Quantitative Analysis

FRM一级培训讲义-基础班

讲师：周　琪
金程教育资深培训师
地点：■上海 □北京 □深圳

Topic Weightings in FRM Part Ⅰ

Session NO.	Content	Weightings
Study Session 1	Foundations of Risk Management	20
Study Session 2	Quantitative Analysis	20
Study Session 3	Financial Markets and Products	30
Study Session 4	Valuation and Risk Models	30

Framework

1. Probability
2. Basic Statistics
3. Distributions
4. Hypothesis Tests and Confidence Intervals
5. Linear Regression with One Regressor & Multiple Regressors
6. Modeling cycles
7. Modeling and Forecasting Trend
8. Modeling and Forecasting Seasonality
9. Simulation Methods
10. Estimating Volatilities and Correlations
11. Correlations and Copulas

Probability

Framework

1. Basis Concept
   - Random experiment
   - Outcome & Event
   - Population & Sample
   - Random variable
2. Probability & Probability algorithm (概率运算法则)
   - Concept of Probability
   - Multiplication rule、Addition rule
   - Total Probability Formula、Bayes' Formula
3. Probability Function & Cumulative Distribution Function
   - Discrete random variables、Continuous random variables

Terms

- **Random experiment (随机试验)**
  - An observation or measurement process with multiple but uncertain outcomes.
- **Outcome (结果)**
  - The result of a single trial. For example, if we roll two dices, an outcome might be 3 and 4; a different outcome might be 5 and 2.
- **Event (事件)**
  - The result that reflects none, one, or more outcomes in the sample space. Events can be simple or compound. An event is a subset of the sample space. If we roll two dices, an example of an event might be rolling 7 in total.
    - ✓ Mutually exclusive events (互斥事件): Events that cannot both happen at the same time.
    - ✓ Exhaustive events (完备事件): Those include all possible outcomes.

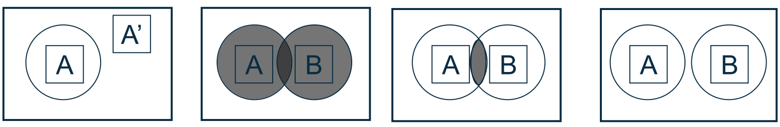
Probability

- **Probability of an event**

$$P(A) = \frac{\text{number of outcomes favorable to A}}{\text{total number of outcomes}}$$

- **Two defining properties of probability**
  - $0 \le P(E) \le 1$
  - If E_1, E_2, ……, E_n is mutually exclusive and exhaustive, then:

    $P(E_1) + P(E_2) + …… + P(E_n) = 1$
  - Venn Diagrams



Probability

- **Joint Probability**
  - The probability that the random variables (in this case, both random variables) take on certain values simultaneously, P(AB).
- **Unconditional Probability (边际概率, a.k.a marginal probability)**
  - The expected value of the variable without any restrictions (or lacking any prior information), P(A).
- **Conditional Probability (条件概率)**
  - An expected value for the variable conditional on prior information or some restriction (e.g., the value of a correlated variable). The conditional expectation of B, conditional on A , is given by P(B|A).

Probability and Probability Distributions

- **Unconditional probability: P(A), P(B)**
- **Conditional probability: P(A|B)**

$$P(A|B) = \frac{P(AB)}{P(B)};\ P(B) > 0$$

$$P(B|A) = \frac{P(AB)}{P(A)};\ P(A) > 0$$

Probability and Probability Distributions

- **Joint probability: P(AB)**
  - Multiplication rule
    - ✓ $P(AB) = P(A|B) \times P(B) = P(B|A) \times P(A)$
  - If A and B are mutually exclusive events, then:
    - ✓ $P(AB) = P(A|B) = P(B|A) = 0$
- **Probability that at least one of two events will occur:**
  - Addition rule
    - ✓ $P(A \text{ or } B) = P(A) + P(B) - P(AB)$
  - If A and B are mutually exclusive events, then:
    - ✓ $P(A \text{ or } B) = P(A) + P(B)$

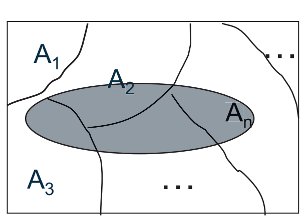
Probability and Probability Distributions

- **The occurrence of A has no influence of on the occurrence of B.**
  - $P(A|B) = P(A)$ or $P(B|A) = P(B)$
  - $P(AB) = P(A) \times P(B)$
  - $P(A \text{ or } B) = P(A) + P(B) - P(AB)$
- **Independence and Mutually Exclusive are quite different.**
  - If exclusive, must not independence;
  - Cause exclusive means if A occur, B can not occur, A influents B.
    - ✓ $P(AB) = P(A) \times P(B)$

Probability and Probability Distributions

- **Total Probability Formula**
  - If an event A must result in one of the mutually exclusive events $A_1, A_2, A_3, \ldots\ldots, An$, then

$$\boxed{P(A) = P(A_1)P(A|A_1) + P(A_2)P(A|A_2) + \ldots + P(A_n)P(A|A_n)}$$

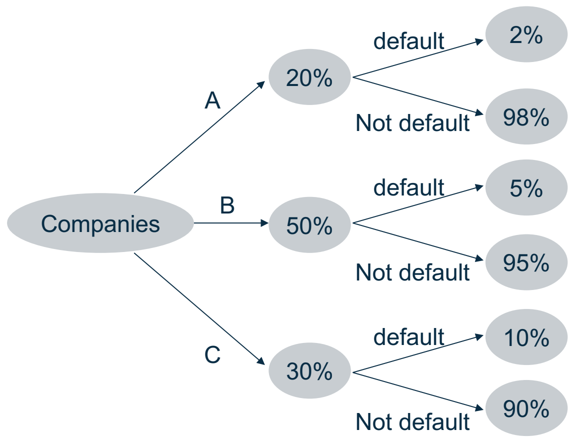


(1) $A_i A_j = \Phi \ (i \neq j)$

(2) $\bigcup_{i=1}^{n} A_i = \Omega$

Probability and Probability Distributions

- **Example**



X: Company's choice – {A,B,C}
Y: Whether company will default – {0,1}

Conditional Probability:
P(Y=1|X=A)=?

Unconditional Probability:
P(Y=1)=?

Probability and Probability Distributions

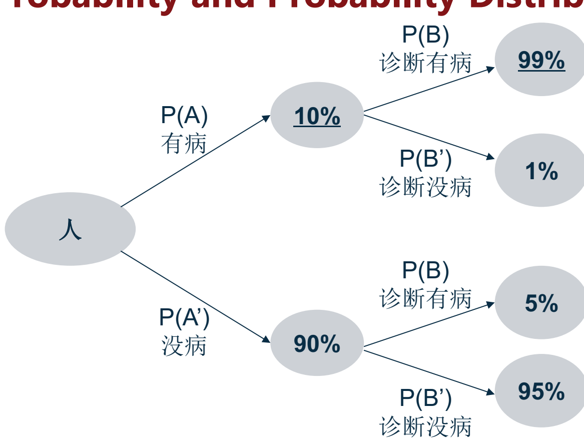
- **Bayes' Theorem**

$$\boxed{P(A|B)=\frac{P(B|A)}{P(B)}\times P(A)}$$ ← Prior Probability

- **Example**
  - 一个人有病的概率是10%，没病的概率是90%。在有病的情况下机器诊断出有病的概率是99%，诊断出没病的概率是1%；在没病的情况下机器诊断出有病的概率是5%，诊断出没病的概率是95%。若机器诊断出有病的情况下人真的有病的概率是多少？

	机器说有病	机器说没病
如果人真有病	0.99	0.01
如果人真没病	0.05	0.95

Probability and Probability Distributions



$$\boxed{P(A|B)=\frac{99\%\times10\%}{99\%\times10\%+5\%\times90\%}=68.75\%}$$

Probability and Probability Distributions

- **Probability Distribution**
  - Describe the probabilities of all the possible outcomes for a random variable.
- **Discrete and continuous random variables**
  - Discrete random variables: the number of possible outcomes can be counted, and for each possible outcome, there is a measurable and positive probability.
  - Continuous variables: the number of possible outcomes is infinite, even if lower and upper bounds exist.
    - ✓ $P(x) = 0$ even though x can occur.
    - ✓ $P(x_1 < X < x_2)$

Probability and Probability Distributions

- **Probability function: $p(x) = P(X = x)$**
  - For discrete random variables
  - $0 \le p(x) \le 1$
  - $\Sigma p(x) = 1$
- **Probability density function (p.d.f): f(x)**
  - For continuous random variable commonly
- **Cumulative probability function (c.p.f): F(x)**
  - Discrete
    - ✓ $\boxed{F(x) = P(X \le x)}$
  - Continuous
    - ✓ $\boxed{F(x) = \int_{-\infty}^{x} f(u)du}$

Probability and Probability Distributions

- **Random Variables and Their Probability Distributions**
  - Probability Distribution of a Discrete Random Variable
    - ✓ Probability Mass Function (PMF) or Probability Function (PF)

$$\boxed{f(X = x_i) = P(X = x_i),\ i = 1,2,3\ldots}$$

- ✓ Properties of the PMF

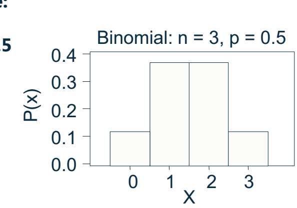
$$\boxed{1.\ f(X = x_i) = 0,\ x \ne x_i}$$

$$\boxed{2.\ 0 \le f(x_i) \le 1}$$

$$\boxed{3.\ \sum_x f(x_i) = 1}$$

For example:
Binomial
n = 3 p = 0.5

x	P(x)
0	0.125
1	0.375
2	0.375
3	0.125
	1.000

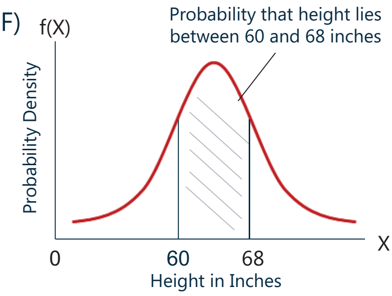


Probability and Probability Distributions

➢ Probability Distribution of a Continuous Random Variable

- Probability density function (PDF)

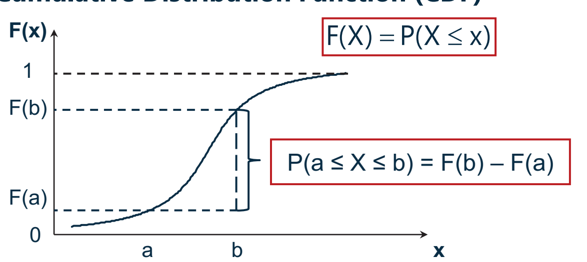
$$\boxed{P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)dx}$$

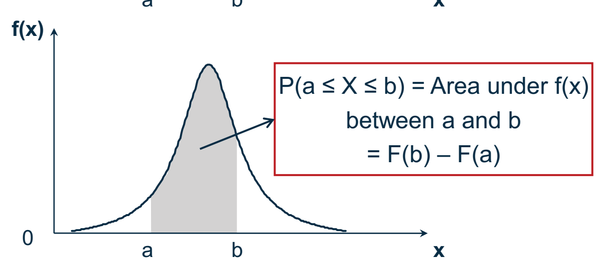


- A PDF has the following properties:
  - ✓ The total area under the curve f(x) is 1.
  - ✓ $P(x_1 < X < x_2)$ is the area under the curve between x_1 and x_2.
  - ✓ $P(x_1 \le X \le x_2) = P(x_1 < X \le x_2) = P(x_1 \le X < x_2) = P(x_1 < X < x_2)$

Probability and Probability Distributions

➢ Cumulative Distribution Function (CDF)





Probability and Probability Distributions

➢ Properties of CDF

- $F(-\infty) = 0$ and $F(+\infty) = 1$
- F(X) is a non-decreasing function such that if $x_2 > x_1$ then $F(x_2) \ge F(x_1)$.
- $P(X \ge k) = 1 - F(k)$
- $P(x_1 \le X \le x_2) = F(x_2) - F(x_1)$

Probability and Probability Distributions

- **Multivariate probability density function**
  - We take X from 1 or 2 with the same probability. We take Y from [1,X] with the same probability.

Y	X		
	1	2	Total
1	0.50	0.25	0.75
2	0.00	0.25	0.25
Total	0.50	0.50	1.00

- **Definition: f(X,Y) = P(X=x and Y=y)**
- **Properties of the bivariate or joint probability mass function (PMF)**
  - f(X,Y) ≥ 0 for all pairs of X and Y. This is because all probabilities are nonnegative.
  - ∑∑f(X,Y) =1.

www.gfedu.net 金程教育 GOLDEN FUTURE

Probability and Probability Distributions

- **Marginal probability function**

Marginal probability distribution of X and Y			
Value of X	f(x)	Value of Y	f(y)
1	0.50	1	0.75
2	0.50	2	0.25
Sum	1.00		1.00

- **Definition of marginal probability function**

$$\boxed{f(X)=\sum_{y} f(X,Y) \text{ for all } X}$$

$$\boxed{f(Y)=\sum_{x} f(X,Y) \text{ for all } Y}$$

www.gfedu.net 金程教育 GOLDEN FUTURE

Probability and Probability Distributions

- **Statistical Independence**
  - Definition of Statistical Independence: f(X,Y) = f(X)f(Y)

Y \ X	1	2	3	f(Y)
1	1/9	1/9	1/9	3/9
2	1/9	1/9	1/9	3/9
3	1/9	1/9	1/9	3/9
f(X)	3/9	3/9	3/9	1

Example 1

- The joint probability distribution of random variables X and Y is given by f(x,y) = kxy for x = 1,2,3, y = 1,2,3 and k is a positive constant, what is the probability that X+Y will exceed 5?

  A. 1/9

  B. 1/4

  C. 1/36

  D. Cannot be determined.

- Correct Answer：B

Example 2

- Half of the mortgages in a portfolio are considered subprime. The principal balance of half of the subprime mortgages and one-quarter of the non-subprime mortgages exceeds the value of the property used as collateral. If you randomly select a mortgage from the portfolio for review and its principle balance exceeds the value of the collateral, what is the probability that it is a subprime mortgage?

  A. 1/4

  B. 1/3

  C. 1/2

  D. 2/3

- Correct Answer：D

Basic Statistics

Framework

Central Moment

- **Moment**
  - The k-th moment of X is defined as: $m_K = E(X^K)$
  - If k = 1, then m_1 = E[X], it is the mean.
- **Central moment**
  - The k-th central moment of X is defined as: $\mu_K = E\left[(X-\mu)^K\right]$
  - Central moments are measured relative to the mean.
  - If k = 1, the first central moment is equal to 0.
  - If k = 2, the second central moment is the variance.
  - If k = 3, then the third central moment divided by the cube of the standard deviation is the skewness.
  - If k = 4, then the fourth central moment divided by the square of the variance is the kurtosis.

Expected Value

- **Expected Value**
  - A measure of central tendency – the first moment.

$$\boxed{E(X)=\sum P(x_i)x_i = P(x_1)x_1 + P(x_2)x_2 + \ldots + P(x_n)x_n} \qquad \boxed{E(X)=\int xf(x)dx}$$

- **Properties of Expected Value**
  - If b is a constant, E(b) = b.
  - If a is a constant, E(aX) = aE(X).
  - If a and b are constants, then E(aX + b) = aE(X) + E(b) = aE(X) + b.
  - $E(X^2) \neq [E(X)]^2$
  - E(X + Y) = E(X) + E(Y)
  - In general, E(XY) ≠ E(X)E(Y); If X and Y are independent random variables, then E(XY) = E(X)E(Y).

Variance

Variance

- A measure of dispersion – the second moment.

$$\sigma^2 = E(X-\mu)^2$$

- Above formula is the definition of variance. To compute the variance, we use the following formula:

$$\sigma^2 = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$$

- Measures how noisy or unpredictable that random variable is.
- The positive square root of σ_x^2, σ_x , is known as the standard deviation, also called volatility.

Variance

Properties of Variance

- If c is constant, then: $\sigma^2(c) = 0$.
- If a is constant, then: $\sigma^2(aX) = a^2\sigma^2(X)$.
- If b is a constant, then: $\sigma^2(X + b) = \sigma^2(X)$.
- If a and b are constant, then: $\sigma^2(aX + b) = a^2\sigma^2(X)$.
- If X and Y are independent random variables and a and b are constants, then $\sigma^2(aX + bY) = a^2\sigma^2(X) + b^2\sigma^2(Y)$.
- The relationship between expectation and variance:

$$\sigma^2(X) = E(X^2) - [E(X)]^2$$

Sample Mean

Sample Mean

- The sample mean of a random variable, X, is defined as:

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n}$$

- The sample mean is known as an estimator of E(X), which we can now call the population mean.
- An estimate of the population is simply the numerical value taken by an estimator.

Sample Variance

➢ Sample Variance

- The sample variance, denoted by S_x^2 which is an estimator of σ_x^2, which we can now call the population variance. The sample variance is defined as:

$$\boxed{S_x^2 = \sum_{i=1}^{n} \frac{\left(X_i - \overline{X}\right)^2}{n-1}}$$

- The expression (n – 1) is known as the degrees of freedom.
- If the sample size is reasonably large, we can divide by n instead of (n – 1).
- S_X (the positive square root of S_x^2), is called the sample standard deviation.

Sample Mean and Variance

	Population	Sample
Mean	$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$	$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
Variance	$\sigma^2 = \frac{\sum_{i=1}^{N}\left(X_i - \mu\right)^2}{N}$	$s^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1}$
Standard Deviation	σ	s

Chebyshev's Inequality

➢ Chebyshev's Inequality

- For any set of observations (samples or population), the proportion of the values that lie within k standard deviations of the mean is at least $1 - 1/k^2$, $k > 1$.

$$\boxed{P\left(\left|X - \mu\right| \le k\sigma\right) \ge 1 - \frac{1}{k^2},\ k > 1}$$

- This relationship applies regardless of the shape of the distribution.

Lin within	k	Standard deviations of the mean	
	2		$\ge 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{3}{4} = 75\%$
	3		$\ge 1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{8}{9} = 89\%$
	4		$\ge 1 - \frac{1}{4^2} = 1 - \frac{1}{16} = \frac{15}{16} = 94\%$

Covariance

- **Covariance**

$$\boxed{Cov(X,Y)=E\left[\left(X-E(X)\right)\left(Y-E(Y)\right)\right]=E(XY)-E(X)E(Y)}$$

  - Covariance measures how one random variable moves with another random variable.
  - Covariance ranges from negative infinity to positive infinity.

