# Lecture 2
# Introduction: Review, Returns and Data

 1

## This Class – Organization

• All the information and material is on my webpage:
https://www.bauer.uh.edu/rsusmel/4397/4397.htm

• Textbook: **Introductory Econometrics for Finance**, 4th edition or older, by Chris Brooks.

• **Exams**
- Midterms: **September 28** and **November 9** (tentative)
- Final: According to UH Schedule (**December 12**, 5PM-8PM)
- Research Project (Paper): **November 2**
- Case Presentations (presentation and discussion of project): After project, to be scheduled during Office Hours
- Homework: **Aug 31**, **Sep 14**, **Sep 26** & **Nov 30** (or **Dec 5**, depending on progress of class)

2

## This Class – Organization

- **Final Grade**

A weighted average:

Midterms: **40%**

Final : **30%**

Homework: **10%**

Class Project: **15%**

Presentation: **5%**

- **R Program**

After this class, **install R** in your machine. Previous students had a strong preference for **R Studio**.

Next class, we will introduce R and run some simple R programs. More advance programs will be run throughout the semester.

3

## This Class – Comments from Previous Classes

- **Class is very technical**

We will go over many stats concepts and definitions, some derivations and lots of formulas. But, we will apply each topic to a finance setting.

- **Class covers a lot of material**

We will cover as much as possible of Chris Brooks's textbook. Last year we were able to cover 7 chapters (& previous years, 8 chapters).

- **Instructor (me) goes fast**

Questions are a great way to slow me down. Ask questions. All questions and interruptions are appreciated

4

## This Class – Comments from Previous Classes

- **More comments from previous classes**

"It was difficult to keep awake in the class."

"Very technical course."

"Very organized lectures and course."

"I had problems with R almost all semester. Only at the end, I was able to understand what was going on."

"Learned a lot. Good course. Enjoyed the exams. One of the good courses (in program)."

"This course is <u>much</u> too quantitative."

"We covered too much info too fast."

"This is one of the few courses that I feel I've truly earned what I'm paying the university."

"He fried my brain."

5

## This class – Overview

- This is an applied technical class, with some econometric theory and many stats concepts, followed by related financial applications.

- We will review many math and statistical topics.

- Some technical material may be new to you, for example Linear Algebra. The new material is introduced to simplify exposition. You will not be required to have a deep understanding of the new material, but you should be able to follow the intuition.

- This is not a programming class, but we will use R to estimate models. I will cover some of the basics in class and I will run in class all the programs you need to run.

- For some students, the class will be dry ("*He fried my brain*," a student said in 2020.)

6

## This class – Main Applied Topics

- How do we measure returns and risks of financial assets?
- Is the equity premium (excess returns of stocks over bonds) really that high?
- Can we measure left-tail (unusual/extreme negative) risk?
- How do we determine a good model for financial assets? Is the CAPM a good model for stock returns? What about Fama-French?
- Can we explain asset returns?
- How can one explain variations in stock returns across various stocks?
- Are asset returns predictable? In the short run? In the long run?
- Are markets efficient?
- Does the risk of an asset vary with time? What are the implications? How can one model time-varying risk?

7

## This class – Main Technical Topics

- Understanding Distributions and Moments
- Testing and Confidence Intervals
- Bootstrap
- Linear Regression
- Testing Hypothesis in the Classical Linear (Regression) Model
- Finding a Good Statistical and Financial Model
- Forecasting
- Time Series Models
- Efficiency & EMH. (Application of Many Concepts)
- Time-varying Volatility (if time allows).
- Integrated Time Series and Co-integration (if time allows).

8

## This class – Goals

- Goal of the class: Students should be comfortable with applied regressions, testing financial hypothesis and forecasting.

- Secondary goal: Get students familiar with R & running R programs.

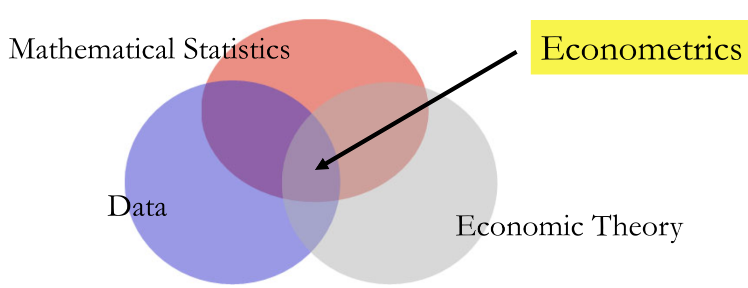
9

## What is Econometrics?

- Ragnar Frisch, Econometrica Vol.1 No. 1 (1933) revisited

"Experience has shown that each of these three view-points, that of *statistics*, *economic theory*, and *mathematics*, is a necessary, but not by itself a sufficient, condition for a real understanding of the quantitative relations in modern economic life.

It is the unification of all three aspects that is powerful. And it is this unification that constitutes econometrics."




10

## What is Econometrics?

- **Economic Theory**:
  - The CAPM: $E[r_i - r_f] = \beta_i E[(r_M - r_f)]$

- **Mathematical Statistics**:
  - Method to estimate CAPM. For example,

    Linear regression: $r_i - r_f = \alpha_i + \beta_i (r_M - r_f) + \varepsilon_i$
  - Properties of $\mathbf{b_i}$ (the LS estimator of $\beta_i$).
  - Properties of CAPM tests. For example, a *t-test* for $H_0$: $\alpha_i = 0$.

- **Data**:
  - Collect data: $r_i$, $r_f$, and $r_M$
  - Typical problems: Missing data, Measurement errors, Survivorship bias, Auto- and Cross-correlated returns, Time-varying moments.

11

## What is Financial Econometrics?

- Financial Econometrics is applied econometrics to financial data. That is, we study the statistical tools that are needed to analyze and address the specific types of questions and modeling challenges that appear in analyzing financial data.

- Always keep in mind that almost in all cases, financial data is not "*experimental data*." We have no control over the data. We have to learn how to deal with the usual problems in financial data.

