Quantitative Analysis of Financial Markets

Session 2: Simple Hypothesis Testing

Benjamin Ee

Week 2

Briefing on class project

- All groups will present during regularly scheduled class timings on class 10 [*final class*]
- Please submit a project proposal by the Saturday of week 4 to allow for written feedback
- Around 15 minutes presentation time followed by 5 minutes of Q&A

Potential project topics

Groups are free to think of their own topics. Some ideas:

- **Alpha generation / portfolio construction:** E.g. multi-factor models

- **Risk management**: E.g.
  - **Portfolio tail risk hedging** (e.g. see https://www.bloomberg.com/news/articles/2020-04-08/taleb-advised-universa-tail-risk-fund-returned-3-600-in-march)
  - **Forecasting bond defaults** with a probit / logit model
  - **Forecasting credit ratings** with a ordered probit/logit model,
  - **Forecasting value at risk** using various more finely tuned distributions [what happens to value at risk when we introduce leptokurtosis and skewness into a normal setting?]

- **GDP / inflation / unemployment forecasting** [time series modelling]

Unique / unconventional projects will have a higher score. Plus, **your fellow classmates will be able to learn many new things** from your work

Summary project deadlines

In your groups:

- **Week 4 (Saturday)**: Discuss and email project proposal by this date
- **Week 4 (Sunday)**: I will provide written feedback on or around this date, and try to meet groups following week via Zoom
- **Class 10**: Presentation date

Overall, we want ideas and implementations which are:

- Grading is equal weighted across below 3 points
- Economically / commercially interesting question. Unique/creative scores extra points. Although there are some suggestions on the previous slide, you are ultimately responsible for thinking of an interesting topic yourself
- Quantitatively rigorous methodology.
- Insightful discussion of results [grading will be based more on 'what you learn from the quantitative results' than whether you get a significant outcome]

Today:

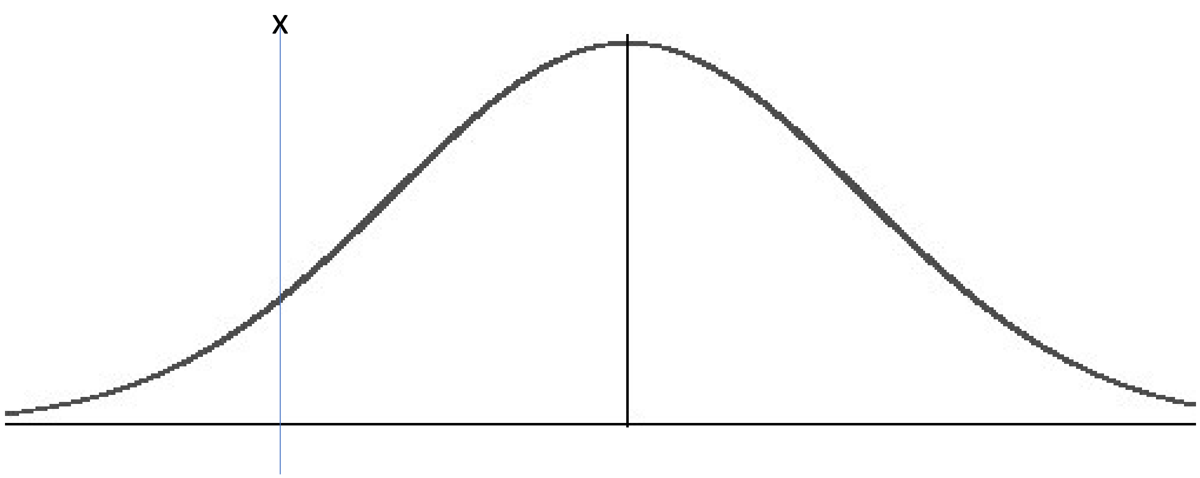
Hypothesis Testing and Statistics

Hypothesis Testing

1. **Standardizing** a single variable
2. How to work with **linear combinations** of variables
3. **Hypothesis tests**
    1. **Population**
    2. **Sample**

Single variable standardization

Motivating question



Variable x is drawn from a statistical distribution **with normal probability density function**

a. Is x **significantly different from 0**?

b. Risk management: **What must x be so that probability of being <=x is (say) 5%?**

This would depend on:

i. **Where x is** in relation to ‘center’ of distribution [mean]

ii. **How ‘thin’ or ‘spread out’ the distribution** is [variance], and also ‘higher order’ metrics such as how thick/thin the ‘tails’ are

Example: asset returns

- Simple return over one period (eg. 5 minutes, one day, one week, one month)

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

- If P_t is observed at t, then the resulting R_t is said to be ex post return.
- If only P_{t-1} is known but P_t is not observed yet, then R_t is said to be *ex ante return*.
- The ex ante return R_t is a random variable.
- Expected value of R_t :

$$\mu := \mathbb{E}(R_t), \qquad \forall t$$

- Variance of R_t :

$$\sigma^2 := \mathbb{V}(R_t) = \mathbb{E}\left((R_t - \mu)^2\right) = \mathbb{E}(R_t^2) - \mu^2, \qquad \forall t$$

Example statistical distribution: Normal

- A very common assumption of finance is that the returns are normally distributed. Lets go with this assumption for now, and we will revisit later.

$$r \overset{d}{\sim} N(\mu, \sigma^2).$$

- The probability density function f(r) is

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

- The mean and variance are, respectively,

$$\begin{aligned} \mathbb{E}(r) &= \int_{-\infty}^{\infty} r f(r)\, dr = \mu; \\ \mathbb{V}(r) &= \int_{-\infty}^{\infty} (r-\mu)^2 f(r)\, dr = \sigma^2. \end{aligned}$$

Standard Normal Distribution

- For convenience, define

$$z := \frac{r_t - \mu}{\sigma}, \qquad f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

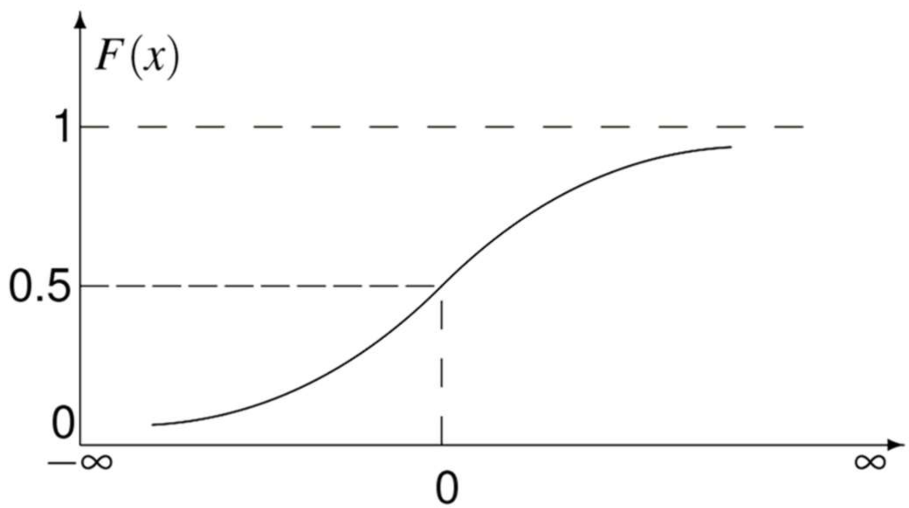
- The probability density function f(z) is the well-known bell-shaped curve with mean 0 and variance 1

Cumulative Distribution Function F(x)