- **Properties of Covariance**
  - If X and Y are independent random variables, their covariance is zero.

$$Cov(X,\ X)=E\left[\left(X-E(X)\right)\left(X-E(X)\right)\right]=\sigma^2(X)$$

$$Cov(a+bX,\ cY)=Cov(a,cY)+Cov(bX,cY)=b\times c\times Cov(X,Y)$$

  - The relationship between covariance and variance:

$$\sigma^2(X \pm Y) = \sigma^2(X) + \sigma^2(Y) \pm 2Cov(X,Y)$$

Correlation Coefficient

- **Correlation coefficient**

$$\boxed{\rho_{XY}=\frac{Cov(X,Y)}{\sigma_x\sigma_y}}$$

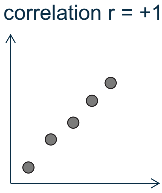
- **Properties of Correlation coefficient**
  - Correlation has no units, ranges from -1 to +1.
  - Correlation measures the linear relationship between two random variables.
  - If two variables are independent, their covariance is zero, therefore, the correlation coefficient will be zero. The converse, however, is not true. For example, $Y = X^2$.
  - Variances of correlated variables:

$$\sigma^2(X \pm Y) = \sigma^2(X) + \sigma^2(Y) \pm 2\rho\sigma(X)\sigma(Y)$$

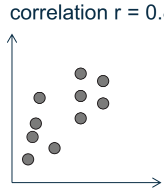
Correlation Coefficient

Correlation coefficient	Interpretation
r = +1	perfect positive correlation
0 < r < +1	positive linear correlation
r = 0	no linear correlation
-1 < r < 0	negative linear correlation
r = -1	perfect negative correlation

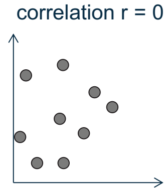
perfect positive correlation r = +1



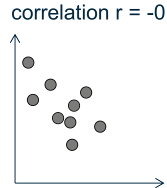
perfect positive correlation r = 0.8



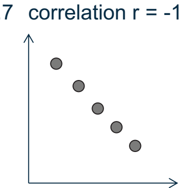
perfect positive correlation r = 0



perfect positive correlation r = -0.7



perfect positive correlation r = -1



Measures of Portfolio

$$E(R_P) = \sum_{i=1}^{N} w_i E(R_i) = w_1 E(R_1) + w_2 E(R_2) + \cdots + w_n E(R_N)$$

$$\sigma^2(R_P) = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \mathrm{Cov}(R_i, R_j)$$

$$w_i = \frac{\text{market value of investment in asset i}}{\text{market value of the portfolio}}$$

Skewness

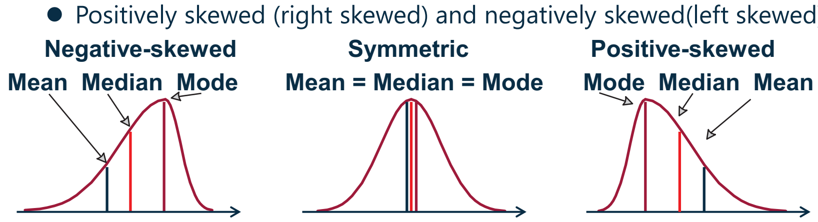
➢ **Skewness**

- A measure of asymmetry of a PDF – the third moment.

$$S = \frac{E(X-\mu_x)^3}{\sigma_x^3} = \frac{\text{third moment about mean}}{\text{cube of standard deviation}}$$

➢ **Symmetrical and nonsymmetrical distributions**

- Positively skewed (right skewed) and negatively skewed(left skewed)



- Positive skewed: Mode < median < mean, having a right fat tail
- Negative skewed: Mode > media > mean, having a left fat tail

Kurtosis

➢ **Kurtosis**

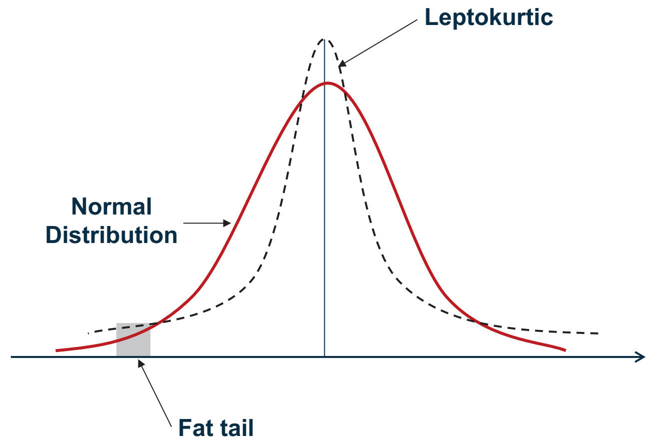
- A measure of tallness or flatness of a PDF – the fourth moment.

$$K = \frac{E(X-\mu_x)^4}{\left[E(X-\mu_x)^2\right]^2} = \frac{\text{fourth moment}}{\text{square of second moment}}$$

- For a normal distribution, the K value is 3.
- Excess kurtosis = kurtosis – 3

	leptokurtic	mesokurtic	platykurtic
Kurtosis	> 3	= 3	< 3
Excess kurtosis	> 0	= 0	< 0
Tails (assuming same variance)	fat tail	normal	thin tail

Kurtosis



A leptokurtic distribution has more frequent extremely large deviations from the mean than a normal distribution.

Coskewness and Cokurtosis

- **Coskewness and Cokurtosis**
  - The third cross central moment is referred to as coskewness.
  - The fourth cross central moment is referred to as cokurtosis.
- Risk models with time-varying volatility or time-varying correlation can display a wide range of behaviors with very few free parameters.
  - **Copulas** can also be used to describe complex interactions between variables that go beyond covariances, and have become popular in risk management in recent years.

Coskewness and Cokurtosis

- **Example**
  - Assume four series of fund returns (A、B、C、D) where the mean, standard deviation, skew, and kurtosis are all the same, but only the order of returns is different:

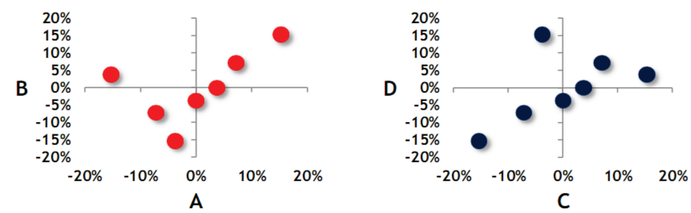
time	A	B	C	D
1	0.0%	-3.8%	-15.3%	-15.3%
2	-3.8%	-15.3%	-7.2%	-7.2%
3	-15.3%	3.8%	0.0%	-3.8%
4	-7.2%	-7.2%	-3.8%	15.3%
5	3.8%	0.0%	3.8%	0.0%
6	7.2%	7.2%	7.2%	7.2%
7	15.3%	15.3%	15.3%	3.8%

time	A + B	C + D
1	-1.9%	-15.3%
2	-9.5%	-7.2%
3	-5.8%	-1.9%
4	-7.2%	5.8%
5	1.9%	1.9%
6	7.2%	7.2%
7	15.3%	9.5%

  - The two portfolios (A + B and C + D) have the same mean and standard deviation, but the skews of the portfolios are different.

Coskewness and Cokurtosis

- Scatterplots show the difference between B versus A and D versus C:
  - ✓ A and B: their best positive returns occur during the same time period, but their worst negative returns occur in different periods. This causes the distribution of points to be skewed toward the top-right of the chart.
  - ✓ C and D: their worst negative returns occur in the same period, but their best positive returns occur in different periods. In the second chart, the points are skewed toward the bottom-left of the chart.



Coskewness and Cokurtosis

- The reason the above charts look different or <u>the reason the returns of the two portfolios are different, is because the coskewness between the portfolios is different.</u>

	A and B	C and D
S_{xxy}	0.99	-0.58
S_{xyy}	0.58	-0.99

➢ **Notices**

- The nontrivial coskewness of two variables: S_{XXY} and S_{XYY}
  - ✓ For example $\boxed{S_{XXY} = \dfrac{E\left[(X-\mu_X)^2(Y-\mu_Y)\right]}{\sigma_X^2\sigma_Y}}$
- The nontrivial cokurtosis of two variables: K_{XXXY}、 K_{XXYY} and K_{XYYY}
  - ✓ For example $\boxed{K_{XXXY} = \dfrac{E\left[(X-\mu_X)^3(Y-\mu_Y)\right]}{\sigma_X^3\sigma_Y}}$

Best Linear Unbiased Estimator (BLUE)

➢ **BLUE**

- Another property of a point estimate is <u>linearity</u>. A point estimate should be a linear estimator (i.e., it can be used as a linear function of the sample data). If the estimator is the best available (i.e., has the minimum variance), exhibits linearity, and is unbiased, it is said to be the <u>best linear unbiased estimator (BLUE)</u>.
- All of the estimators that we produced in this chapter for the mean, variance, covariance, skewness, and kurtosis <u>are either BLUE or the ratio of BLUE estimators.</u>

Example 1

- Suppose that A and B are random variables, each follows a standard normal distribution, and the covariance between A and B is 0.35. What is the variance of (3A + 2B)?
  - A. 14.47
  - B. 17.20
  - C. 9.20
  - D. 15.10
- Correct Answer:B

Example 2

- Given that x and y are random variables, and a, b, c and d are constant, which one of the following definitions is wrong?
  - A. $E(ax + by + c) = aE(x) + bE(y) + c$, if x and y are correlated.
  - B. $\sigma^2(ax + by + c) = \sigma^2(ax + by) + c$, if x and y are correlated.
  - C. $Cov(ax + by, cx + dy) = ac\sigma^2(x) + bd\sigma^2(y) + (ad + bc)Cov(x, y)$, if x and y are correlated.
  - D. $\sigma^2(x - y) = \sigma^2(x + y) = \sigma^2(x) + \sigma^2(y)$, if x and y are uncorrelated.
- Correct Answer：B

Example 3

- Which one of the following statements about the correlation coefficient is false?
  - A. It always ranges from -1 to +1.
  - B. A correlation coefficient of zero means that two random variables are independent.
  - C. It is a measure of linear relationship between two random variables.
  - D. It can be calculated by scaling the covariance between two random variables.
- Correct Answer：B

Distributions

Framework

1. Discrete Probability Distribution
   - Bernoulli Distribution
   - Binomial Distribution
   - Poisson Distribution
2. Continuous Probability Distribution
   - Continuous Uniform Distribution
   - Normal Distribution
   - Lognormal Distribution
3. Other Commonly used Probability Distributions
   - Chi-Square Distribution
   - t Distribution
   - F Distribution
4. Parametric and Nonparametric Distributions
   - Mixture Distribution

Binomial Distribution

- Bernoulli Distribution
- P(X = 1) = p   P(X = 0) = 1 – p

Binomial Distribution

- The probability of x successes in n trails

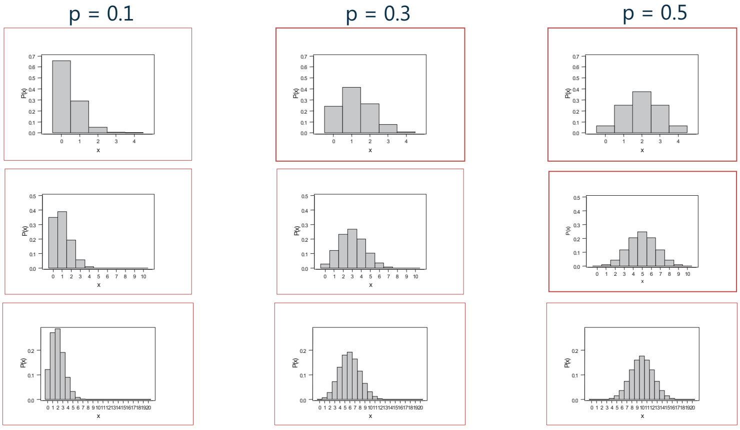
$$\boxed{p(x) = P(X = x) = C_n^x p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}}$$

Expectations and variances

	Expectation	Variance
Bernoulli random variable	p	p(1 – p)
Binomial random variable	np	np(1 – p)

Some Important Probability Distributions

➢ The Binomial Distribution – Overview



➢ Binomial distributions become more symmetric as n increases and as p = 0.5.

Poisson Distribution

➢ **Poisson Distribution**

- When there are a large number of trials but a small probability of success, Binomial calculations become impractical.
- If we substitute λ/n for p, and let n very large, the Binomial Distribution becomes the Poisson Distribution.

$$\boxed{p(k) = P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (\lambda = np)}$$

  ✓ X refers to the number of success per unit.

  ✓ λ indicates the rate of occurrence of the random events; i.e., it tells us how many events occur on average per unit of time.

- Example:

  ✓ The number of fish caught in a day; the number of potholes on a 1 km stretch of road; the number of persons appeared in a shopping mall; the number of phone calls in a day.

Poisson Distribution

➢ **Properties**

- E(X) = D(X) = λ
- The sum of independent Poisson variables is a further Poisson variable with mean equal to the sum of the individual means.
- The Poisson Distribution is the limiting case of the Binomial Distribution as n goes to infinity and p goes to zero, while np = λ remains fixed. In addition, when λ is large the Poisson Distribution is well approximated by the Normal Distribution with mean and variance of λ, through the central limit theorem.

Some Important Probability Distributions

- **Example**
  - A company receives three complaints per day on average. What is the probability of receiving more than one complaint on a particular day?
  - $\lambda = 3$
    - ✓"more than one" means that k = 2 or 3 or 4 or …
    - ✓P('more than one') = P(2) + P(3) + P(4) + …
    - ✓P('more than one') = 1 – {P(0) + P(1)}
    - ✓$P(0) = e^{-3} \times 3^0 / 0! = 0.0498$
    - ✓$P(1) = e^{-3} \times 3^1 / 1! = 0.1494$
    - ✓P(0) + P(1) = 0.1992
    - ✓P('more than one') = 1 – {P(0) + P(1)} = 1 – 0.1992 = 0.8008

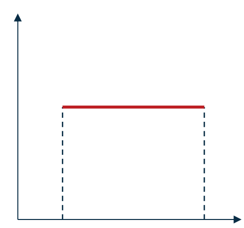
Continuous Uniform Distribution

- **Probability density function**

$$f(x) = \begin{cases} \dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

- **Cumulative distribution function**

$$F(x) = \begin{cases} 0 & \text{for } x \le a \\ \dfrac{x-a}{b-a} & \text{for } a < x < b \\ 1 & \text{for } x \ge b \end{cases}$$



Continuous Uniform Distribution

- **Properties**
  - E(X) = (a + b)/2
  - $D(X) = (b - a)^2/12$
  - For all $a \le x_1 < x_2 \le b$, we have:

$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x)dx = \frac{x_2 - x_1}{b-a}$$

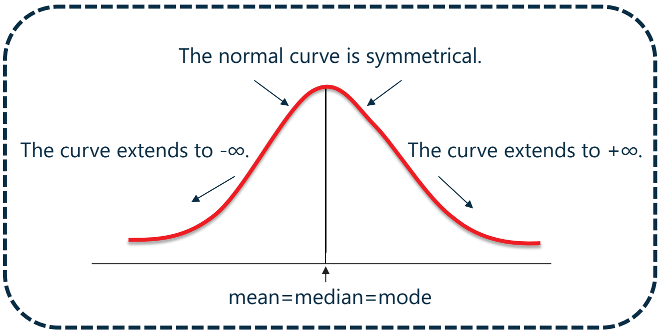
- **Example**
  - The random variable X with density function f(x) = k/3 for 2 ≤ x ≤ 8, and 0 otherwise. Calculate its mean.

Normal Distribution

Normal Distribution

- As n increases, the binomial distribution approaches Normal Distribution.

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{1}{2\sigma^2}(x-u)^2}$$



Normal Distribution

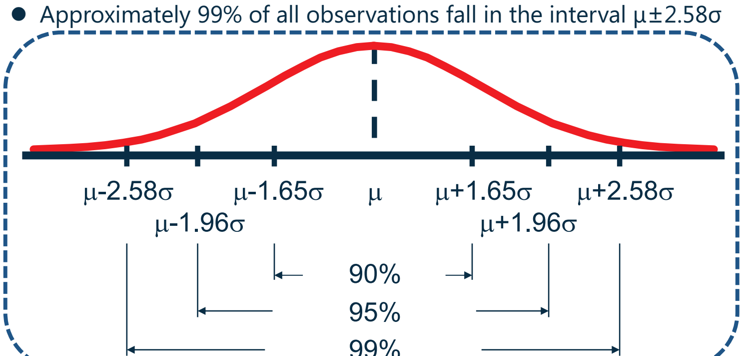
Properties

- $X \sim N(\mu, \sigma^2)$, fully described by its two parameters μ and σ^2.
- Bell-shaped, symmetrical distribution: skewness = 0; kurtosis = 3.
- A linear combination (function) of two (or more) normally distribution random variables is itself normally distributed.
- The tails get thin and go to zero but extend infinitely, asymptotic.

Normal Distribution

The confidence intervals

- Approximately 68% of all observations fall in the interval $\mu\pm\sigma$
- Approximately 90% of all observations fall in the interval $\mu\pm1.65\sigma$
- Approximately 95% of all observations fall in the interval $\mu\pm1.96\sigma$
- Approximately 99% of all observations fall in the interval $\mu\pm2.58\sigma$



The Standard Normal Distribution

- **The standard normal distribution**
  - N(0,1) or Z
  - Standardization: if X ~ N(μ, σ^2), then $\boxed{Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)}$
  - Z-table
- **How we use the standard normal distribution to compute various probabilities?**
  - Example: X ~ N(70, 9), compute the probability of X ≤ 64.12.
    - ✓ $Z = \dfrac{X-\mu}{\sigma} = \dfrac{64.12-70}{3} \approx -1.96$
    - ✓ P(Z ≤ -1.96) = 0.0250
  - Question 1: compute the probability of X ≥ 75.9.
  - Question 2: compute the probability of 64.12 ≤ X ≤ 75.9.

