STAT 520, Fall 2015: Exam I

This exam was posted online Tuesday October 13, 2015 at noon. Complete the exam, scan your solutions, and email to Yawei Liang (yliang@email.sc.edu) by noon Wednesday October 14, as usual. You are to work completely independly on this exam; however you may use notes, your textbook, Google, etc.

1. Let $\{Y_t\}$ be a stationary process with constant mean $E(Y_t) = \mu$ and autocorrelation function ρ_k. Define $\overline{Y} = \frac{1}{n}\sum_{t=1}^{n} Y_t$ to be the sample mean of $Y_1, Y_2, \ldots, Y_n$. Recall that

$$\text{var}(\overline{Y}) = \frac{\gamma_0}{n}\left[1 + 2\sum_{k=1}^{n-1}\left(1 - \frac{k}{n}\right)\rho_k\right],$$

where $\gamma_0 = \text{var}(Y_t)$. Find $\text{var}(\overline{Y})$ when $\{Y_t\}$ is ARMA(1,1) with parameters ϕ and θ. Find one pair of values (ϕ, θ) that makes $\text{var}(\bar{Y})$ smaller than that for white noise when $n = 3$. By what fraction is the variance reduced?

The model is

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1}.$$

Page 78 in Cryer and Chan gives

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

So

$$\text{var}(\bar{Y}) = \frac{\gamma_0}{n}\left[1 + 2\sum_{k=1}^{n-1}(1 - \tfrac{k}{n})\frac{(1-\theta\phi)(\phi-\theta)}{1-2\phi\theta+\theta^2}\phi^k\right].$$

For $n = 3$ this reduces to

$$\text{var}(\bar{Y}) = \frac{\gamma_0}{n}\left[1 + 2\frac{(1-\theta\phi)(\phi-\theta)}{1-2\phi\theta+\theta^2}(\tfrac{2}{3}\phi + \tfrac{1}{3}\phi^2)\right].$$

We need the part in square brackets $[\cdots]$ to be less than one to do better than white noise. I just used trial and error to find the pair $(\phi, \theta) = (0.5, 0.9)$. R code to verify:

```
> p=0.5; t=0.9
> 1+2*(1-t*p)*(p-t)*(2*p/3+p^2/3)/(1-2*p*t+t^2)
[1] 0.7985348
```

2. Let $\{e_t\}$ be mean-zero white noise, as usual, and consider the series

$$Y_t = 1.25Y_{t-1} - 0.125Y_{t-2} - 0.125Y_{t-3} + e_t + 0.25e_{t-1}.$$

(a) Write the series as $\phi(B)(1-B)^d Y_t = \theta(B)e_t$. That is, find $\phi(B)$, $\theta(B)$, and d so that the differenced series is stationary and invertible. You may use an online root finder for help if you want. Hint: as shown in class, root finders will rewrite the equation $1+c_1B+c_2B^2+c_3B^3=0$ as $a(B-r_1)(B-r_2)(B-r_3)=0$. You need to divide both sides of this latter equation by $ar_1r_2r_3$ and multiply by $(-1)^3=-1$ to get it in the form $(1-B/r_1)(1-B/r_2)(1-B/r_3)=0$.

This was trickier than I intended as a root is shared by the AR and MA characteristic polynomials; I graded this generously! The model is written initially

$$(1-1.25B+0.125B^2+0.125B^3)Y_t = (1+0.25B)e_t.$$

Using an online root finder gives

$$(1+B/4)(1-B/2)(1-B)Y_t = (1+B/4)e_t.$$

So at first glance $\{Y_t\}$ is ARIMA (2,1,1). However, the root -4 appears on both sides, so we must cancel this out obtaining the simpler series

$$(1-B/2)(1-B)Y_t = e_t,$$

an ARI(1,1).

(b) Obtain the population autocorrelation function (ACF) for the appropriately differenced series $W_t = \nabla^d Y_t$. You can use `ARMAacf` in R and report a plot.