- Typical applications of econometric tools to finance:
  - Describe data. For example, expected returns & volatility.
  - Test hypothesis. For example, are stocks riskier than bonds?
  - Build and test models. For example, the different Fama-French factor Models.

12

## What is Financial Econometrics?

• In general, in finance we deal with **trade-offs**. The usual trade-off: Risk & Return.

– How do we measure risk and return?

– Can we predict them?

– How do we measure the trade-off?

– How much should I be compensated for taking a given risk?

• Thus, we will be concerned with quantifying rewards and risks associated with uncertain outcomes.

13

## What is Financial Econometrics?

• Trade-off application: Fund Management

A fund manager has to allocate money across potentially many different investment alternatives to form portfolios.

At the time of the investment, the fund manager does not know what the return will be on each investment opportunity. (Returns are random variables.)

However, the fund manager can still make good investment decisions.

Q: How? By quantifying the uncertainty associated with all the investment alternatives. For this purpose, the fund manager needs a **model** for the returns of all the different investment alternatives.

From the model, we get: expected returns, variances & covariances.

14

## What is Financial Econometrics?

**This Lecture**

We will review some basic concept of Probability and Statistics:

- Population & Sample
- Sample Statistics & Estimators
- Random Variable
- Distribution Functions
- Descriptive Statistics: Moments
- Law of Large Numbers (LLN)
- Central Limit Theorem (CLT)
- Sampling Distributions

15

## Review – Population and Sample

**Definition:** Population

A population is the totality of the elements under study. We are interested in learning something about this population.

**Examples:** Number of alligators in Texas, percentage of unemployed workers in cities in the U.S., total return of all stocks in the U.S., 10-year Japanese government bond yields from 1960-2022.

Usually, a complete enumeration of all the values in the population is impractical or impossible. Thus, the descriptive statistics describing/generating the population –i.e., the *population parameters*– will be considered unknown.

Note: A Random Variable (RV) $X$ defined over a population is called the population RV. The population RV generates the data. We call the population RV the "*Data Generating Process*," or DGP.

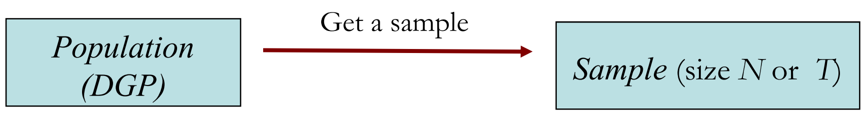
16

## Review – Population and Sample

**Definition:** Sample

The *sample* is a (manageable) subset of elements of the population.

**Example:** The total returns of the stocks on the S&P 500 index.




Samples are collected to learn about the population. The process of collecting information from a sample is referred to as *sampling*.

Definition: Random Sample
A *random sample* is a sample where the probability that any individual member from the population being selected as part of the sample is exactly the same as any other individual member of the population.

17

## Review – Population and Sample

**Example:** The total returns of the stocks on the S&P 500 index is *not* a random sample.

- In general, in finance and economics, we do not deal with random samples. The collected observations will have issues that make the sample not a true random sample.

Remark: In mathematical terms, given a random variable $X$ with distribution $F$, a *random sample* of length $n$ is a set of $n$ independent, identically distributed (*i.i.d.*) random variables with distribution $F$.

- We will estimate population parameters using sample analogues: mean, sample mean; variance, sample variance; $\boldsymbol{\beta}$, $\mathbf{b}$; etc.

18

## Review – Samples and Types of Data

- The samples we collect are classified in three groups:

- **Time Series Data**: Collected over time on one or more variables, with a particular *frequency* of observation. Example: we record for 10 years the monthly S&P 500 returns, or 10' IBM returns.

Usual notation: $x_t$, $t = 1, 2, \ldots, T$.

- **Cross-sectional Data**: Collected on one or more variables collected at a single point in time. Example: today we record all closing returns for the members of the S&P 500 index.

Usual notation: $x_i$, $i = 1, 2, \ldots, N$.

- **Panel Data**: Cross-sectional data collected over time. Example: the CRSP database collects daily prices of all U.S. traded stocks since 1962.

Usual notation: $x_{i,t}$, $i = 1, 2, \ldots, N$ & $t = 1, 2, \ldots, T$.

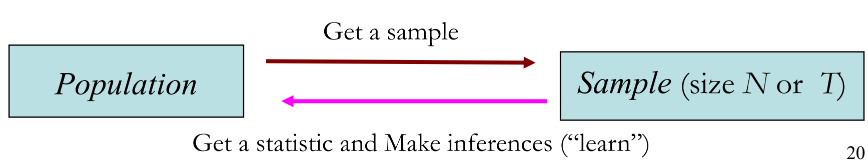
19

## Review – Statistics

- Typical situation in statistics: We want to make inferences about an unknown population parameter $\theta$ using a sample: $\{X_1, X_2, \ldots, X_N\}$.

- We summarize the information in the sample with a *statistic*, which is a function of the sample.

- Any statistic summarizes the data or reduces the information in the sample to a single number. To make inferences, we use the information in the statistic instead of the entire sample.




20

## Review – Sample Statistic

- A *statistic* (singular) is a single measure of some attribute of a sample (for example, its arithmetic mean value). It is calculated by applying a function (statistical algorithm) to the values of the items comprising the sample, which are known together as a set of data.

**Definition:** Statistic

A *statistic* is a function of the observable random variable(s), which does not contain any unknown parameters.

**Examples:** sample mean ($\bar{X}$), sample variance ($s^2$), minimum, median, $(x_1 + x_N)/2$, etc.

Note: A statistic is distinct from a population parameter. A statistic will be used to estimate a population parameter. In this case, the statistic is called an *estimator*.