The Standard Normal Distribution

Reference Table: Let Z be a standard normal random variable.

z	P(Z<z)	z	P(Z<z)	z	P(Z<z)	z	P(Z<z)	z	P(Z<z)	z	P(Z<z)
-3	0.0013	-2.50	0.0062	-2.00	0.0228	-1.50	0.0668	-1.00	0.1587	-0.50	0.3085
-2.99	0.0014	-2.49	0.0064	-1.99	0.0233	-1.49	0.0681	-0.99	0.1611	-0.49	0.3121
-2.98	0.0014	-2.48	0.0066	-1.98	0.0239	-1.48	0.0694	-0.98	0.1635	-0.48	0.3156
-2.97	0.0015	-2.47	0.0068	-1.97	0.0244	-1.47	0.0708	-0.97	0.1660	-0.47	0.3192
-2.96	0.0015	-2.46	0.0069	-1.96	0.0250	-1.46	0.0721	-0.96	0.1685	-0.46	0.3228
-2.95	0.0016	-2.45	0.0071	-1.95	0.0256	-1.45	0.0735	-0.95	0.1711	-0.45	0.3264
-2.94	0.0016	-2.44	0.0073	-1.94	0.0262	-1.44	0.0749	-0.94	0.1736	-0.44	0.3300
-2.93	0.0017	-2.43	0.0075	-1.93	0.0268	-1.43	0.0764	-0.93	0.1762	-0.43	0.3336
-2.92	0.0018	-2.42	0.0078	-1.92	0.0274	-1.42	0.0778	-0.92	0.1788	-0.42	0.3372
-2.91	0.0018	-2.41	0.0080	-1.91	0.0281	-1.41	0.0793	-0.91	0.1814	-0.41	0.3409
-2.9	0.0019	-2.40	0.0082	-1.90	0.0287	-1.40	0.0808	-0.90	0.1841	-0.40	0.3446
-2.89	0.0019	-2.39	0.0084	-1.89	0.0294	-1.39	0.0823	-0.89	0.1867	-0.39	0.3483
-2.88	0.0020	-2.38	0.0087	-1.88	0.0301	-1.38	0.0838	-0.88	0.1894	-0.38	0.3520
-2.87	0.0021	-2.37	0.0089	-1.87	0.0307	-1.37	0.0853	-0.87	0.1922	-0.37	0.3557
-2.86	0.0021	-2.36	0.0091	-1.86	0.0314	-1.36	0.0869	-0.86	0.1949	-0.36	0.3594
-2.85	0.0022	-2.35	0.0094	-1.85	0.0322	-1.35	0.0885	-0.85	0.1977	-0.35	0.3632
-2.84	0.0023	-2.34	0.0096	-1.84	0.0329	-1.34	0.0901	-0.84	0.2005	-0.34	0.3669
-2.83	0.0023	-2.33	0.0099	-1.83	0.0336	-1.33	0.0918	-0.83	0.2033	-0.33	0.3707
-2.82	0.0024	-2.32	0.0102	-1.82	0.0344	-1.32	0.0934	-0.82	0.2061	-0.32	0.3745
-2.81	0.0025	-2.31	0.0104	-1.81	0.0351	-1.31	0.0951	-0.81	0.2090	-0.31	0.3783
-2.8	0.0026	-2.30	0.0107	-1.80	0.0359	-1.30	0.0968	-0.80	0.2119	-0.30	0.3821
-2.79	0.0026	-2.29	0.0110	-1.79	0.0367	-1.29	0.0985	-0.79	0.2148	-0.29	0.3859
-2.78	0.0027	-2.28	0.0113	-1.78	0.0375	-1.28	0.1003	-0.78	0.2177	-0.28	0.3897
-2.77	0.0028	-2.27	0.0116	-1.77	0.0384	-1.27	0.1020	-0.77	0.2206	-0.27	0.3936
-2.76	0.0029	-2.26	0.0119	-1.76	0.0392	-1.26	0.1038	-0.76	0.2236	-0.26	0.3974
-2.75	0.0030	-2.25	0.0122	-1.75	0.0401	-1.25	0.1056	-0.75	0.2266	-0.25	0.4013

Example 1

- Let Z be a standard normal random variable, and event X is defined to happen if either Z takes a value between -0.5 and +0.5 or Z takes any value greater then 1.5. What is the probability of event X happening if N(0.5) = 0.6915 and N(-1.5) = 0.0668, where N(.) is the cumulative distribution function of a standard normal variable?
  - A. 0.2583
  - B. 0.3753
  - C. 0.4498
  - D. 0.7583
- Correct answer：C

Example 2

- Which of the following statement about the normal distribution is not accurate?
  - A. Kurtosis equals three.
  - B. Skewness equals one.
  - C. The entire distribution can be characterized by two moments, mean and variance.
  - D. The normal density function has the following expression:

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

- Correct Answer：B

Example 3

- Which type of distribution produces the lowest probability for a variable to exceed a special extreme value which is greater than the mean, assuming the distribution all have the same mean and variance?
  - A. A leptokurtic distribution with a kurtosis of 4.
  - B. A leptokurtic distribution with a kurtosis of 8.
  - C. A normal distribution.
  - D. A platykurtic distribution.
- Correct Answer：D

Example 4

- A \$50 million prudent fund (PF) is merged with a \$200 million aggressive fund (AF). The return of PF~N(0.03, 0.07^2) and the return of AF~ N(0.07, 0.15^2). Senior manager asked you to estimate the likelihood that the returns of the combined portfolio will exceed 26%. Assuming the returns are independent, what is the probability that the return will exceed 26%?
  - A. 1.0%
  - B. 2.5%
  - C. 5.0%
  - D. 10.0%
- Correct Answer：C

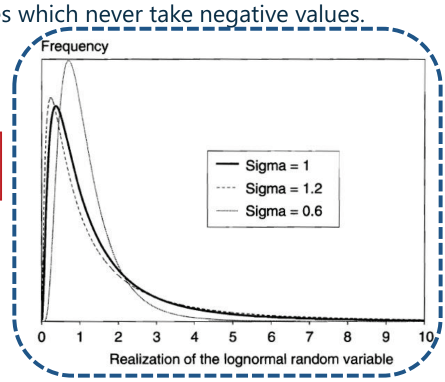
Lognormal Distribution

Lognormal Distribution

- The Black-Scholes Model assumes that the price of the underlying asset is lognormally distributed.
- If lnX is normal, then X is lognormal; if a variable is lognormal, its natural log is normal.
- It is useful for modeling asset prices which never take negative values.
- Right skewed.
- Bounded from below by zero.

$$f(x)=\frac{1}{x\sqrt{2\pi\sigma^2}}\exp\left[-\frac{1}{2\sigma^2}(\ln x-\mu)^2\right], x>0$$

$$\ln X \sim N(\mu,\sigma^2)$$
$$E(X)=\exp\left(\mu+\frac{1}{2}\sigma^2\right)$$
$$D(X)=\exp(2\mu+2\sigma^2)-\exp(2\mu+\sigma^2)$$



Chi-Square Distribution

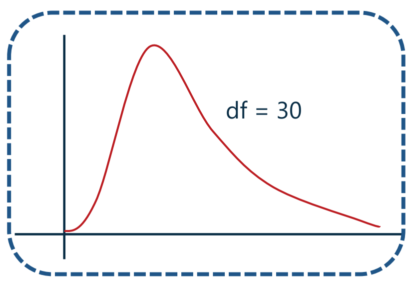
Chi-Square (χ^2) Distribution

- Chi-Square test statistic, χ^2, with n – 1 degrees of freedom, is computed as:

$$\chi^2_{n-1}=\frac{(n-1)s^2}{\sigma_0^2}$$

- Notice

$$\sum Z_i^2 = Z_1^2+Z_2^2+\cdots+Z_k^2 \sim \chi^2_{(k)}$$



Chi-Square Distribution

Properties

- The Chi-Square distribution take only positive value and ranges from 0 to infinity (after all, it is the distribution of a squared quantity).
- The Chi-Square distribution is a positive skewed distribution, the degree of the skewness depending on the d.f.
  - ✓ For comparatively few d.f. the distribution is highly skewed to the right, but as the d.f. increase, the distribution becomes increasingly symmetrical and approaches the normal distribution.
- E(X) = k, D(X) = 2k, where k is the d.f.
- If Z_1 and Z_2 are two independent Chi-Square variables with k_1 and k_2 d.f., then their sum (Z_1 + Z_2) is also a Chi-Square variable with d.f. = (k_1 + k_2).

t Distribution

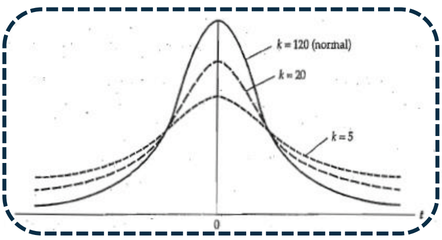
- **t Distribution (student's t distribution)**
  - Recall that, $Z=\dfrac{\overline{X}-\mu_X}{\sigma_X/\sqrt{n}}\sim N(0,1)$, both μ_X and σ_x^2 are known.
  - Suppose we only know μ_X and estimate σ_x^2 by its (sample) estimator $S_X^2=\dfrac{\sum(X_i-\overline{X})^2}{n-1}$, we obtain a new variable.

$$\boxed{t=\frac{\overline{X}-\mu_X}{S_x/\sqrt{n}}\sim t_{n-1}}$$

- **Explain the d.f. (degrees of freedom)**
  - Before we compute the S_x^2 (and hence S_x), we must first compute $\overline{X}$. But since we use the same sample to compute $\overline{X}$, we have (n-1), not n, independent observations to compute S_x^2, so to speak, we lose 1d.f.

t Distribution

- **Property**
  - Symmetric.
  - The mean of t distribution is zero, and its variance n/(n – 2).
  - The variance of t distribution is <u>larger</u> than the variance of the standard normal distribution, so t distribution is <u>flatter</u> than the normal distribution, but as n increases, the variance of t distribution approaches the variance of the standard normal distribution, namely 1.



T-distribution table

one-tailed	0.1	0.05	0.025	0.01	0.005
two-tailed	0.2	0.1	0.05	0.02	0.01
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
31	1.309	1.696	2.040	2.453	2.744
32	1.309	1.694	2.037	2.449	2.738
33	1.308	1.692	2.035	2.445	2.733
34	1.307	1.091	2.032	2.441	2.728
35	1.306	1.690	2.030	2.438	2.724
40	1.303	1.684	2.021	2.423	2.704
50	1.299	1.676	2.009	2.403	2.678
60	1.296	1.671	2.000	2.390	2.660
70	1.294	1.667	1.994	2.381	2.648
80	1.292	1.664	1.990	2.374	2.639
90	1.291	1.662	1.987	2.368	2.632
100	1.290	1.660	1.984	2.364	2.626
无穷大	1.282	1.645	1.960	2.326	2.576

F-Distribution

- **F-Distribution**
  - If U_1 and U_2 are two independent Chi-Squared distributions with k_1 and k_2 degrees of freedom, respectively, then X:

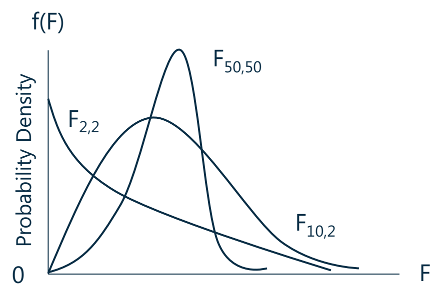
$$\boxed{X = \frac{U_1/k_1}{U_2/k_2} \sim F(k_1, k_2)}$$

  follows an F-distribution with parameters k_1 and k_2.
  - As d.f. increase, the F-Distribution approaches Normal Distribution.
  - If X is a random variable with a t distribution with k degrees of freedom, then X_2 has an F-Distribution with 1 and k degrees of freedom:

$$\boxed{X^2 \sim F(1,\ k)}$$

F-Distribution

- **F-Distribution**



The F distribution for various d.f.

- **Properties**
  - Skewed to the right and also ranges between 0 and infinity.
  - Approaches the Normal Distribution as k_1 and k_2, the d.f. become large.

Example 1

- The annual marginal probability of default of a bond is 15% in year 1 and 20% in year 2. What is the probability of the bond surviving (i.e. no default) to the end of two years?

  A. 68%

  B. 65%

  C. 80%

  D. 85%

- Correct Answer： A
  - Probability (no default) = (1 – 15%) × (1 – 20%) = 68%

Example 2

- On a multiple choice exam with four choices for each of six questions, what is the probability that a student gets less than two questions correct simply by guessing?
  - A. 0.46%
  - B. 23.73%
  - C. 35.60%
  - D. 53.39%

- Correct Answer：D
  - $p(X=0) = (3/4)^6 = 17.80\%$
  - $p(X=1) = 6\times(1/4)\times(3/4)^5 = 35.59\%$
  - The probability of getting less than two questions correct is $p(X=0) + p(X=1) = 53.39\%$.

Example 3

- A call center receives an average of two phone calls per hour. The probability that they will receive 20 calls in an 8-hour day is closest to:
  - A. 5.59%
  - B. 16.56%
  - C. 3.66%
  - D. 6.40%

- Correct Answer： A
  - To solve this question, we first need to realize that the expected number of phone calls in an 8-hour day is 16. Using the Poisson distribution, we solve for the probability that X will be 20.
  - $P(X=20)=\dfrac{16^{20}e^{-16}}{20!}=5.59\%$

Example 4

- If we say that commodity price follow a lognormal distribution, we mean that over time:
  - A. The natural logarithm of the price is normally distributed.
  - B. The change in the price is normally distributed.
  - C. The change in the natural logarithm of the price is normally distributed over time.
  - D. The reciprocal of the price is normally distributed.

- Correct Answer： C
  - A random variable has a lognormal distribution if its logarithm is itself normally distributed.

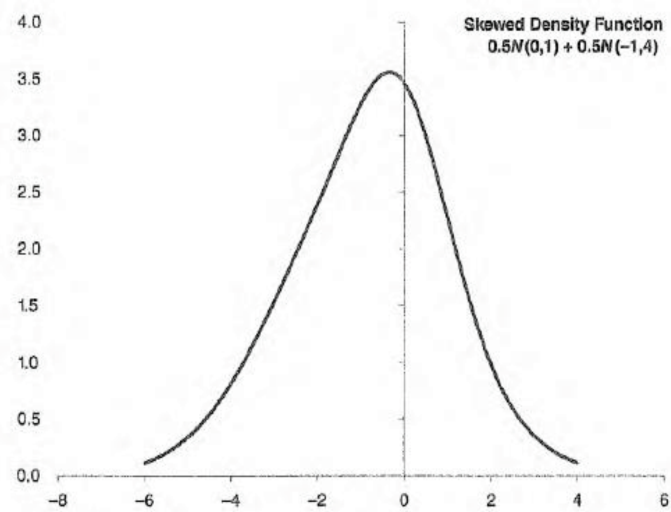
Mixture distribution

- The distribution that results from a weighted average distribution of density functions is known as a mixture distribution. More generally, we can create a distribution:

$$f(x) = \sum_{i=1}^{n} w_i f_i(x) \quad s.t. \sum_{i=1}^{n} w_i = 1$$

  - where the various $f_i(x)$'s are known as the component distributions, and the w_i's are known as the mixing proportions or weights.

Mixture distribution



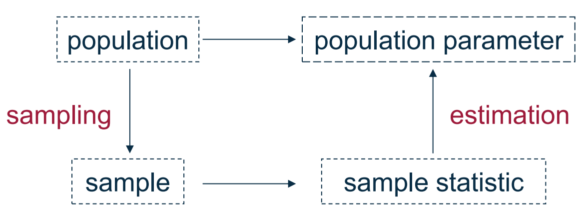
Hypothesis Tests and Confidence Intervals

Framework

1. Sampling and Estimation
   - Point Estimation、Confidence Interval Estimate
   - The Central Limit Theorem
   - Properties of point estimators
2. Hypothesis Tests
   - The basis of Hypothesis
   - The application of Hypothesis

Sample and Population

- **Sampling and Estimation**
  - Descriptive statistics: Summarize the important characteristics of large data sets.
  - Inferential statistics: Make forecasts, estimates, or judgments about a large set of data on the basis of the statistical characteristics of a smaller set (a sample).



Statistical Inference: Estimation and Hypothesis Testing

- **Sampling and estimation**
  - Simple random sampling
  - Stratified random sampling: to separate the population into smaller groups based on one or more distinguishing characteristics. Stratum and cells = M × N.
- **Sampling error**
  - sampling error of the mean = sample mean – population mean
- **The sample statistic itself is a random variable and has a probability distribution.**

The Central Limit Theorem

- **The Central Limit Theorem (CLT)**
  - If $X_1, X_2, \cdots, X_n$ a random sample from any population (i.e., probability distribution) with mean μ_X and σ_x^2 , the sample mean $\overline{X}$ tends to be normally distributed with mean μ_X and σ_x^2/n variance as the sample size increases indefinitely (technically, infinitely) (≥ 30).
  - Of course, if the X_i happen to be from the normal population, the sample mean follows the normal distribution regardless of the sample size.
  - Standard Error (SE) of mean $\overline{X}$: $\boxed{SE(\overline{X}) = \frac{s}{\sqrt{n}}}$
    - ✓ However, the population's standard deviation is almost never known. Instead, we use the standard deviation of the sample mean.

Properties of point estimators

- **Unbiasedness**
  - The mean of the estimators coincides with the true parameter value.
  - e.g. $E(\overline{X}) = \mu_X$
- **Efficiency**
  - An unbiased estimator is also efficient if the variance of its sampling distribution is smaller than all the other unbiased estimators of the parameter you are trying to estimate.
  - e.g. $\overline{X} \sim N(\mu_X, \sigma^2/n)$
- **Consistency**
  - The accuracy of the parameter estimate increases as the sample size increases (see the standard error).
  - e.g. $\overline{X} = \sum \frac{X_i}{n}$ $\quad X^* = \sum \frac{X_i}{n+1}$

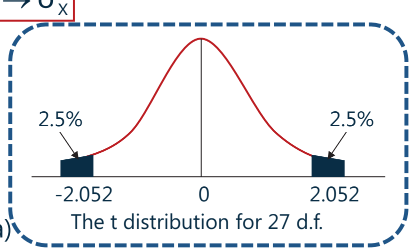
Point Estimation and Confidence Interval Estimate

- **Point Estimation**
  - Using a single numerical value to estimate the parameter of the population.

parameter

$\boxed{\overline{X} \rightarrow \mu_X \qquad S^2 \rightarrow \sigma_X^2}$

Estimator



- **Confidence Interval Estimate**
  - Level of significance (alpha)
  - Degree of confidence (1 – alpha)
  - Confidence Interval = [Point Estimate +/- (reliability factor) × standard error]

Confidence Intervals

- **Confidence Interval Estimation**
- **The population has a normal distribution with <u>a known variance</u>.**
  - Confidence interval: $\boxed{\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$

Point estimate   Reliability factor   Standard error

- **The population has a normal distribution with <u>a unknown variance</u>.**
  - Confidence interval: $\boxed{\overline{X} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}}$

When sampling form a:	small sample (n < 30)	larger sample (n ≥ 30)
Normal distribution with known variance	z-Statistic	z-Statistic
Normal distribution with unknown variance	t-Statistic	t-Statistic or z-Statistic
Nonnormal distribution with known variance	not available	z-Statistic
Nonnormal distribution with unknown variance	not available	t-Statistic or z-Statistic

Statistical Inference: Estimation and Hypothesis Testing

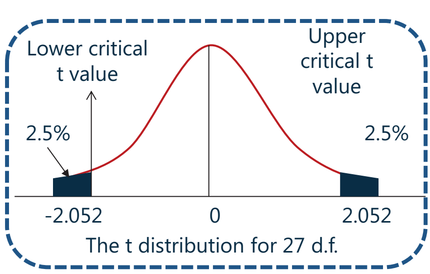
- Estimation and Hypothesis Testing: Twin Branches Of Statistical Inference

Price to earning (P/E) ratios of 28 companies on the New York stock exchange (NYSE)			
Company	P/E	Company	P/E
AA	27.96	INTC	36.02
AXP	22.90	IBM	22.94
T	8.30	JPM	12.10
BA	49.78	JNJ	22.43
CAT	24.88	MCD	22.13
C	14.55	MRK	16.48
KO	26.22	MSFT	33.75
DD	28.21	MMM	26.05
EK	34.71	MO	12.21
XOM	12.99	PG	24.49
GE	21.89	SBC	14.87
GM	9.86	UTX	14.87
HD	20.26	WMT	27.84
HON	23.36	DIS	37.10
Mean = 23.25 Variance = 90.13 Standard deviation = 9.49			

Statistical Inference: Estimation and Hypothesis Testing

- **Example: Confidence Interval Estimation**
  - Statistic

$$t = \frac{\overline{X} - \mu_x}{S_x/\sqrt{n}} \sim t(n-1)$$



  - Critical Value

$$P(-2.052 \le t \le 2.052) = 0.95$$

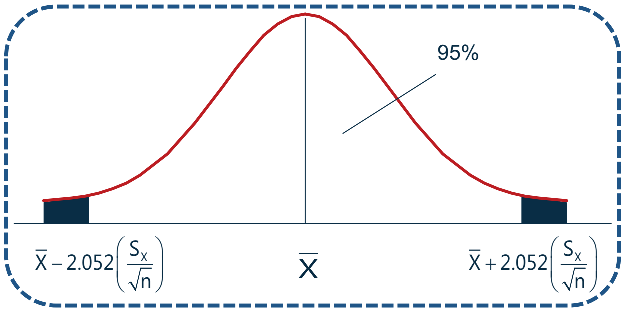
  - Obtain a Random Interval

$$P\left(\overline{X} - 2.052\frac{S_X}{\sqrt{n}} \le \mu_X \le \overline{X} + 2.052\frac{S_X}{\sqrt{n}}\right) = 0.95$$

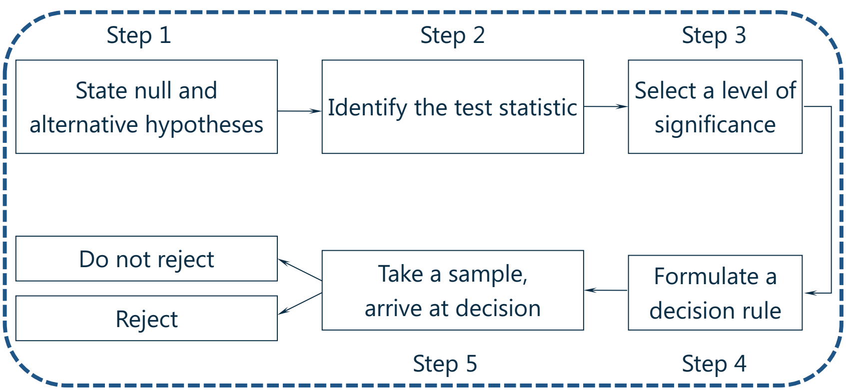
- **Important point: one should say that the probability is 0.95 that random interval contains the true μ_X.**

Statistical Inference: Estimation and Hypothesis Testing

- Using sample estimator, obtaining a confidence interval.
  - ✓ Returning to our P/E example, we have n = 28, $\bar{X} = 23.25$, and S_X = 9.49. We obtain $19.57 \le \mu_X \le 26.93$ as the 95% confidence interval for μ_X.



