Asset Returns - Univariate Distributions

STAT 421/621 - Applied Time Series and Forecasting
Fall 2014

Roberto Bertolusso

Department of Statistics
Rice University

Outline

Random Variable and Distribution of Probability

- A **random variable** (RV), X, maps the sample space to R (or R^d for multivariate random variables, or random vectors.)
- **Distribution of probability**:
  - Cumulative Distribution Function (CDF):

$$F(x) = P(X \leq x)$$

  - Probability Density Function (pdf), for continuous variables:

$$f(x) = dF(x)/dx$$

  - Probability Mass Function (pmf), for discrete variables:

$$p(x) = P(X = x)$$

Expectation of a Continuous RV

RICE

- Given any function $g(X)$ of the continuous random variable X with pdf $f(x)$, its expected value is defined as:

$$\mathsf{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

Moments of a Continuous RV

- If $g(X) = X^l$, its expectation is called the lth moment of the continuous random variable X:

$$m'_l = \mathsf{E}[X^l] = \int_{-\infty}^{\infty} x^l f(x) dx$$

- The first moment is the mean or expectation:

$$\mu_x = m'_1 = \mathsf{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$$

Central Moments of a Continuous RV

RICE

- If $g(X) = (X - \mu_x)^l$, its expectation is called the lth central moment of X:

$$m_l = \mathsf{E}[(X - \mu_x)^l] = \int_{-\infty}^{\infty} (x - \mu_x)^l f(x)dx$$

- The second central moment of X is the variance:

$$\sigma_x^2 = \mathsf{E}[(X - \mu_x)^2] = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x)dx$$

- NOTE: The first two moments uniquely determine a normal distribution.

Skewness and Kurtosis

RICE

- The (standardized) skewness measures the symmetry of X with respect to its mean:

$$S(x) = \frac{\mathsf{E}[(X-\mu_x)^3]}{\sigma_x^3}$$

A symmetric RV has skewness = 0.

- The (standardized) kurtosis measures the tail behavior of X:

$$K(x) = \frac{\mathsf{E}[(X-\mu_x)^4]}{\sigma_x^4}$$

The excess kurtosis is $K(x)-3$, as 3 is the kurtosis of a normal random variable.

Estimation of Moments from a Random Sample

Given a collection $\{X_1, \ldots, X_T\}$ of random variables, they form a random sample of size T if they are mutually independent and identically distributed ($f(X_i)$ is the same for all $i = 1, \ldots, T$).
Let $\{x_1, \ldots, x_T\}$ be one realization of it.

- The estimation of the sample mean is:

$$\hat{\mu}_x = \frac{1}{T} \sum_{i=1}^{T} x_i$$

- The estimation of the sample variance is:

$$\hat{\sigma}_x^2 = \frac{1}{T-1} \sum_{i=1}^{T} (x_i - \hat{\mu}_x)^2$$

Estimation of Moments from a Random Sample

- The estimation of the sample skewness is:

$$\hat{S}(x) = \frac{1}{(T-1)\hat{\sigma}_x^3} \sum_{i=1}^{T} (x_i - \hat{\mu}_x)^3$$

- The estimation of the sample kurtosis is:

$$\hat{K}(x) = \frac{1}{(T-1)\hat{\sigma}_x^4} \sum_{i=1}^{T} (x_i - \hat{\mu}_x)^4$$

Asset Returns

RICE

Let P_t be the price of an asset at time index t. Then:

- One period simple return:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1$$

- Continuously compounded return:

$$r_t = \ln(R_t + 1) = \ln \frac{P_t}{P_{t-1}} = \ln P_t - \ln P_{t-1} = p_t - p_{t-1}$$

Asset Returns: Dividend Payment

RICE

If the company pays a dividend D_t, then:

- One period simple return:

$$R_t = \frac{P_t + D_t}{P_{t-1}} - 1$$

- Continuously compounded return:

$$r_t = \ln(P_t + D_t) - \ln P_{t-1}$$

Excess Returns

Used to measure the performance of an asset to, for example, a reference index (such as S&P500).

- One period simple excess return:

$$Z_t = R_t - R_{0t}$$

- Log excess return:

$$z_t = r_t - r_{0t}$$

where R_{0t} and r_{0t} are the simple and log returns of the reference asset or index, respectively.

Why do we use Asset Returns?

RICE

1. Return of an asset is a complete and scale-free summary of the investment opportunity.
2. It has more attractive statistical properties than price.

Asset Returns: Relevance of Moments

RICE

- The mean is related with the long-term return.
- The variance is related to the risk.
- The skewness has implications in risk management and on holding long or short positions.
- The kurtosis is related to reliability of estimated parameters, statistical tests, and forecasts.

