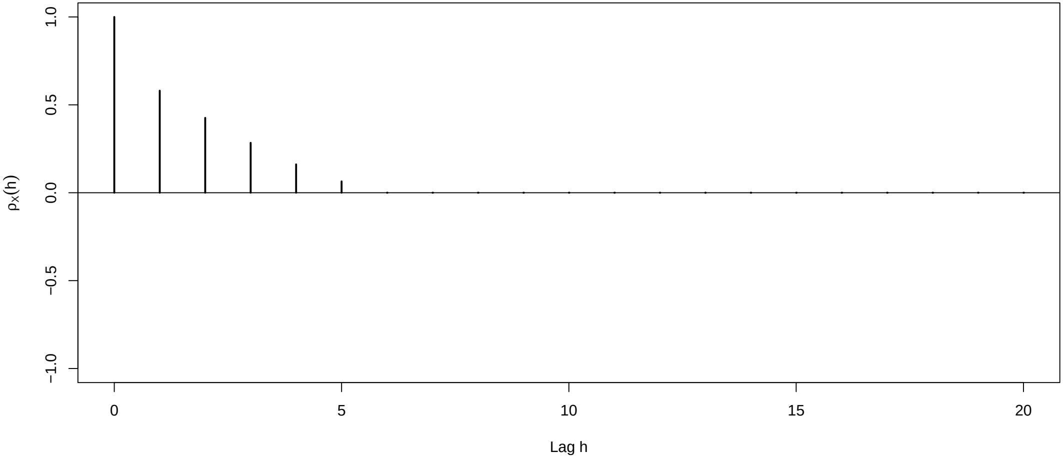
MA(2), $\theta_1 = -0.6$, $\theta_2 = -0.3$

```r
plot.theoretical.acf(ma=c(-0.6,-0.3))
```



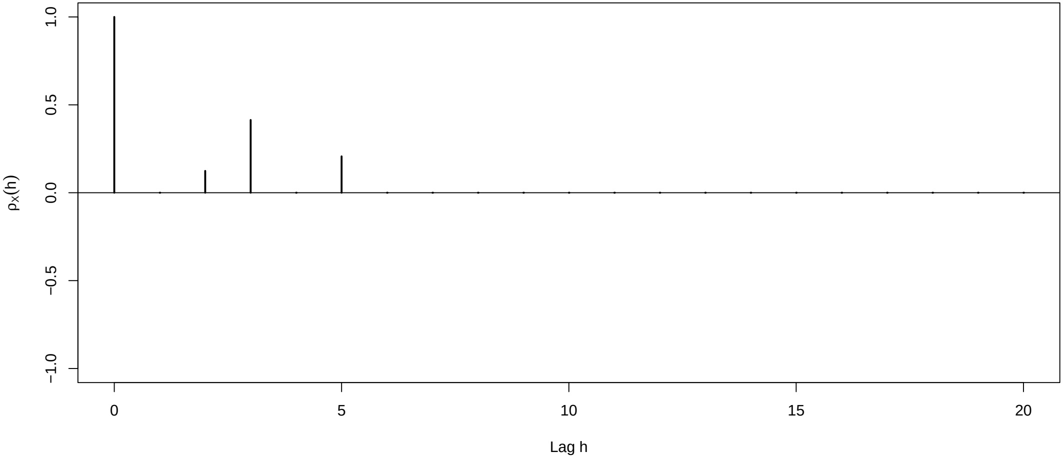
MA plays nice: MA(5), $\theta_1 = 0.5$, $\theta_2 = 0.4$, $\theta_3 = 0.3$, $\theta_4 = 0.2$, $\theta_5 = 0.1$

```
plot.theoretical.acf(ma=c(0.5,0.4,0.3,0.2,0.1))
```



