Foundation of Probability Theory

Professor Yongmiao Hong
Cornell University

May 22, 2019

CONTENTS

Random Experiments

Recall fundamental axioms of econometrics:

- Axiom A: An economic system can be viewed as a random experiment governed by some probability law.

- Axiom B: Any economic phenomenon (often in form of data) can be viewed as an out-come of this random experiment. The random experiment is called a “data generating process".

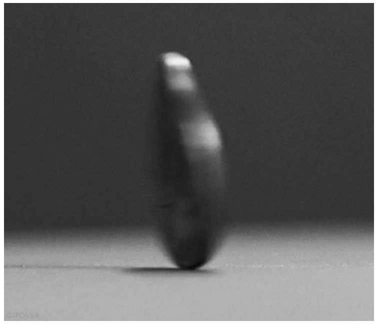
Random Experiments

Question: How to characterize a random system or stochastic process?

Definition 1. [Random Experiment]

A random experiment is a mechanism which has at least two possible outcomes.

When a random experiment is performed, one and only one outcome will occur, but which outcome to occur is unknown in advance.



Random Experiments

Remarks:

- The word "experiment" means a process of observation or measurement in a broad sense. It is not necessarily a real experiment as encountered in (e.g.) physics.

Random Experiments

Remarks:

- There are two essential elements of a random experiment:

The set of all possible outcomes;

The likelihood with which each outcome will occur.

Random Experiments

Remarks:

- The purpose of mathematical statistics is to provide mathematical models for random experiments of interest.
- Once a model for such an experiment is provided and the theory worked out in detail, the statistician may, within this framework, make inference about the probability law of the random experiment.

CONTENTS

Basic Concepts of Probability

Definition 2. [**Sample Space**]

The possible outcomes of a random experiment are called "basic outcomes", and **the set of all basic outcomes** constitutes "the sample space", which is denoted by *S.*

Flipping a Coin

SAMPLE SPACE

{Head, Tail}

Uniform

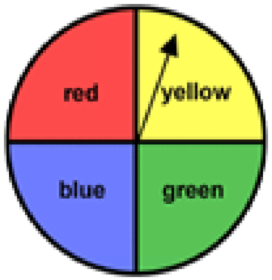
Rolling a Six Sided Dice

SAMPLE SPACE

{1, 2, 3, 4, 5, 6}

Uniform

Spinning a 4 color spinner



SAMPLE SPACE

{Red, Yellow, Green, Blue}

Uniform

Rolling a Weighted Dice

SAMPLE SPACE

{4, 5, 6}

Not Uniform

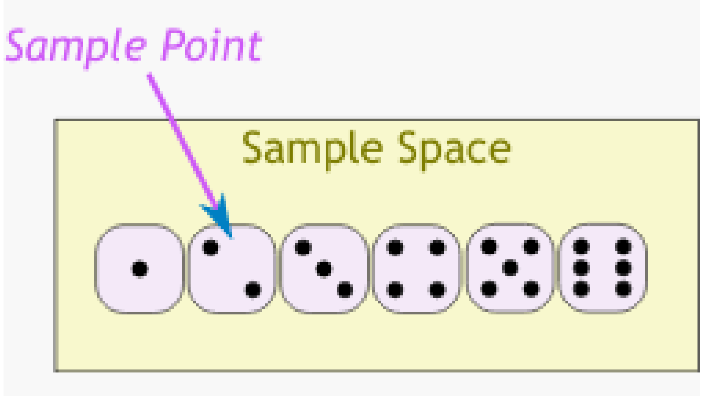
Basic Concepts of Probability

- When an experiment is performed, the realization of the experiment will be one (and only one) outcome in the sample space.
- If the experiment is performed a number of times, a different outcome may occur each time or some outcomes may repeat.

Basic Concepts of Probability

Remarks:

- A sample space S is sometimes called an outcome space. Each outcome in S is called an element of S, or simply a sample point.



- It is important to note that for a random experiment, one knows the set of all possible basic outcomes, but one does not know which outcome will arise before performing the random experiment.

