


## Lab Report #2: Sums, Mixtures, & Certainty Equivalents

Revised: September 20, 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:** Brief answers follow, but see also the attached Matlab program; download the pdf, open, click on pushpin: 

1. *Sums.* An easy way to put together time series data is to add independent increments. If the one-period return in period  $i$  is  $x_i$ , then the return over  $n$  periods might be expressed by  $y = x_1 + x_2 + \cdots + x_n$ . If the  $x_i$ 's are independent, then we've seen that the cgf of the sum  $y$  is the sum of the cgf's of the  $x$ 's. And if the  $x_i$  have the same distribution, the cgf of  $y$  is proportional to that of  $x$ :  $k_y(s) = nk_x(s)$ .

- (a) Suppose the  $x_i$ 's are independent Bernoulli random variables:

$$x_i = \begin{cases} 0 & \text{with probability } 1 - \omega \\ 1 & \text{with probability } \omega. \end{cases}$$

What is the cgf of each  $x_i$ ? What is the cgf of  $y$ ?

- (b) Suppose the  $x_i$ 's are independent Poisson random variables:  $x_i = j$  for  $j = 0, 1, \dots$  with probability  $e^{-\omega}\omega^j/j!$ . What is the cgf of each  $x_i$ ? What is the cgf of  $y$ ?
- (c) For which of these examples does  $y$  have the same distribution as the  $x_i$ 's?
- (d) Now do the reverse and consider time intervals shorter than one period. If  $x$  has cgf  $k_x(s)$ , then we might guess  $y$  has cgf  $k_y(s) = \delta k_x(s)$  for some fraction  $\delta$ . The question is whether the cgf is associated with a legitimate distribution. We'll ignore that question and simply ask whether we recognize the cgf. Can you identify the distribution of  $y$  if  $x$  is Poisson? If  $x$  is Bernoulli?

### Solution:

- (a) The mgf for each  $x_i$  is  $h_i(s) = (1 - \omega) + \omega e^s$ , so the cgf is its log:  $h_i(s) = \log[(1 - \omega) + \omega e^s]$ . Both are in the notes. The cgf of  $y$  is  $n$  times this:  $h_y(s) = n \log[(1 - \omega) + \omega e^s]$ . Some may recognize this as the cgf of a binomial random variable.
- (b) The cgf of a Poisson random variable is  $k_i(s) = \omega(e^s - 1)$ . This is in the notes, too. If we sum  $n$  of these, the cgf is  $k_y(s) = n\omega(e^s - 1)$ . You can see that this has the Poisson form, too, with  $n\omega$  replacing  $\omega$ .

- (c) The Poisson.
- (d) This works fine for the Poisson, not the Bernoulli. Think about it: for a Bernoulli, an event either happens once or not at all. It's not possible for it to happen half a time. For the Poisson, any number of events can happen. If we change the time interval, we simply scale the intensity parameter  $\omega$  up and down.

2. *Sums and mixtures.* More fun with cumulant generating functions (cgfs). The ingredients include two independent normal random variables:  $x_1 \sim \mathcal{N}(0, 1)$  and  $x_2 \sim \mathcal{N}(0, \delta)$ . Do the calculations in Matlab and submit your code with your answer.

- (a) *Sum.* Consider the random variable  $y = x_1 + x_2$ . What is its cgf? Its first four cumulants? What are its measures of skewness and excess kurtosis,  $\gamma_1$  and  $\gamma_2$ ?
- (b) *Mixture.* Now consider the mixture  $z$ :

$$z = \begin{cases} x_1 & \text{with probability } 1 - \omega \\ x_2 & \text{with probability } \omega. \end{cases}$$

What is its cgf? What are its first four cumulants?

- (c) *Mixture, continued.* What are  $z$ 's skewness and kurtosis,  $\gamma_1$  and  $\gamma_2$ ? If  $\omega = 0.05$ , what value of  $\delta$  gives us  $\gamma_2 = 1$ ? In words: What does this do to the distribution?

### Solution:

- (a) The cgf of the sum is the sum of the cgf's:

$$k_y(s) = k_1(s) + k_2(s) = s(0 + 0) + s^2(1 + \delta)/2.$$

The form of the cgf should be familiar, we derived it in class. See the notes if you have questions.

The cumulants follow from differentiating:  $\kappa_1 = 0$  (the mean),  $\kappa_2 = 1 + \delta$  (the variance),  $\kappa_3 = 0$ , and  $\kappa_4 = 0$ . The last two are zero, because the cgf is quadratic in  $s$ , so all the derivatives after the second one are zero. Therefore  $\gamma_1 = \gamma_2 = 0$  as well.

- (b) Mixtures are a little different. The mgf is the weighted average of the mgf's of the components:

$$h_y(s) = (1 - \omega)h_1(s) + \omega h_2(s) = (1 - \omega)e^{s^2/2} + \omega e^{s^2\delta/2}.$$

The cgf is the log of this. The first four cumulants are  $\kappa_1 = 0$  (the mean),  $\kappa_2 = 1 - \omega + \omega\delta$  (the variance),  $\kappa_3 = 0$ , and  $\kappa_4 = 3(1 - \omega)\omega(\delta - 1)^2$ . So now we're getting excess kurtosis.