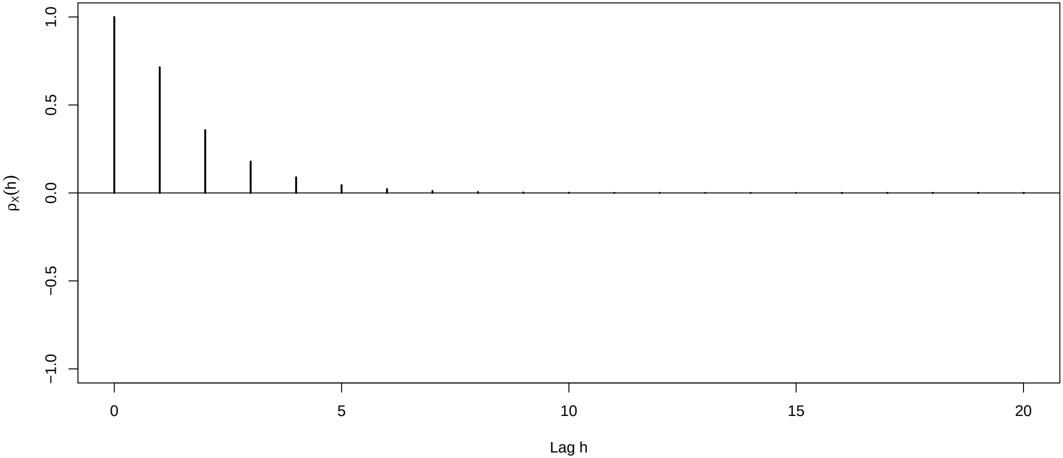
Well, usually: MA(5), $\theta_1 = \theta_2 = \theta_4 = 0$, $\theta_3 = 0.6$, $\theta_5 = 0.3$

```r
plot.theoretical.acf(ma=c(0,0,0.6,0,0.3))
```



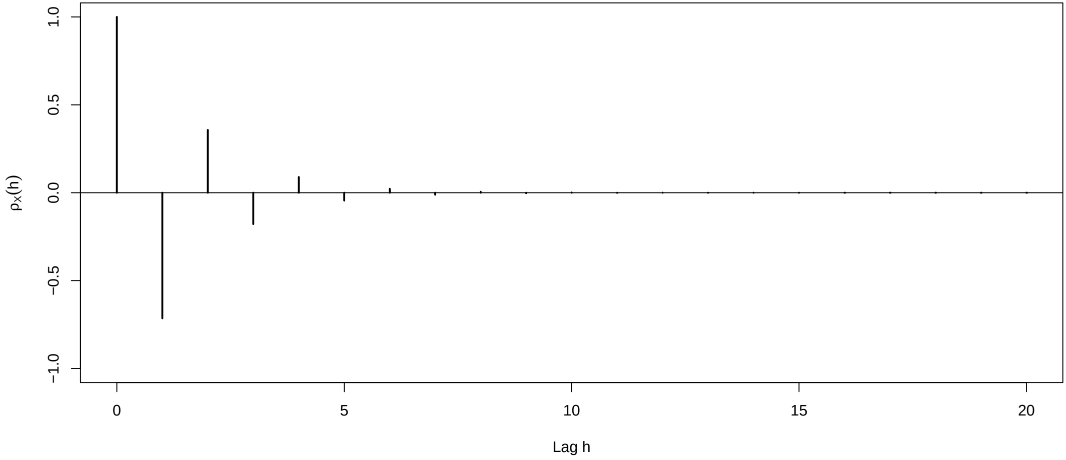
ARMA(1,1) $\theta = 0.5$, $\phi = 0.5$

```
plot.theoretical.acf(ma=0.5,ar=0.5)
```



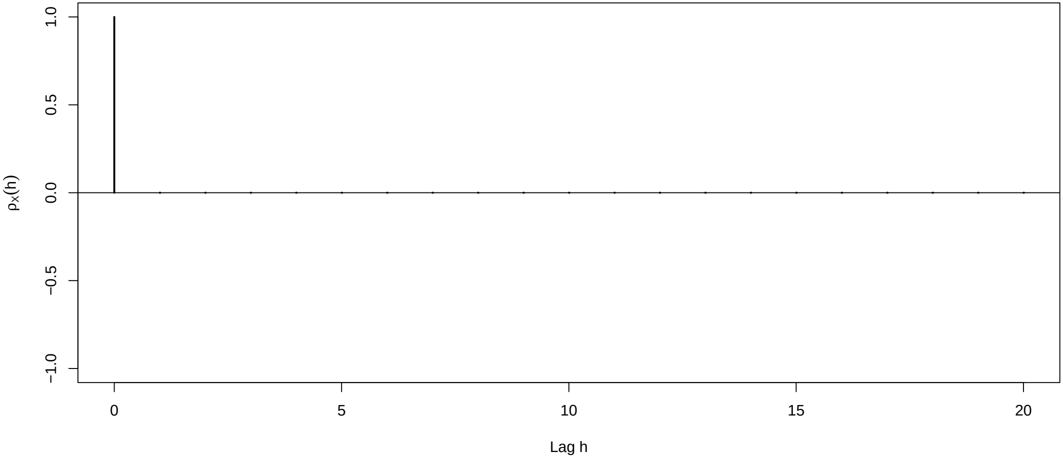
ARMA(1,1) $\theta = -0.5$, $\phi = -0.5$

```
plot.theoretical.acf(ma=-0.5,ar=-0.5)
```