Direct computation using the Yule-Walker equations yields $\rho_k = 2^{-k}$ for $\{\nabla Y_t\}$. Otherwise we can use R. Treating $\{Y_t\}$ (improperly) as ARIMA(2,1,1), the model can be rewritten

$$(1-0.25B-0.125B^2)(1-B)Y_t = (1+0.25B)e_t.$$

We multiply characteristic polynomials in the R code all by -1 to obtain the ACF for $W_t = \nabla Y_t$:

```
plot(ARMAacf(ar=c(0.25,0.125),ma=c(0.25),lag.max=10))
```

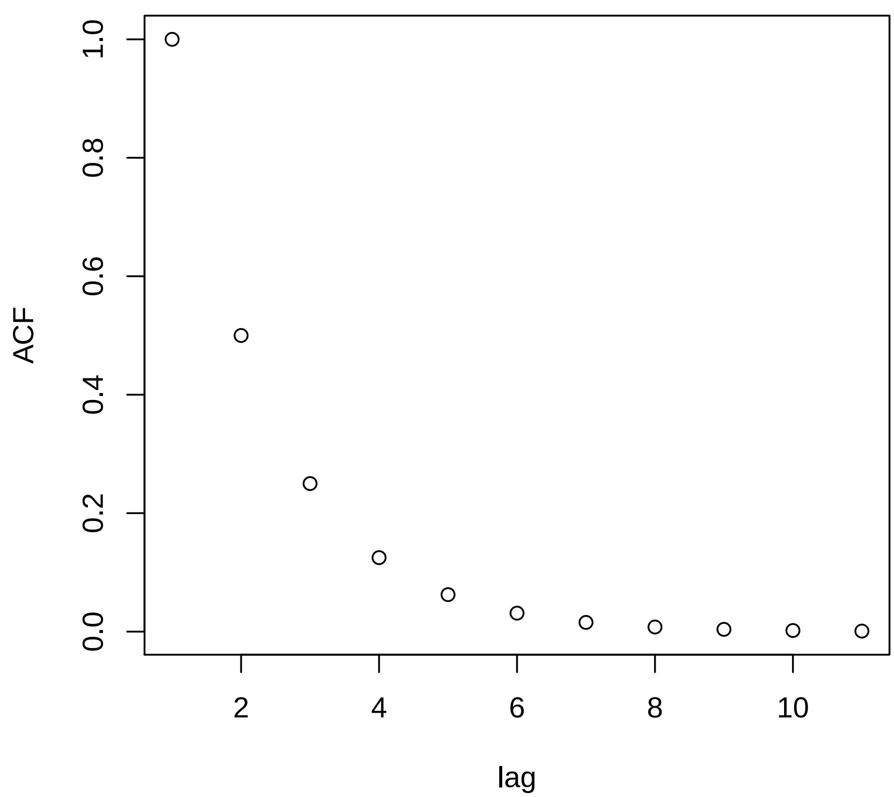
Treating $\{Y_t\}$ (properly) as ARI(1,1), the model can be rewritten

$$(1-0.5B)(1-B)Y_t = e_t.$$

The code is simply

```
plot(ARMAacf(ar=c(0.5),ma=c(),lag.max=10))
```

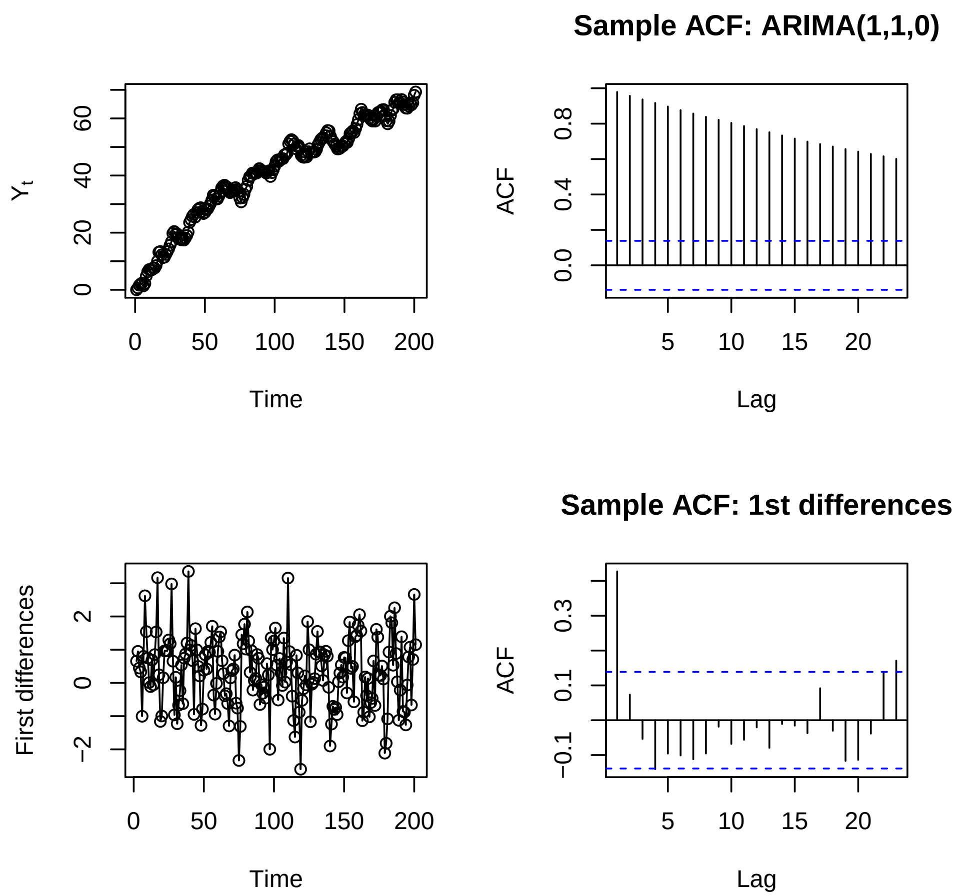
Note that either way, *the ACF function is correctly given and simulated from by R.*