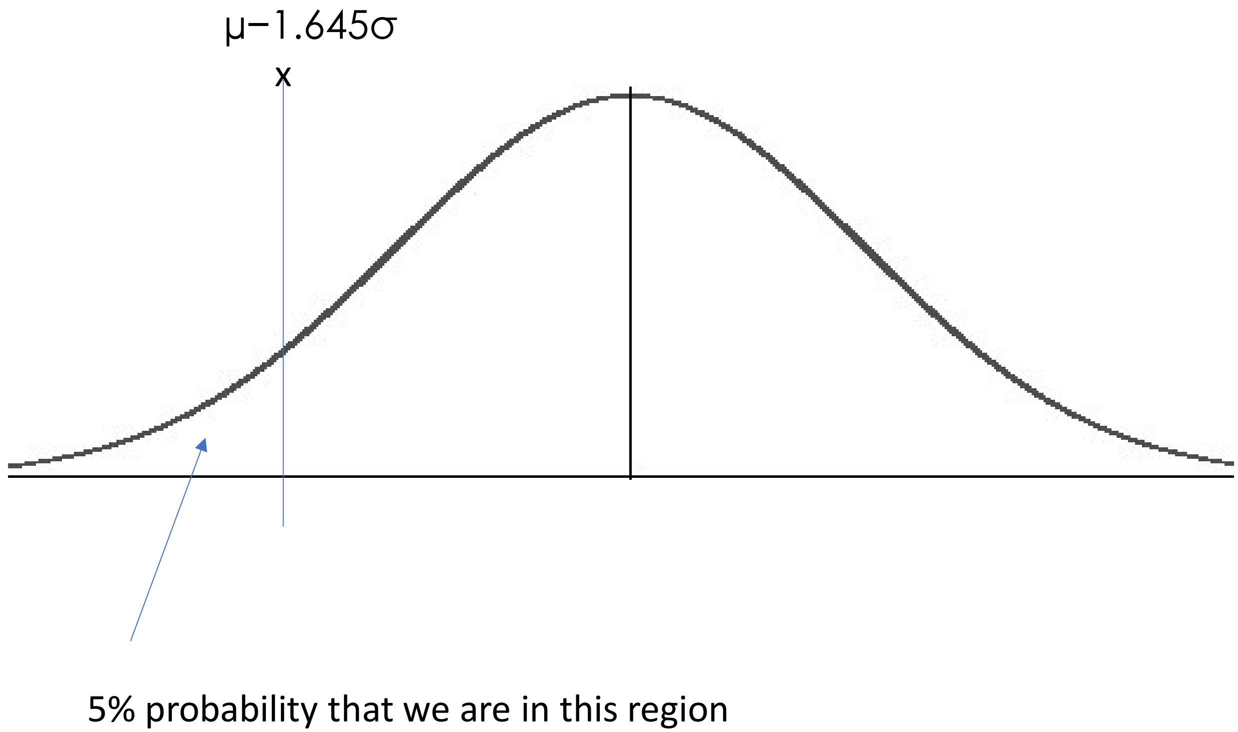
- What is the probability that z <−1.645?

$$F(-1.645) := \mathbb{P}(z < -1.645) = \int_{-\infty}^{-1.645} f(z)\, dz = 0.05$$

- Thus there is 5% probability that $r_t < \mu - 1.645\sigma$



Motivating question revisited: Value at Risk



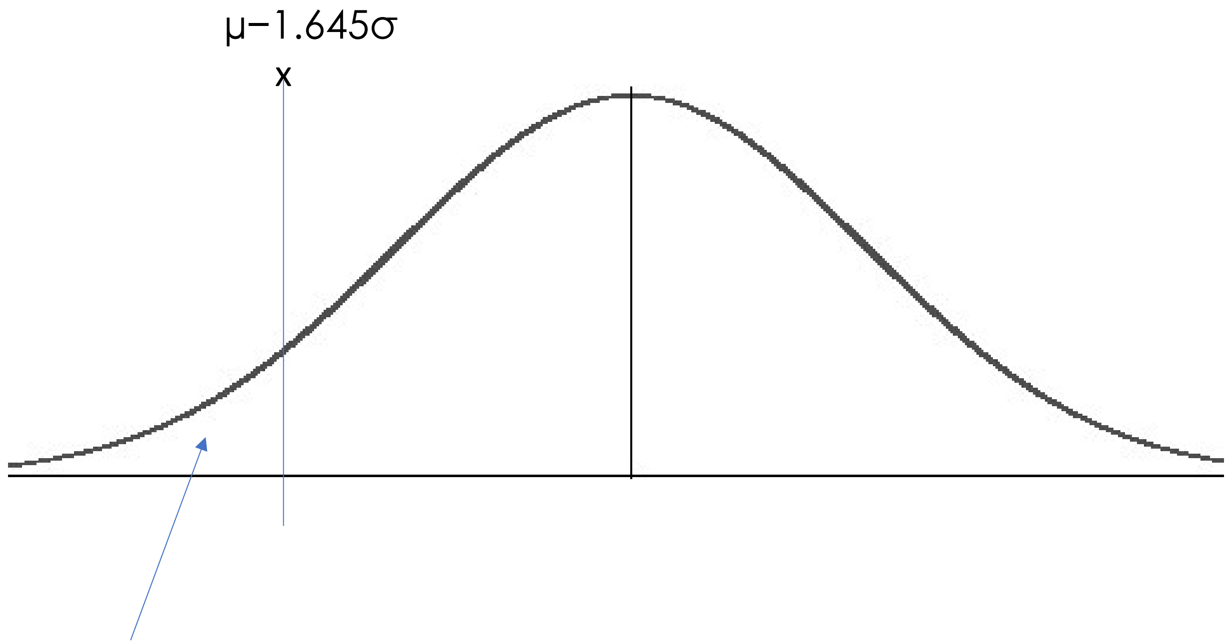
There is a 5% chance that returns will be at **x = u - 1.645σ or lower.**

1-day 95% Value at Risk (VaR) = x%

x is generally a **negative number**

This means "there is a **95% chance that daily loss will not exceed x%"**

Motivating question revisited: Value at Risk



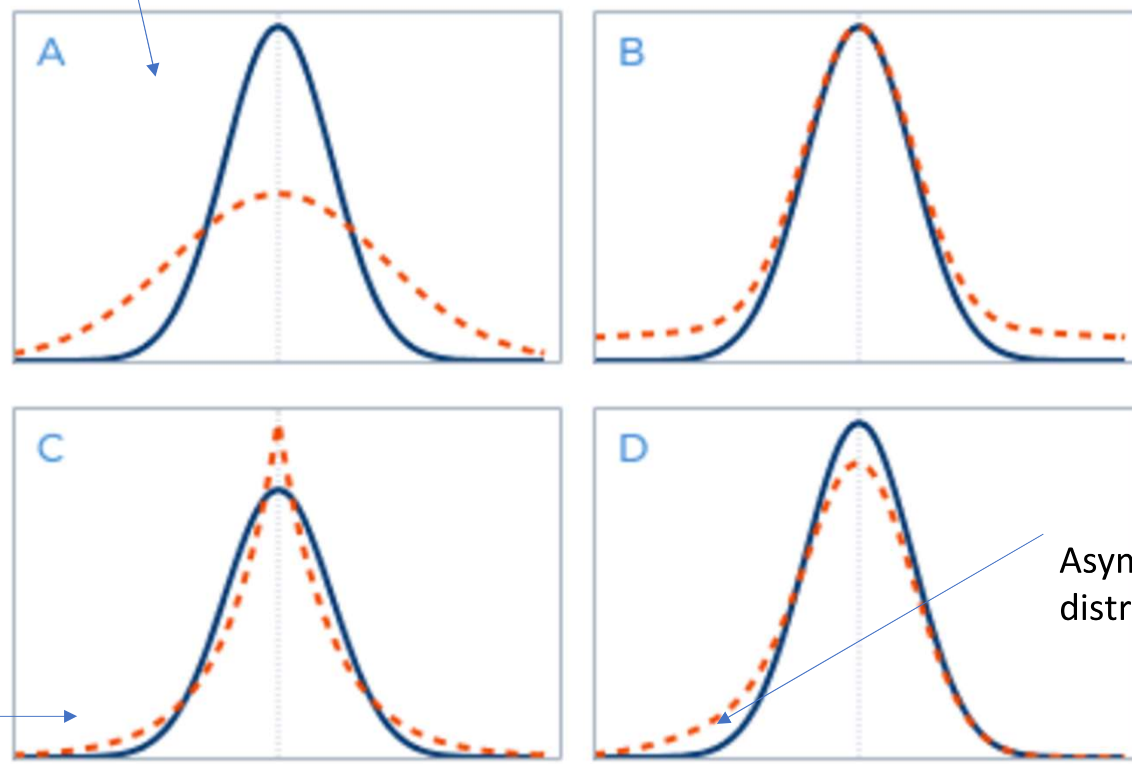
5% probability that we are in this region

Additional questions:

1. How **realistic is the assumption of normality** for real world applications?

How realistic is the assumption of normality for asset returns?

Normal distribution with higher s.d.