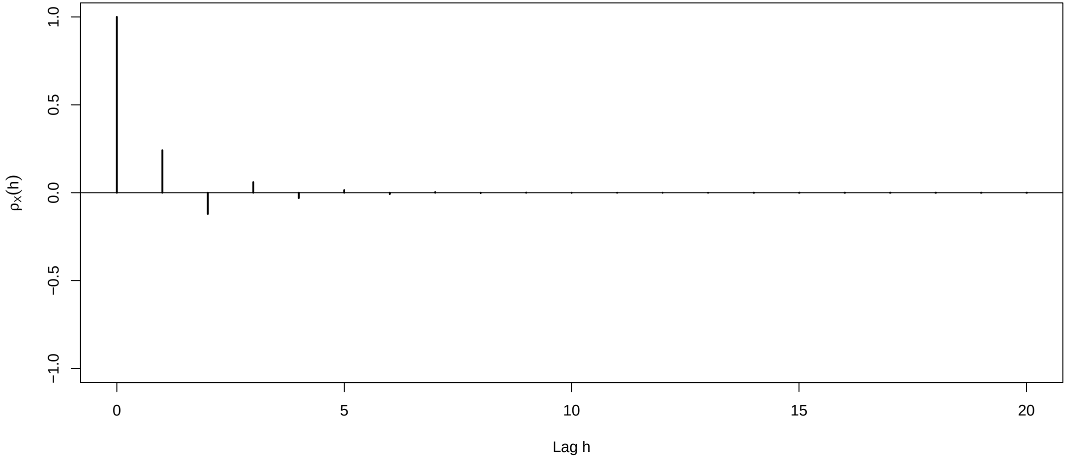
ARMA(1,1) $\theta = 0.5$, $\phi = -0.5\ldots$ What happened?

```
plot.theoretical.acf(ma=0.5,ar=-0.5)
```



ARMA(1,1) $\theta = 0.9$, $\phi = -0.5$

```r
plot.theoretical.acf(ma=0.9,ar=-0.5)
```



ARMA: a flexible way to model ACF's

- ARMA models can fit many forms of the ACF, but this brings up two questions:

1. What we've seen are (essentially) the expected values of the autocorrelations. What is reasonable uncertainty around these values?
2. It seems to be much easier to distiguish different values of q (the MA order) than it does p (the AR order). In other words, some AR(1) ACFs look similar to AR(2) ACFs. How can we differentiate, especially considering p can be larger than 2?

- We will look into #1 today and #2 over the next couple lectures.

Change gears: Approximate distribution of sample autocorrelations

Address uncertainty (i.e. variance)

- We've been looking at the expected values of the ACF, but how much uncertainty is there?

- Remember that for the white noise model of residuals, we could simply plot the sample autocorrelations $r_1, \ldots, r_k$ (denoted as correlogram) and check whether this looks like i.i.d. Gaussian with variance $1/n$.

- Recall that theorem: Under some general conditions, if $X_1, \ldots, X_n$ is white noise, then for any fixed lag k and n large enough, the sample autocorrelations $r_1, r_2, \ldots, r_k$ are approximately independent normally distribution with mean zero and variance $1/n$, that is,

$$\sqrt{n}\begin{pmatrix} r_1 \\ \vdots \\ r_k \end{pmatrix} \to N(0, I) \qquad \text{as } n \to \infty, \tag{1}$$

where I denotes the $k \times k$ identity matrix.

Distribution of ARMA(p,q) Autocorrelations

- We can do something similar for ARMA(p,q) as the following theorem shows.
- The proof is not easy and can be found in Brockwell and Davis, 1991, Theorem 7.2.2. See also Remark 1, 2.