21

## Review – Sample Statistic: Estimators

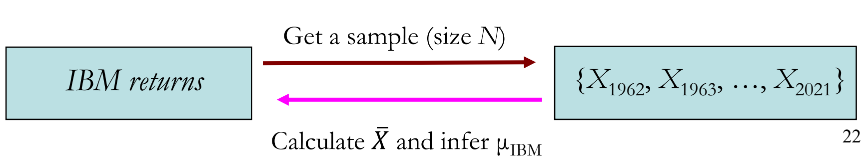
- Sample Statistics are used to estimate population parameters

**Example:** $\bar{X}$ is an estimate of the population mean, μ.

Notation: Population parameters: Greek letters (μ, σ, θ, etc.)

Estimators: A hat over the Greek letter ($\hat{\theta}$).

Suppose we want to learn about the mean of IBM annual returns, $\mu_{IBM}$. From the population, we get a sample of size $N$: $\{X_{1962}, X_{1963}, \ldots, X_{N=2021}\}$. Then, we compute a statistic, $\bar{X}$. As we will see later, on average $\bar{X}$ is a good estimator of μ.




22

## Review – Sample Statistic: Estimators

- The definition of a sample statistic is very general. For example, by definition $(x_1 + x_N)/2$ is a statistic; we could claim that it estimates the population mean of the variable *X*. However, this is probably not a good estimate.

- We would like our estimators to have certain desirable properties.

23

## Review – Sample Statistic: Properties

- Some simple properties for estimators, $\hat{\theta}$:

- $\hat{\theta}$ is *unbiased* estimator of $\theta$ if $E[\hat{\theta}] = \theta$.

- $\hat{\theta}$ is *most efficient* if the variance of the estimator is minimized.

- $\hat{\theta}$ is *BUE*, or *Best Unbiased Estimator,* if it is the estimator with the smallest variance among all unbiased estimates.

- $\hat{\theta}$ is *consistent* if as the sample size, *N*, increases to $\infty$, $\hat{\theta}_N$ converges to $\theta$. We write $\hat{\theta}_N \xrightarrow{p} \theta$.

- $\hat{\theta}$ is *asymptotically normal* if as the sample size, *N*, increases to $\infty$, $\hat{\theta}_N$ ,often standardized or transformed, converges in distribution to a Normal distribution. We write $\hat{\theta}_N \xrightarrow{d} N(\theta, Var(\hat{\theta}_N))$.

24

## Review – Random Variable

• In probability, a *random variable* (RV), or *stochastic variable*, is described informally as a variable whose values depend on outcomes of an *experiment*. (Experiment: Act/process with an unknown outcome).

**Examples:**

1. We throw two coins and count the number of heads.
2. We define X = 1 if the economy grows two consecutive quarters and X = 0, otherwise. (This is an example of a *Bernouille* (or *indicator*) RV.)
3. We read comments from IBM's CEO and compute IBM's return.
4. We count the days in a week the stock market has a positive return.
5. We look at a CEO and write his/her highest education degree.

• For some RVs, it is easy to enumerate all possible outcomes. For instance, for the fourth example above: {0, 1, 2, 3, 4, 5}. The set of all possible outcomes is called *sample space,* usually denoted by Ω.

25

## Review – Random Variable

• In general, a RV is a *function* whose domain is the sample space. It produces numbers. (The name "random variable" is confusing; it is just a function!)

Definitions & Notation:

Ω the sample space: The set of possible outcomes from an experiment.

An event *A*: A set containing outcomes from the sample space.

Σ: Collection of all possible events involving outcomes from Ω.

P: A function assigning a number between [0,1] to each event in Σ.

RV **X**: A function takes Ω into $\chi$ (numbers) and induces $P_X$ from P.

P is the probability measure over the sample space, Ω, and $P_X$ is the probability measure over $\chi$, the range of the random variable.

26

## Review – Random Variables – Example

**Example:** The weekly sign of stock returns of two unrelated firms: Positive (U: up) or negative (D: down).

The sample space is Ω = [{U,D}; {D,U}; {D,D} & {U,U}].

Possible events (*A*):

- Both firms have the same signed return: {U,U} & {D,D}.
- At least one firm has positive returns: {U,U}; {D,U} & {U,D}.
- The first firm has positive returns: {U,U} & {U,D}

Collection of all possible events: Σ = [Φ, {U,U}, {U,D}, {D,U}, {D,D}, {UU, UD}, {UU, DU}, {UU, DD}, {DD, DU}, {DD, UD}, {DU, DD}, {UU, DU, UD}, {UD, DU, DD}, {UU, UD, DU, DD}]

Define RV: **X** = *"Number of Up cycles."* Recall, **X** takes Ω into $\chi$,

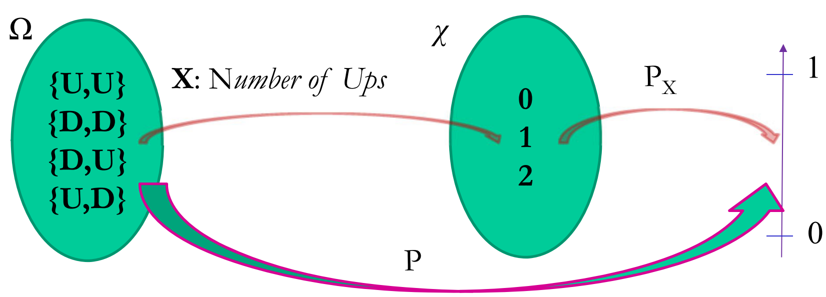
$\chi$ = {0; 1; 2} and $\Sigma_\chi$ = {Φ; {0}; {1}; {2}; {0;1}; {0;2}; {1;2}; {0;1;2}} 27

## Review – Random Variables – Example

**Example (continuation):**

Then, **X**: $\Omega \rightarrow \chi$

Then, we associate the elements in $\chi$ with a probability, $P_X$.




