Assignment 2: More Simulations

Due: Sept 27th, 2016 at 9:59 am

Question 1: Send a monopoly to jail

In the board game monopoly, a player tosses two dices together and the sum of the two dices points is the number of steps he or she should go. For example a player gets a 2 and a 3, then he or she should go 5 steps forward. If the two dices give the same number (for example two 2), which is called a "double", after moving forward (4 steps), the player gets another chance to toss and move forward again. However, if someone tosses 3 consecutive doubles, he or she will be sent to jail.

Consider that you are tossing two fair dices for many times. You won't stop until you get 3 doubles. After you stop, take down the number of tosses you use. Use a loop to repeat this at least 1000 times. Calculate the average of the 1000 results, that is, how many tosses is needed on average, to be able to get three consecutive doubles.

Question 2: Checking Distributions

The function `hist()` can be used to plot a histogram of frequency (by default) or a histogram of density (by setting the argument `freq` to be `FALSE`).

1. Fit normal density
   - Generate 1000 standard normal random numbers and save them to **x**.
   - Use `hist()` to draw its density.
   - Create a vector **xfit** using `seq()` function. Use the following values for the arguments of `seq()`: `from=min(x)`, `to = max(x)` and `by`=0.1.
   - Run `lines(xfit,dnorm(xfit,mean=0,sd=1))`.
   - You should see that the theoretical density curve is roughly going through the edge of the data density histogram. Try to increase your data points to make the fit better.
2. Create a vector `y=exp(x)`. Now `y` should be ramdom numbers drawn from a log-normal distribution. Fit `y` using the density function of log-normal distribution (`dlnorm`). Create a vector `y2` directly using `rlnorm` and fit it again.
3. Create random numbers from gamma distribution (rgamma) and fit the density using the same method.
4. In the above question 1, try to add another density line by using `lines(xfit,dnorm(xfit,mean=mean(x),`
   Compare it with the theoretical density line. Can you see any difference? Increase the data points and try again. Try to explain the difference between the two lines.

Question 3: Returns

- Download data "pepci.csv" and save it into your working directory.
- Extract recent 3-year daily prices, recent 5-year weekly prices, recent 10-year weekly prices and 20-year quarterly prices.
- For each of the four time series, calculate simple returns and log returns for both overlapping and non-overlapping types.
- Draw histograms of the all returns. Are they normal?
- Feel free to use any package in this question.

Question 4: Intro. to HFT

In high frequency trading, the data format is different from other intraday data. Instead of open, high, low, close, you would see bid price, ask price and trade price. Bid price is the price trader want to sell at market, ask price is the price trader want to buy from market, trade price is the price at which stock are sold or bought.

Read AAPL.O.csv and do the following questions;

1. Create a table which contains: Date.L., Time.L., Type, Price, Volume, Bid.price, Bid.Size, Ask.Price and Ask.Size
2. Create a column and name it as Trading.Time. For this column, you have to combine Date.L. and Time.L together, then put this column to the table you created in step 1
3. Subtract all the rows, which marked as Quote in Type column. Name this table as AAPL.Quote. Also create a table which only contains Trade, name this table as AAPL.Trade
4. Calculate time difference for each row and save it as numeric vector, unit should be minute
5. Calculate spread between bid price and ask price. Make a histogram for spread
6. Make a line chart for AAPL.Trade, use time in x-axis, volume in y-axis.