(c) Simulate $n = 200$ from the series $\{Y_t\}$ assuming $e_t \overset{iid}{\sim} N(0, \sigma_e^2)$ where $\sigma_e = 1$ and plot it along with the sample ACF. Do the same for the differenced series $\{\nabla Y_t\}$. R code to simulate from an ARIMA(1,1,1) is in Chapter 5; you could modify this.

```
library(TSA)
par(mfrow=c(2,2))
y.sim=arima.sim(list(order=c(2,1,1),ar=c(0.25,0.125),ma=c(0.25)),n=200)
y.sim=arima.sim(list(order=c(1,1,0),ar=c(0.5),ma=c()),n=200) # also works
```

```
plot(y.sim,ylab=expression(Y[t]),xlab="Time",type="o")
acf(y.sim,main="Sample ACF: ARIMA(1,1,0)")
plot(diff(y.sim),ylab="First differences",xlab="Time",type="o")
acf(diff(y.sim),main="Sample ACF: 1st differences")
```



3. Let $\{Y_t\}$ be an AR(1) process with $|\phi| < 1$. Find the autocorrelation function for $W_t = \nabla Y_t = Y_t - Y_{t-1}$ in terms of ϕ and σ_e^2. What does this simplify to at lag $k = 0$, i.e. what is $\text{var}(W_t)$?

This is problem 4.6 in Cryer and Chan.

For AR(1) with $|\phi| < 1$ we know

$$\gamma_k = \frac{\sigma_e^2}{1-\phi^2}\phi^k, \text{ for } k \geq 0.$$

Then

$$\begin{aligned}
\text{cov}(W_t, W_{t-k}) &= \text{cov}(Y_t - Y_{t-1}, Y_{t-k} - Y_{t-k-1}) \\
&= \text{cov}(Y_t, Y_{t-k}) + \text{cov}(Y_t, -Y_{t-k-1}) + \text{cov}(-Y_{t-1}, Y_{t-k}) + \text{cov}(-Y_{t-1}, -Y_{t-k-1}) \\
&= \gamma_k - \gamma_{k+1} - \gamma_{k-1} + \gamma_k \\
&= 2\gamma_k - \gamma_{k+1} - \gamma_{k-1} \\
&= \frac{\sigma_e^2}{1-\phi^2}(2\phi^k - \phi^{k+1} - \phi^{k-1}) \\
&= \frac{\sigma_e^2\phi^{k-1}}{1-\phi^2}(2\phi - \phi^2 - 1) \\
&= \frac{-\sigma_e^2\phi^{k-1}}{(1-\phi)(1+\phi)}(1-\phi)^2 \\
&= \frac{-(1-\phi)\sigma_e^2\phi^{k-1}}{(1+\phi)}
\end{aligned}$$

At lag $k = 0$, this reduces to

$$\frac{-(1-\phi)\sigma_e^2}{(1+\phi)\phi},$$

which can be negative! Since variances cannot be negative, there is something wrong. The problem is that when we plug in $k = 0$ above, we actually end up with