Hypothesis Testing



Hypothesis Testing

- **Hypothesis**
  - Statistical assessment of a statement or idea regarding a population parameter.
- **The null hypothesis (H_0) and alternative hypothesis (H_a)**
- **One-tailed test vs. Two-tailed test**
  - One-tailed test

    $H_0: \mu \ge 0 \quad H_a: \mu < 0$ | $H_0: \mu \le 0 \quad H_a: \mu > 0$

  - Two-tailed test

    $H_0: \mu = 0 \quad H_a: \mu \ne 0$
- **Critical Value**
  - The distribution of test statistic (z, t, χ^2, F)
  - Significance level (α)
  - One-tailed or two-tailed test

Statistical Inference: Estimation and Hypothesis Testing

➢ **The Test of Significance Approach to Hypothesis Testing**

- Steps:

① State the null hypothesis H_0 and alternative hypothesis H_1

$H0: \mu_X = 18.5$, $H1: \mu_X \neq 18.5$

② Select the test statistic and determine the distribution

Test statistic = (sample statistic – hypothesized value)/(standard error of the sample statistic)

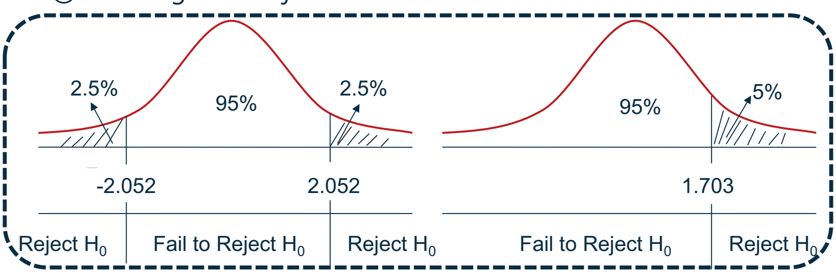
$$t = \frac{(\bar{X} - \mu_X)}{S_X / \sqrt{n}} \sim t_{(n-1)}$$

③ Choose the level of significance α(5%)

④ Obtain critical t value $t_{\frac{\alpha}{2}(n-1)} = 2.052$

Statistical Inference: Estimation and Hypothesis Testing

⑤ The region of rejection



✓Reject H_0 if |test statistic| > critical value.

✓Fail to reject H_0 if |test statistic| < critical value.

✓We can never say "accept" H_0.

✓State the conclusion: μ_X is (not) significantly different from μ_0.

Statistical Inference: Estimation and Hypothesis Testing

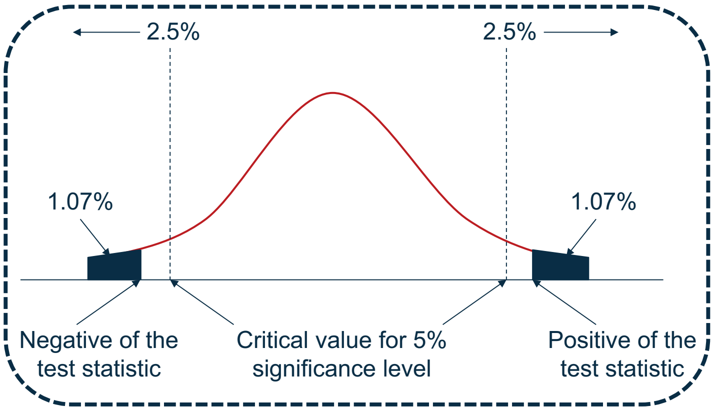
➢ **The Rule of P Value**

- P value is the smallest level of significance for which the null hypothesis can be rejected.
- Return to our P/E example,

$$t = \frac{(\bar{X} - \mu_X)}{S_X / \sqrt{n}} = \frac{(23.25 - 18.5)}{9.49 / \sqrt{28}} = 2.65 \sim t_{\frac{p}{2}(27)} \rightarrow P \approx 0.015$$

- Reject H_0 if the p-value is less than the significance level of the hypothesis test.
- Do not reject H_0 if the p-value is greater than the significance level.

P-value testing



- If P-value < alpha, we reject null hypothesis.

Test of Single Population Mean

- **H_0: μ = μ_0**
  - z-test vs. t-test

	Normal population, n < 30	n ≥ 30
Known population variance (σ^2)	z-test	z-test
Unknown population variance	t-test	t-test or z-test

- z-statistic

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

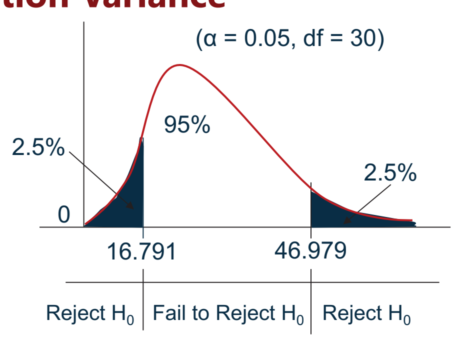
- t-statistic

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

Test of Single Population Variance

- **H_0: $\sigma^2 = \sigma_0^2$**
  - The Chi-Square test

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} \qquad df = n-1$$



- Where:
  - ✓ n = sample size
  - ✓ s^2 = sample variance
  - ✓ σ_0^2 = hypothesized value for the population variance

Chi-Square Distribution Table

	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928

Test of Variances Difference

- **H_0: $\sigma_1^2 = \sigma_2^2$**
  - The F-test

- $$\boxed{F = \frac{s_1^2}{s_2^2} \quad df_1 = n_1 - 1 \quad df_2 = n_2 - 1}$$
  - Always put the larger variance in the numerator ($s_1^2 > s_2^2$).
  - The rejection region is always the right-side tail, no matter the test is one-tailed or two-tailed.

F-Distribution Table

$\alpha = 0.05$ 附表 5(续)

m_2 \ m_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.95	1.90	1.85	1.80	1.75	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13	2.06	1.97	1.93	1.88	1.84	1.79	1.73	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.91	1.87	1.82	1.77	1.71	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.10	2.03	1.94	1.90	1.85	1.81	1.75	1.70	1.64

Summary of Hypothesis Testing

Test type	Assumptions	H_0	Test-statistic	distribution
Mean hypothesis testing	Normally distributed population, known population variance	$\mu = 0$	$z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$	N(0,1)
	Normally distributed population, unknown population variance	$\mu = 0$	$t = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}}$	t(n-1)
Variance hypothesis testing	Normally distributed population	$\sigma^2 = \sigma_0^2$	$\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$	$\chi^2(n-1)$
	Two independent normally distributed populations	$\sigma_1^2 = \sigma_2^2$	$F = s_1^2 / s_2^2$	$F(n_1 - 1, n_2 - 1)$

Type I and Type II Errors

- **Type I error**
  - Reject the null hypothesis when it's actually true.
- **Type II error**
  - Fail to reject the null hypothesis when it's actually false.
- **Significance level (α)**
  - The probability of making a Type I error:
    Significance level = P(Type I error)
- **Power of a test**
  - The probability of correctly rejecting the null hypothesis when it is false:

    Power of a test = 1 – P(Type II error)

Type I and Type II Errors

	True Condition	
Decision	H_0 is true	H_0 is false
Do not reject H_0	Correct decision	Incorrect decision **Type II error**
Reject H_0	Incorrect decision **Type I error** Significance level, α, = P(Type I error)	Correct decision Power of the test = 1 – P(Type II error)

Example 1

- Consider a stock with an initial price of \$100. Its price one year from now is given by $S = 100 \times e^r$, where the rate of return r is normally distributed with a mean of 0.1 and a standard deviation of 0.2. With 95% confidence, after rounding, S will be between:
  - A. \$67.57 and \$147.99
  - B. \$70.80 and \$149.20
  - C. \$74.68 and \$163.56
  - D. \$102.18 and \$119.53

- Correct Answer： C
  - The 95% confidence interval for r is -0.292 to 0.492:
    - ✓ $r = 0.1 - 0.2 \times 1.96 = -0.2920$ or $r = 0.1 + (0.2 \times 1.96) = 0.4920$
  - The 95% confidence interval for S is \$74.68 to \$163.56:
    - ✓ $S = 100 \times e^{-0.292} = 74.68$ or $S = 100 \times e^{0.492} = 163.56$

Example 2

- According to the Basel back-testing framework guidelines, penalties start to apply is there are five or more exceptions during the previous year. The Type I error rate of this test is 11 percent. If the true coverage is 97 percent of exceptions instead of the required 99 percent, the power of the test is 87 percent. This implies that there is a(an):
  - A. 89% probability regulators will reject the correct model.
  - B. 11% probability regulators will reject the incorrect model.
  - C. 87% probability regulators will not reject the correct model.
  - D. 13% probability regulators will not reject the incorrect model.

- Correct Answer：D

Example 3

- Which of the following statements regarding hypothesis testing is incorrect?
  - A. Type II error refers to the failure to reject the null hypothesis when it is actually false.
  - B. Hypothesis testing is used to make inferences about the parameters of a given population on the basis of statistics computed for a sample that is drawn from that population.
  - C. All else being equal, the decrease in the chance of making a Type I error comes at the cost of increasing the probability of making a Type II error.
  - D. The p-value decision rule is to reject the null hypothesis if the p-value is greater than the significance level.

- Correct Answer：D

Example 4

- What does a hypothesis test at the 5% significance level mean?
  - A. P(not reject H_0 | H_0 is true) = 0.05
  - B. P(not reject H_0 | H_0 is false) = 0.05
  - C. P(reject H_0 | H_0 is true) = 0.05
  - D. P(reject H_0 | H_0 is false) = 0.05
- Correct Answer：C

Example 5

- Sunstar is a mutual fund with a stated objective of controlling volatility, as measured by the standard deviation of monthly returns. Given the information below, you are asked to test the hypothesis that the volatility of Sunstar's return is equal to 5%. Mean monthly return: 2.5%; Monthly standard deviation: 4.9%; Number of observations: 30.

Chi-square table

df\p	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010
26	12.19815	13.84390	15.37916	17.29188	35.56317	38.88514	41.92317	45.64168
27	12.87850	14.57338	16.15140	18.11390	36.74122	40.11327	43.19451	46.96294
28	13.56471	15.30786	16.92788	18.93924	37.91592	41.33714	44.46079	48.27824
29	14.25645	16.04707	17.70837	19.76774	39.08747	42.55697	45.72229	49.58788
30	14.95346	16.79077	18.49266	20.59923	40.25602	43.77297	46.97924	50.89218

t-table

df\p	0.40	0.25	0.10	0.05	0.025	0.01
26	0.25596	0.68404	1.31497	1.70562	2.05553	2.47863
27	0.25586	0.68369	1.31370	1.70329	2.05183	2.47266
28	0.25577	0.68335	1.31253	1.70113	2.04841	2.46714
29	0.25568	0.68304	1.31143	1.69913	2.04523	2.46202
30	0.25561	0.68276	1.31042	1.69726	2.04227	2.45726

Example 5

- What is the correct test should be used and what is the correct conclusion at 5% level of significance?
  - A. Chi-square test, reject the hypothesis that volatility is 5%.
  - B. Chi-square test, do not reject the hypothesis that volatility is 5%.
  - C. t test, reject the hypothesis that volatility is 5%.
  - D. t test, do not reject the hypothesis that volatility is 5%.
- Correct Answer：B

Linear Regression-Linear Regression with One Regressor

Framework

1. Basis Knowledge
   - Dependent and Independent Variable
   - OLS
2. Confidence Interval & Hypothesis Testing
   - Analysis of Variance (ANOVA) Table
   - Measures of Fit
   - Confidence Interval
   - Hypothesis Testing
3. Predicted Values
4. Heteroskedasticity and serial correlation

Dependent and Independent Variable

- **Regression analysis**
  - Regression analysis is concerned with the study of the relationship between one variable called the dependent or explained variable and one or more other variables called independent or explanatory variables.

Dependent variable

$$Y = b_0 + b_1 X + \varepsilon$$

Independent variable

  - The objectives of regression analysis: to predict or forecast dependent variable.

Interpretation of Regression Coefficients

- **Interpretation of regression function**
  - An estimated slope coefficient of 2 would indicate that the dependent variable will change two units for every 1 unit change in the independent variable.
  - The intercept term of 2% can be interpreted to mean that the independent variable is zero, the dependent variable is 2%.

Interpretation of Regression Coefficients

- **Interpretation of regression coefficients**
  - The estimated intercept coefficient ($\hat{b}_0$) is interpreted as the value of Y when X is equal to zero.
  - The estimated slope coefficient ($\hat{b}_1$) defines the sensitivity of Y to a change in X. The estimated slope coefficient ($\hat{b}_1$) equals covariance divided by variance of X.

$$Y = \underbrace{b_0}_{\boxed{\text{Intercept coefficient}}} + \overbrace{b_1}^{\boxed{\text{Slope coefficient}}} X + \underbrace{\varepsilon}_{\boxed{\text{Error term}}}$$

Ordinary Least Squares (OLS)

- **Ordinary least squares (OLS)**
  - The OLS estimation is the process of estimating the population parameter b_i using the corresponding b_i value, which minimizes the square residual (i.e., the error terms).
  - The OLS sample coefficients are those that:

$$\boxed{\text{minimize } \sum \varepsilon_i^2 = \sum [Y_i - (\hat{b}_0 + \hat{b}_1 \times X_1)]^2}$$

$$\boxed{\hat{b}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{Cov(X,Y)}{Var(X)}} \qquad \boxed{\hat{b}_0 = \bar{Y} - \hat{b}_1 \bar{X}}$$

  - The estimated intercept coefficient ($\hat{b}_0$): the point ($\bar{X}$, $\bar{Y}$) is on the regression line.

Ordinary Least Squares Regression

➤ **Example: Sample of Returns and Corresponding Lockup periods**

	X_i	Y_i	$X_i-\bar{X}$	$Y_i-\bar{Y}$	Cov(X,Y)	Var(X)	e_i	$\sum(Y_i-\hat{Y}_i)^2$	$\sum(Y_i-\bar{Y})^2$	$\hat{Y}_i$	$\sum(\hat{Y}_i-\bar{Y})^2$
	5	10	-2.5	-6	15	6.25	-1	1	36	11	25
	6	12	-1.5	-4	6	2.25	-1	1	16	13	9
	7	19	-0.5	3	-1.5	0.25	4	16	9	15	1
	8	16	0.5	0	0	0.25	-1	1	0	17	1
	9	18	1.5	2	3	2.25	-1	1	4	19	9
	10	21	2.5	5	12.5	6.25	0	0	25	21	25
sum	**45**	**96**	**0**	**0**	**35**	**17.5**	**0**	**20**	**90**	**96**	**70**
average	**7.5**	**16**									

- b_1 = 35/17.5 = 2; b_0 = 16 - 2×7.5 = 1;
- The sample regression function is: $Y_i = 1 + 2 \times X_i + e_i$.
- According to the data, on average a hedge fund with a lockup period of 6 years will have a 2% higher return than a hedge fund with a 5-year lockup period.

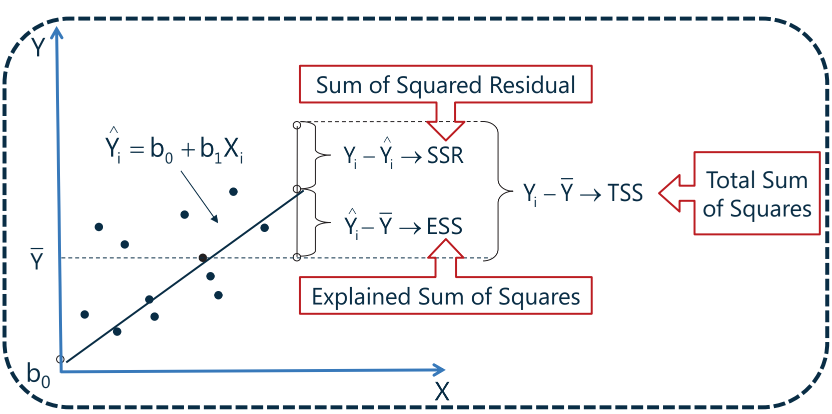
The Basics of Simple Linear Regression

➤ **The assumptions of simple linear regression**

- X and Y have a linear relationship.
- X is not random, and the condition that X is uncorrelated with the error term can substitute the condition that X is not random.
- The expected value of the error term is zero (i.e., $E(\varepsilon_i) = 0$).
- The variance of the error term is constant (i.e., the error terms are homoskedastic).
- The error term is uncorrelated across observations (i.e., $E(\varepsilon_i\varepsilon_j) = 0$ for all $i \neq j$).
- The error term follows normal distribution.

Analysis of Variance (ANOVA) Table

➤ **Components**



The Standard Error of the Regression

Total sum of squares = explained sum of squares + sum of squared residuals

$$\sum(Y_i-\bar{Y})^2 = \sum(\hat{Y}-\bar{Y})^2 + \sum(Y_i-\hat{Y})^2$$
$$TSS = ESS + SSR$$

- SER is a measure of the spread of the observations around the regression line, measured in the units of the dependent variable.

$$SER=\sqrt{\frac{\sum(Y_i-\hat{Y}_i)^2}{n-2}}$$

- The **standard error of the regression (SER)** is an estimator of the standard deviation of the regression error u_i
- The SER measures the "fit" of the regression line. The smaller the standard error, the better it fits. if the relationship is very strong, SER will be low. (relative to total variability)

Analysis of Variance (ANOVA) Table

- **Analysis of Variance (ANOVA) Table**

	df	SS	MSS
Regression	k = 1	ESS	ESS/k
Residual	n – 2	SSR	RSS/(n – 2)
Total	n – 1	TSS	-

- **Notes**
  - Total sum of squares (TSS) is also known as sum of squares total (SST).
  - Explained sum of squares (ESS) is also known as regression sum of squares (RSS).
  - Sum of squares residual (SSR) is also known as sum of squares errors (SSE).
  - Standard error of regression (SER) is also known as standard error of estimate (SEE).

