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.

Solution: The solutions to this assignment are computed in the Matlab program listed at the end and posted here. If you run the program yourself, the output will give you most of what's asked for. The answers below go beyond this.

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 $\omega = 1/2$?

Solution: This the Bernoulli multiplied by δ, so it illustrates the impact (or lack thereof) of scaling.

(a) Since ω and $1-\omega$ are nonnegative and sum to one, we're ok.

(b) Mean: $\delta\omega$. Variance: $\delta^2\omega(1-\omega)$. Standard deviation: square root of variance.

(c) The mgf is $h(s) = 1-\omega+\omega e^{s\delta}$. The cgf is $k(s) = \log h(s)$. Everything so far should look familiar from class.

(d) The cumulants are

$$\begin{aligned}
\kappa_1 &= \delta\omega \\
\kappa_2 &= \delta^2\omega(1-\omega) \\
\kappa_3 &= \delta^3\omega(1-\omega)(1-2\omega) \\
\kappa_4 &= \delta^4\omega(1-\omega)[1-6\omega(1-\omega)].
\end{aligned}$$

Note the scaling: κ_j includes δ^j. Other than scaling, we've seen the first two before. The third one tells us that skewness depends on the sign of δ and whether ω is greater or less than one half (graph the probabilities against x if this isn't clear). The fourth one depends on $\omega(1-\omega)$. At $\omega = 1/2$, this terms reaches its max of $1/4$, so the overall term is negative, which generates negative excess kurtosis. As ω moves toward zero or one, this term shrinks.

(e) You'll note that Matlab doesn't do the obvious cancellation of δ's. Once you do, you have

$$\begin{aligned}
\gamma_1 &= \mathrm{sgn}(\delta)(1-2\omega)/[\omega(1-\omega)]^{1/2} \\
\gamma_2 &= 1/[\omega(1-\omega)] - 6.
\end{aligned}$$

There's a subtle issue with γ_1: the magnitude of δ doesn't matter, but its sign ("sgn") does. That shows up in the ratio of $[\delta^2]^{3/2}$ to δ^3 (think about this a minute). What about excess kurtosis? At $\omega = 1/2$, $\gamma_2 = -2$, so there's excess kurtosis. Loosely speaking, the Bernoulli has thinner tails than the normal. As ω approaches zero or one, we reverse that. As we increase/decrease ω, we get a distribution with increasing skewness and excess kurtosis.

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?

Solution:

(a) The moments are close to what the distribution implies: mean one, variance (and standard deviation) one, skewness zero, and kurtosis three. The small differences reflect sampling variability. You can see this by increasing the sample size, which typically makes the sample moments closer to the "population" moments. The one difference is in the variance and standard deviation, which are computed by dividing the sum of squared deviations by the number of observations minus one, rather than just the number of observations. It's a small difference, to be sure.

(b) The `skewness` and `kurtosis` commands give exactly the same answers as our own calculations, which assures us that they do what we want them to do.

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

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

[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?

Solution: See the Matlab program. Note that there's negative skewness and positive excess kurtosis in all of these series. Industrial production and employment are highly correlated (0.8137), the S&P 500 less so (correlation 0.3390 with IP, 0.1674 with employment). Part of this is timing. If we shifted the S&P 500 back six months, the correlations would be larger.

I recommend, in addition, that you look at their histograms, which is a more direct indication of their distribution. See the Matlab program.

Matlab program

```
%  hw1_s13.m
%  Matlab program for Lab Report #1, Spring 2013
%  Written by:  Dave Backus, February 2013
format compact
clear all

%%
disp(' ')
disp('Answers Lab Report #1')
disp('*** I put pauses in the program.  If Matlab stops, hit return')

disp(' ')
disp('---------------------------------------------------------------')
disp('Question 1')
syms p delta omega s

mgf = 1-omega+omega*exp(s*delta)
cgf = log(mgf)

disp(' ')
kappa1 = subs(diff(cgf,s,1),s,0)    %  mean
kappa2 = subs(diff(cgf,s,2),s,0)    % variance
kappa3 = subs(diff(cgf,s,3),s,0)
factor(kappa3)  % sometimes this cleans up the expression; see also simplify
kappa4 = subs(diff(cgf,s,4),s,0)
factor(kappa4)

disp(' ')
gamma1 = kappa3/kappa2^(3/2)
gamma2 = kappa4/kappa2^2

pause

%%
disp(' ')
disp('---------------------------------------------------------------')
disp('Question 2')
disp('Moments of simulated normal data')
disp('Compare our calculations to Matlab functions')
disp(' ')
clear all

% set parameters
mu = 0;
```

```
sigma = 1;
nobs = 1000;

x = normrnd(mu,sigma,nobs,1);

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

disp(' ')
var_x = var(x)
compare = moments(2)

disp(' ')
std_x = std(x)
compare = sqrt(moments(2))

disp(' ')
skw_x = skewness(x)
compare = moments(3)/moments(2)^(3/2)

disp(' ')
krt_x = kurtosis(x)
compare = moments(4)/moments(2)^2

pause

%%
disp(' ')
disp('------------------------------------------------------------------')
disp('Question 3')
clear all

% read data from spreadsheet:  INDPRO, PAYEMS, SP500
data = xlsread('FRED_hw1_13.xls','Data');

% year-on-year growth rates
diffdata = log(data(13:end,:)) - log(data(1:end-12,:));

disp('Basic moments')
disp('Variables:  IP, EMP, SP500')
means = mean(diffdata)
stdev = std(diffdata)
skew = skewness(diffdata)
exkurt = kurtosis(diffdata) - 3
pause(1)
```

```
disp(' ')
disp('Correlations')
corrcoef(diffdata)

% histogram to see skewness and kurtosis
subplot(3,1,1), hist(diffdata(:,1))
subplot(3,1,2), hist(diffdata(:,2))
subplot(3,1,3), hist(diffdata(:,3))
pause(1)

return
```