Example: Getting Apple Prices from Yahoo

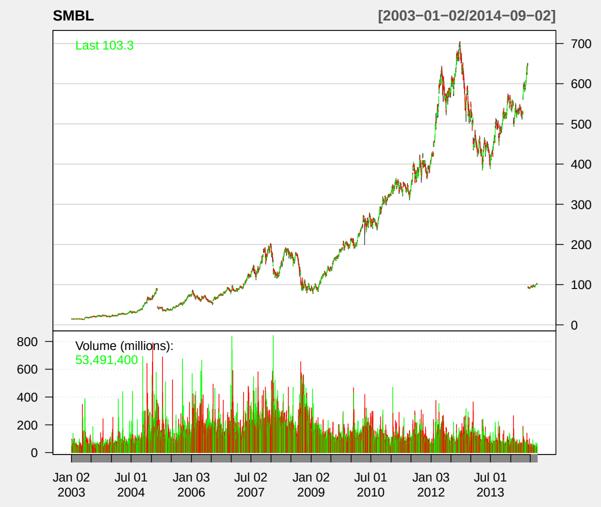
RICE

```
## Load required library
library(quantmod)

## Get prices using function getSymbols
SMBL <- getSymbols("AAPL", from="2003/01/01", auto.assign=FALSE)
## Show last 6 trading days
tail(SMBL)
```

```
##            AAPL.Open AAPL.High AAPL.Low AAPL.Close
## 2014-08-25     101.8     102.2    101.3      101.5
## 2014-08-26     101.4     101.5    100.9      100.9
## 2014-08-27     101.0     102.6    100.7      102.1
## 2014-08-28     101.6     102.8    101.6      102.2
## 2014-08-29     102.9     102.9    102.2      102.5
## 2014-09-02     103.1     103.7    102.7      103.3
##            AAPL.Volume AAPL.Adjusted
## 2014-08-25    40144700         101.5
## 2014-08-26    33119800         100.9
## 2014-08-27    46827400         102.1
## 2014-08-28    68389800         102.2
## 2014-08-29    44567000         102.5
## 2014-09-02    53491400         103.3
```

Example: Plotting Daily Adjusted Closing Price

```
chartSeries(SMBL, theme="white", up.col="green", dn.col="red")
```



The adjusted closing price and trading volume are plotted with green denoting upward changes in the price and red the downward changes.

Example: Calculating Returns

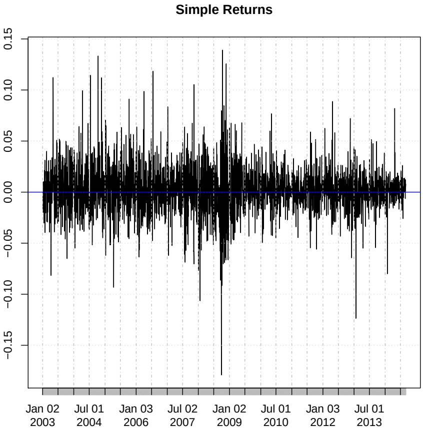
```
## Use function that calculates daily returns
drSMBL <- dailyReturn(SMBL[,6])
head(drSMBL)
##              daily.returns
## 2003-01-02          0.0000
## 2003-01-03          0.0000
## 2003-01-06          0.0000
## 2003-01-07          0.0000
## 2003-01-08         -0.0198
## 2003-01-09          0.0101
## NOTE: dailyReturn creates an initial value of 0

## Verify by direct calculation
dapr <- as.vector(SMBL[,6])
dapr[2:6] / dapr[1:5] - 1  ## Initial simple returns
## [1]  0.0000  0.0000  0.0000 -0.0198  0.0101
## Log (natural) returns
ldrSMBL <- log(drSMBL + 1)
```

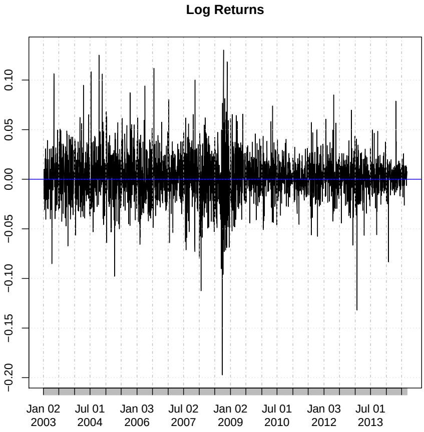
Example: Plotting Returns

RICE

```
plot(drSMBL,
     main="Simple Returns")
abline(h=0, col="blue")
```

```
plot(ldrSMBL,
     main="Log Returns")
abline(h=0, col="blue")
```