Measures of Fit

- **R^2 (the Coefficient of Determination)**
  - A measure of the "goodness of fit" of the regression. It is interpreted as a percentage of variation in the dependent variable explained by the independent variable. Its limits are $0 \le R^2 \le 1$.

$$R^2=\frac{ESS}{TSS}=1-\frac{SSR}{TSS}$$

  - In a simple two-variable regression, the square root of R^2 is the correlation coefficient (r) between X_i and Y_i.

$$r^2=R^2 \rightarrow r=\pm\sqrt{R^2}$$

Measures of Fit

➢ **The Different between the R^2 and Correlation Coefficient**

- The correlation coefficient indicates the sign of the relationship between two variables, whereas the coefficient of determination does not.
- The coefficient of determination can apply to an equation with several independent variables, and it implies a explanatory power, while the correlation coefficient only applies to two variables and does not imply explanatory between the variables.

Confidence Interval

➢ The confidence interval for the regression coefficient, b_1, is calculated as:

$$\boxed{\hat{b}_1 \pm (t_c \times s_{\hat{b}_1}), \text{or } [\hat{b}_1 - (t_c \times s_{\hat{b}_1}) < b_1 < \hat{b}_1 + (t_c \times s_{\hat{b}_1})]}$$

- t_c: The critical two-tailed t value of the selected confidence level with an appropriate number of degrees of freedom is equal to the number of sample observations minus 2 (i.e., n-2).
- s_{b_1} :The standard error of the regression coefficient.

➢ $s_{\hat{b}_1}$ is the function of SER: As the SER rises, $s_{\hat{b}_1}$ also increases, and the confidence interval is widened. The SER measures the variability of the regression line data, the larger the data variable, the smaller the

Hypothesis Testing

➢ **Regression Coefficient Hypothesis Testing**

- The hypothesis that if the true slope is zero (b_1 = 0). The appropriate test structure for zero and alternative assumptions is:

$$H_0: b_1 = 0 \quad H_a: b_1 \neq 0$$

- A t-test may also be used to test the hypothesis that the true slope coefficient b_1, is equal to some hypothesized value. Letting b_1 be the point estimate for b_1, the appropriate test statistic with n – 2 degrees of freedom is:

$$\boxed{t = \frac{\hat{b}_1 - b_1}{s_{\hat{b}_1}} \sim t(n-2)}$$

- The decision rule for tests of significance for regression coefficients is:

Reject H_0 if $t > +t_{critical}$ or $t < -t_{critical}$

Example

- Calculating the confidence interval for a regression coefficient
  - The estimated slope coefficient, B_1, from WPO regression is 0.64 with a standard error equal to 0.26. Assuming that the sample had 36 observations, calculate the 95% confidence interval for B_1.
- Answer：
  - The confidence interval at 95% for b_1 is:

$$0.64 \pm (2.03)(0.26) = 0.64 \pm 0.53 = 0.11 \text{ to } 1.17$$

  - Because this confidence interval does not include 0, we can conclude that the slope coefficient is significantly different from 0.

$$t = \frac{b_1 - B_1}{s_{b_1}} = \frac{0.64 - 0}{0.26} = 2.46$$

  - Because $t > t_{critical}$ (i.e., 2.46 > 2.03), we reject the null hypothesis and conclude that the slope is different from zero.

Linear Regression-Linear Regression with Multiple Regressors

Framework

1. Basis Knowledge
   - Interpreting the Multiple Regression Results
   - OLS
2. Hypothesis test & Confidence Intervals
   - Hypothesis Testing
   - R^2 and Adjusted R^2
   - Analysis of Variance (ANOVA) Table
3. Multicollinearity & Omitted Variable Bias
4. Dummy Variables

The Basics of Multiple Regression

- Multiple regression is regression analysis with more than one independent variable.
  - The multiple linear regression model

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \cdots + b_k X_{ki} + \varepsilon_i$$

  - Predicted value of the dependent variable

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 \hat{X}_1 + \hat{b}_2 \hat{X}_2 + \cdots + \hat{b}_k \hat{X}_k$$

  - OLS Estimator

$$\text{minimize } \sum \varepsilon_i^2 = \sum \left[Y_i - (b_0 + \sum_{i=1}^{n} b_i \times X_i) \right]^2$$

Interpreting the Multiple Regression Results

- When independent variables are all equal to zero, dependent variable equals to the intercept term β_0
- The slope β_1 is the change in dependent variable associated with a unit change in independent variable, keeping other independent variables unchanged. This is why the slope coefficients in the multiple regression are sometimes referred to as local slope coefficients.

Multiple Regression Assumptions

- **The assumptions of the multiple linear regression**
  - There is a linear relationship between dependent and independent variables.
  - The independent variable is not random (Or X is not related to the error item). There is no precise linear relationship between any two or more independent variables.
  - The expected value of the error term is zero (i.e., $E(\varepsilon_i)=0$) .
  - The variance of the error term is constant (i.e., the error terms are homoskedastic).
  - The error term is uncorrelated across observations (i.e., $E(\varepsilon_i\varepsilon_j)=0$ for all $i \neq j$).
  - The error term is normally distributed.

Analysis of Variance (ANOVA) Table

- Analysis of Variance (ANOVA) Table

	df	SS	MSS
Regression	k	ESS	ESS/k
Residual	n – k – 1	SSR	RSS/(n – k – 1)
Total	n – 1	TSS	-

- Specially, in the single regression, the ANOVA table is:

	df	SS	MSS
Regression	k = 1	ESS	ESS/k
Residual	n – 2	SSR	RSS/(n – 2)
Total	n – 1	TSS	-

Interpreting Regression Results – ANOVA

Figure 2: ANOVA Table

R-squared	0.963
Adj R-squared	0.926
Standard error	1.439
Observations	6

	Degrees of Freedom	SS	MSS	F
Explained	2	88.064	44.032	21.261
Residual	3	6.213	2.071	
Total	5	94.277		

Variables	Coeff	Std. Error	t-stat	P-value	Lower 95%	Upper 95%
Intercept	-4.176	3.299	-1.266	0.270	-14.6734	6.3214
Lockup	2.375	0.337	7.047	0.009	1.3003	3.45
Experience	1.986	0.754	2.634	0.076	-0.4132	4.3853

Hypothesis test for a Partial Slope Coefficient

- Hypothesis test for a Partial Slope Coefficient
  - H_0: $b_j = 0$ ($j = 1, 2, \ldots, k$)

$$\boxed{t = \frac{\hat{b}_j}{s_{\hat{b}_j}} \sim t(n-k-1)}$$

  - Regression coefficient confidence interval

$$\boxed{\hat{b}_j \pm \left(t_c \times s_{\hat{b}_j} \right)}$$

Hypothesis Testing of Regression Coefficients

$$t = \frac{b_j - B_j}{s_{b_j}} = \frac{\text{estimated regression coefficient} - \text{hypothesized value}}{\text{coefficient standard error or } b_j}$$

- **The t-statistic has n-k-1 degrees of freedom.**
- Example: 10% significance level, 40 observations.

	Coefficient	Standard Error	t-statistic	p-value
Intercept	-10.60%	1.542%	6.87	<0.0001
PR	0.35	0.014	25	<0.0001
YCS	0.21	0.38	0.55	

- (1) (2) (3)

$$\begin{aligned} H_0 &: PR = 0 \\ H_1 &: PR \neq 0 \end{aligned} \qquad \begin{aligned} H_0 &: PR = 0.2 \\ H_1 &: PR \neq 0.2 \end{aligned} \qquad \begin{aligned} H_0 &: \text{intercept} \geq -10.0\% \\ H_1 &: \text{intercept} < -10.0\% \end{aligned}$$

Joint Hypothesis Testing

- **Joint Hypothesis Testing**
  - An F-test is used to test whether at least one slope coefficient is significantly different from zero.

  H_0: $b_1 = b_2 = b_3 = \ldots = b_k = 0$; H_a: at least one $b_j \neq 0$ (j = 1 to k)
- **F-Statistic**

$$F = \frac{ESS/k}{SSR/(n-k-1)}$$

  - The F-test here is always a one-tailed test.
  - The test assesses the effectiveness of the model as a whole in explaining the dependent variable.
  - Decision rule: reject H_0, if $F_{(\text{test-statistic})} > F_{c(\text{critical value})}$.

R^2 and Adjusted R^2

- **Standard Error of Regression (SER)**

$$SER = \sqrt{\frac{SSR}{n-k-1}}$$

- **R^2 and Adjusted R^2**
  - In multiple regression, the R^2 increases whenever a regressor (independent variable) is added, unless the estimated efficient on the added regressor is exactly zero.
  - The adjusted R^2 is a modified version of the R^2 that does not necessarily increase with a new independent variable is added. Adjusted R^2 is given by:

$$\text{Adjusted } R^2 = 1 - \frac{SSR/n-k-1}{TSS/n-1} = 1 - \frac{n-1}{n-k-1}\frac{SSR}{TSS}$$

  ✓ Adjusted $R^2 \leq R^2$; adjusted R^2 may be less than zero.

R^2 and Adjusted R^2

- **The reason of adjusted R^2**
  - To further analyze the importance of an added variable to a regression, we can compute an adjusted R^2.
    - ✓ Mathematically, if the variable with any explanatory power is added to the regression, the determined coefficient (R^2**) will increase**, <u>even if the marginal contribution of the new variable is not statistically significant</u>.
    - ✓ A relatively high R^2 may reflect the influence of a large set of independent variables, rather than the extent to which the set interprets the dependent variable.
    - ✓ This phenomenon is often referred to as overestimate regression.

Example

- Calculating R^2 and adjusted R^2
  - The analyst runs a regression of monthly return on the five independent variables within 60 months. The sum of squares is 570, the sum of the squared errors is 180. Calculate R^2 and adjust the R^2.
  - Assuming that the analyst now adds four independent variables for the regression, R^2 increases to 70.0%. Identify the models that analysts are most likely to like.
- Answer:

$$R^2 = \frac{570-180}{570} = 68.42\%$$

$$R_a^2 = 1 - \left(\frac{60-1}{60-5-1}\right)\left(1-R^2\right) = 65.5\%$$

$$R_a'^2 = 1 - \left(\frac{60-1}{60-9-1}\right)\left(1-R'^2\right) = 64.6\%$$

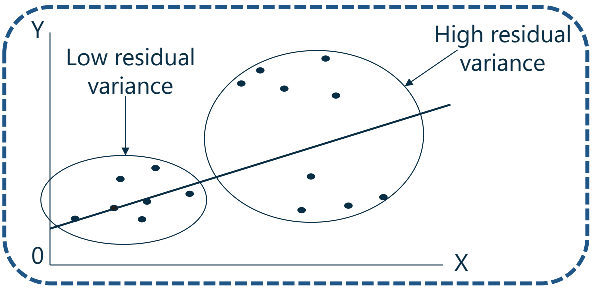
Dummy Variables

- **Dummy Variables**
  - For observe most of the independent variables can have a wide range of values. However, there are times when the argument is binary - it is "on" or "off." Independent variables fall into this category is called a **dummy variable**, usually used to quantify the impact of qualitative events.
  - <u>Dummy variables are assigned a value of "0" or "1".</u>
  - The coefficient on dummy variables indicates **the difference** in the dependent variable for the category represented by the dummy variable and the average due of the dependent variable for all classes except the dummy variable class.
    - ✓ For example, testing the slope coefficient for rise January dummy variable would indicate whether, and by how much, security returns are different in January as compared to the other months.

Homoskedasticity and Heteroskedasticity

Homoskedasticity and Heteroskedasticity

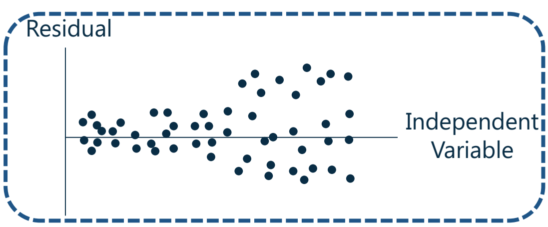
- The error term ε_i is homoskedasticity if the variance of the conditional distribution of εi given X_i is **constant** for i = 1, ..., n and in particular does not depend on X_i.
- Otherwise the error term is **heteroskedastic**.



Homoskedasticity and Heteroskedasticity

Detecting Heteroskedasticity

- As shown in the figure below, the residuals and the scatter plot of the independent variables can show the relationship between the observations.



- The residual graph in the graph illustrates the existence of the conditional heteroscedasticity. Note that with the increase of independent variables, regression residuals how incremental changes.

Homoskedasticity and Heteroskedasticity

Effect of Heteroskedaticity on Regression Analysis

- The standard errors are usually not reliable estimate.
- The coefficient estimates (the b_1) aren't influenced.
- If the standard error is too small, but the coefficient estimates itself is not affected, the t-statistic will become too large, no statistically significant null hypothesis is rejected too often.
  - ✓ The opposite will be true if the standard errors are too large.

Serial Correlation (autocorrelation)

- **Serial correlation (autocorrelation)**
  - Serial correlation (autocorrelation) refers to the situation that the error terms are correlated with one another.
  - Serial correlation is often found in time series data.
  - Positive serial correlation exists when a positive regression error in one time period increases the probability of observing regression error for the next time period.
  - Negative serial correlation occurs when a positive error in one period increases the probability of observing a negative error in the next period.

Multicollinearity

- **Multicollinearity**
  - Multicollinearity refers to the case where two or more independent variables are highly interrelated..
  - In practice, multicollinearity is often a matter of degree rather than of absence or presence.
- **Two methods to detect multicollinearity**
  - t-test indicate that none of the individual coefficients is significantly different than zero, while the F-test indicates overall significance and the R^2 is high.
  - The absolute value of the correlation between any two sample independent variable is greater than 0.7(i.e., $|r| > 0.7$).
- **Methods to correct multicollinearity**
  - Omit one or more of the correlated independent variables.

Omitted Variable Bias

- **Omitted Variable Bias**
  - Omitted variable bias is the bias in the OLS estimator that arises when one or more included regressors are correlated with an omitted variable.
  - For omitted variable bias to arise, two things must be true:
    - ✓ At least one of the included regressors must be correlated with the omitted variable.
    - ✓ Omitted variables must be the determinant of the dependent variable Y.

Example 1

➢ A factor analysis of the dividend-adjusted returns of ABC Ltd.'s stock price was undertaken to determine which economic factors contributed to its performance. The regression was performed on 460 observations. The results are as follows:

Predictor	Coefficient	Standard Error of Coefficient
Intercept	-0.0243	0.005772
All share index	0.0256	0.017655
Industrial index	0.0469	0.006398
Financial index	0.0012	0.001412

Sum of Squared Regression (SSR)	12,466.47
Sum of Squared Errors (SSE)	1,013.22
Sum of Squared Total (SST)	13,479.69

Which one of the following options correctly describes which variables are significant at the 5% level, and the R^2 statistic, respectively?

	Significant Variables at 5% level	R^2 statistic
A.	Intercept; Industrial index	0.924834
B.	Intercept; Industrial index	0.075166
C.	All share index; Industrial Index	0.924834
D.	All share index; Industrial Index	0.075166

Example 1

➢ Correct Answer: A

- The following table shows the test statistics for each of the four variables, calculated by dividing the variable coefficient by the standard error. The variable is significant if the absolute value of the t-test is greater than the critical value from the student's t-distribution for 456 degrees of freedom (which is very close to the z-statistic since the number of observations is so high), i.e. 1.96.

Predictor	T-stat	Significant
Intercept	-4.21	Yes
All share index	1.45	No
Industrial index	7.33	Yes
Financial index	0.85	No

Example 2

➢ A regression of a stock's return (in percent) on an industry index's return (in percent) provides the following results:

	Coefficient	Standard Error
Intercept	2.1	2.01
Industry index	1.9	0.31
	Degrees of Freedom	**SS**
Explained	1	92.648
Residual	3	24.512
Total	4	117.160

Which of the following statements regarding the regression is incorrect?

A. The correlation coefficient between the X and Y variables is 0.889.
B. The industry index coefficient is significant at the 99% confidence interval.
C. If the return on the industry index is 4%, the stock's expected return is 9.7%.
D. The variability of industry returns explains 21% of the variation of company returns.

Example 2

- **Answer: D**
  - $\rho^2 = R^2 = 92.648/117.160 = 79\%$, the variability of industry returns explains 79% of the variation of company.
  - t-stat (industry index) = 1.9/0.31 = 6.13, so the coefficient of industry index is significant.
  - R = 2.1% + 1.9% × R(industry index) = 2.1% + 1.9% × 4 = 9.7%

Example 3

- In an OLS regression, t-tests are used to determine the statistical significance of:
  - A. the individual parameter estimates
  - B. the regression
  - C. a set of parameters
  - D. the error term
- Correct Answer：A

Example 4

- According to the datas in the table ,and calculate the F-statistic and critical F-value is closed to:
  - A. 46.61, 2.37
  - B. 67.15, 2.53
  - C. 64.84, 2.76
  - D. 54.03, 3.24

	df	**SS**	**MSS**
Regression	3	ESS	2000
Residual	46	SSR	69.78
Total	49	TSS	-

- Correct Answer：B

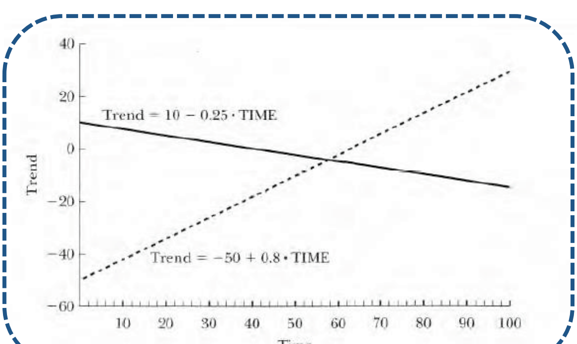
Forecasting Trends

Framework

1. Modeling and Forecasting Trend
2. Modeling and Forecasting Seasonality
3. Characterizing Cycles
4. Modeling Cycles

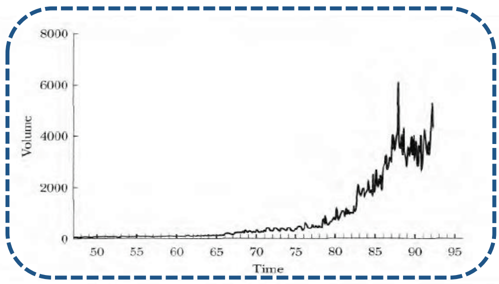
Modeling and Forecasting Trend

- **Linear trend**: The trend in which appears roughly linear, meaning that it increases or decreases like **a straight line**.



Modeling and Forecasting Trend

- **Non-linear Trend**
- Sometimes trend appears **nonlinear**, or curved, as, for example, when a variable increases **at an increasing or decreasing rate**. Ultimately, we don't require that trends be linear, only that they be smooth. Next figure shows the monthly volume of shares traded on the New York Stock Exchange (NYSE). Volume increases at **an increasing rate**; the trend is therefore **nonlinear**.



