Lab Report #1: Moments & Cumulants

Revised: February 7, 2013

Due at the start of class. You may speak to others, but whatever you hand in should be your own work. Please include your Matlab code.

1. *Cumulants of Bernoulli random variables.* Consider a random variable x that equals δ (an arbitrary number) with probability ω (a number between zero and one) and 0 with probability $1-\omega$. We'll use its cumulant generating function (cgf) to find its first four cumulants, representing, respectively, its mean, variance, skewness, and kurtosis. Feel free to use Matlab — I would, it's easier.

   (a) Verify that this is a legitimate probability distribution.

   (b) What is the mean? The variance? The standard deviation?

   (c) Derive the moment generating function. (If you're confused about this, apply the definition.) What is the cumulant generating function?

   (d) Differentiate the cgf to find the first four cumulants, labelled κ_1 through κ_4. What are the mean and variance?

   (e) Derive the standard measures of skewness and excess kurtosis:

   $$\begin{aligned}\gamma_1 &= \kappa_3/(\kappa_2)^{3/2} \quad \text{(skewness)} \\ \gamma_2 &= \kappa_4/(\kappa_2)^2 \quad \text{(excess kurtosis)}\end{aligned}$$

   How do they depend on ω? δ? What is excess kurtosis when $\omega = 1/2$?

2. *Sample moments.* It's often helpful to experiment with artificial test problems, so that we know when we make a mistake. Here we compute sample moments of data generated by the computer and verify what we think they are.

   We'll start by generating some artificial data:

```
format compact
mu = 0;
sigma = 1;
nobs = 1000;
rng('default');     % sets "seed" so you can replicate the output
x = randn(nobs,1);
```

   These commands generate "pseudo-random" numbers from a standard normal distribution and put them in the vector x. (Standard normal means normal with mean equal to zero and variance equal to one.) They serve as data for what follows.

   (a) Our first check is to see if the sample moments correspond, at least approximately, to our knowledge of normal random variables. For example, use the commands:

```
xbar = mean(x)
moments = mean([(x-xbar).^2 (x-xbar).^3 (x-xbar).^4])
```

What do you get? What are the first four sample moments? How do they compare to the mean and variance of the specified distribution?

(b) Our second check is on the Matlab commands `mean(x)`, `std(x)`, `skewness(x)`, and `kurtosis(x)`. How do they compare to the sample moments you computed earlier? Are they exactly the same, almost the same, or completely different?

3. *Volatility and correlation in business cycles.* We're going to use macroeconomic data from FRED, the St Louis Fed's convenient online data repository, to get a summary picture of US business cycles. (Search "FRED St Louis".) Data preparation should follow these steps or the equivalent:

   - The idea. Download monthly data, 1960 to the present, for the following series: Industrial Production (series INDPRO), Employment (All Employees: Total Nonfarm, series PAYEMS), and the S&P 500 Index (SP500). The first two series serve as monthly measures of economic activity. Variation in their growth rates reflects the business cycle. The last one is an important asset price, and its growth rate is the (approximate) return on equity overall.
   - The mechanics. It's probably easier to use FRED's Excel add-in, but here's what I do. (i) Go to FRED: http://research.stlouisfed.org/fred2/. (ii) Type INDPRO in the search box in the upper right corner. A graph will appear. (iii) Click on Edit Graph below the left side of the figure. (iv) Click on Add Data Series below the figure. (v) Enter PAYEMS in the search box. (vi) Repeat (iv,v) with SP500. Once you have it, you'll notice a drop-down menu indicating the frequency. Choose Monthly. To its right, choose Average as the Aggregation Method. (vii) Click on Download Data in Graph above the left side of the figure to create an Excel Spreadsheet. (viii) Open the spreadsheet, delete everything before 1960, including the headings, and save it in a directory that's convenient for Matlab to read from.
   - Read your data into Matlab and construct year-on-year growth rates using the formula

     $$g_t = \log x_t - \log x_{t-12}.$$

     [The function `log` here is the natural log. We refer to growth rates computed this way as continuously-compounded. Take that as given for now, but ask me in class if you'd like an explanation.] If the data are in an array called `data`, with each column a different variable, you can compute growth rates with

     ```
     diffdata = log(data(13:end,:)) - log(data(1:end-12,:))
     ```

     [The expressions `data(i,j)` have two forms here. The first index is a range (eg, `x:y`), where `end` means the last one. The second index (`:`) means use all the available columns.]

(a) For each variable (that is, its growth rate), compute the standard deviation, skewness, and excess kurtosis. (The commands `std`, `skewness`, and `kurtosis` are helpful here.) Which variables have the highest "volatility" (standard deviation)? Do any of them seem "nonnormal" (non-Gaussian) to you?

(b) For the same variables, compute correlations with employment growth. (The command `corrcoef` may come in handy.) Which variable has the highest correlation?