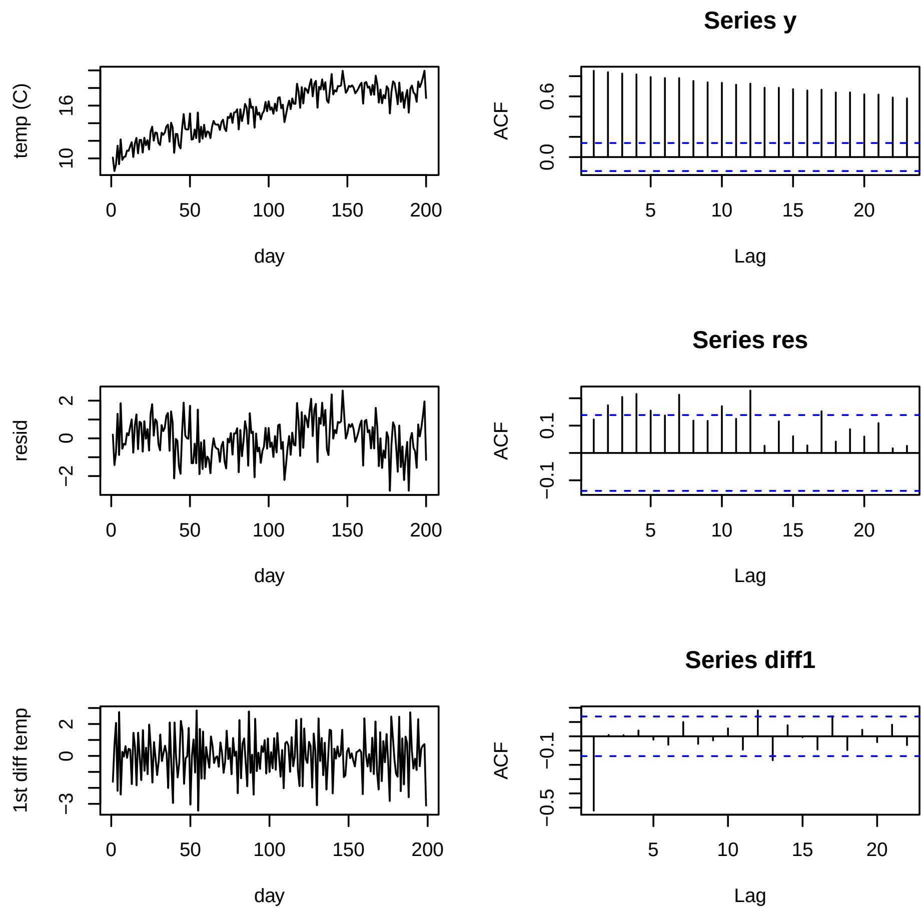
$$\begin{aligned}
\text{cov}(W_t, W_{t-0}) &= 2\gamma_0 - \gamma_1 - \gamma_{-1} \\
&= 2\gamma_0 - \gamma_1 - \gamma_1 \\
&= 2\gamma_0 - 2\gamma_1 \\
&= 2\frac{\sigma_e^2}{1-\phi^2} - 2\frac{\sigma_e^2}{1-\phi^2}\phi \\
&= 2\sigma_e^2\left[\frac{1}{1-\phi^2} - \frac{\phi}{1-\phi^2}\right] = \frac{2\sigma_e^2}{1+\phi}.
\end{aligned}$$

4. An underground temperature probe was placed about a half kilometer away from geothermal borehole in Iceland. Degrees Celsius Y_t were recorded daily for 200 days in 2014. This will read the data into R:

```
library(TSA)
y=ts(read.table("http://people.stat.sc.edu/hansont/stat520/thermal.txt"),start=1)
```

Here is code to complete parts (a), (c), and (d):

```
library(TSA)
par(mfrow=c(3,2))
y=ts(read.table("http://people.stat.sc.edu/hansont/stat520/thermal.txt"),start=1)
plot(y,xlab="day",ylab="temp (C)")
acf(y)
t=time(y); tsq=time(y)^2
f=lm(y~t+tsq)
res=resid(f)
plot(t,res,xlab="day",ylab="resid",type="l")
acf(res)
diff1=diff(res,1)
plot(diff1,xlab="day",ylab="1st diff temp",type="l")
acf(diff1)
```