Modeling and Forecasting Trend

- **Quadratic trend models** can potentially capture nonlinearities such as those observed in the volume series. Such trends are quadratic, as opposed to linear, functions of time,

$$T_t = \beta_0 + \beta_1 TIME_t + \beta_2 TIME_t^{\,2}$$

- **Polynomial trend:**

$$T_t = \beta_0 + \beta_1 TIME_t + \beta_2 TIME_t^{\,2} + \beta_3 TIME_t^{\,3} + \ldots + \beta_n TIME_t^{\,n}$$

- **Exponential trend, or log-linear trend**, and is very common in business, finance, and economics. That's because economic variables often display roughly **constant growth rates** (for example, 3% per year). If trend is characterized by constant growth at rate β_1 , then we can write

$$T_t = \beta_0 e^{\beta_1 TIME_t}$$

Modeling and Forecasting Trend

- One of the ways of selecting best fit model is by estimating the Mean Squared Error (MSE) of the model. The model with least MSE would be chosen for fitting the data series. MSE is computed as:

$$\boxed{MSE = \frac{\sum_{t=1}^{T} e_t^2}{T}}$$

- As degree of freedom represents the choice of freely selecting the variables during the model fitting exercise therefore, to reduce the MSE bias, the degree of freedom must be deducted from the sample size to arrive at adjusted MSE, commonly referred as S^2.

$$\boxed{S^2 = \frac{\sum_{t=1}^{T} e_t^2}{T-k}}$$

Modeling and Forecasting Trend

- The Akaike information criterion (AIC) is a measure of the relative quality of statistical models for a given set of data. Given a collection of models for the data, AIC estimates the quality of each model, relative to each of the other models. Hence, AIC provides a means for model selection.
- AIC is founded on information theory: it offers a relative estimate of the information lost when a given model is used to represent the process that generates the data. In doing so, it deals with the trade-off between the goodness of fit of the model and the complexity of the model.

$$AIC = e^{\left(\frac{2k}{T}\right)} \boxed{\frac{\sum_{t=1}^{T} e_t^2}{T}} \longrightarrow \text{MSE}$$

Modeling and Forecasting Trend

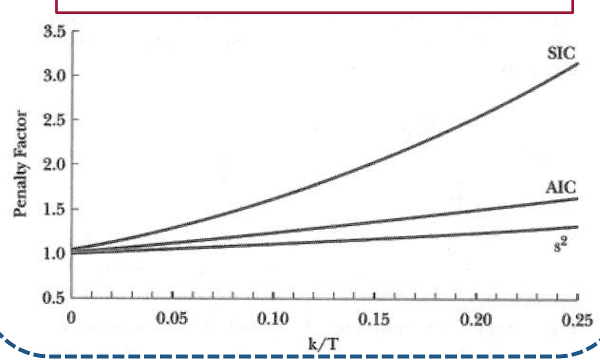
- The model with the lower value of SIC is the one to be preferred. The SIC is an increasing function of unexplained variation in the dependent variable and the number of explanatory variables. Hence, lower SIC implies either fewer explanatory variables, better fit, or both.

$$\boxed{SIC = T^{\left(\frac{k}{T}\right)} \frac{\sum_{t=1}^{T} e_t^2}{T}}$$

Modeling and Forecasting Trend

- The SIC （ the most consistency ceritria ） generally penalizes free parameters more strongly than does the Akaike information criterion, though it depends on the size of T and relative magnitude of T and k.

Degrees-of-freedom penalties, various model selection criteria.



Modeling and Forecasting Seasonality

- **The Sources Of Seasonality**
  - Any **technology** that involves the weather, such as production of agricultural commodities, is likely to be seasonal as well.
  - **Preferences** may also be linked to the calendar. People want to do more vacation travel in the summer, which tends to increase both the price and quantity of summertime gasoline sales.
  - **Social institutions** that are linked to the calendar, such as holidays.
- **A key technique for modeling seasonality is regression on seasonal dummies.**

Modeling and Forecasting Seasonality

- **The pure seasonal dummy model is**

$$y_t = \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t$$

- **Trend may be included as well, in which case the model is**

$$y_t = \beta_1 TIME_t + \sum_{i=1}^{s} \gamma_i D_{it} + \varepsilon_t$$

- **Expand the seasonality model: calendar effects.**
  - **Holiday variation** refers to the fact that some holidays' dates change over time.
  - **Trading-day variation** refers to the fact that different months contain different numbers of trading days or business days.

$$y_t = \beta_1 TIME_t + \sum_{i=1}^{s} \gamma_i D_{it} + \sum_{i=1}^{v1} \delta_i^{HD} HDV_{it} + \sum_{i=1}^{v2} \delta_i^{TD} TDV_{it} + \varepsilon_t$$

Modeling cycle

- **What is cycle?**
  - As we mentioned the cycles, we generally considered the all-encompassing notion of cyclicality into two parts:
    - ✓ Can be captured by trend or seasonal
    - ✓ Cannot be captured by trend or seasonal
- **Charactering cycle**
- **Modeling cycle**

Characterizing Cycles

- **Why we need covariance stationary?**
  - If the underlying probabilistic structure of the series were changing over time, there would be no way to predict the future accurately on the basis of the past, **because the laws governing the future would differ from those governing the past**.
  - If we want to forecast a series, at minimum we require the mean and covariance to be stable and finite over time, which we call it **covariance stationary**.
  - In this chapter all we mentioned covariance stationarity is called **second-order stationarity** or **weak stationarity** .It means that a series whose mean and variance and covariance are stable and finite all the time but the skewness and kurtosis are not necessary.

Characterizing Cycles

- **Covariance stationary**
  - In practice, many economic, business, financial, and government series are not covariance stationary.
    - ✓ For example, many series that are clearly nonstationary in levels appear covariance stationary in growth rates.

Characterizing Cycles

- The autocovariance is just the covariance between y_t and $y_{t-\tau}$, as the series is covariance stationary, so the autocorrelation will only depend on τ, and have no relationship with t , so the function
  can be written as follow:

$$\gamma = (\tau) = \operatorname{cov}(y_t, y_{t-\tau}) = E(y_t - u)(y_{t-\tau} - u)$$

- The autocorrelations are just the "simple" or "regular" correlations between
y_t and $y_{t-\tau}$ so the function can be written as follow:

$$\rho(\tau) = \frac{\operatorname{cov}(y_t, y_{t-\tau})}{\sqrt{\operatorname{var}(y_t)}\sqrt{\operatorname{var}(y_{t-\tau})}} = \frac{\gamma(\tau)}{\sqrt{\gamma(0)}\sqrt{\gamma(0)}} = \frac{\gamma(\tau)}{\gamma(0)}$$

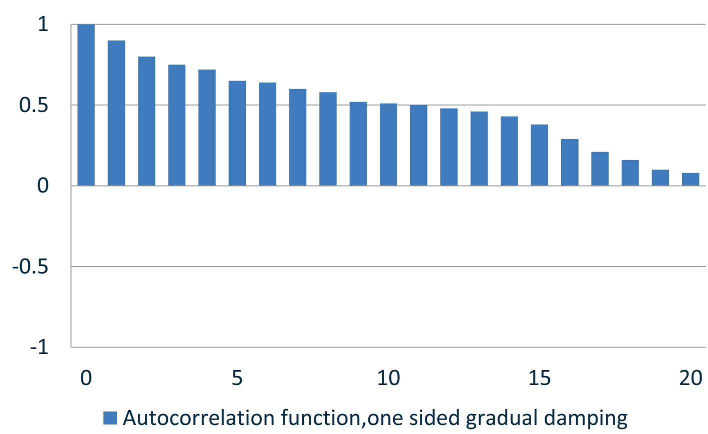
Characterizing Cycles

- The partial autocorrelations function measures the relationship between y_t and $y_{t-\tau}$ removed the effects of $y_{t-1}\cdots y_{t-\tau}$,in other words, the partial autocorrelations function measures the relationship only between y_t and $y_{t-\tau}$.

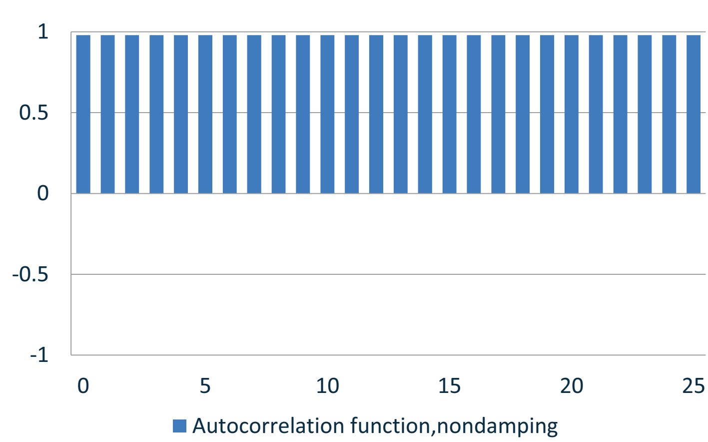
$$y_t = b_0 + b_1 y_{t-1} + b_2 y_{t-2} + \cdots + b_n y_{t-n} + \varepsilon_t$$
$$p(\tau) = b_\tau$$

Characterizing Cycles

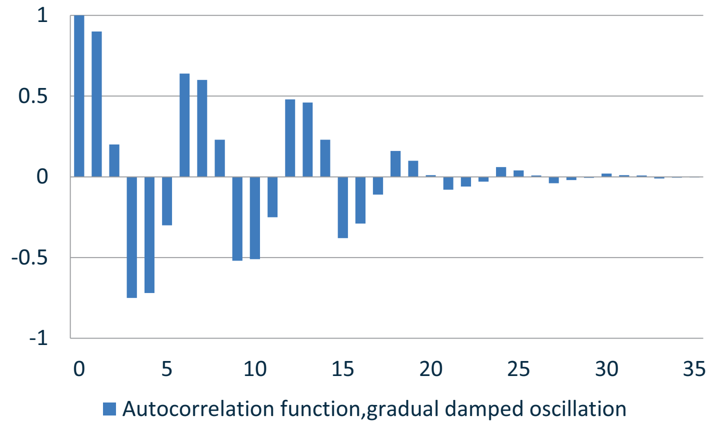
- The four characteristics of autocorrelation function(Graphs)



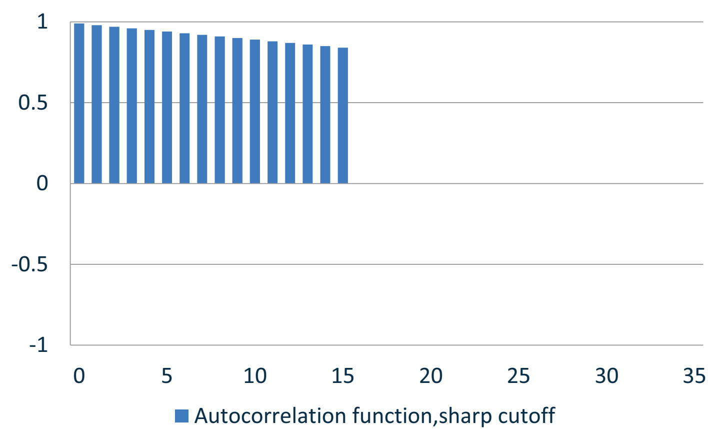
Characterizing Cycles



Characterizing Cycles



Characterizing Cycles



Characterizing Cycles

- **White noise**
  - In general, if there is a process that have zero mean, constant variance, and no serial correlation, the process is called **zero-mean white noise, or simply white noise**. Sometimes for short we write:

$$\varepsilon_t \sim WN(0, \sigma^2)$$

$$y_t = \varepsilon_t$$

hence:

$$y_t \sim WN(0, \sigma^2)$$

Characterizing Cycles

- **Independent white noise**
  - There is a point to be mentioned that ε_t and y_t are serially uncorrelated, they are not necessarily normally distributed. If y is serially independent, then we say that y is **independent white noise.** We write:

$$y_t \overset{iid}{\sim} (0, \sigma^2)$$

  - Another name for independent white noise is **strong white noise,** in contrast to standard serially uncorrelated, **weak white noise**.

Characterizing Cycles

- **Normal white noise**
  - If y is serially uncorrelated and normal distributed, of course, y also is a serially independent, we will say that y is **normal white noise** or **Gaussion white noise.**

Characterizing Cycles

- **Lag operators:**
  - The lag operator can be explained very simply that it just "operates" on a series by lagging it, just described it as follow:

$$Ly_t = y_{t-1}$$

Similarly,

$$L^2 y_t = L(L(y_t)) = L(y_{t-1}) = y_{t-2}$$

Characterizing Cycles

- We 'll also operate on a series not with the lag operator but with a **polynomial in the lag operator.** A lag operator polynomial of degree *m* is just a linear function of powers of *L*, up through the mth power, like this:

$$B(L) = b_0 + b_1 L + b_2 L^2 + \ldots\ldots + b_m L^m$$

Characterizing Cycles

- **Wold's theorem(Wold's representation)**
  - Let $\{y_t\}$ be any zero-mean covariance-stationary process.

$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i} \qquad \varepsilon_t \sim WN(0, \sigma^2)$$

    - ✓ where the b_i are coefficients with $b_0 = 1$ and $\sum_{i=0}^{\infty} b_i^2 < \infty$
    - ✓ In short, the correct "model" for any covariance stationary series is some infinite distributed lag of white noise, called the **Wold's representation.** The ε_t are often called **innovations.**

Characterizing Cycles

- **General linear process**
  - Wold's theorem tells us that when form ulating forecasting models for covariance stationary time series, we need only consider models of the form :

$$y_t = B(L)\varepsilon_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i} \qquad \varepsilon_t \sim WN(0, \sigma^2)$$

    - ✓ where the b_i are coefficients with $b_0 = 1$ and $\sum_{i=0}^{\infty} b_i^2 < \infty$.
  - We call this the **general linear process**, "general" because any covariance stationary series can be written that way, and "linear" because the Wold representation expresses the series as a linear function of its innovations.

Characterizing Cycles

- **Rational distributed lags**
  - The Wold's representation points to the crucial importance of models with **infinite distributed lags**. But it is not suitable for practical cases, so we transfer the infinite polynomials to finite polynomials. Such polynomials are called rational polynomials, and distributed lags constructed from them are called **rational distributed lags**.
    - ✓ Where the numerator polynomial is of degree q ,

$$\Theta(L)=\sum_{i=0}^{q}\theta_i L^i$$

Characterizing Cycles

- Hypothesis test for white noise
- Hypothesis testing for one autocorrelation coefficient is not enough to conclude covariance stationary, instead, **a joint hypothesis test for all the autocorrelation coefficients to be zero is needed**.
- If the autocorrelation coefficients follow a normal distribution, then the sum of squared autocorrelation coefficients follow a chi-squared distribution.
  - **Box-Pierce Q-Statistic & Ljung-Box Q-Statistic**

Characterizing Cycles

- **Box-Pierce Q-Statistic & Ljung-Box Q-Statistic**
  - H_0: The series are white noise and hence autocorrelation of the series is zero.
  - For Box-Pierce Q-statistic, the formula used is:

$$\boxed{Q_{BP}=T\sum_{\tau=1}^{m}\hat{\rho}^2(\tau)}$$

  - Whereas in case of Ljung-Box Q-statistic, the test statistic is derived as:

$$\boxed{Q_{LP}=T(T+2)\sum_{\tau=1}^{m}\left(\frac{1}{T-\tau}\right)\hat{\rho}^2(\tau)}$$

    - ✓ T = Sample size
    - ✓ m = the maximum lag under observation
  - Reject the null when Q-statistic is large.(look up in the χ^2 table)

Modeling Cycle

- **Describe the types of cyclical models**
  - MA
  - AR
  - ARMA

Modeling Cycle

- **MA(1) Model**

$$\boxed{y_t = \varepsilon_t + \theta\varepsilon_{t-1} = (1+\theta L)\varepsilon_t;\ \varepsilon_t \sim WN(0,\sigma^2)}$$

  - The current value of the observed series is expressed as a function of current and lagged unobservable shocks.
  - It is the very special case of Wold's representation.

- **MA(q) Model**

$$\boxed{\begin{array}{l} y_t = \varepsilon_t + \theta\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q} = \Theta(L)\varepsilon_t;\ \varepsilon_t \sim WN(0,\sigma^2) \\ \Theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q \end{array}}$$

Modeling Cycle

- Characteristic of MA(1)
  - If the coefficient $|\theta| < 1$, then MA(1) process is **a convergent process**.
  - The structure of the MA(1) process, in which only the first lag of the shock appears on the right, forces it to **have a very short memory**, and hence weak dynamics, regardless of the parameter value.
  - MA(1) process with parameter θ = 0.95 almost share the same persistence with the process with a parameter of θ = 0.4.
  - Autocorrelation graph appears to **have sharp cutoff**.
  - Partial Autocorrelation graph appears to have **gradual damped oscillation.**

Modeling Cycle

- **Autocorrelation of MA process**

$$\rho(\tau)=\begin{cases}\dfrac{\theta}{1\text{-}\theta^2},\ \tau=1\\ 0,\ otherwise\end{cases}$$

  - The key feature of the autocorrelation function here is **the sharp cutoff.**
- **Partial Autocorrelation of MA process**

$$y_t=\varepsilon_t+\theta y_{t-1}-\theta^2 y_{t-2}+\theta^3 y_{t-3}-\cdots$$

$$p_\tau=-(-\theta)^\tau$$

- **Partial Autocorrelation graph appears to have gradual damped oscillation.** If θ > 0, then the pattern of decay will be one of damped oscillation; otherwise, the decay will be one-sided.

Modeling Cycle

- **AR(1) Model**

$$\boxed{\begin{aligned}&y_t=\varphi y_{t-1}+\varepsilon_t;\ \varepsilon_t\sim WN(0,\sigma^2)\\&(1-\varphi L)y_t=\varepsilon_t\end{aligned}}$$

- **The relationship between AR and MA model**
  - The AR model described a relationship between y_t and y_{t-i}
  - The MA model described a relationship between y_t and ε_t ,which ε_t is a white noise process.
- AR(p) Model

$$\boxed{\begin{aligned}&y_t=\varphi_1 y_{t-1}+\varphi_2 y_{t-2}+\ldots+\varphi_p y_{t-p}+\varepsilon_t;\ \varepsilon_t\sim WN(0,\sigma^2)\\&\Phi(L)y_t=(1-\varphi_1 L-\varphi_2 L^2+\ldots-\varphi_p L^p)y_t=\varepsilon_t\end{aligned}}$$

Modeling Cycle

- **Characteristic of AR(1)**
  - The AR(1) model is capable of capturing much more persistent dynamics than is the MA(1).
  - Autocorrelation graph appears to have **gradual damped oscillation or one sided gradual damping.**
    - ✓ |φ| < 1 is the condition for covariance stationary in the AR(1). If φ is positive, the autocorrelation decay is one-sided. If φ is negative, the decay involves back-and-forth oscillations.
  - Partial Autocorrelation appears to **have sharp cutoff.**

Modeling Cycle

- **Autocorrelation of AR process**

$$\rho(\tau)=\varphi^{\tau}$$

  - If φ is positive, the autocorrelation decay is one-sided. If φ is negative, the decay involves back-and-forth oscillations.

- **Partial autocorrelation of AR process**

$$p_{\tau}=\begin{cases}\varphi, & \tau=1\\ 0, & otherwise\end{cases}$$

  - Partial Autocorrelation appears to have sharp cutoff.

Modeling Cycle

- **ARMA(p, q) Model**
  - ARMA models are often both highly accurate and highly parsimonious.

$$\boxed{y_t=\varphi_1 y_{t-1}+\varphi_2 y_{t-2}+\dots+\varphi_p y_{t-p}+\theta\varepsilon_{t-1}+\dots+\theta_q\varepsilon_{t-q}+\varepsilon_t;\varepsilon_t\sim WN\left(0,\sigma^2\right)}$$

  - Regardless of autocorrelation or partial autocorrelation, their graphs all appear to be gradual damped.

Simulation Methods

Why Use Simulation?

➢ **Advantage of simulation**

- Simplifies matters. Simulation enables us to evaluate a function of a random variable. We assume the normal distribution of r, and get the distribution of C in closed form. If r don't follow a normal distribution, or if the function of input and output is more complex. Simulation will simplifies matters.
- Visualize the probability distribution resulting from compounding probability distributions for multiple input variables.
- Incorporate correlations between input variables.
- Low-cost tool for checking the effect of changing a strategy on an output variable of interest.

Monte Carlo Simulation

➢ **What is Monte Carlo Simulation?**

- Monte Carlo was derived from the name of a famous casino established in 1862 in the south of France (actually, in Monaco).
- Monte Carlo Simulation is a statistical simulation method.