In this example, $\chi$ = {0; 1; 2}

$\Sigma_\chi$ = {Φ; {0}; {1}; {2}; {0;1}; {0;2}; {1;2}; {0;1;2}} 28

## Review – Random Variables – Example

**Example (continuation): X** takes Ω into $\chi$ & induces $P_X$ from P.

Assuming equal probabilities for U & D, P[U] = P[D] = ½ :

Prob. of 0 Ups = $P_X[0]$ = P[{DD}] = ¼

Prob. of 1 Ups = $P_X[1]$ = P[{UD; DU}] = ½

Prob. of 2 Ups = $P_X[2]$ = P[{UU}] = ¼

Prob. of 0 or 1 Ups = $P_X[\{0; 1\}]$ = P[{DD; DU; UD}] = ¾

Prob. of 0 or 2 Ups = $P_X[\{0; 2\}]$ = P[{DD; UU}] = ½

Prob. of 1 or 2 Ups = $P_X[\{1; 2\}]$ = P[{UU; DU; DD}] = ¾

Prob. of 0, 1, or 2 Ups = $P_X[\{0; 1; 2\}]$ = P[{UU; DU; UD; DD}] = 1

Prob. of "nothing" = $P_X[\Phi]$ = P[Φ] = 0

The empty set is simply needed to complete the σ-algebra (a technical point). Its interpretation is not important since P[Φ] = 0.

29

## Review – Random Variable: Remarks

**Remarks**:

- A random variable is a convenient way to express the elements of Ω as numbers rather than abstract elements of sets.
- A random variable ***X*** is a function.
- It is a numerical quantity whose value is determined by a random experiment.
- It takes single elements in outcome set Ω, which can be abstract elements, and maps them to points in R.

• We put some mathematical structure (pdf, CDF) to the concept of RV to describe what is more/less likely to happen to the (randomly determined) values.

30

## Review – PDF for a Continuous RV

**Definition**: Suppose that $X$ is a random variable. Let $f(x)$ denote a function defined for $-\infty < x < \infty$ with the following properties:

1. $f(x) \geq 0$.
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.
3. $P[a \leq X \leq b] = \int_a^b f(x)dx$

- Then, $f(x)$ is called the *probability density function* (pdf) of $X$. The RV $X$ is called *continuous*. We use the pdf to describe the behavior of $X$.

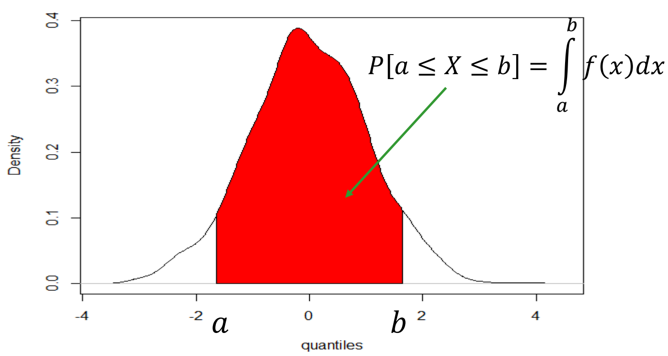
- Analogous definition applies for a discrete RV, with $p(x)$ notation:
  1. $0 \leq p(x) \leq 1$
  2. $\sum_{i=1}^{\infty} p(x_i) = 1$
  3. $P[a \leq X \leq b] = \sum_{a \leq x \leq b} p(x)$

31

## Review – PDF for a Continuous RV

- The pdf is non-negative and integrates to $\int_{-\infty}^{\infty} f(x)dx = 1$.

PDF for a Continuous RV: Area between *a* & *b*



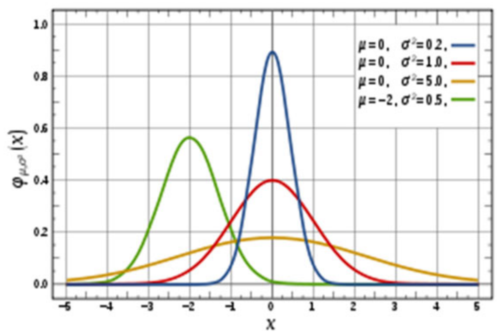

- Remark: We use the pdf to describe the behavior of the RV (discrete or continuous).

32

## Review – Popular PDFs: Normal Distribution

• A RV $X$ is said to have a *normal distribution* with parameters $\mu$ *(mean)* and $\sigma^2$ *(variance)* if $X$ is a continuous RV with pdf $f(x)$:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$




Note: Described by two parameters: $\mu$ and $\sigma^2$. We write $X \sim N(\mu, \sigma^2)$

33

## Review – Popular PDFs: Normal Distribution

• When $\mu = 0$ and $\sigma^2 = 1$, we call the distribution *standard normal.* We write $X \sim N(0, 1)$. This is the distribution that is tabulated.

The normal distribution is often used to describe or approximate any variable that tends to cluster around the mean. It is the most assumed distribution in economics and finance: rates of return, growth rates, IQ scores, observational errors, etc.

• The central limit theorem (CLT) provides a justification for the normality assumption when the sample size, *N,* is large.

Notation: PDF: $X \sim N(\mu, \sigma^2)$
CDF: $\Phi(x)$

34

## Review – Popular PDFs: Gamma Distribution

- Let the continuous RV $X$ have density function):

$$f(x) = \begin{cases} \dfrac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where $\alpha, \lambda > 0$ and $\Gamma(\alpha)$ is the gamma function evaluated at $\alpha$.

Then, $X$ is said to have a *Gamma distribution* with parameters $\alpha$ and $\lambda$, denoted as $X \sim \text{Gamma}(\alpha, \lambda)$ or $\Gamma(\alpha, \lambda)$.

It is a family of distributions, with special cases:

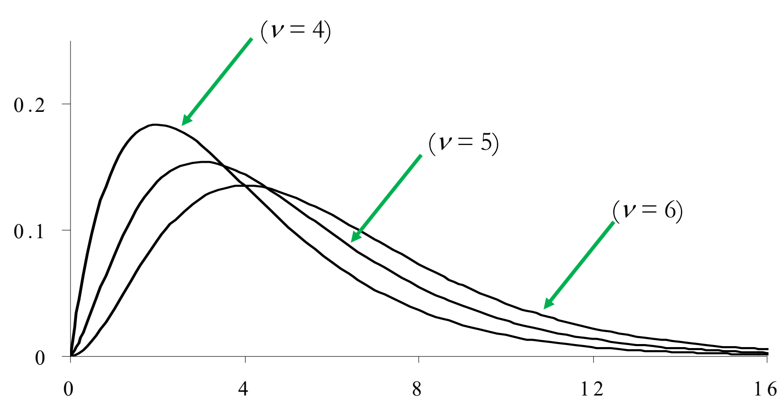
– Exponential Distribution, or Exp($\lambda$): $\alpha = 1$.