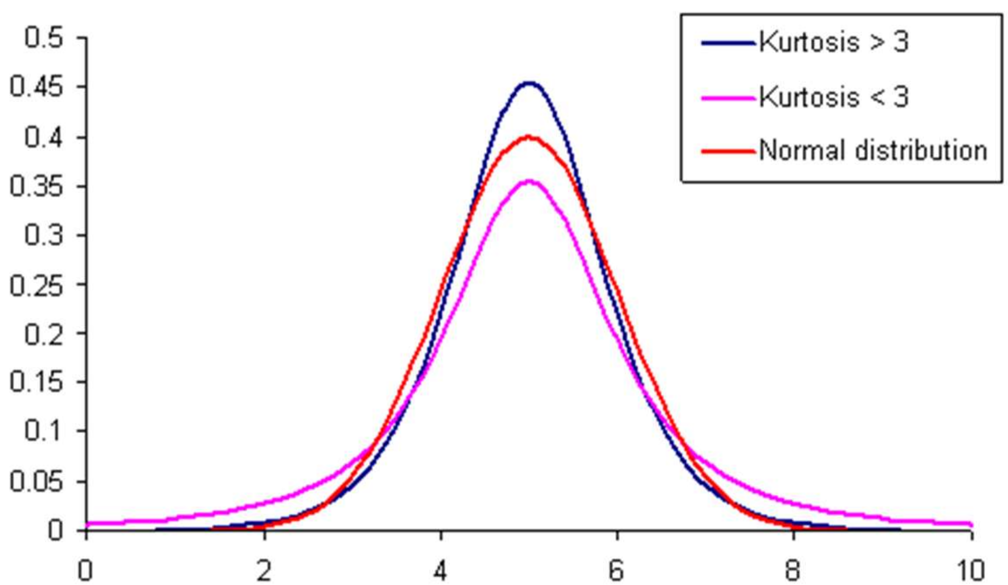
“Fatter tails”

Asymmetrical distribution

Source: https://www.transtrend.com/en/insights/riding-kurtosis/

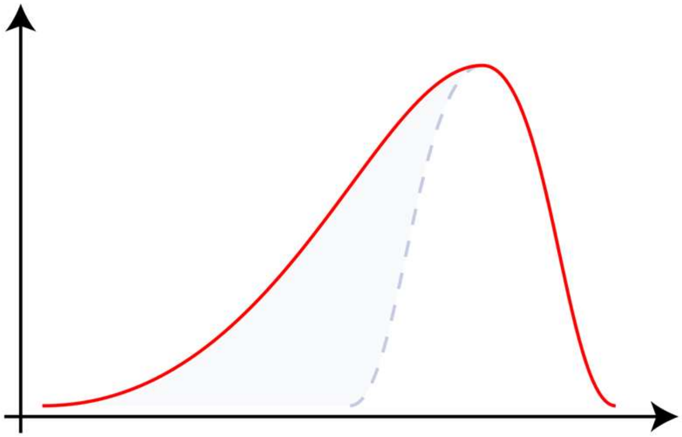
Fatter Tails: Positive Kurtosis (4^{th} moment)

Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution.

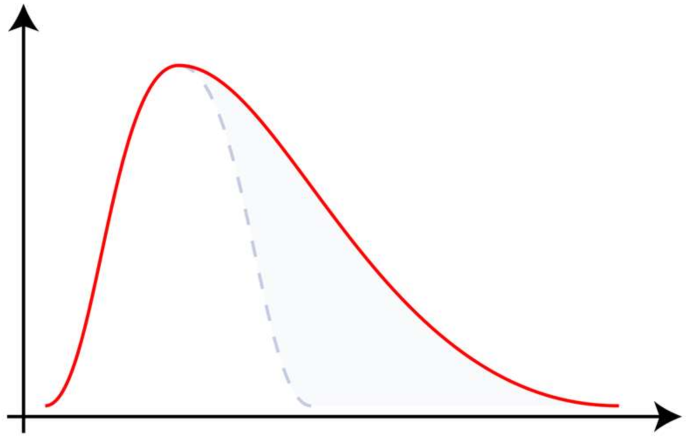


Asymmetrical distribution: Positive and negative skewness for normal distribution (3^{rd} moment)

Skewness is a measure of symmetry, or more precisely, the lack of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point



Negative Skew

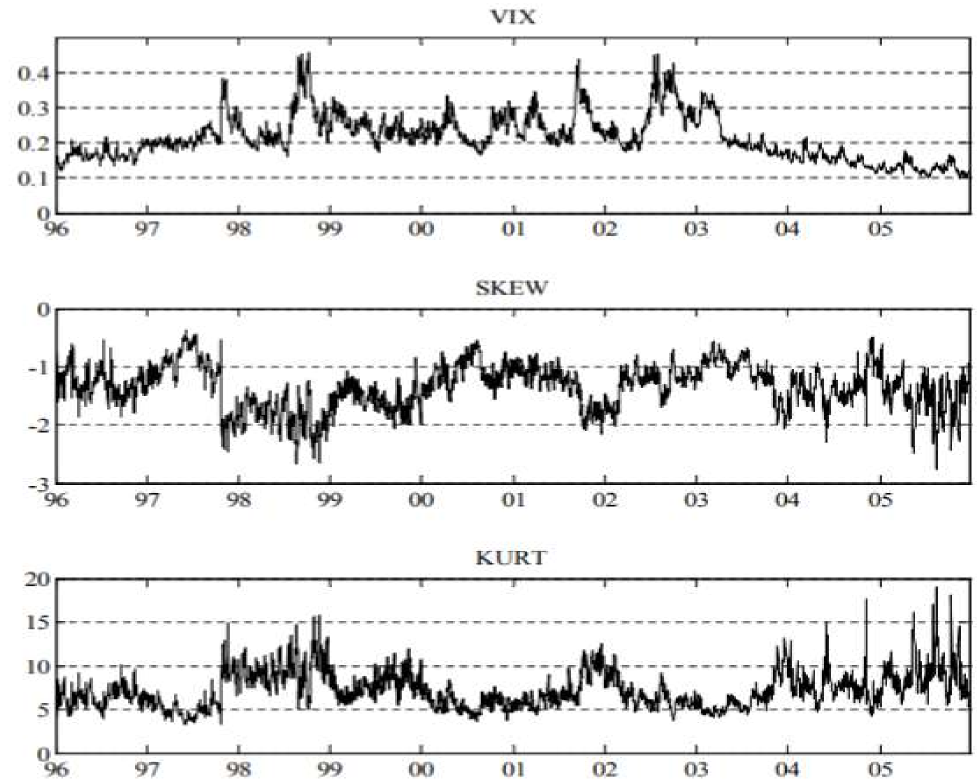


Positive Skew

Empirical asset returns appear to be negatively skewed with positive kurtosis

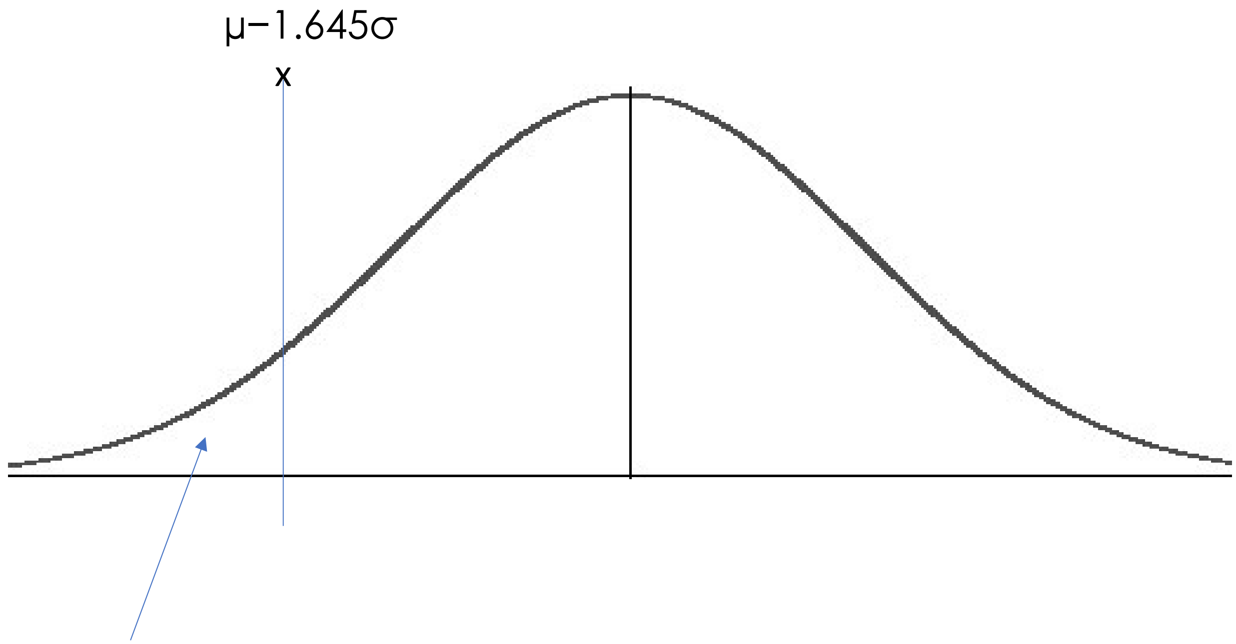
Figure 1
Daily Option Implied Moments of S&P 500 Index Returns

We plot daily option implied volatility, skewness, and kurtosis of the S&P 500 index return between 1996 and 2005. The model-free methodology developed in Carr and Madan (2001), Bakshi and Madan (2000), and Bakshi, Kapadia, and Madan (2003) is applied to extract the option implied moments using option data available in the OptionMetrics Ivy DB. See the Appendix for details of the methodology and implementation.