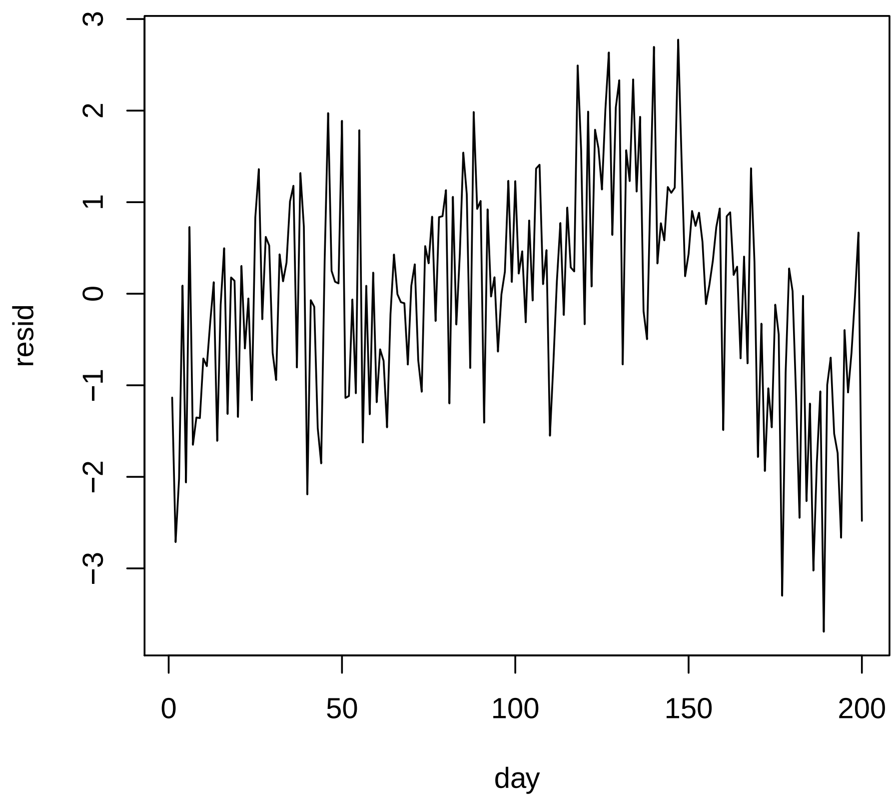
(a) Plot the temperature Y_t versus time t and accompanying sample ACF. Does the data appear to be stationary? Are there any pronounced trends? Describe.

Temperate slowly increases and then levels off. The variance appears constant, but not the mean, so the series does not appear to be stationary.

(b) Fit a straight-line regression model to the data for detrending purposes and plot the residuals $r_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 t$ versus time t. Does a line appear to fit okay?

```
par(mfrow=c(1,1))
```

```
t=time(y);
f1=lm(y~t)
res1=resid(f1)
plot(t,res1,xlab="day",ylab="resid",type="l")
```



The residual plots shows a great deal of curvature; no a line does not provide adequate fit.

(c) Fit a quadratic regression model to the data for detrending purposes and plot the residuals $r_t = Y_t - \hat{\beta}_0 - \hat{\beta}_1 t - \hat{\beta}_2 t^2$ versus time t, along with the sample ACF of the residuals. Do the residuals appear to be stationary?

This is the 2nd row in the first figure with six plots. There is still some overall structure to the residuals, even after removing a quadratic trend; the residuals do not appear to be stationary.

(d) Now plot the first difference of the residuals $\nabla r_t = r_t - r_{t-1}$ from part (c) versus time t along with the sample ACF. Comment on whether stationarity is finally achieved.

Yes, the mean and variance seem roughly constant over time; stationarity is much more likely for the differenced residuals. Also note that the sample ACF dies down very slowly for $\{Y_t\}$, a sure indication of non-stationarity. The sample ACF for the residuals from a quadratic fit $\{r_t\}$ also shows significant lags out past $k = 10$, also a sign of non-stationarity. The differenced residuals $\{\nabla r_t\}$, however, have a sample ACF with only one significant lag at $k = 1$.

(e) Based on the ACF from part (d), what might a plausible model for $w_t = \nabla r_t$ be? Why?

An MA(1) model has $\rho_k = 0$ for $k \geq 2$, which is what the sample ACF appears to indicate.