Monte-Carlo, Monaco

Monte carlo casino, Monaco

Monte Carlo Simulation

➢ **Monte Carlo Simulation**

- Monte Carlo Simulation are central to financial engineering and risk management. It can take probability distribution assumptions on the uncertainties as inputs, and generate scenarios and record what happens to variables of interest over these scenarios.
- It allows financial engineering to price complex financial instruments; allow risk managers to build the distribution of portfolios that are too complex to model analytically.

Monte Carlo Simulation

- **Simulations with one random variable**
  - Wiener process

    $$\boxed{\Delta z \sim N(0,\Delta t),\ \text{if } \varepsilon \sim N(0,1),\ \text{then } \Delta z = \varepsilon\sqrt{\Delta t}}$$

  - Generalized wiener process

    $$\boxed{\Delta x = a\Delta t + b\Delta z}$$

  - Ito process

    $$\boxed{\Delta x = a(x,t)\Delta t + b(x,t)\Delta z}$$

  - Geometric Brownian motion

    $$\boxed{\Delta S = \mu S\Delta t + \sigma S\Delta z}$$

Monte Carlo Simulation

- **Simulating a price path**
  - Geometric Brownian Motion (GBM) Model

    $$\boxed{dS_t = \mu_t S_t dt + \sigma_t S_t dz}$$

    $$\boxed{dz \sim N(0,dt),\ \text{if } \varepsilon \sim N(0,1),\ \text{then } dz = \varepsilon\sqrt{dt}}$$

  - Where:
    - ✓ St = asset price
    - ✓ dS_t = infinitesimally small price changes
    - ✓ μ_t = constant instantaneous drift term
    - ✓ σ_t = constant instantaneous volatility
    - ✓ d_z = normally distributed random variable (mean = 0, variance = d_t)
  - GBM model is widely used for stock prices and currencies.

Monte Carlo Simulation

- The assumption of GBM is the normal distribution for dS/S = dln(S), that means S follows a lognormal distribution.

- **Example**
  - The process of simulation:

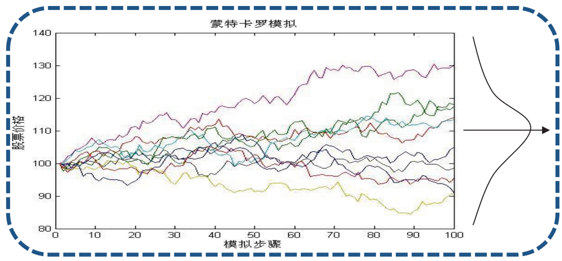
    $$\boxed{\Delta S_t = S_t(\mu\Delta t + \sigma\varepsilon\sqrt{\Delta t})}$$

  - Where ε is now a standard normal random variable.
  - Assumption: μ = 0, σ = 0.1, S_t = 100, the result is following (divide the stock price moving into 100 steps, so Δt = 1/100):

Monte Carlo Simulation

- **Simulating a price path**

Step i	Previous Price S_{t+i-1}	Random Variable ε_i	Increment ΔS	Current Price S_{t+i}
1	100.00	0.199	0.199	100.20
2	100.20	1.665	1.668	101.87
…	…	…	…	…
100	92.47	-1.153	-1.153	91.32



Evaluations of Monte Carlo Simulation

- **Advantages**
  - Flexibility, so they allow financial engineering <u>to price complex financial instruments</u>;
  - Allow risk managers to build the distribution of portfolios that are too complex to model analytically.

Evaluations of Monte Carlo Simulation

- **Disadvantages**
  - Simulation is good tool for <u>modeling uncertainty</u>, but the outcome is only as good as the inputs we provide to our models: the shape of the distribution, the parameters, and the pricing functions matters.
  - <u>As time interval shrinks, the volatility shrinks as well</u>. This implies that large discontinuities cannot occur over short intervals. But in reality, some assets experience discrete jumps.
  - <u>Price increments are assumed to have a normal distribution.</u> But in reality, price changes have fatter tails and variance of returns can change.

How to select a probability distribution

- Look at a historical distribution of past returns and assume that the future will behave in the same way. A simple approach is to draw randomly from historical scenarios to create for future realizations.(a.k.a. bootstrapping).
- Assume a particular probability distribution for future returns, and use historical data to estimate parameters, like mean and deviation.
- Use historical data to find a distribution for returns that provides the best fit to the data. Goodness-of-fit test: Chi-Square hypothesis, Kolmogorov-Smirnov (K-S) test, Anderson-Darling (A-D)test and root-mean-squared-error (RMSE).
- Construct a probability distribution based on subjective guess. Constructing a probability distribution based on your subjective guess about how the uncertain variable in your model will behave.

Bootstrapping Method

- **Bootstrapping** is related to simulation, but with one crucial difference. With simulation, the data are constructed completely artificially. Bootstrapping, on the other hand, is used to obtain a description of the properties of empirical estimators by using the sample data points themselves, and it involves sampling repeatedly with replacement from the actual data.
- The **advantage** of bootstrapping over the use of analytical results is that it allows the researcher to make inferences without making strong distributional assumptions, since the distribution employed will be that of the actual data.
- **Situations where the bootstrap will be ineffective:**
  - Outliers in the Data
  - Non-Independent Data: Use of the bootstrap implicitly assumes that the data are independent of one another.

Variance Reduction Techniques

- **Sampling Variation**
  - The sampling variation in a Monte Carlo study is measured by the standard error estimate, denoted S_x:

$$\boxed{S_x = \sqrt{\frac{\text{var}(x)}{N}}}$$

    - ✓ where var(x) is the variance of the estimates of the quantity of interest over the N replications.
  - It can be seen from this equation that to reduce the Monte Carlo standard error by a factor of 10, the number of replications must be increased by a factor of 100.
  - Consequently, in order to achieve acceptable accuracy, the number of replications may have to be set at an unfeasibly high level. An alternative way to reduce Monte Carlo sampling error is to use a variance reduction technique.

Variance Reduction Techniques

- **Antithetic Variates**
  - The antithetic variate technique involves taking the complement of a set of random numbers and running a parallel simulation on those. For example, if the driving stochastic force is a set of TN(0, 1) draws, denoted u_t, for each replication, an additional replication with errors given by $-u_t$ is also used.
  - It can be shown that the Monte Carlo standard error is reduced when antithetic variates are used.

Variance Reduction Techniques

- Simple Illustration: suppose that the average value of the parameter of interest across two sets of Monte Carlo replications is given by:

$$\bar{x} = (x_1 + x_2)/2$$

  - ✓ where x1 and x2 are the average parameter values for replications sets 1 and 2 respectively, the variance will be given by:

$$\text{var}(\bar{x}) = \frac{1}{4}\left[\text{var}(x_1) + \text{var}(x_2) + 2\,\text{cov}(x_1, x_2)\right]$$

  - ✓ If no antithetic variates are used, the two sets of Monte Carlo replications will be independent, so that their covariance will be zero, i.e.
  - ✓ The use of antithetic variates would lead the covariance to be negative, and then the Monte Carlo sampling error to be reduced.

$$\text{var}(\bar{x}) = \frac{1}{4}\left[\text{var}(x_1) + \text{var}(x_2)\right]$$

Variance Reduction Techniques

- **Control Variates**
  - The control variates help to reduce the Monte Carlo variation owing to particular sets of random draws by <u>using the same draws on a related problem whose solution is known.</u> The application of control variates involves employing a variable similar to that used in the simulation, but whose properties are known prior to the simulation.
  - Denote the variable whose properties are known by y, and that whose properties are under simulation by x. The simulation is conducted on x and also on y, with the same sets of random number draws being employed in both cases. Denoting the simulation estimates of x and y be $\hat{x}$ and y, respectively, a new estimate of x can be derived from:

$$x^* = y + (\hat{x} - y)$$

Variance Reduction Techniques

- **Example**
  - A researcher may be interested in pricing an arithmetic Asian option using simulation. At the time of writing, an analytical model is not yet available for pricing such options.
  - A control variate price could be obtained by finding the price via simulation of a similar derivative whose value is known analytically.
    - ✓ Asian and vanilla options would be priced using simulation, as shown below, with the simulated price given by P_A and P_{BS}*, respectively. The price of the vanilla option, P_{BS} is also calculated using an analytical formula, such as Black-Scholes. The new estimate of the Asian option price, P_A*, would then be given by:

$$P_A^* = (P_A - P_{BS}) + P_{BS}^*$$

Variance Reduction Techniques

- It is worth noting that control variates succeed in reducing the Monte Carlo sampling error only if the control and simulation problems are very closely related. Take the variance of both sides:

$$\mathrm{var}(x^*) = \mathrm{var}\left[y + (\hat{x} - y)\right]$$

- VAR(y) = 0 since y is a quantity which is known analytically and is therefore not subject to sampling variation. The condition that must hold for the Monte Carlo sampling variance to lower with control variates than without is that VAR(x*) is less than VAR($\hat{x}$). Then:

$$\boxed{\begin{aligned} &\mathrm{var}(y) - 2\mathrm{cov}(\hat{x}, y) < 0 \\ &\mathrm{Corr}(\hat{x}, y) > \frac{1}{2}\sqrt{\frac{\mathrm{var}(y)}{\mathrm{var}(\hat{x})}} \end{aligned}}$$

Variance Reduction Techniques

- **Random Number Re-Usage across Experiments**
  - Using the same sets of draws across experiments can greatly reduce the variability of the difference in the estimates across those experiments. However, the accuracy of the actual estimates in each case will not be increased, of course.
  - Another possibility involves taking long series of draws and then slicing them up into several smaller sets to be used in different experiments.

Random Number Generation

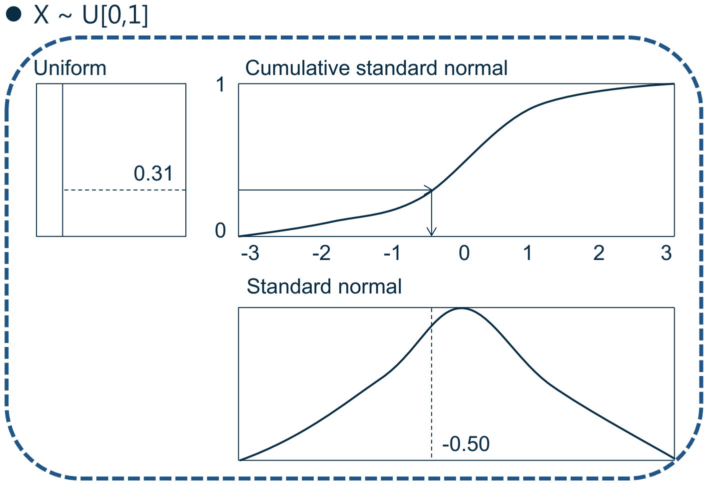
- **Pseudorandom number generators**
  - Good uniform random numbers on [0,1] is critical for simulation. Truly random number generation is difficult and time consuming.
  - Use random number generation algorithms that produce streams of numbers that appear to be random. In fact, they are a result of a clearly defined series of calculation steps.
  - It starts with a number called seed and $x_n = g(x_{n-1})$, so if the same seed is used in several simulation, each simulation sequence will contain exactly the same numbers, which is helpful for running fair comparison between different strategies evaluated under uncertainty.

Random Number Generation

- **Standards for effective pseudorandom number generator**
  - The numbers in the generated sequence are uniformly distributed between 0 and 1. This can be tested by Chi-Square or K-S test.
  - The sequence has a long cycle (takes many iterations before repeating).
  - Numbers are not autocorrelated, can be tested by D-W test.

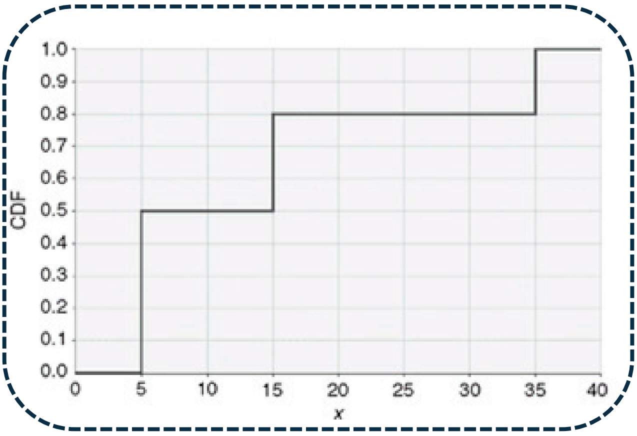
Random Number Generation

- **Inverse transform method**
  - X ~ U[0,1]



Random Number Generation

- **If it is a discrete distribution:**
  - 5: 50%
  - 15: 30%
  - 35: 20%



Example 1

- You simulate the price path of stock HHF using a geometric Brownian motion model with the following parameters:

  Drift: $\mu = 0$ Volatility: $\sigma = 0.2$ Time step: $\Delta t = 0.01$

  Assuming that S_t is the price of the stock at time t, if $S_0 = 50$ and the simulated standard normal random variables in the first two steps are $\varepsilon_1 = -0.521$ and $\varepsilon_2 = 1.22$, respectively, by what percent will the stock price change in the second step of the simulation?

  A. -1.04%

  B. 0.43%

  C. 1.12%

  D. 2.45%

- Correct Answer：D

Example 2

- Which of the following statements is true regarding the bootstrap simulation method used in VaR estimation?

  I. Bootstrapping uses actual market data.

  II. The bootstrapping method always uses a time horizon based on the time scale of the historical data.

  III. Bootstrapping is based on synthesis of normally distributed random numbers.

  A. I only

  B. II only

  C. I and III

  D. I, II and III

Example 2

- Correct Answer: A
  - Bootstrapping uses actual market data. Bootstrapping can be done with data that uses the same time line as the one of interest or shorter term data. Monte Carlo is based on a synthesis of normally distributed random numbers.

Example 3

- A risk manager has been requested to provide some indication of accuracy of a Monte Carlo simulation. Using 1,000 replications of a normally distributed variable S, the relative error in the one-day 99% VaR is 5%. Under these conditions:

  A. Using 1,000 replications of a long option position on S should create a larger relative error.

  B. Using 10,000 replications should create a larger relative error.

  C. Using another set of 1,000 replications will create an exact measure of 5.0% for relative error.

  D. Using 1,000 replications of a short option position on S should create a larger relative error.

Example 3

- Correct Answer: D
  - Short option positions have long left tails, which makes it more difficult to estimate a left-tailed quantile precisely. Accuracy with independent draws increases with the square root of K. Thus increasing the number of replications should shrink the standard error, so answer B is incorrect.

Estimating Volatilities and Correlations

Estimating Volatilities

- **Estimating Volatility**
  - Define σ_n as the volatility of a market variable on day n, as estimated at the end of day n – 1. σ_n^2 as the variance rate.
  - Define S_i as the value of the market variable at the end of day i.
  - Define u_i as the continuously compounded return during day i (between the end of day i – 1 and the end of day i).

$$u_i = \ln\frac{S_i}{S_{i-1}} \qquad \bar{u} = \frac{1}{m}\sum_{i=1}^{m} u_{n-i}$$

$$\sigma_n^2 = \frac{1}{m-1}\sum_{i=1}^{m}(u_{n-i}-\bar{u})^2$$

Estimating Volatilities

- For the purpose of monitoring daily volatility, we give the following changes:
  - Define u_i as the percentage change in the market variable between the end of day i – 1 and the end of day i.

$$u_i = \frac{S_i - S_{i-1}}{S_{i-1}}$$

  - $\bar{u}$ is assumed to be zero.
  - m – 1 is replaced by m.
- Then we can get a simple formula for the variance rate.

$$\sigma_n^2 = \frac{1}{m}\sum_{i=1}^{m} u_{n-i}^2$$

Estimating Volatilities

- **Weighting Schemes**
  - The above formula gives equal weight to u_{n-1}^2, u_{n-2}^2, ……, u_{n-m}^2 .
  - Our objective is to estimate the current level of volatility, so we give more weight to recent data.

$$\boxed{\sigma_n^2 = \sum_{i=1}^{m} \alpha_i u_{n-i}^2} \quad \boxed{\alpha_i < \alpha_j \text{ where } i > j} \quad \boxed{\sum_{i=1}^{m} \alpha_i = 1}$$

  - If the objective is to generate a greater influence on recent observations, then the α's will decline in value for older observations.

Estimating Volatilities

- **ARCH Model**
  - Adding a long-run average variance rate and be given a weight.

$$\boxed{\sigma_n^2 = \gamma V_L + \sum_{i=1}^{m} \alpha_i u_{n-i}^2} \quad \boxed{\gamma + \sum_{i=1}^{m} \alpha_i = 1}$$

    ✓ Where V_L is the long-run variance rate and γ is the weight assigned to V_L.
  - Defining $\omega = \gamma V_L$, then the model can be written:

$$\boxed{\sigma_n^2 = \omega + \sum_{i=1}^{m} \alpha_i u_{n-i}^2}$$

Estimating Volatilities

- **EWMA Model**
  - In an exponentially weighted moving average model, as time goes by, the weight assigned to α_i is exponentially decreasing.

$$\boxed{\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1-\lambda) u_{n-1}^2}$$

  - The estimate, σ_n , of the volatility for day n (made at the end of day n-1) is calculated from σ_{n-1} (the estimate made at the end of day n-2 of the volatility for day n-1) and u_{n-1} (the most recent daily percentage change).
  - High values of λ will minimize the effect of daily percentage returns, whereas low values of λ will tend to increase the effect of daily percentage returns on the current volatility estimate.

Estimating Volatilities

- Example
  - Suppose that λ is 0.90, the volatility estimated for a market variable for day n – 1 is 1% per day, and during day n – 1 the market variable increased by 2%.

$$\sigma_{n-1}^2 = 0.01^2 = 0.0001$$

$$\mu_{n-1}^2 = 0.02^2 = 0.0004$$

$$\sigma_n^2 = 0.9 \times 0.0001 + 0.1 \times 0.0004 = 0.00013$$

Estimating Volatilities

- **Using EWMA To Forecast Future Volatility**

$$\sigma_n^2 = \lambda\sigma_{n-1}^2 + (1-\lambda)u_{n-1}^2$$

$$\sigma_n^2 = \lambda\left[\lambda\sigma_{n-2}^2 + (1-\lambda)u_{n-2}^2\right] + (1-\lambda)u_{n-1}^2 = (1-\lambda)\left(u_{n-1}^2 + \lambda u_{n-2}^2\right) + \lambda^2\sigma_{n-2}^2$$

$$\sigma_n^2 = (1-\lambda)\left(u_{n-1}^2 + \lambda u_{n-2}^2\right) + \lambda^2\sigma_{n-2}^2 = (1-\lambda)\left(u_{n-1}^2 + \lambda u_{n-2}^2 + \lambda^2 u_{n-3}^2\right) + \lambda^3\sigma_{n-3}^2$$

$$\cdots\cdots$$

$$\sigma_n^2 = (1-\lambda)\sum_{i=1}^{m}\lambda^{i-1}u_{n-i}^2 + \lambda^m\sigma_{n-m}^2$$

- **从现在算起，i天前收益率平方的权重记为α_i，那么$\alpha_i = (1-\lambda)\lambda_{i-1}$.**

Estimating Volatilities

- Given λ of 0.94, under an infinite series, what is the weight assigned to the seventh prior daily squared return?
  A. 4.68%
  B. 4.40%
  C. 4.14%
  D. 3.89%

- Correct Answer: C

$$\text{weight} = 0.94^6 \times (1-0.94) = 4.14\%$$

Estimating Volatilities

- **The Attractive Feature Of EWMA Approach**
  - Relatively little data needs to be stored.
  - We need only remember the current estimate of the variance rate and the most recent observation on the value of the market variable.
  - Tracks volatility changes. The value of λ governs how responsive the estimate of the daily volatility is to the most recent daily percentage change.
  - Risk Metrics uses λ = 0.94 for daily volatility forecasting.