Market skewness risk and the cross section of stock returns, Journal of Financial Economics, Volume 107, Issue 1,2013, Pages 46-68

Motivating question: Value at Risk (revisited)



5% probability that we are in this region **according to normal distribution**

With positive kurtosis (fatter tails) and negative skew, **is *true* VaR higher or lower** than computed based on normal distribution?

Linear Combination of 2 Variables

Motivating question: Value at Risk (continued)

- **Previously**: we had a single asset, and we want to estimate the loss given a tail event ("Value at Risk")

- What if we **have 2 assets, combined into a portfolio**, and we want to estimate severity of a tail event for entire portfolio?

- Continue to assume that each individual asset return **follows the same statistical distribution**

Linear Combination of Variables: Mean and Variance

Let *a*, *b* and *c* be constant. Let *X* and *Y* be two random variables (e.g. returns of 2 assets), with means μ_X and μ_Y, respectively. Also, the corresponding variances are σ_X^2 and σ_Y^2.

Then,

$$\mathbb{E}\left(aX+bY+c\right)=a\mathbb{E}\left(X\right)+b\mathbb{E}\left(Y\right)+c \quad (1)$$

$$\mathbb{V}\left(X\right)=\mathbb{E}\left(X^2\right)-\mu_X^2 \quad (2)$$

$$\mathbb{V}\left(aX+b\right)=a^2\,\mathbb{V}\left(X\right) \quad (3)$$

$$\mathbb{V}\left(aX+bY+c\right)=a^2\,\mathbb{V}\left(X\right)+b^2\,\mathbb{V}\left(Y\right)+2ab\,\mathbb{C}\left(X,Y\right) \quad (4)$$

Think of "**aX + bY + c**" as a portfolio with a units of X, b units of Y, and c units of "cash"

Linear Combination of Variables: Mean and Variance

Proposition 1

1. Suppose **X** and **Y** form a pair random variables with means $\mathbf{\mu_X := E(X)}$ and $\mathbf{\mu_Y := E(Y)}$, respectively. Also, suppose **a** and **b** are two constants.

   Then, $\mathbb{V}(aX+bY) = a^2\mathbb{V}(X) + b^2\mathbb{V}(Y) + 2ab\mathbb{C}(X,Y)$.

Proof

1. $\mathbb{V}(aX+bY) = \mathbb{E}\left((aX+bY)^2\right) - (a\mu_X + b\mu_Y)^2$.
2. Expanding the two quadratic term and collecting the expanded terms accordingly,

   we obtain $a^2\mathbb{E}(X^2) - a^2\mu_X^2 + b^2\mathbb{E}(Y^2) - b^2\mu_Y^2 + 2ab\mathbb{E}(XY) - 2ab\mu_X\mu_Y$.

   $$\Longrightarrow a^2\left(\mathbb{E}(X^2) - \mu_X^2\right) + b^2\left(\mathbb{E}(Y^2) - \mu_Y^2\right) + 2ab\left(\mathbb{E}(XY) - \mu_X\mu_Y\right).$$

Note: If X and Y are independent, then

$$\mathbb{V}(aX+bY) = a^2\mathbb{V}(X) + b^2\mathbb{V}(Y).$$

More on Covariance

Definition 3: Covariance

- Covariance is a generalized version of variance. It is defined as **$C(X,Y) = E((X-\mu_X)(Y-\mu_Y))$**
- Variance is a special case: **$C(X,X) = \sigma_{XX} = V(X)$.**
- Whereas variance is strictly positive, covariance can be positive, negative, and zero.
- If **X** and **Y** are independent, then it must be that **$C(X,Y) = 0$.**
- If **$C(X,Y)$** = 0, it is not necessarily true that **X** and **Y** are independent.

Correlation: Normalized Covariance

- Normalization of covariance σ_{XY} gives rise to correlation, which is written as $\rho_{XY} := \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$.
- Correlation has the nice property that $-1 \leqq \rho \leqq 1.$
- If two variables have a correlation of +1 (-1), then we say they are perfectly correlated (anti-correlated).
- If one random variable causes the other random variable, or that both variables share a common underlying driver, then they are highly correlated.
- But high correlation does not necessarily imply causation of one variable on the other.
- If two variables are uncorrelated, it does not necessarily follow that they are unrelated.
- So what does correlation tell you?

Covariance and Correlation

- Consider the return on M1 $R_{X,t}$, and on Starhub $R_{Y,t}$.
- The respective mean and variance are μ_X, σ_X^2 for M1 and μ_Y, σ_Y^2 for Starhub.

- The covariance between $R_{X,t}$ and $R_{Y,t}$ is

$$\begin{aligned}\sigma_{XY} := \mathbb{C}\left(R_{X,t}\,,\,R_{Y,t}\right) &= \mathbb{E}\Big(\left(R_{X,t}-\mu_X\right)\left(R_{Y,t}-\mu_Y\right)\Big)\\ &= \mathbb{E}\left(R_{X,t}\,R_{Y,t}\right)-\mu_X\,\mu_Y\,.\end{aligned}$$

- The correlation between $R_{X,t}$ and $R_{Y,t}$ is

$$\rho_{XY} := \frac{\sigma_{XY}}{\sigma_X\,\sigma_Y}$$

Example

- The covariance between the return on gold and the return on silver is 0.04. The volatility of return on gold is 60%, and the volatility of return on silver is 30%. What is the correlation between gold return and silver return?

  Answer: ____________________

Case study:

Portfolio formation (2 assets)

- We are constructing a portfolio with 2 assets.
- Asset A has average return 5% per year, while Asset B has average return 10% per year
- Asset A has volatility of 10% per year, while Asset B has volatility of 20% per year.
- Correlation of returns between Asset A and B is 50%
- What is the expected return and volatility of a portfolio with 50% in asset A and 50% in asset B?
- Note: In finance, "volatility" means standard deviation, not variance

Case study:

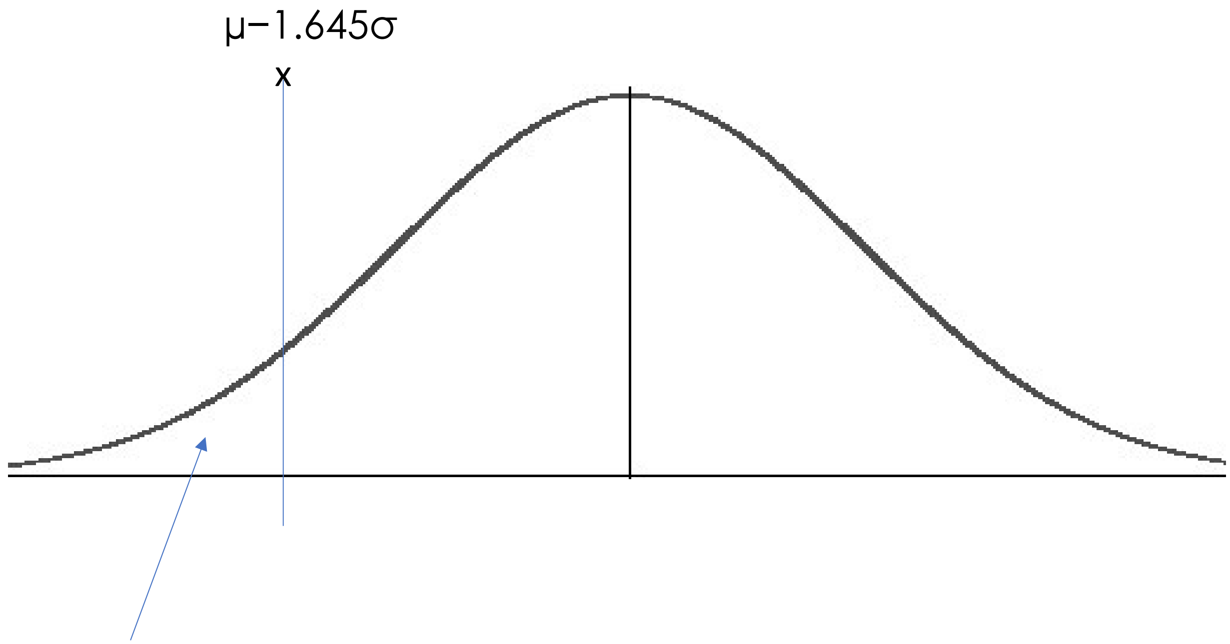
Portfolio formation (2 assets)