– Chi-square Distribution, or $\chi^2_\nu$: $\alpha = \nu/2$ and $\lambda = ½$.

35

## Review – Popular PDFs: Gamma Distribution

- The Chi-square distribution, $\chi^2_\nu$, will appear a lot, since it is derived from a sum of $\nu$ independent square standard normals. Below we plot the $\chi^2_\nu$ distribution with parameter $\nu$, called degrees of freedom:




Note: When $\nu$ is large, the $\chi^2_\nu$ converges to a N($\nu$, 2$\nu$).

36

## Review – Popular PDFs: Other Distributions

- Other distributions we will use: t-distribution and F-distribution.

**(1) t-distribution**: A ratio of a standard normal and the square root of a Chi-squared distribution divided by its degrees of freedom. That is, let Y ~ N(0, 1) and W ~ $\chi^2_\nu$, then

$$t = \frac{Y}{\sqrt{W/\nu}} \sim t_\nu.$$

It looks like a normal distribution, but with thicker tails. As $\nu$ increases, the t-distribution converges to a N(0, 1) distribution.

**(2) F-distribution**: A ratio of two independent $\chi^2$ distributions, divided by their degrees of freedom. That is, let $Z_1 \sim \chi^2_{\nu_1}$ and $Z_2 \sim \chi^2_{\nu_2}$, then

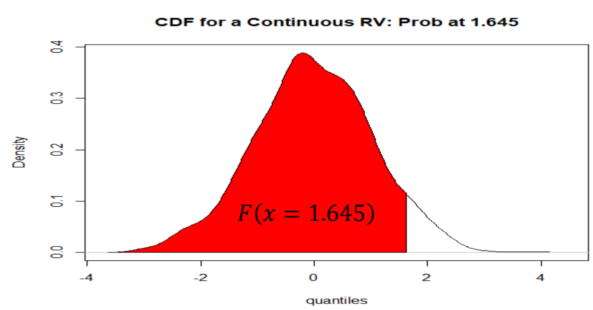
$$F = \frac{Z_1/\nu_1}{Z_2/\nu_2} \sim F_{\nu,\nu_2}$$

37

## Review – CDF for a Continuous RV

- If $X$ is a continuous random variable with probability density function, $f(x)$, the *cumulative distribution function* (CDF) of $X$ is given by:

$$F(x) = P[X \le x] = \int_{-\infty}^{x} f(t)\, dt.$$




- Note: The FTC implies: $F'(x) = \frac{dF(x)}{dx} = f(x)$

38

## Review – The Empirical Distribution

• The empirical distribution (ED) of a dataset is simply the distribution that we observe in the data.

The ED is a discrete distribution that gives equal weight to each data point, assigning a $1/N$ probability to each of the original $N$ observations.

We form a cumulative distribution function, $F^*$, that is a step function that jumps up by $1/N$ at each of the $N$ data point:

$$F^*(x) = \sum_{i=1}^{N} I(x_i \leq x),$$

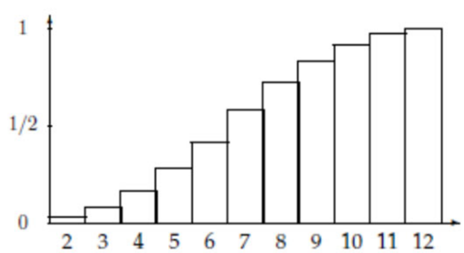
where $I(.)$ is the indicator function:

$$I(x_i \leq x) = 1, \qquad \text{if } x_i \leq x$$
$$I(x_i \leq x) = 0, \qquad \text{if } x_i > x$$

39

## Review – The Empirical Distribution

**Example:** We throw 100 times two dice and sum the results. The CDF is given below:



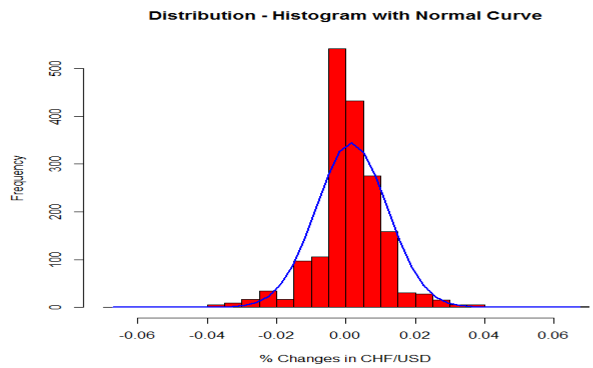

• In general, we use a histogram to describe the ED of a dataset.

Important result: Let $F$ be the true distribution of the data and $F^*$ be the ED of the data. As $N \to \infty$, the Law of large numbers (LLN) tells us that $F^*$ becomes a good approximation of $F$.

40

## Review – Histogram of a RV

**Example:** We use a histogram to estimate the distribution of a RV. Let
$X$ = Percentage changes in the CHF/USD exchange rate = $e_f$
Data: Monthly from January 1971 to June 2020 ($N$=594 observations).




Note: We overlay a Normal density (blue line) over the histogram.

41

## Review – Moments of Random Variables

• The moments of a random variable $X$ are used to describe the behavior of the RV (discrete or continuous).

**Definition**: $K^{th}$ Moment
Let $X$ be a RV (discrete or continuous), then the *$k^{th}$ moment of* $X$ is:

$$\mu_k = E(X^k) = \begin{cases} \sum_x x^k p(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} x^k f(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

• The first moment of $X$, $\mu = \mu_1 = E(X)$ is the center of gravity of the distribution of $X$.