Basic Statistics and Features of Returns

RICE

```r
library(fBasics)

## Using function basicStats
basicStats(drSMBL)
```

```
##              daily.returns
## nobs             2.937e+03
## NAs              0.000e+00
## Minimum         -1.791e-01
## Maximum          1.392e-01
## 1. Quartile     -1.007e-02
## 3. Quartile      1.351e-02
## Mean             1.838e-03
## Median           1.087e-03
## Sum              5.398e+00
## SE Mean          4.220e-04
## LCL Mean         1.010e-03
## UCL Mean         2.666e-03
## Variance         5.240e-04
## Stdev            2.289e-02
## Skewness         1.740e-01
## Kurtosis         4.654e+00
```

```r
drSMBL <- as.vector(drSMBL)

## Using individual commands
(nobs <- length(drSMBL))
```

```
## [1] 2937
```

```r
mean(drSMBL)
```

```
## [1] 0.001838
```

```r
summary(drSMBL)
```

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## -0.17900 -0.01010  0.00109  0.00184  0.01350  0.13900
```

```r
var(drSMBL)
```

```
## [1] 0.0005239
```

```r
sd(drSMBL)/sqrt(nobs)  ## SE Mean
```

```
## [1] 0.0004224
```

```r
skewness(drSMBL)
```

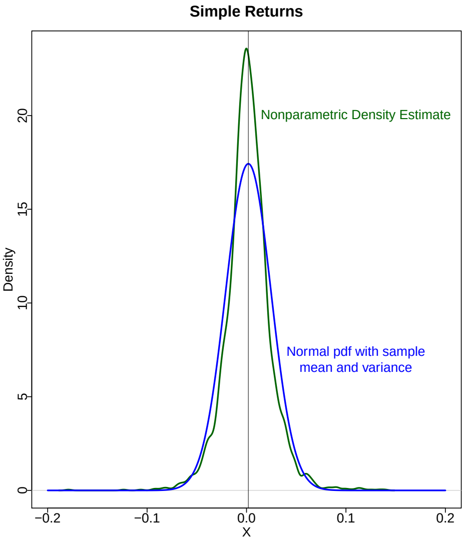
```
## [1] 0.174
## attr(,"method")
## [1] "moment"
```

```r
kurtosis(drSMBL)
```

```
## [1] 4.654
## attr(,"method")
## [1] "excess"
```

Basic Statistics and Features of Returns

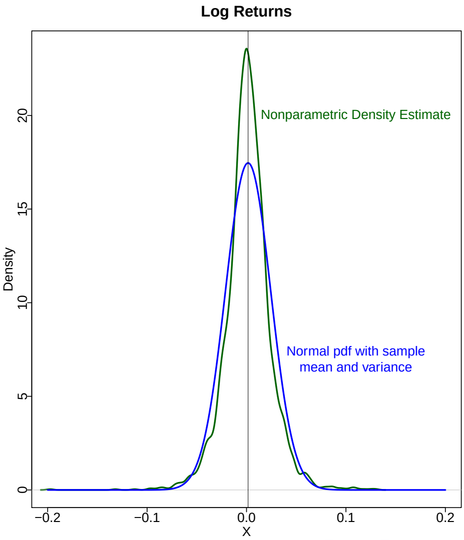
```r
## Density plots
npdensity <- density(drSMBL)
plot(npdensity, type="l",
     xlim=c(-.20,.20),xlab="X",
     main="Simple Returns",
     col="darkgreen",lwd=2)

X <- seq(-.20,.20,.001)
tdensity <- dnorm(X, mean(drSMBL), sd(drSMBL))
lines(X,tdensity,
      col="blue",lwd=2)

abline(v=mean(drSMBL),lwd=.5)

text(.11,20,
     "Nonparametric Density Estimate",
     col="darkgreen")

text(.11,7,
     "Normal pdf with sample\nmean and variance",
     col="blue")
```



Basic Statistics and Features of Returns

RICE

```r
## Density plots
npdensity <- density(ldrSMBL)
plot(npdensity,type="l",
     xlim=c(-.20,.20),xlab="X",
     main="Log Returns",
     col="darkgreen",lwd=2)