- Expected portfolio return:
  - 0.5Asset(A) return + 0.5Asset(B) return
  - = 0.5(5%) + 0.5(10%)
  - = 7.5%
- Expected portfolio volatility:
  - $(0.5^2$Asset(A) variance + 0.5^2Asset(B) variance + 2(0.5)(0.5)Cov(A,B)$)^{0.5}$
  - $=(0.5^2\text{Var}(A) + 0.5^2\text{Var}(B) + 2(0.5)(0.5)\text{Corr}(A,B)\text{Volatility}(A)\text{Volatility}(B))^{0.5}$
  - $=(0.5^2(0.1)^2 + 0.5^2(0.2)^2 + 2(0.5)(0.5)(0.5)(0.1)(0.2))^{0.5}$
  - 13.2%

What happens to portfolio volatility as correlation falls?

- Note that the portfolio remains 50% in asset A and 50% in asset B
- Hence, portfolio expected returns remain unchanged at 7.5% regardless of Corr(A,B)

Motivating question: Value at Risk (revisited)



5% probability that we are in this region

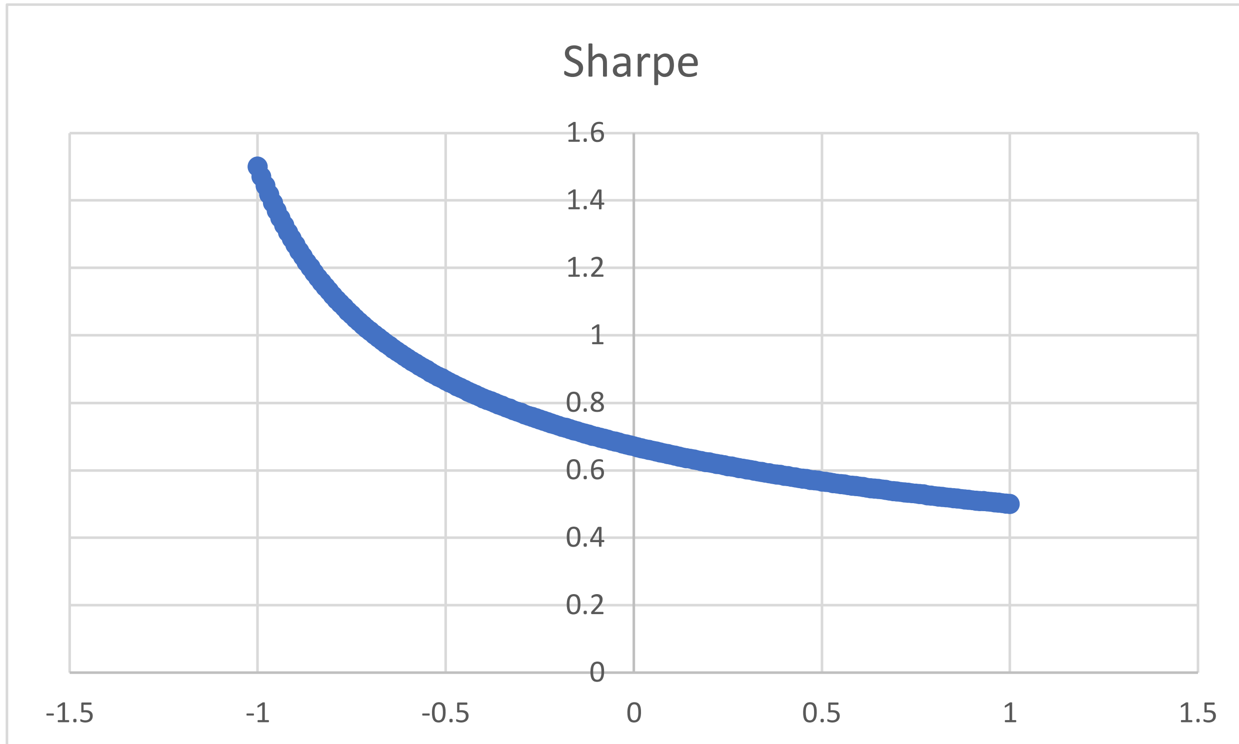
As correlation between A and B decreases **what happens to σ?** [volatility of combined portfolio return]

What happens to **u – 1.645 σ**?

Does **VaR increase or decrease?**

As a portfolio manager or risk manager, **do we prefer higher or lower correlation between the daily returns of A and B?**

Portfolio sharpe ratio is highest when correlation is the lowest



Note: Sharpe = Returns / Volatility (we ignore benchmark)

Hypothesis Testing: Linear Combination of multiple variables

Variance of arbitrary linear combination of multiple random variables

Not examinable

$$
\begin{aligned}
\operatorname{var}\left(\sum_{i=1}^{n} a_i X_i\right) &= E\left[\left(\sum_{i=1}^{n} a_i X_i\right)^2\right] - \left(E\left[\sum_{i=1}^{n} a_i X_i\right]\right)^2 \\
&= E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j X_i X_j\right] - \left(E\left[\sum_{i=1}^{n} a_i X_i\right]\right)^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E[X_i X_j] - \left(\sum_{i=1}^{n} a_i E[X_i]\right)^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E[X_i X_j] - \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E[X_i]E[X_j] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \left(E[X_i X_j] - E[X_i]E[X_j]\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \operatorname{cov}(X_i, X_j) \\
&= \sum_{i=1}^{n} a_i^2 \operatorname{var}(X_i) + 2\sum_{i=1}^{n}\sum_{j:j>i}^{n} a_i a_j \operatorname{cov}(X_i, X_j)
\end{aligned}
$$

Source: Wikipedia

Bilinearity

Not examinable but this relationship will be used in the derivation of regression formulas

- **Bilinearity** is a rule for working with covariance of “linear combination of variables”

- **Let W, X, Y, Z be random variables**, and a, b, c, d be scalar constants

  Cov(aW + bX, cY + dZ)
  = ac**Cov**(W,Y) + ad**Cov**(W,Z) + bc**Cov**(X,Y) + bd**Cov**(X,Z)

Portfolio combination (3 assets)

Illustrative example: will not be detailed during class discussion

- Using the same assumptions on asset A and B as with slide 27, we add asset C with annual returns of 20% and volatility of 30%
- Correlation(A,C) is 40% and Correlation(B,C) is 30%
- We equal weigh all assets in the portfolio (1/3 weight to each asset)
- What is the resulting portfolio return and volatility?

Portfolio combination (3 assets)

Illustrative example: will not be detailed during class discussion

- Returns: (1/3)5% + (1/3)10% + (1/3)20% = 11.67%

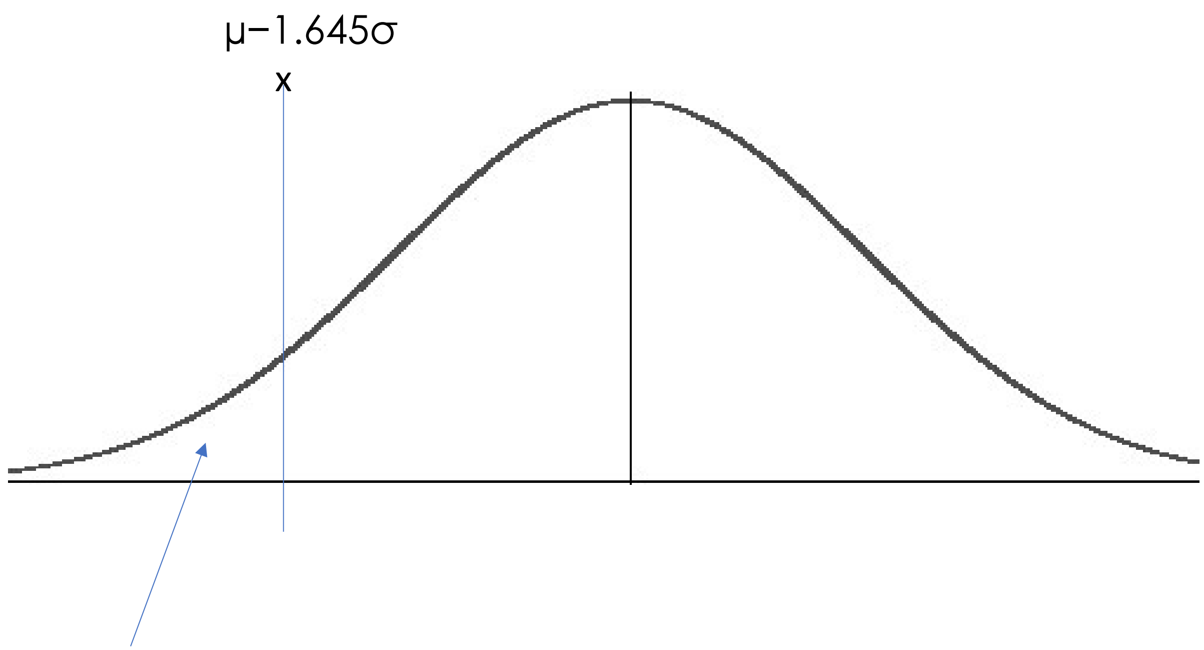
- Volatility:
  1. $=(0.33^2\text{Asset(A) variance} + 0.33^2\text{Asset(B) variance} + 0.33^2\text{Asset(C) variance} + 2(0.33)(0.33)\text{Cov(A,B)} + 2(0.33)(0.33)\text{Cov(A,C)} + 2(0.33)(0.33)\text{Cov(B,C)})^{0.5}$