• The higher moments give different information regarding the shape of the distribution of $X$.

42

## Review – Moments of a RV

**Definition:** Central Moments

Let $X$ be a RV (discrete or continuous). Then, the $k^{th}$ *central moment* of $X$ is defined to be:

$$\mu_k^0 = E[(X-\mu)^k] = \begin{cases} \sum_x (x-\mu)^k p(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} (x-\mu)^k f(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

where $\mu = \mu_1 = E(X) =$ the first moment of $X$.

- The central moments describe how the probability distribution is distributed about the center of gravity, $\mu$.

43

## Review – Moments of a RV

- The first central moments is given by:

$$\mu_1^0 = E[X-\mu] = 0$$

The second central moment depends on the *spread* of the probability distribution of $X$ about $\mu$. It is called the variance of $X$ and is denoted by the symbol $\sigma^2 = \text{var}(X)$:

$$\mu_2^0 = E[(X-\mu)^2] = \text{Var}[X] = \sigma^2$$

The square root of var($X$) is called the *standard deviation* of $X$ and is denoted by the symbol $\sigma = \text{SD}(X)$. We also refer to it as *volatility*:

$$\sqrt{\mu_2^0} = \sqrt{E[(X-\mu)^2]} = \sigma$$

44

## Review – Moments of a RV: Skewness

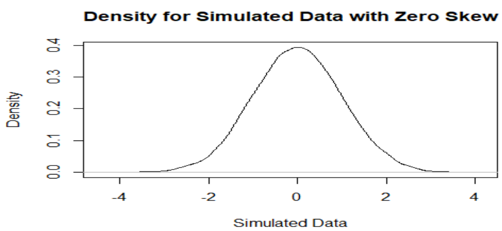
- The third central moment: $\mu_3^0 = E[(X-\mu)^3]$

$\mu_3^0$ contains information about the *skewness* of a distribution.

- A popular measure of skewness: $\gamma_1 = \frac{\mu_3^0}{\sigma^3} = \frac{\mu_3^0}{(\mu_2^0)^{\frac{3}{2}}}$

- Distribution according to skewness:

1) Symmetric distribution

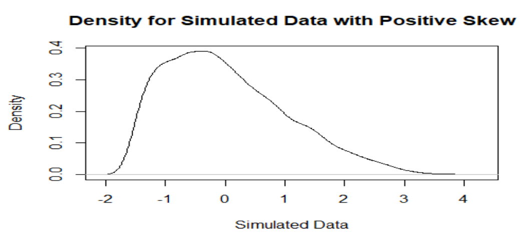



$\mu_3^0 = 0, \gamma_1 = 0$

45

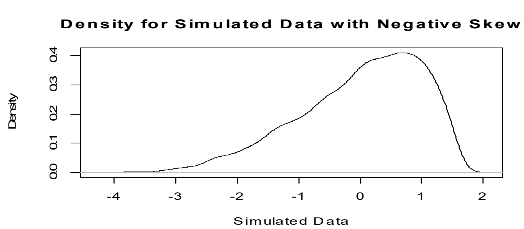
## Review – Moments of a RV: Skewness

2) Positively (right-) skewed distribution (with mode < median < mean)




$\mu_3^0 > 0, \gamma_1 > 0$

3) Negatively (left-) skewed distribution (with mode > median > mean)




$\mu_3^0 < 0, \gamma_1 < 0$

46

## Review – Moments of a RV: Skewness

- Skewness and Economics

For returns:
- Zero skew means symmetrical gains and losses.
- Positive skew suggests many small losses and few rich returns.
- Negative skew indicates lots of minor wins offset by rare major losses.

- In financial markets, stock returns at the firm level show positive skewness, but at the aggregate (index) level show negative skewness.

- From horse race betting and from U.S. state lotteries there is evidence supporting the contention that gamblers are not necessarily risk-lovers but **skewness-lovers**: Long shots are overbet (positive skewness loved!).

47

## Review – Moments of a RV: Kurtosis

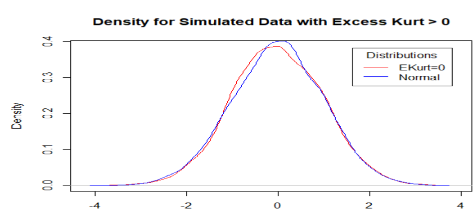
- The fourth central moment: $\mu_4^0 = E[(X-\mu)^4]$.

$\mu_4^0$ is a measure of the *shape* of a distribution. The property of shape measured by this moment is called *kurtosis,* usually estimated by $\kappa = \frac{\mu_4^0}{\sigma^4}$.

- The *measure of (excess) kurtosis*: $\gamma_2 = \frac{\mu_4^0}{\sigma^4} - 3 = \frac{\mu_4^0}{(\mu_2^0)^2} - 3$.

- Distributions:

1) Mesokurtic distribution ($\gamma_2 = 0$ or κ=3, like the normal distribution)

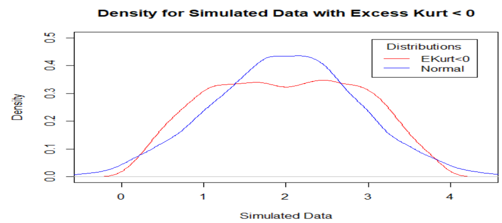



48

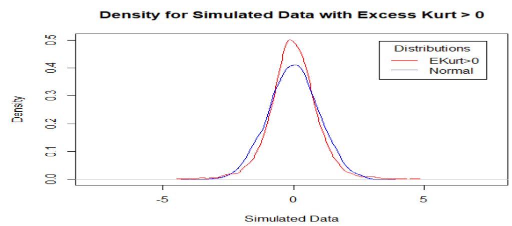
 

## Review – Moments of a RV: Kurtosis

2) Platykurtic distribution ($\gamma_2 < 0$)