- (c) Skewness  $\gamma_1 = 0$  but  $\gamma_2 = \kappa_4/(\kappa_2)^2$  (it doesn't simplify much further). If we set  $\omega = 0.05$  and  $\delta = 4.05$ , we set  $\gamma_2 = 0.9980$ , which is pretty close to our target.

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".) Data preparation should follow these steps or the equivalent:

- Read this file into Matlab:

[http://pages.stern.nyu.edu/~dbackus/233/FRED\\_hw1.xls](http://pages.stern.nyu.edu/~dbackus/233/FRED_hw1.xls)

The same file is attached to the pdf version of this document: download the pdf, open, click on pushpin: 

The file contains monthly data, 1960 to the present, for 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 an approximate aggregate equity return.

- Compute 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 of rows (eg, `x:y`), where `end` means the last one. The second index `(:)` means use all available columns.]

- (a) For each variable in `diffdata`, compute the standard deviation, skewness, and excess kurtosis. (The commands `std`, `skewness`, and `kurtosis` are helpful here. Note that they can be applied directly to `diffdata`, giving you moments for all three series at once.) Do any of these variables seem "non-normal" to you?
- (b) For the same data, compute correlations with employment growth. (The command `corrcoef` may come in handy.) Do the correlations make sense to you?

**Solution:** The relevant Matlab output includes

```
Variables:  IP, EMP, SP500
means =
    0.0173    0.0274    0.0628
stdev =
    0.0197    0.0482    0.1603
skew =
   -0.7814   -1.1311   -0.9085
```

```
exkurt =
    0.7550    2.0286    1.2750
```

Correlations

```
ans =
    1.0000    0.8122    0.1428
    0.8122    1.0000    0.3004
    0.1428    0.3004    1.0000
```

Note that there's negative skewness and positive excess kurtosis in all of these series. Industrial production and employment are highly correlated (0.8122), the S&P 500 less so (correlation 0.3004 with IP, 0.1428 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.

4. *Asymmetric risk (optional, extra credit)*. The idea is to explore the impact of asymmetric risk (think skewness) on the assessment of risk of a power utility agent. The agent has utility  $u(c) = c^{1-\alpha}/(1-\alpha)$  with risk aversion parameter  $\alpha = 5$ . The log of consumption is Bernoulli:  $\log c = x = \beta + \delta z$  with

$$z = \begin{cases} 0 & \text{with probability } 1 - \omega \\ 1 & \text{with probability } \omega \end{cases}$$

and  $\delta > 0$ . We'll change the degree of asymmetry by varying  $\omega$ . As we do, we'll adjust  $(\beta, \delta)$  to maintain the mean and standard deviation of  $\log c$  at 1.0 and 0.25, respectively. Do the calculations in Matlab and submit your code with your answer.

- What is the mgf of  $x$ ? The cgf?
- What are the first four cumulants of  $x$ ? Skewness and kurtosis? How does skewness depend on  $\omega$ ?
- With  $\omega = 0.25$ , what values of  $(\beta, \delta)$  give us the requested mean and standard deviation of log consumption?
- With these values, what is the expected value of consumption,  $E(c) = E(e^x)$ ? What is expected utility? What is the certainty equivalent, the constant value of consumption that generates the same level of utility?
- What is the risk penalty?
- Now set  $\omega = 0.75$  and repeat the process. How does the risk penalty change? Can you guess why?

### Solution:

- The mgf has two parts, one corresponding to  $x = 0$ , the other to  $x = 1$ :

$$h(s) = (1 - \omega)e^{\beta s} + \omega e^{(\beta + \delta)s}.$$

The cgf is its log:  $k(s) = \log h(s)$ .

- (b) The first four cumulants are

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

Skewness and kurtosis are  $\gamma_1 = \kappa_3/(\kappa_2)^{3/2}$  and  $\gamma_2 = \kappa_4/(\kappa_2)^2$ , but do not simplify much further.

- (c) The idea is to vary  $\omega$  but adjust  $(\beta, \delta)$  to keep the mean and variance fixed. Here we set  $\omega = 0.25$ . To get the right mean and standard deviation, we set

$$\begin{aligned}\delta &= 2.5/[\omega(1 - \omega)]^{1/2} = 0.5774 \\ \beta &= 100 - \beta\delta = 0.8557.\end{aligned}$$

- (d) Now consumption and utility. Mean consumption is  $h(1) = (1 - \omega)e^\beta + \omega e^{\beta + \delta} = 2.8125$ . Expected utility is (this is a trick, see the notes for why it works)  $h(1 - \alpha)/(1 - \alpha) = 0.0253/(1 - \alpha)$ . The certainty equivalent is the constant consumption that gives the same utility:

$$\mu^{1 - \alpha} = 0.0253 \Rightarrow \mu = 2.5079.$$

Ok, where does that leave us? Mean consumption is 2.81. The certainty equivalent is 2.51, so there's a penalty for risk: we'd take 0.30 less, on average, to eliminate the risk. The risk penalty is  $rp = \log(\bar{c}/\mu) = 0.1146$  — and 11% penalty for risk.

- (e) Since  $\omega < 1/2$ , skewness is positive above. (See the expression for  $\kappa_3$ .) Suppose we reversed it, making  $\omega = 0.75 > 1/2$  and introducing negative skewness. What happens? We might guess that the risk penalty rises, because power utility agents don't like negative skewness. If you go through the calculations (preferably in Matlab), you find that the risk penalty becomes 0.1798, so it goes up quite a bit.