  2. $=(0.33^2\text{Asset(A) variance} + 0.33^2\text{Asset(B) variance} + 0.33^2\text{Asset(C) variance} + 2(0.33)(0.33)\text{Corr(A,B)Var(A)Var(B)} + 2(0.33)(0.33)\text{Corr(A,C)Var(A)Var(C)} + 2(0.33)(0.33)\text{Corr(B,C)Var(B)Var(C)})^{0.5}$

  3. $=(0.33^2 0.1^2 + 0.33^2 0.2^2 + 0.33^2 0.3^2 + 2(0.33)(0.33)(0.5)0.1^2 0.2^2 + 2(0.33)(0.33)(0.4)0.1^2 0.3^2 + 2(0.33)(0.33)(0.3)0.2^2 0.3^2)^{0.5}$

Hypothesis Tests: Population versus sample

Motivating question: Value at Risk (revisited)



5% probability that we are in this region

In practice, we will need to translate (u-1.645σ) into a **real number**

This enables us to get an **actual risk exposure value**

How **do we know what u and σ are?**

u and σ are **unobserved population parameters**

What is a statistical population

Statistical population is the set of all possible elements that are of interest for a statistical analysis.

Example 1

The time series of split-adjusted daily stock prices of Dell Inc. since IPO on June 22, 1988 till taken private on October 29, 2013.

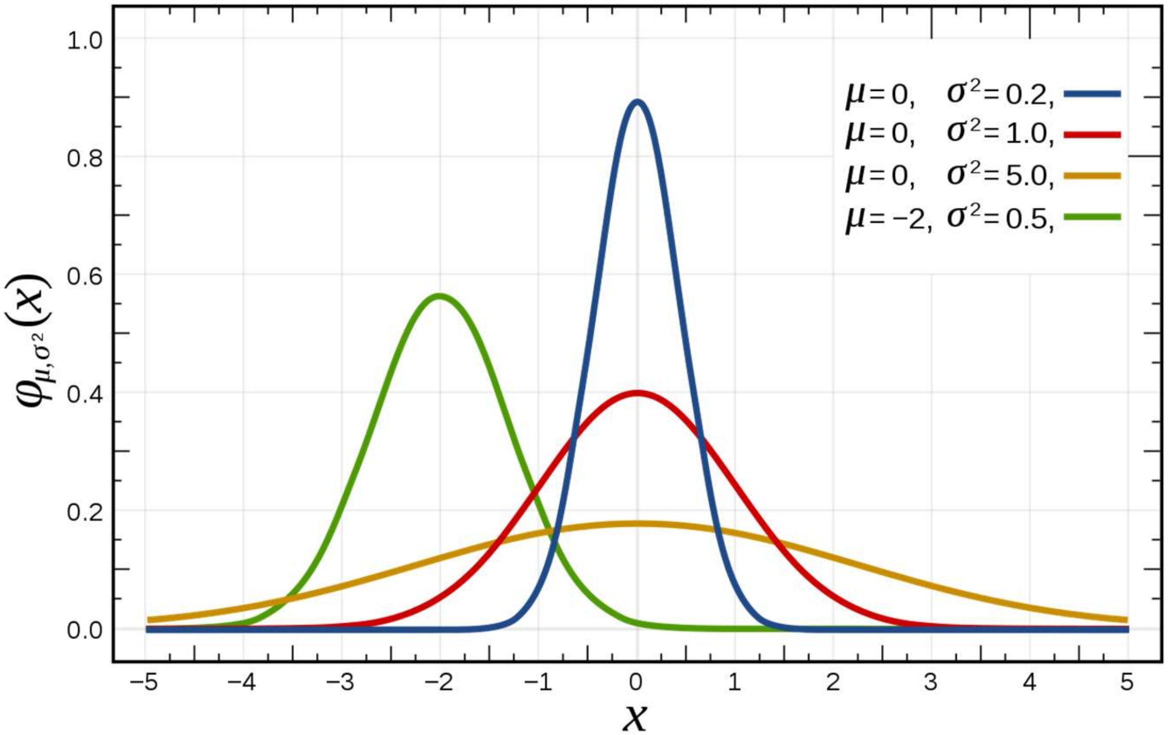
Example 2

The cross section of daily returns of all component stocks of Nikkei 225 index on October 12, 2019.

Population versus Sample

- A statistic sample (versus a population) is a finite subset of the population which is usually **directly observed**
- Random variable: **X**
- Mean of a statistical **population**: **E(X) = μ**
- Variance of a statistical **population**: **V(X) := E ((X − μ)² =: σ²**
- An example of sample average **estimate**: $\widehat{\mu} := \frac{1}{n}\sum_{i=1}^{n} x_i$
- An example of sample variance **estimate**: $\widehat{\sigma^2} := \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \widehat{\mu})^2$

Mean and variance for normal distribution



The normal distribution is characterized by 2 population parameters, u and σ

If you know these (generally unobserved) numbers, they can determine the distribution's shape

Differences between population and sample statistics

- Population parameters (e.g. mean, variance, etc) are usually unobserved.
- Although population parameters are unobserved, what we can observe are datasets
- If we assume that the dataset follows a certain parametric distribution (say “normal”), then we can try to estimate the parameters of that distribution from the data

Properties of estimators

- Generally, an estimator is function of data

- For example, sample mean is often used as estimate of population mean, and is computed as follows:

$$\frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right)$$

- Estimators are themselves also random variables, and will have mean and variances, etc

Desired properties of estimators

- **Bias**: Bias is the difference between the expected value of the estimator and the true (unobserved) parameter value being estimated. A zero bias estimator is an unbiased estimator

- **Consistency**: Estimators are typically computed over a finite sample size n. Consistency refers to how, as n increases, the estimated value gets closer and closer to the true parameter value

- **Efficiency [will not be covered in detail for this course]**: For an unbiased estimator, efficiency refers to how much its precision is lower than the theoretically highest possible precision. A more efficient estimator needs fewer observations to achieve a given level of variance (precision).

Unbiasedness

- A statistic $\psi(\boldsymbol{X})$ is an unbiased estimator of **θ** if
$$\mathbb{E}\big(\psi(\boldsymbol{X})\big) = \boldsymbol{\theta}\,.$$
- If $\mathbb{E}\big(\psi(\boldsymbol{X})\big) \neq \boldsymbol{\theta}$, then the estimator is said to be biased,
- The bias is simply the difference: $\mathbb{E}\big(\psi(\boldsymbol{X})\big) - \boldsymbol{\theta}\,.$
- For convenience, we write $\widehat{\boldsymbol{\theta}} := \psi(\boldsymbol{X})$

Bias of an Estimator

Definition 4

In statistics, the bias (or bias function) of an estimator $\widehat{\theta}$ is the difference between this estimator's expected value **E**$(\widehat{\theta})$ and the true value **θ** of the parameter being estimated.

Bias : = **E** $(\widehat{\theta})$ **- θ**

Proposition 2
Sample mean is an unbiased estimator of population mean, i.e., **E** $(\widehat{X}) = \mu$.