Basic Concepts of Probability

Example 1: [**Throwing a Coin**]

Head(H) Tail(T)

Two possible outcomes:

+

The sample space is $S = \{H, T\}$

Basic Concepts of Probability

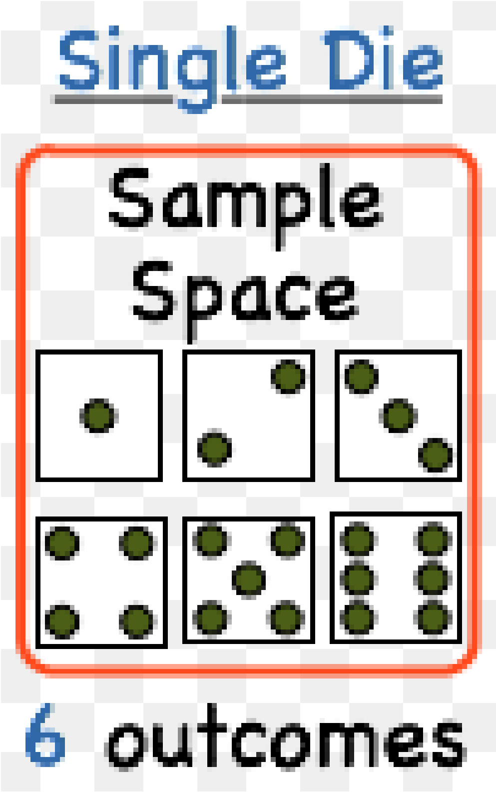
Example 2: [**Direction of Changes**]

Let $Y = 1$ if the U.S. Gross Domestic Product (GDP) growth rate is positive and $Y = 0$ if the U.S. GDP growth rate is negative.

Basic Concepts of Probability

Example 3: [**Rolling a Die**]

The basic outcomes are the numbers 1, 2, 3, 4, 5, 6.

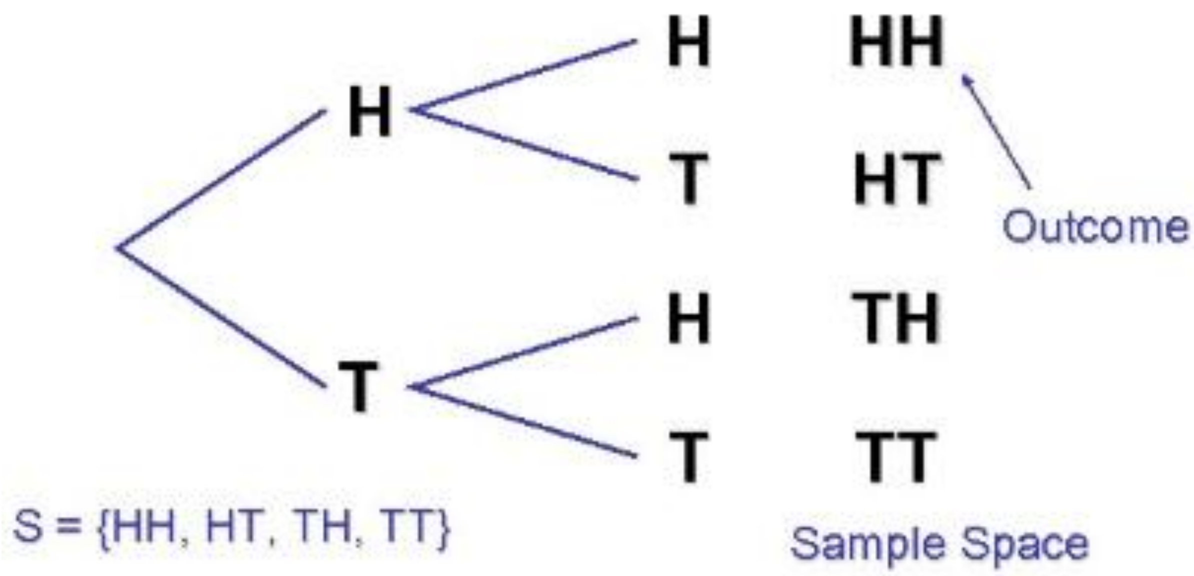


$$S = \{1,2,3,4,5,6\}$$

Basic Concepts of Probability

Example 4: [**Throwing Two Coins**]