Theorem: Bartlett's Formula

Under some general conditions on the white noise process $\{W_t\}$, if $\{X_t\}$ is an causal and invertible ARMA process $\phi(B)X_t = \theta(B)W_t$, then for any fixed lag k and n large enough the sample autocorrelations $(r_1, r_2, \ldots, r_k)$ are approximately multivariate normal distributed with mean $(\rho_X(1), \ldots, \rho_X(k))$ and covariance matrix W/n with (i,j)th entry equal to $W_{ij} =$

$$\sum_{m=1}^{\infty} [(\rho_X(m+i) + \rho_X(m-i) - 2\rho_X(i)\rho_X(m)) \\ * (\rho_X(m+j) + \rho_X(m-j) - 2\rho_X(j)\rho_X(m))]$$

that is,

$$\sqrt{n}\left(\begin{pmatrix} r_1 \\ \vdots \\ r_k \end{pmatrix} - \begin{pmatrix} \rho_X(1) \\ \vdots \\ \rho_X(k) \end{pmatrix}\right) \to N(0, W) \qquad \text{as } n \to \infty.$$

Equivalence for ARMA(0,0)

- Does Bartlett's formula for white noise $X_t = W_t$ yield the same results as the white noise ACF Theorem?
- Check on the board (yes it does)

ARMA(p,q)

- In particular, for each individual r_i, Bartlett's formula yields that under an ARMA(p,q) model

$$P(|r_i - \rho_X(i)| \geq 1.96\sqrt{W_{ii}/n}) \approx 5\%.$$

- Analog for the expected number of $r_i's$ which lie outside of the $\rho_X(i) \pm 1.96\sqrt{W_{ii}/n}$ band we find that

$$E\left(\#\left\{i = 1, \ldots, k \ : \ |r_i - \rho_X(i)| \geq 1.96\sqrt{W_{ii}/n}\right\}\right) \approx k \cdot 5\%.$$

- Word of caution: unless $W_{ii} = 1$, the blue lines on the acf() plot in R aren't the interval for ARMA as $1.96\frac{1}{\sqrt{n}} \neq 1.96\sqrt{W_{ii}/n}$.

Example: MA(1)

- Suppose $X_t = W_t + \theta W_{t-1}$.
- We have seen that $\rho_X(1) = \theta/(1+\theta^2)$ and $\rho_X(h) = 0$ for higher lags h.
- Bartlett's formula says that the variance of r_i is approximately W_{ii}/n where

$$W_{ii} = \sum_{m=1}^{\infty} \left(\rho(m+i) + \rho(m-i) - 2\rho(i)\rho(m)\right)^2 .$$

- What's the variance of r_1? Let's find W_{ii} on the board

Example: MA(1)

- For $i = 1$ i.e., when we consider the first order sample autocorrelation, this formula gives
$$var(r_1) \approx W_{11}/n = (1 - 3\rho^2(1) + 4\rho^4(1))/n < 1/n$$
.
- This last inequality requires noting that for any θ we have $\rho^2(1) = \frac{\theta^2}{(1+\theta^2)^2} \leq 1/4$.
- In other words, r_1 for MA(1) is less variable than r_1 for white noise.

Example: MA(1)

- For higher values of i, the formula gives

$$W_{ii} = \sum_{m=1}^{\infty} \rho^2(m-i) = 1 + 2\rho^2(1) > 1.$$

- In other words, r_k for $k \geq 2$ are more variable for MA(1) than for white noise.
- Thus we can expect to see more r_k's sticking out the horizontal blue lines for MA(1).

Empirical Strategy

A general strategy to find out whether ARMA(p,q) is a good model for data is as follows:

1. Plot the correlogram $(r_1, \ldots, r_k)$.
2. Compare this with the theoretical ACF $\rho_X(h)$ of the ARMA(p,q) model.
3. Keep in mind the variability of the r_k's given by Bartlett's formula

Empirical Strategy

- For example, when the sample autocorrelations after lag q drop off and lie between the band for MA(q) given by Bartlett's formula, the MA(q) model might be appropriate.
- For AR(p) models the ACF does not drop to zero for large lags. Thus, it is more difficult to choose the order of an appropriate AR process for data by looking at the sample ACF.
- We will later introduce the *partial autocorrelation function (PACF)*, which will be more helpful for AR(p) models: the PACF of an AR(p) model is zero for lags strictly larger than p.