Proof

$$\mathbb{E}\left(\widehat{X}\right) = \frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}(n\mu) = \mu$$

Consistency

- In practice, unbiased estimators workable on small samples are rare.
- A sequence of estimators $\boldsymbol{\theta_n}$**(X)** of $\boldsymbol{\theta}$ from sample **X** of size *n* is said to be a consistent estimator if $\theta_n \xrightarrow{\mathbb{P}} \theta$ as $n \longrightarrow \infty$ or $\operatorname{plim} \theta_n = \theta$.
- That is, $\boldsymbol{\theta_n}$ converges in probability to $\boldsymbol{\theta}$; for any arbitrary $\boldsymbol{\varepsilon > 0}$,

$$\lim_{n \longrightarrow \infty} \mathbb{P}\left(\left|\theta_n - \theta\right| < \varepsilon\right) = 1\,.$$

More on the Consistency of an Estimator

Definition 5

- An estimator is said to be consistent if its difference with the true value θ. i.e., error, becomes smaller and insignificant, as the sample size grows larger and larger.
- The convergence is in probability, i.e., the absolute difference between the estimate and the true value mean being greater then some arbitrarily small margin ε has zero probability, as the sample size increases to **∞.**

$$\lim_{n \to \infty} \mathbb{P}\left(|\widehat{\theta} - \theta| > \varepsilon\right) = 0.$$

- Implication: the more data you collect, a consistent estimator will be close to the real population parameter you're trying to measure.

Examples

- Consider another estimator of mean $\check{X} = \frac{1}{n-1}\sum_{i=1}^{n} X_i + \frac{X_n}{n}$

(a) Is this estimator unbiased?

(b) Is this estimator consistent?

- Consider the estimator $\mathbf{X_7}$ of $\boldsymbol{\mu}$. [i.e. just the 7th item in the list]

(a) Is this estimator unbiased?

(b) Is this estimator consistent?

Estimator for Sample Mean

- Given a set of observations{**R1,R2,... ,Rn**}, how should one estimate **μ** and **σ²**? We treat the observed value as a particular outcome or realization of the random variable $\mathbf{R_t}$ at each ***t***:

$$\overline{R} = \frac{1}{n}\sum_{t=1}^{n} R_t$$

- The sample mean R itself is a random variable with mean and variance, assuming identical distribution,

$$\begin{aligned}\mathbb{E}(\overline{R}) &= \frac{1}{n}\sum_{t=1}^{n}\mathbb{E}(R_t) = \mu \\ \mathbb{V}(\overline{R}) &= \frac{1}{n^2}\sum_{t=1}^{n}\mathbb{V}(R_t) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}\end{aligned}$$

- The last equation (on variance of the **estimator**) is relevant for determining whether the estimator is efficient or not

Unbiased Sample Variance Estimator

First with $\widehat{\mu}$ being the sample mean, and **μ** the population mean

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\widehat{\mu})^2=\frac{1}{n-1}\sum_{i=1}^{n}\Big((X_i-\mu)-(\widehat{\mu}-\mu)\Big)^2$$

$$=\frac{1}{n-1}\sum_{i=1}^{n}\Big((X_i-\mu)^2-2(\widehat{\mu}-\mu)(X_i-\mu)+(\widehat{\mu}-\mu)^2\Big)$$

$$=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2-\frac{2}{n-1}(\widehat{\mu}-\mu)\sum_{i=1}^{n}(X_i-\mu)+\frac{1}{n-1}(\widehat{\mu}-\mu)^2\cdot n$$

$$=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2-\frac{2n}{n-1}(\widehat{\mu}-\mu)^2+\frac{n}{n-1}(\widehat{\mu}-\mu)^2$$

$$=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\mu)^2-\frac{n}{n-1}(\widehat{\mu}-\mu)^2.$$

Unbiased Sample Variance Estimator

Taking expectation on the sample variance estimator,

$$\mathbb{E}\left(\widehat{\sigma}^2\right) = \mathbb{E}\left(\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\mu\right)^2 - \frac{n}{n-1}\left(\widehat{\mu}-\mu\right)^2\right)$$

The first term is

$$\frac{1}{n-1}\mathbb{E}\left(\frac{n}{n}\sum_{i=1}^{n}\left(X_i-\mu\right)^2\right) = \frac{n}{n-1}\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\mu\right)^2\right) = \frac{n}{n-1}\sigma^2$$

For the second term, we note that $\mathbb{E}\left(\left(\widehat{\mu}-\mu\right)^2\right) = \frac{\sigma^2}{n}$. It follows that

$$\mathbb{E}\left(\widehat{\sigma}^2\right) = \frac{n}{n-1}\sigma^2 - \frac{n}{n-1}\cdot\frac{1}{n}\sigma^2 = \frac{n-1}{n-1}\sigma^2 = \sigma^2.$$

Caution:

- Following two things are **not the same**

  - **Variance of sample mean estimator**
    - For a point estimate of (unobserved) population mean, how uncertain is that estimate?

  - **Sample estimate of population variance**
    - This gives a point estimate of (unobserved) population variance

- What is **each used for?**
  - To test if our estimate of the sample mean is "reliable", we should technically **compare it with variance of sample mean estimate**
  - But variance of sample mean estimator requires population variance. **This is unobserved**
  - So we **substitute population variance with sample estimate of population variance**
  - Equations will be **detailed next few slides**

Hypothesis Testing

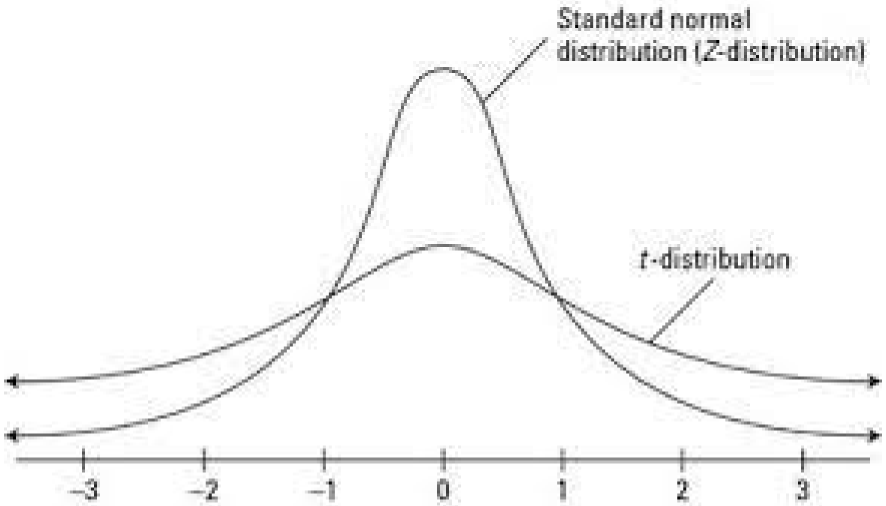
Estimation of Sample Variance and *t* Statistic

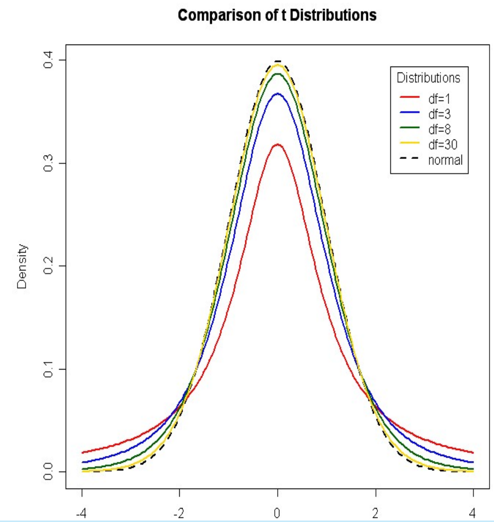
- Unbiased sample variance is $$s^2 = \frac{1}{n-1}\sum_{t=1}^{n}\left(R_t - \overline{R}\right)^2$$

- The ratio of the sample variance with population variance $$V := (n-1)\frac{s^2}{\sigma^2} \overset{d}{\sim} \chi^2_{n-1}$$

- Application $$\overline{R} \overset{d}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) \quad \text{or} \quad \frac{\overline{R}-\mu}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\sqrt{n}\left(\overline{R}-\mu\right)}{\sigma} \overset{d}{\sim} N(0,1)$$

$$\Longrightarrow \quad \frac{\frac{\sqrt{n}\left(\overline{R}-\mu\right)}{\sigma}}{\sqrt{\frac{V}{n-1}}} = \frac{\frac{\sqrt{n}\left(\overline{R}-\mu\right)}{\sigma}}{\sqrt{\frac{s^2}{\sigma^2}}} = \frac{\sqrt{n}\left(\overline{R}-\mu\right)}{s} \overset{d}{\sim} t_{n-1}$$