The sample space



$$S = \{(H, H), (H, T), (T, H), (T, T)\}$$

Basic Concepts of Probability

Example 5

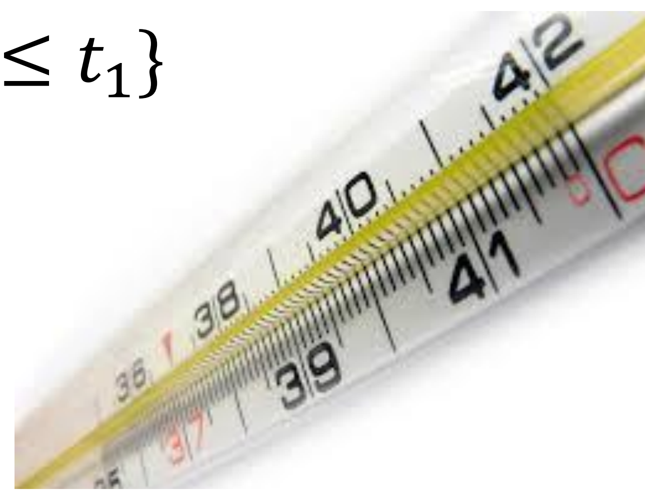
Suppose t_0 is the lowest temperature in an area, and t_1 is the highest temperature of the area.

Let T denotes the possible temperature of the area.

Then the sample space of T is

$$S = \{t \in \mathbb{R} : t_0 \leq t \leq t_1\}$$

where $\mathbb{R}$ *denotes the real line*.



Basic Concepts of Probability

Remarks:

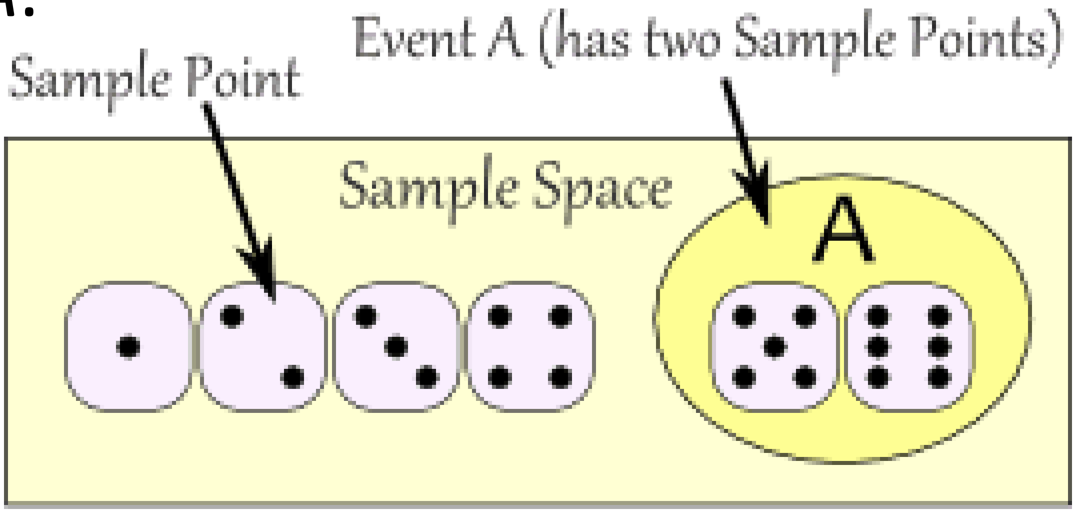
- A sample space S can be countable or uncountable.
- The distinction between a countable sample space and an uncountable sample space dictates the ways in which probabilities will be assigned.

Basic Concepts of Probability

Definition 3. [Event]

An event A is a collection of basic outcomes from the sample space S that share certain common features or equivalently obey certain restrictions.

The event A is said to occur if the random experiment gives rise to one (and only one) of the constituent basic outcomes in A.