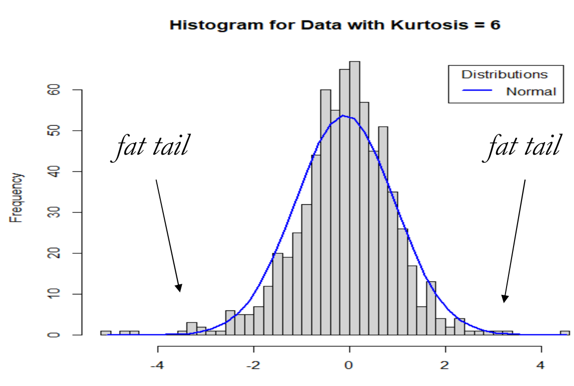

3) Leptokurtic distribution ($\gamma_2 > 0$, usual shape for asset returns)




49

## Review – Moments of a RV: Kurtosis

- Typical financial returns series has $\gamma_2 > 0$. Below, I simulate a series with $\mu$=0, $\sigma$=1, $\gamma_1$=0 & kurtosis = 6 ($\gamma_2$=3), overlaid with a standard normal distribution. Fat tails are seen on both sides of the distribution.




50

## Review – Moments and Expected Values

• Note that moments are defined by expected values. We define the expected value of a function of a continuous RV X, $g(X)$, as

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

• If $X$ is *discrete* with probability function $p(x)$

$$E[g(X)] = \sum_x g(x)p(x) = \sum_i g(x_i)p(x_i)$$

**Examples**: $g(x) = (x-\mu)^2 \quad \Rightarrow \text{E}[g(x)] = \text{E}[(x-\mu)^2]$

$g(x) = (x-\mu)^k \quad \Rightarrow \text{E}[g(x)] = \text{E}(x-\mu)^k]$

• We estimate expected values with sample averages. The Law of Large Numbers (LLN) tells us they are *consistent* estimators of expected values.

51

## Review – Estimating Moments

• We estimate expected values with sample averages. For example, the first moment, $\mu$, & the second central moment, $\sigma^2$, are estimated by:

$$\bar{X} = \frac{\sum_{i=1}^{N} x_i}{N}$$

$$s^2 = \frac{\sum_{i=1}^{N}(x_i - \bar{X})^2}{N-1}$$ (*N*-1 adjustment needed for $\text{E}[s^2] = \sigma^2$)

• They are both *unbiased* estimators of their respective population moments That is,

$\text{E}[\bar{X}] = \mu$

$\text{E}[s^2] = \sigma^2$ "$\mu$ & $\sigma^2$population parameter"

<u>Note</u>: Unbiased estimator = "On average, we get the population parameter."

52

## Review – Law of Large Numbers (LLN)

• Long history: Gerolamo Cardano (1501-1576) stated it without proof. Jacob Bernoulli published a rigorous proof in 1713.

**Theorem (Weak LLN)**

Let $X_1, \ldots, X_N$ be $N$ mutually independent random variables each having mean $\mu$ and a finite $\sigma$ -i.e, the sequence $\{X_N\}$ is *i.i.d.*

Let $\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$.

Then for any $\delta > 0$ (no matter how small)

$$P[|\bar{X} - \mu| < \delta] = P[\mu - \delta < \bar{X} < \mu + \delta] \rightarrow 1, \qquad \text{as } N \rightarrow \infty$$

• There are many versions of the LLN. It is a general result: A sample average as the sample size goes to infinite tends to its expected value. Also written as $\bar{X}_N \overset{p}{\rightarrow} \mu$. (convergence in probability)

53

## Review – Central Limit Theorem (CLT)

• Let $X_1, \ldots, X_N$ be a sequence of *i.i.d.* RVs with finite mean $\mu$, and finite variance $\sigma^2$. Then, as $N$ increases, $\bar{X}_N$, the sample mean, approaches the normal distribution with mean $\mu$ and variance $\sigma^2/N$.

This theorem is sometimes stated as $\frac{\sqrt{N}(\bar{X} - \mu)}{\sigma} \overset{d}{\rightarrow} N(0,1)$

where $\overset{d}{\rightarrow}$ means "the limiting distribution (asymptotic distribution) is" (or *convergence in distribution*).

• Many versions of the CLT. This one is the *Lindeberg-Lévy CLT*.

• The CLT gives only an asymptotic distribution. We usually take it as an approximation for a finite number of observations. In these cases, the notation goes from $\overset{d}{\rightarrow}$ to $\overset{a}{\rightarrow}$.

54

## Review – Sampling Distributions: $\bar{X}$

- All statistics, T($X$), are functions of RVs and, thus, they have a distribution. Depending on the sample, we can observe different values for T($X$), thus, the finite sample distribution of T($X$) is called the *sampling distribution*.

For the sample mean $\bar{X}$, if the $X_i$'s are normally distributed, then the sampling distribution is normal with mean μ and variance $\sigma^2/N$. Or

$$\bar{X} \sim \text{N}(\mu, \sigma^2/N).$$

Note: If the data is not normal, the CLT is used to approximate the sampling distribution by the asymptotic one, usually after some manipulations. Again, in those cases, the notation goes from $\xrightarrow{d}$ to $\xrightarrow{a}$.

- The SD of the sampling distribution is called the *standard error* (SE). Then, SE($\bar{X}$) = σ/sqrt($N$).

55

## Review – Sampling Distributions: $\bar{X}$

- Summary for $\bar{X}$:

Sampling distribution: $\bar{X} \sim \text{N}(\mu, \sigma^2/N)$.

Mean: $\text{E}[\bar{X}] = \mu$

Variance: $\text{Var}[\bar{X}] = \sigma^2/N$.