Normal distribution versus t-distribution

More on t-statistic

- After getting a point estimate value (say for the population mean), the next step is usually to determine if **this point estimate value is different from “random noise”**

- We can do this from the **numerical values of the estimator’s mean and variance**. Example:

  - **Point estimate of population mean is 1.0, but variance of estimate for population mean is 100**

  - **Point estimate of population mean is 1.0, while variance of estimate for population mean is 0.01**

  - Based on your intuition, **which estimate of the population’s mean is more ‘reliable’?**

Empirically, t-distribution approaches normal distribution as data size increases. This is also consistent with Central Limit Theorem which predicts the same thing: distribution of estimator approaches normal

- Suppose the random variables $\mathbf{X_t}$ for $\mathbf{t}$ = 1,2,…,n are i.i.d.
- Law of Large Number (LLN)

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} X_t \xrightarrow{\mathbb{P}} \mu = \mathbb{E}(X_t)$$

Stating Central Limit Theorem (CLT) without proof:

- For sufficiently large sample size n, given $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$,

$$Y := \frac{\frac{1}{n}\sum_{t=1}^{n} X_t - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\frac{1}{n}\sum_{t=1}^{n} X_t = \mu + \frac{\sigma}{\sqrt{n}} Y \overset{d}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Recall that we can transform this normal variable to the Standard Normal Distribution

- For convenience, define

$$z := \frac{r_t - \mu}{\sigma}, \qquad f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

- The probability density function f(z) is the well-known bell-shaped curve with mean 0 and variance 1

Summary:

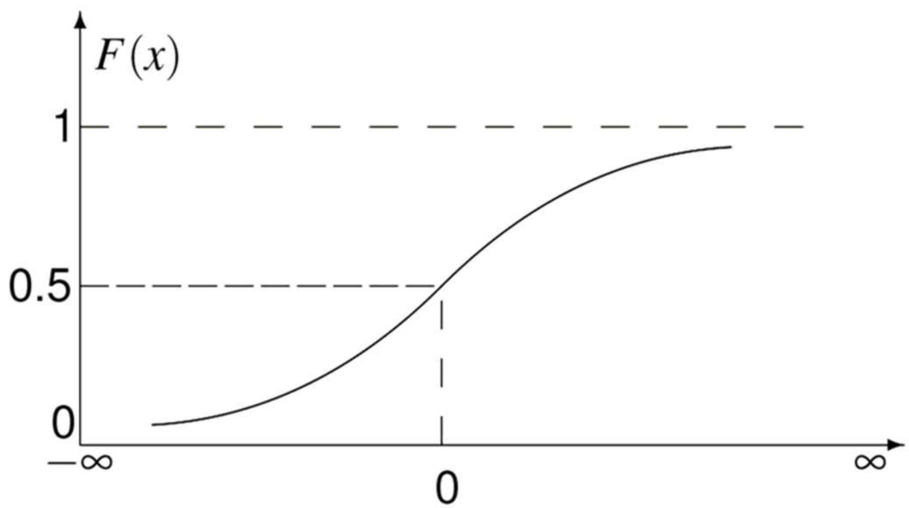
1. **For 'small' datasets, we should test our hypothesis using t-distribution and t-statistic**
2. **For 'large' datasets, we apply CLT and use the normal distribution to simplify**
3. **Next few slides assume that we are applying CLT and hence, normal distribution**

If we use CLT, Normal Cumulative Distribution Function F(x) can be used to define our ‘cut-offs’ for hypotheses testing

- What is the probability that z <−1.645?

$$F(-1.645) := \mathbb{P}(z < -1.645) = \int_{-\infty}^{-1.645} f(z)\, dz = 0.05$$

- Thus there is 5% probability that $r_t < \mu - 1.645\sigma$



Confidence Interval

Suppose data are randomly sampled from $X \overset{d}{\sim} N(\mu, \sigma^2)$ such that for an a > 0

$$\mathbb{P}(-a \leq t_{n-1} \leq +a) = 95\%.$$

Given the formula for the t statistic, (10),

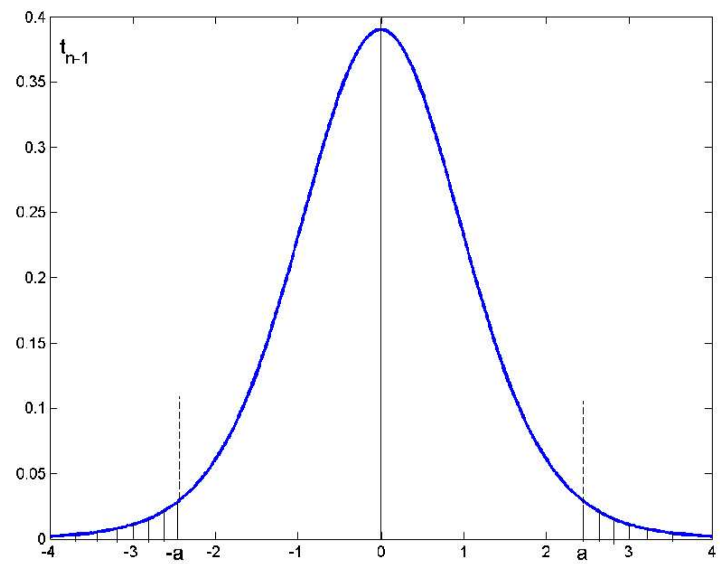
$$\mathbb{P}\left(-a \leq \frac{\sqrt{n}(\overline{X} - \mu)}{s} \leq a\right) = 0.95.$$

Thus the probability of μ falling within the confidence interval is 95%:

$$\mathbb{P}\left(\overline{X} - a\frac{s}{\sqrt{n}} \leq \mu \leq \overline{X} + a\frac{s}{\sqrt{n}}\right) = 95\%.$$

Illustration of Critical Regions

- The critical regions correspond to the shaded areas.
- The sum of the shaded area is the probability of rejecting H0 when it is true. This probability is known as the significance level.



Two-Tail versus One-Tail

If the hypotheses are, say, H0 :μ = 7% and HA :μ != 7%, then the decision rule is based on two-tail test. The critical region comprises of the left and right tails of the t_{n-1} pdf.

When the theory rules out, say, μ > 7%, the hypotheses become H0 :μ = 7% and HA :μ < 7%, then the decision rule is based on one-tail test.

The critical region is only the left side, for when μ <7%, then t_{n-1} will become larger. Thus at the one-tail 5% significance level, the critical region is{t_{n-1},95% > 1.671}for n = 61, where {t_{n-1},95% > 1.671} is the 95-th percentile of the t distribution.

P Value

P value is the observed significance level, which is the probability of getting a value of the t statistic that is extreme or more extreme than the observed value of t statistic.

Example H_0 = 7% against H_1 != 7%, n = 25, R = 10%, s = 5%, t =$\sqrt{25}(10-7)/5 = 3$ The P value is P =P | t(24) | > 3.

Type I and Type II Errors

If H_0 is true but is rejected, Type I error is committed.

< If H_0 is false but is “accepted,” Type II error is committed.

Result of the Test	Reality: H_0 is true	Reality: H_0 is false
Reject H_0	Type I error	Correct inference
Do not reject H_0	Correct inference	Type II error

Hypothesis and Test Statistic

- Suppose the population mean **μ** is, say**, 7%**. The null hypothesis is $\mathbf{H_0}$ = **7%.** The alternative hypothesis is $\mathbf{H_A} \neq$ **7%**
- A statistical test of the hypothesis is a decision rule that either rejects or does not reject the null $\mathbf{H_0}$
- Defined as $\{t_{n-1} < -a \text{ or } t_{n-1} > +a\}$**,** $a > 0$, the critical region is the set of values that leads to the rejection of $\mathbf{H_0}$.
- The statistical rule on $\mathbf{H_0}: \mu = 7\%$**,** $\mathbf{H_A}: \mu \neq 7\%$**,** is that if the **t-** distributed test statistic falls within the critical region, then $\mathbf{H_0}$ is rejected. Otherwise $\mathbf{H_0}$ cannot be rejected.

Inference Recap

The probability of committing a Type I error when H0 is true is called the

____

The probability of the population t-statistic exceeding the t-statistic obtained from the test sample is known as the p value.

If the p value < test significance level, reject H_0; otherwise H_0 cannot be rejected.

In practice, the probability of Type I error is fixed and the significance level set at e.g. 10%, 5%, or 1%.

Why is the probability of committing a Type 1 error typically set at a low number?