Stats Scribe 5/4

Final Review time: Monday 5/11 9:30-11 am

Wednesday 5/13 2-5pm

Final: - do not memorize R – understand the logical structure of models (ex. Bootstrapping) and know pseudo-code (step by step process of running a do or for function)

- will be cumulative in the sense that ideas from the first part of the course cross over into material learned in the second half of the course

- there may be questions similar to ones on the midterm

3 potential problems of classification:

1. What's max expected play?
2. What is value at risk of a given portfolio? (with respect to a level (ex. 5% interest rate) and a horizon (ex. 2 weeks)
3. How do we use Monte Carlo simulation to incorporate future sources of uncertainty?

Value at risk: alpha percentile of the profit/loss distribution

- When you look at a histogram, the value where you get alpha = 5% (5% of the graph is to the right or left of this point depending on which side of the distribution you're evaluating) is the number defined as value at risk.
- To simulate the profit/loss distribution
  - Bootstrap with resampling of the data
  - Then simulate 1 notional future after another
  - Similar to buying stocks in a portfolio and letting your holdings run however you resample your past set of interest rates with interest rate from your normal sampling interest formula; this answers the question: what are your new holdings at the end of day 2 with new interest rates
  - End after 'n' number of days and find your new total wealth
  - Subtract new total wealth from original total wealth to find the value of your portfolio
  - VaR calculation-only thing that matters is wealth; alpha quantile of the profit loss distribution (way we calculate VaR for a portfolio is that we start with initial wealth (Wo); simulate the probability distribution of future wealth (Wf) and then calculate profit/loss distribution (delta=Wf-Wo)
  - find me the point that has the specified fraction to the left (quantile-50% would be 50% of data on right and left)

- in this problem the alpha is 5% which means that you are looking to find the probable loss of the profit loss distribution at 5%
- random variable-we describe uncertainty about with probability distribution
- calculate the alpha quantile of P(delta
- how do you interpret the 5% in VaR?
- because delta = Wf - W0 and wf is rand but w0 is fixed, delta is rand because of w
- the point at which you cut off from 5% to 95% on your distrubition is your Va
- Wf=total wealth that is being returned everytime and then plot histogram of result
- calculate profit loss distribution=sim1result-10000(initial wealth
- estimate VaR by finding quantile of alpha distribution (negative numbers but express positively)
- 5% value at risk is x number, that means that theres no more than 5% chance of losing x number or more; if you think about losses, on 95% of days you lose no more than x number
- VaR-it's not something you calculate everytime during the simulation just at the end like expected value; these are properties of probability distributions and once you've simulated probability distribution then you can summarize the features (VaR, expected values, variances, and shapes that are all done at the end once we've simulated relavant portfolios)
- define new random variable as function of the old one
- higher the n is the lower the error the expected value of a function of a random variable is not the same as calculating the expected value and applying the function (Jensen's Inequality-only time that this is true is for a linear function)

Problem 2 of homework 10:

- every single day I sample draw from the past (bootstrapping)
- Interest rate=r(t)
  current holdings=h(t)
  Gain/loss=r(t)*h(t)
  New Holdings:h(t)+h(t)*r(t)
- Key thing is to rebalance which is redistributing wealth so you have equal distribution of wealth in each stock
- holdings=weights=totalwealth
- for loop-chain these computations together (computations run 250 times in a row); that chains the events together which can only be done in a for loop (sequentiality); B2 refers to B1, B3 refers to B2, B4 refers to B3
- at end of simulated trading year total wealth=11690.85 (if that were my entire wealth, at the end of the year how happy am I?)
- to find utility, take the log(totalwealth)=9.366562 (this is only useful for comparison and the value itself isn't very interpretable); utility is own preference relation, utility function, which boils down to E(u(P))>E(u(Q)) or vice versa
- this is the foundation; being able to do the computation of one day is the foundation (if you can make something out of day one and day two

Problem 1 of homework 10:

- model needed to count for trends, seasonal variations, changes based on weekend days, and the quadratic effects
  lm1=PeakDemand ~ TimeIndex + factor(Month) + DailyTemp + I(DailyTemp^2) + Sat + Sun, data= peakdemand
- pretend mean daily temp will be daily future temp
- create time index creating day 1,2, etc. but staring from end of first data set
- plot old peak demand and then fit values of new peak demand dataset and the forecast but forecast looks way too smooth because ignoring
- sources of uncertainty by pretending that mean is future daily temperature
- actual peak demand will have same error
- simulated random set of residuals to take into account noise of the forecast
- limitations-assuming that past is a good representation that future is going to look like
- how do we deal with uncertainty of temperature? : simulate new temperatures-use temp50year means and standard deviations and create simulated temperature distribution for two years which takes into account uncertainty about temperature and need to feed that uncertainty into the prediction
- will now insert new simulated temperatures into the model to add noise to the forecast and take into account future uncertainty (taking fitted values that take into account future uncertainty and the uncertainty of the residuals but now the forecast includes more noise)
- source of uncertainty is critical when answering business question of whether to invest in more electricity
- and then count how many were bigger than 5,000 and then simulate 1000 futures in which have randomly drawn temperatures and residuals using Monte Carlo Simulation