Note: If the data is not normal (& $N$ is large), the CLT can be used to approximate the sampling distribution by the asymptotic one:

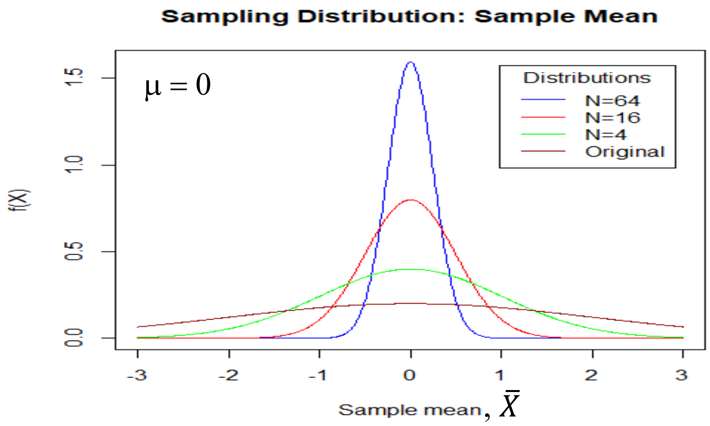
$$\bar{X} \xrightarrow{a} \text{N}(\mu, \sigma^2/N).$$

56

## Review – Sampling Distributions: $\bar{X}$

- Sampling Distribution for the Sample Mean of a normal population:

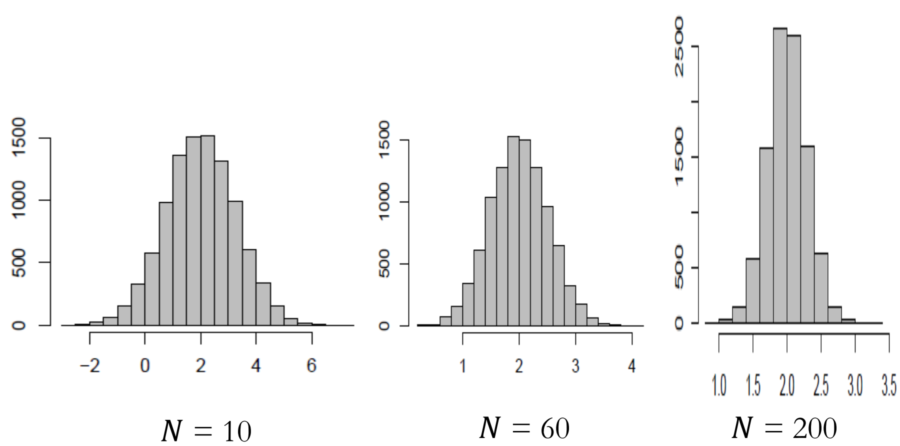
$$\bar{X} \sim N(\mu, \sigma^2/N)$$




Note: As $N \rightarrow \infty$, $\bar{X} \rightarrow \mu$ –i.e., the distribution becomes a spike at $\mu$! 57

## Review – Sampling Distributions: $\bar{X}$

- 10,000 samples, for a N(2, 4) population. Different sample sizes, $N$:




Note: As $N \rightarrow \infty$, $\bar{X}$ becomes more concentrated around $\mu = 2$. In the limit, a spike at $\mu$! 58

## Review – Sampling Distributions: $s^2$

- For the sample variance $s^2$, if the $X_i$'s are normally distributed, the sampling distribution is derived from this result:

$$(N-1)\, s^2/\sigma^2 \sim \chi^2_{N-1}.$$

We use the properties of a $\chi^2_v$ to derive the mean & variance of $s^2$:

**Property 1.** Let $Z \sim \chi^2_v$. Then, $E[Z] = v$.
**Property 2.** Let $Z \sim \chi^2_v$. Then, $Var[Z] = 2 * v$.

**Application:** Let $Z = (N-1)\, s^2/\sigma^2 \sim \chi^2_{N-1}$
From **Property 1**: $E[(N-1)\, s^2/\sigma^2] = N-1$
$\Rightarrow E[s^2] = \sigma^2$

From **Property 2**: $Var[(N-1)\, s^2/\sigma^2] = 2 * (N-1)$
$\Rightarrow Var[s^2] = 2 * \sigma^4/(N-1)$
$\Rightarrow SE(s^2) = SD(s^2) = \sigma^2 * \text{sqrt}[2/(N-1)]$

59

## Review – Sampling Distributions: $s^2$

- Summary for $s^2$ of normal variates:

Sampling distribution: $(N-1)\, s^2/\sigma^2 \sim \chi^2_{N-1}$.

Mean: $E[s^2] = \sigma^2$

Variance: $Var[s^2] = 2 * \sigma^4/(N-1)$.

Note: If the data is not normal (& $N$ is large), the CLT can be used to approximate the sampling distribution by the asymptotic one:

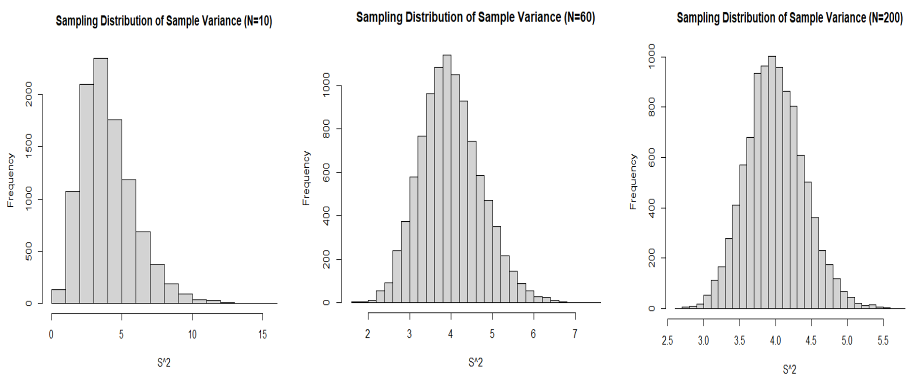
$$s^2 \xrightarrow{a} N(\sigma^2, \sigma^4 * (\kappa - 1)/ N)$$

where $\kappa = \frac{\mu_4^0}{\sigma^4}$ (recall when data is normal, $\kappa = 3$).

60

## Review – Sampling Distributions: $s^2$

- 10,000 samples, for a N(2, 4) population. Different sample sizes, $N$:




$N = 10$ $N = 60$ $N = 200$

Note: As $N \rightarrow \infty$, the distribution of $s^2$ looks more Normal – the CLT at work!

61