Estimating Volatilities

- **GARCH(1, 1) Model**
  - In GARCH(1, 1), σ_n^2 is calculated from a long-run average variance rate VL, as well as from σn-1 and un-1. The equation for GARCH(1, 1) is:

$$\boxed{\sigma_n^2 = \gamma V_L + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2}$$

  - EWMA model is a particular case of GARCH(1, 1), where γ = 0, α = 1 – λ, β = λ.
  - The "(1, 1)" in GARCH(1, 1) indicates that σ_n^2 is based on the most recent observation of u_2 and the most recent estimate of the variance rate.
  - Setting $\omega = \gamma V_L$, the GARCH(1, 1) model can also be written:

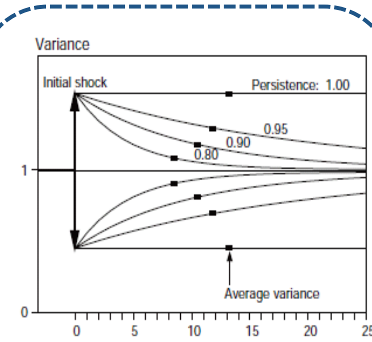
$$\boxed{\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2}$$

Estimating Volatilities

- **GARCH(1, 1) Model**

$$\boxed{\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2}$$

  - Persistence: α + β
  - It defines the speed at which shocks to the variance revert to their long-run values.
  - The higher the persistence (given that it is less than one), the longer it will take to revert to the mean following a shock or large movement.



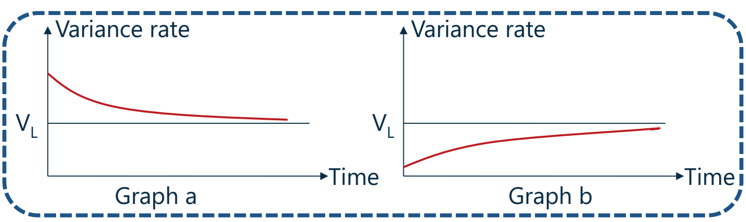
Shocks to a GARCH Process

Estimating Volatilities

- **Example**
  - Suppose that a GARCH(1, 1) model is estimated from daily data as: $\sigma_n^2 = 0.000002 + 0.13u_{n-1}^2 + 0.86\sigma_{n-1}^2$
  - This corresponds to α = 0.13, β = 0.86, ω = 0.000002
  - Because $\gamma = 1 - \alpha - \beta$, it follows that $\gamma = 0.13$
  - Because $\omega = \gamma V_L$ it follows that V_L = 0.0002. This corresponds to a volatility of 0.014 = 1.4% per day.
  - Suppose that the estimate of volatility on day n – 1 is 1.6% per day, and that on day n – 1 the market variable decreased by 1%. Then:
    $$\sigma_n^2 = 0.000002 + 0.13 \times 0.01^2 + 0.86 \times 0.016^2 = 0.00023516$$
  - The new estimate of the volatility is therefore 0.0153 = 1.53% per day.

Estimating Volatilities

- **Choosing between the models**
  - GARCH(1, 1) is theoretically more appealing than EWMA model.
  - In practice, variance rates tend to be mean reverting. The GARCH(1, 1) model incorporates mean reversion, whereas the EWMA model does not.
- **Mean Reversion**



  - (a) current variance rate is above long-term variance rate.
  - (b) current variance rate is below long-term variance rate.

Estimating Volatilities

- **Using GARCH(1, 1) To Forecast Future Volatility**

$$\sigma_n^2 = (1-\alpha-\beta)V_L + \alpha u_{n-1}^2 + \beta\sigma_{n-1}^2$$

$$\sigma_n^2 - V_L = \alpha\left(u_{n-1}^2 - V_L\right) + \beta\left(\sigma_{n-1}^2 - V_L\right)$$

$$\sigma_{n+t}^2 - V_L = \alpha\left(u_{n+t-1}^2 - V_L\right) + \beta\left(\sigma_{n+t-1}^2 - V_L\right)$$

$$\downarrow E\left(u_{n+t-1}^2\right) = \sigma_{n+t-1}^2$$

$$E\left(\sigma_{n+t}^2 - V_L\right) = (\alpha+\beta)E\left(\sigma_{n+t-1}^2 - V_L\right)$$

$\downarrow$ Using this equation repeatedly

$$E\left(\sigma_{n+t}^2 - V_L\right) = (\alpha+\beta)^t\left(\sigma_n^2 - V_L\right)$$

$$\downarrow$$

$$\boxed{E\left(\sigma_{n+t}^2\right) = V_L + (\alpha+\beta)^t\left(\sigma_n^2 - V_L\right)}$$

Estimating Volatilities

- **Example**
  - In the yen-dollar exchange rate example considered earlier $\alpha + \beta = 0.9602$ and $V_L = 0.00004422$. Suppose that our estimate of the current variance rate per day is 0.00006. In 10 days the expected variance rate is:

    $$0.00004422 + 0.9602^{10}(0.00006 - 0.00004422) = 0.00005473$$

  - The expected volatility per day is 0.74%, still well above the long-term volatility of 0.665% per day. However, the expected variance rate in 100 days is:

    $$0.00004422 + 0.9602^{100}(0.00006 - 0.00004422) = 0.00004449$$

  - The expected volatility per day is 0.667%, very close to the long-term volatility.

Estimating Correlations

- **Estimating Correlations**

$$\boxed{\hat{\rho}_{XY} = \frac{Cov_n}{\sigma_{x,n}\sigma_{y,n}}}$$

- For EWMA model

$$\boxed{Cov_n = \lambda Cov_{n-1} + (1-\lambda)x_{n-1}y_{n-1}}$$

- For GARCH(1,1) model

$$\boxed{Cov_n = \omega + \alpha x_{n-1}y_{n-1} + \beta Cov_{n-1}}$$

Example 1

- Suppose that $\lambda = 0.95$ and that the estimate of the correlation between two variable X and Y on day n – 1 is 0.6. Suppose further that the estimate of the volatilities for the X and Y on day n – 1 are 1% and 2%, respectively. From the relationship between correlation and covariance, the estimate of the covariance between the X and Y on day n – 1 is:

$$Cov_{n-1} = 0.6 \times 0.01 \times 0.02 = 0.00012$$

- Suppose that the percentage changes in X and Y on day n – 1 are 0.5% and 2.5%, respectively. The variance and covariance for day n would be updated as follows:

$$\sigma_{x,n}^2 = 0.95 \times 0.01^2 + 0.05 \times 0.005^2 = 0.00009625 \qquad \sigma_{y,n}^2 = 0.95 \times 0.02^2 + 0.05 \times 0.025^2 = 0.00041125$$

$$Cov_n = 0.95 \times 0.00012 + 0.05 \times 0.005 \times 0.025 = 0.00012025$$

$$\hat{\rho}_{XY} = \frac{Cov_n}{\sigma_{x,n}\sigma_{y,n}} = \frac{0.00012025}{\sqrt{0.00009625} \times \sqrt{0.00041125}} = 0.6044$$

Example 2

- Using a daily Risk Metrics EWMA model with a decay factor λ = 0.95 to develop a forecast of the conditional variance, which weight will be applied to the return that is four days old?
  - A. 0.000
  - B. 0.043
  - C. 0.048
  - D. 0.950
- Correct Answer：B

Example 3

- An investment bank uses the Exponentially Weighted Moving Average (EWMA) technique with λ of 0.9 to model the daily volatility of a security. The current estimate of the daily volatility is 1.5%. The closing price of the security is USD 20 yesterday and USD 18 today. Using continuously compounded returns, what is the updated estimate of the volatility?
  - A. 3.62%
  - B. 1.31%
  - C. 2.96%
  - D. 5.44%
- Correct Answer：A

Example 4

- Which of the following GARCH models will take the shortest time to revert to its long-run value?
  - A. $\sigma_n^2 = 0.05 + 0.03u_{n-1}^2 + 0.96\sigma_{n-1}^2$
  - B. $\sigma_n^2 = 0.03 + 0.02u_{n-1}^2 + 0.95\sigma_{n-1}^2$
  - C. $\sigma_n^2 = 0.02 + 0.01u_{n-1}^2 + 0.97\sigma_{n-1}^2$
  - D. $\sigma_n^2 = 0.01 + 0.01u_{n-1}^2 + 0.98\sigma_{n-1}^2$
- Correct Answer： B

Example 5

- The following GARCH(1,1) model is used to forecast the daily return variance of an asset:

$$\sigma_n^2 = 0.000005 + 0.05u_{n-1}^2 + 0.92\sigma_{n-1}^2$$

Suppose the estimate of the volatility today is 5.0% and the asset return is -2.0%. What is the estimate of the long-run average volatility per day?

A. 1.29%

B. 1.73%

C. 1.85%

D. 1.91%

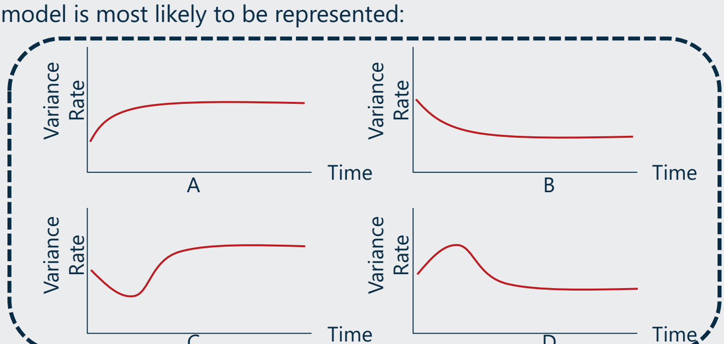
- Correct Answer：A

Example 6

- The following GARCH(1,1) model is used to forecast the daily return variance of an asset:

$$\sigma_n^2 = 0.000005 + 0.05u_{n-1}^2 + 0.92\sigma_{n-1}^2$$

Suppose the estimate of the volatility today is 5.0% and the asset return is -2.0%. The resulting volatility term structure from this GARCH model is most likely to be represented:



Correlations and Copulas

Cholesky Factorization

- **Simulate correlated random variables**

$$\Delta S_{1,t} = S_{1,t-1}\mu_1\Delta t + S_{1,t-1}\sigma_1\varepsilon_{1,t}\sqrt{\Delta t}$$

$$\Delta S_{2,t} = S_{2,t-1}\mu_2\Delta t + S_{2,t-1}\sigma_2\varepsilon_{2,t}\sqrt{\Delta t}$$

- To account for correlations between variables, we start with a set of independent variables η of unit variance, which then are transformed into the ε. In a two variable setting, we construct

$$\boxed{\varepsilon_1 = \eta_1}$$

$$\boxed{\varepsilon_2 = \rho\eta_1 + (1-\rho^2)^{1/2}\eta_2}$$

*

$$V(\varepsilon_1) = V(\varepsilon_2) = 1$$

$$\mathrm{cov}(\varepsilon_1, \varepsilon_2) = \rho$$

$$V(\varepsilon) = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = R$$

Cholesky Factorization

- Cholesky Factorization: $R = TT'$, where T is a lower triangular matrix, R is called "correlation matrix".
- Example: 2 risk factor case

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} a_{11} & a_{21} \\ 0 & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11}^2 & a_{11}a_{21} \\ a_{21}a_{11} & a_{21}^2 + a_{22}^2 \end{bmatrix}$$

- Solution:

$$\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & (1-\rho^2)^{\frac{1}{2}} \end{bmatrix}\begin{bmatrix} 1 & \rho \\ 0 & (1-\rho^2)^{\frac{1}{2}} \end{bmatrix}$$

$$\varepsilon = T\eta$$

$$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \rho & (1-\rho^2)^{\frac{1}{2}} \end{bmatrix}\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix}$$

Copula

- **Drawbacks of using correlation to measure dependence.**
  - Correlation is a good measure of dependence when random variables are distributed as multivariate elliptical (e.g., normal, student's).
  - However, correlation is only defined if variance is finite. There might be a problem for **non-elliptical** distributions. e.g. Levy distribution can have infinite variance.
  - Correlation measures the linear relationship of two variables. If risks are independent → zero correlation, however zero correlation does not imply independence.

Copula

➢ **Introduction of Copula**

- A "copula" is Latin noun which means 'a link, tie or bond'

➢ **More about Copula:**

- When the two variables are independent, the joint density is simply the product of the marginal densities.
- It is rarely the case, however, that financial variables are independent. Dependencies can be modeled by a **function called the copula**, which links, or attaches, marginal distributions into a joint distribution.

Copula

➢ **Copula function**

- Formally, the copula is a function of the marginal distributions F (x), plus some parameters, θ, that are specific to this function (and not to the marginals). In the bivariate case, it has two arguments

$$c_{12}[F_1(x_1),\ F_2(x_2);\ \theta]$$

✓ The link between the joint and marginal distribution is made explicit by Sklar's theorem, which states that, for any joint density, there exists a copula that links the marginal densities

$$f_{12}(x_1, x_2) = f_1(x_1) \times f_2(x_2) \times c_{12}[F_1(x_1), F_2(x_2); \theta]$$

◆ With independence, the copula function is a constant always equal to one.

Copula

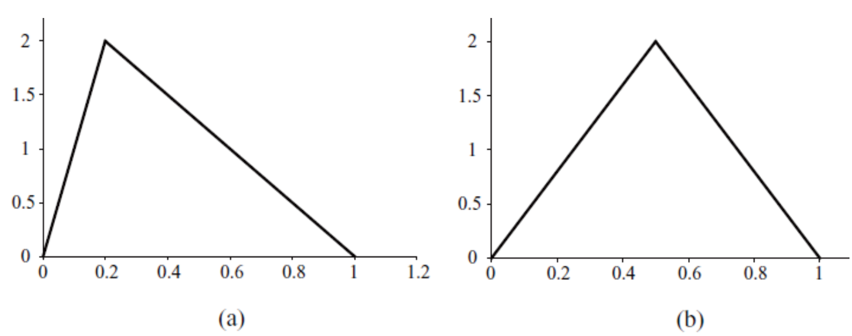


FIGURE 11.2 Triangular Distributions for V_1 and V_2

Copula

TABLE 11.1 Mapping of V_1 Which Has the Triangular Distribution in Figure 11.2(a) to U_1 Which Has a Standard Normal Distribution

V_1 Value	Percentile of Distribution	U_1 Value
0.1	5.00	−1.64
0.2	20.00	−0.84
0.3	38.75	−0.29
0.4	55.00	0.13
0.5	68.75	0.49
0.6	80.00	0.84
0.7	88.75	1.21
0.8	95.00	1.64
0.9	98.75	2.24

TABLE 11.2 Mapping of V_2 Which Has the Triangular Distribution in Figure 11.2(b) to U_2 Which Has a Standard Normal Distribution

V_2 Value	Percentile of Distribution	U_2 Value
0.1	2.00	−2.05
0.2	8.00	−1.41
0.3	18.00	−0.92
0.4	32.00	−0.47
0.5	50.00	0.00
0.6	68.00	0.47
0.7	82.00	0.92
0.8	92.00	1.41
0.9	98.00	2.05

Copula

TABLE 11.3 Cumulative Joint Probability Distribution for V_1 and V_2 in the Gaussian Copula Model

	V_2								
V_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	0.006	0.017	0.028	0.037	0.044	0.048	0.049	0.050	0.050
0.2	0.013	0.043	0.081	0.120	0.156	0.181	0.193	0.198	0.200
0.3	0.017	0.061	0.124	0.197	0.273	0.331	0.364	0.381	0.387
0.4	0.019	0.071	0.149	0.248	0.358	0.449	0.505	0.535	0.548
0.5	0.019	0.076	0.164	0.281	0.417	0.537	0.616	0.663	0.683
0.6	0.020	0.078	0.173	0.301	0.456	0.600	0.701	0.763	0.793
0.7	0.020	0.079	0.177	0.312	0.481	0.642	0.760	0.837	0.877
0.8	0.020	0.080	0.179	0.318	0.494	0.667	0.798	0.887	0.936
0.9	0.020	0.080	0.180	0.320	0.499	0.678	0.816	0.913	0.970

(Correlation parameter = 0.5. Table shows the joint probability that V_1 and V_2 are less than the specified values.)

Copula

- **Various Types of Copulas**
  - Gaussian Copula: most common correlation structure where U1 and U2 are assumed to follow bivariate Normal distribution

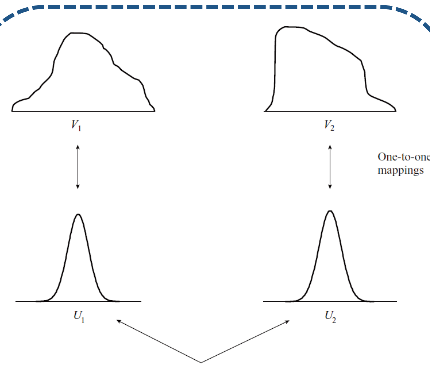


FIGURE 11.3 The Way in Which a Copula Model Defines a Joint Distribution

Copula

- **Various Types of Copulas**
  - Student's t-copula: similar to Gaussian except that U1 and U2 are assumed to follow bivariate student's t-distribution
  - Multivariate copula: copula can be used to define a correlation structure between more than two variables. The simplest example is the multivariate Gaussian Copula.
  - One-factor (factor model) copula: $U_i = \alpha_i F + \sqrt{1-\alpha_i^2} Z_i$

Tail Dependence

- **Tail dependence:**
  - Tail dependence is an important issue because extreme events are often related (i.e., disasters often come in pairs or more)
  - If marginal distributions are continuous, we can define a coefficient of (upper) tail dependence of X and Y as the limit, as α → 1 from below, of

$$\Pr\left[Y > F_y^{-1}(\alpha) \mid Y > F_x^{-1}(\alpha)\right] = \lambda$$

Copula

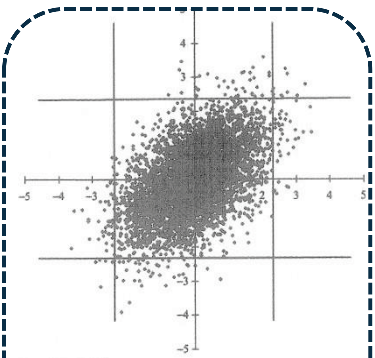


FIGURE 6-4 5,000 random samples from a bivariate normal distribution.

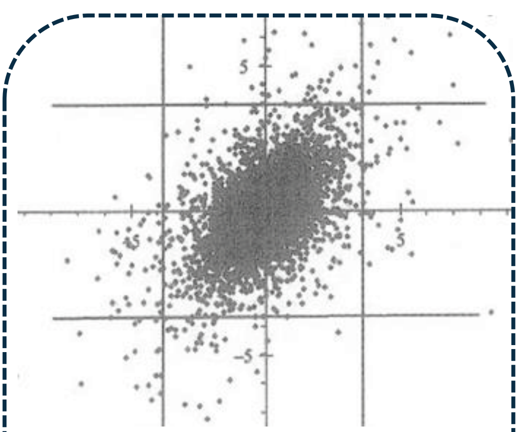


FIGURE 6-5 5,000 random samples from a bivariate Student *t*-distribution with four degrees of freedom.

It's not the end but just beginning.

Thought is already is late, exactly is the earliest time.

感到晚了的时候其实是最快的时候。