STAT 520, Fall 2015: Homework 1

1. The TSA package contains the data set `co2`, monthly carbon dioxide (CO_2) levels in northern Canada from 1/1994 to 12/2004. To load the data in R, type

```
library(TSA)
data(co2)
```

   (a) Construct a time series plot for these data, e.g.

```
plot(co2,ylab="CO2 levels",xlab="Year",type="o")
```

   Describe all systematic patterns you see in the plot.

   (b) To enhance interpretability, add monthly plotting symbols using the following R commands:

```
plot(co2,ylab="CO2 levels",xlab="Year",type='l')
points(y=co2,x=time(co2),pch=as.vector(season(co2)),cex=0.75)
```

   Which months are consistently associated with highest CO_2 levels? the lowest? **Note:** The `cex=0.75` part controls the size of the plotting symbols specified in `pch`.

2. The course web site contains the data set `gasprices`, which lists the average price (dollars/gallon) for one type of gas in the United States. There are $n = 145$ weekly observations collected from 1/5/2009 to 10/10/2011 (**Source:** Rajon Coles, Fall 2011).

   (a) Construct a time series plot for these data. You can read the `gasprices.txt` file directly into R, turn it into a "time series" R object, and plot it via

```
gas=ts(read.table("http://people.stat.sc.edu/hansont/stat520/gasprices.txt"))
plot(gas,ylab="Price ($/gallon)",xlab="Week",type="o")
```

   Qualitatively describe how gas prices change over time.

   (b) Create a scatterplot with the observed series Y_t on the vertical axis and Y_{t-1} on the horizontal axis. This is called a **lag-1 scatterplot.** You can do this using the following R commands:

```
plot(zlag(gas,1),gas,ylab=expression(Y[t]),xlab=expression(Y[t-1]),type='p')
```

   This plot displays the observed data plotted against the lag-1 series; i.e., the scatterplot of the 144 points $(Y_1, Y_2), (Y_2, Y_3), \ldots, (Y_{144}, Y_{145})$. Qualitatively describe the degree of correlation between gas prices from one week to the next. Compute the sample correlation for these data via `cor(gas,zlag(gas,1),use="complete.obs")`; does this number confirm your visual assessment?

   (c) What does this plot look like for larger lags (e.g., 2, 3, 5, 10, 20, etc.)? When compared to the lag-1 series, do the corresponding correlations between Y_t and Y_{t-k} increase or decrease as k increases? Interpret what is meant by this.

3. Suppose that Z_1, Z_2, Z_3 are random variables with

$$\begin{array}{ccc} E(Z_1)=0 & E(Z_2)=1 & E(Z_3)=-1 \\ \text{var}(Z_1)=1 & \text{var}(Z_2)=2 & \text{var}(Z_3)=3 \\ \text{cov}(Z_1,Z_2)=-0.5 & \text{cov}(Z_2,Z_3)=1.5 & \text{cov}(Z_1,Z_3)=0. \end{array}$$

Calculate each of the following:

(a) $E(Z_1 - 6Z_2 - 2Z_3)$

(b) $\text{var}(2Z_1 + Z_3)$

(c) $\text{cov}(3Z_1 - Z_2, Z_2 + 2Z_3)$

(d) $\text{corr}(Z_1 - Z_2, Z_2 + Z_3)$

4. Simulating processes.

(a) Simulate and plot a white noise process $e_t \sim \text{iid } \mathcal{N}(0,1)$ of length $n = 100$ using the following commands in R:

```
wn=rnorm(100,0,1) # n=100, mean=0, standard deviation=1
plot(wn,ylab="White noise process",xlab="Time",type="o")
```

(b) Repeat part (a) under the assumption that

- $e_t \sim \text{iid } t(1)$, also called the Cauchy distribution
- $e_t \sim \text{iid } \chi^2(1)$

To do this, just replace the first line of the code above with `wn=rt(100,1)` and `wn= rchisq(100,1)`, respectively. Comment on the differences among the 3 simulated white noise processes.

(c) For each of your simulated series in parts (a) and (b), create the corresponding random walk process $Y_t = Y_{t-1} + e_t$, for $t = 1, 2, \ldots, 100$; $e_0 = 0$. To do this for the normally distributed white noise series, for example, use the following code:

```
rw=rep(0,100) # vector of 100 zeroes
for(i in 1:100){rw[i]=sum(wn[1:i])} # sum simply sums elements in a vector
plot(rw,ylab="Random walk from normal WN",xlab="Time",type="o")
```

Comment on the differences among the 3 simulated random walk processes.

5. Show $\text{var}(Y) = E(Y^2) - [E(Y)]^2$ starting from the definition $\text{var}(Y) = E\{(Y - E(Y))^2\}$ by expanding $(Y - E(Y))^2$ and taking the expectation of each term. Note that $E(E(Y)) = E(Y)$ because $E(Y)$ is a *number*, not random.

6. Let (X, Y) have the joint density $f_{X,Y}(x, y) = (x + y)$ over $\mathcal{R} = \{(x, y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$, the unit square in the plane.

   (a) Find $E(X)$ and $E(Y)$. Note that by definition $E(X) = \int_0^1 \int_0^1 x f_{X,Y}(x, y) dx dy$.

   (b) Find $\text{var}(X)$ and $\text{var}(Y)$. Note that, e.g., $E(X^2) = \int_0^1 \int_0^1 x^2 f_{X,Y}(x, y) dx dy$.

   (c) What is $E(XY)$? What is $\text{cov}(X, Y)$? Use the computing formula and what you have already done.

   (d) Finally, what is $\text{corr}(X, Y)$? Are X and Y independent? Why or why not?

   Hint: There is an online integral solver at

   `http://www.wolframalpha.com/calculators/integral-calculator/`

   You can obtain, for example, $\int_0^1 \int_0^1 xy(x + y) dx dy$ by typing

   ```
   Integrate[x*y*(x+y),{x,0,1},{y,0,1}]
   ```

   into the box and hitting [Enter]. Feel free to use the online solver for this problem.

Do the following problems from Cryer and Chan (Chapter 2): 2.4, 2.5, 2.11, and 2.12.