X <- seq(-.20,.20,.001)
tdensity <- dnorm(X, mean(ldrSMBL), sd(ldrSMBL))
lines(X,tdensity,
      col="blue",lwd=2)

abline(v=mean(ldrSMBL),lwd=.5)

text(.11,20,
     "Nonparametric Density Estimate",
     col="darkgreen")

text(.11,7,
     "Normal pdf with sample\nmean and variance",
     col="blue")
```



Test of Normality: Skewness

RICE

- Under normality assumption,

$$\hat{S}(x) \sim N(0, 6/T)$$

- We can test $H_0 : \hat{S}(x) = 0$ versus the alternative $H_0 : \hat{S}(x) \neq 0$
- The test statistic is

$$t_{sk} = \frac{\hat{S}(x)}{\sqrt{6/T}}$$

that is rejected at the α significance level if $|t_{sk}| > Z_{1-\alpha/2}$

Test of Normality: Kurtosis

RICE

- Under normality assumption,

$$\hat{K}(x) - 3 \sim N(0, 24/T)$$

- We can test $H_0 : \hat{K}(x) - 3 = 0$ versus the alternative $H_0 : \hat{K}(x) - 3 \neq 0$
- The test statistic is

$$t_{kr} = \frac{\hat{K}(x) - 3}{\sqrt{24/T}}$$

that is rejected at the α significance level if $|t_{kr}| > Z_{1-\alpha/2}$

Test of Normality: Jarque and Bera

RICE

- Jarque and Bera combined the previous two tests.
- The test statistic is

$$JB = \frac{\hat{S}^2(x)}{6/T} + \frac{(\hat{K}(x) - 3)^2}{24/T}$$

which is distributed as a chi-squared with 2 degrees of freedom. Then, the normality assumption is rejected at the α significance level if $JB > \chi^2_2(1-\alpha)$

Example: Jarque and Bera Test

RICE

- Are the log returns normally distributed?

```
## Test if log returns are normal
normalTest(as.vector(ldrSMBL), method="jb")

##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 3086.4822
##   P VALUE:
##     Asymptotic p Value: < 2.2e-16
##
## Description:
##  Wed Sep  3 15:32:38 2014 by user:
```

Example: t test

RICE

- Is the Mean of the simple returns equal to zero?

```
## Test if mean return is == 0
t.test(drSMBL)

##
##  One Sample t-test
##
## data:  drSMBL
## t = 4.352, df = 2936, p-value = 1.395e-05
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.001010 0.002666
## sample estimates:
## mean of x
##  0.001838
```

Important Remainder

RICE

- Remember we are assuming that:
  - *Observations are independent and identically distributed (i.i.d.)*
- NOTE: the examples we will use more than likely will NOT verify this assumption (after all, when studying time series we are interested, for example, in finding dependencies among present and past values of an asset return).

HW 1.1: Create your own basicStats

RICE

```r
## Create functions that replicate the behavior of R functions (such as:
## "mean", "var", "sd", "skewness", "kurtosis", ...) that are given as
## result of the function "basicStats". Of course you cannot use the
## functions supplied by R, but need to create your own,
## using the corresponding mathematical expressions (such as the ones
## in slides 15 and 16. You can use the functions: "length", "sum",
## "order", "sort" and similar (but not "min", "max", "summary").
## Remember that for median and quartiles, there are different
## implementations if the number of observations is odd or even.
## Create a function "myBasicStats" that, using your functions, returns
## the same results as "basicStats".
## Check that your functions are correct using the R supplied ones.
## For example:
mymean <- function(x) {  ## My implementation of the "mean" function
  sum(x)/length(x)
}
mymean(drSMBL)
```

```
## [1] 0.001838
```

```r
mean(drSMBL)             ## Checking with corresponding R function
```

```
## [1] 0.001838
```

HW 1.2: Calculating and Plotting Returns

RICE

```
## Download historical data of 3 of your favorite stock,
## and explore daily and other period returns, using the functions:
dailyReturn(SMBL[,6])
weeklyReturn(SMBL[,6])
monthlyReturn(SMBL[,6])
quarterlyReturn(SMBL[,6])

## Which is the difference between these two? Explore and explain.
annualReturn(SMBL[,6])
yearlyReturn(SMBL[,6])

## For each stock and period plot simple and log returns.
## Create a table similar to table 1.2 of the book (pg. 11)
## using your functions (from 1.1).
## Add tests (p-values).
## (STAT 621) Include monthly log excess return of one of your chosen
## stocks to the following 3 indexes (Dow Jones, S&P500, Nasdaq
## composite) in your analysis.
```

HW 1.3: More on Jarque and Bera Test

RICE

- Jarque and Bera 1980 paper (Jarque, Carlos M., and Anil K. Bera. "Efficient tests for normality, homoscedasticity and serial independence of regression residuals." Economics Letters 6.3 (1980): 255-259.), added, to the previous combination of two tests, a test for serial independence.
- Read the paper (pdf available by following this link from a campus computer).
- Provide a summary explaining this “efficient ’three-directional test’ for residual normality, homoscedasticity and serial independence (NHI)”.
- (STAT 621) Implement the R code of two tests: 1) the serial correlation test alone. 2) the NHI test. Apply your tests to daily log returns of your 3 stocks in 1.2 and interpret results.