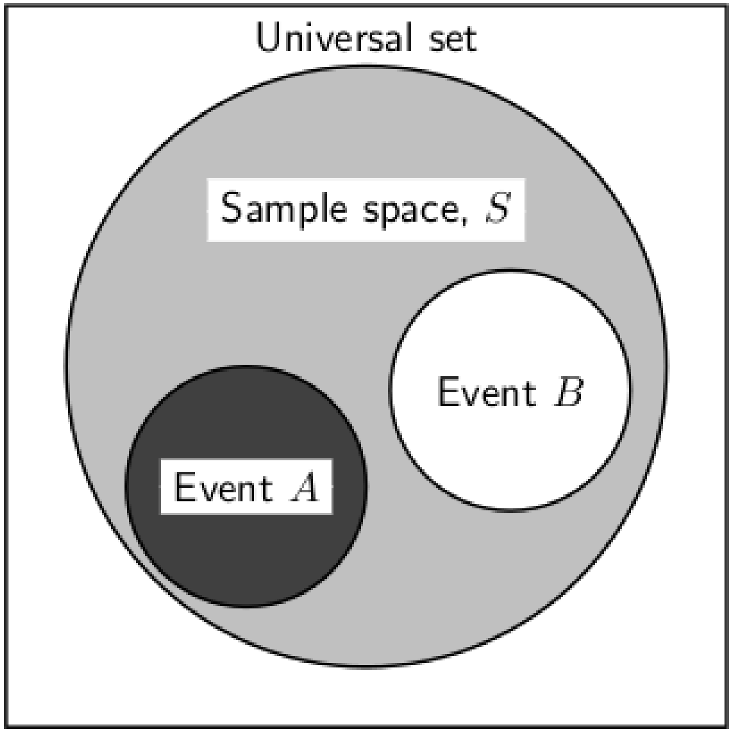


Basic Concepts of Probability

Remark:

- Conceptually speaking

an event is equivalent to a set.



Basic Concepts of Probability

Example 6: [**Rolling a Die**]

Event *A* is defined as "the number resulting is even".

Event *B* is "the number resulting is at least 4".

$A = \{2,4,6\}$

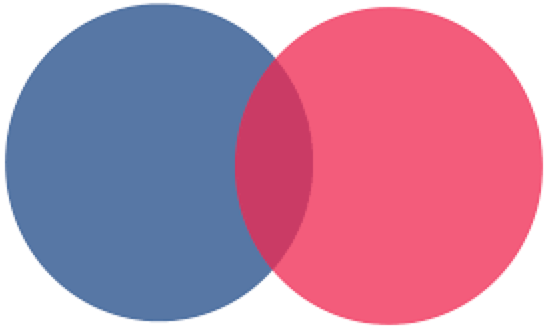
$B = \{4,5,6\}$

Remark: Basic outcome ⊆ Event ⊆ Sample space.

CONTENTS

Review of Set Theory

Geometric representation of sets and their operations: **Venn Diagram**.



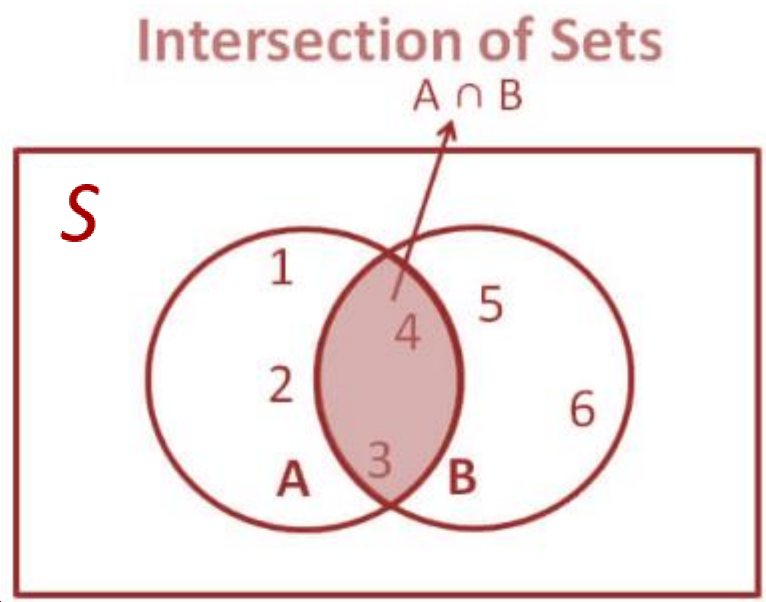
Venn diagram can be used to depict a sample point, a sample space, an event, and related concepts.

Review of Set Theory

Definition 4. [**Intersection**]

Intersection of A and B, denoted $A \cap B$, is the set of basic outcomes in S that belong to both A and B.

The intersection occurs if and only if both events A and B occur.

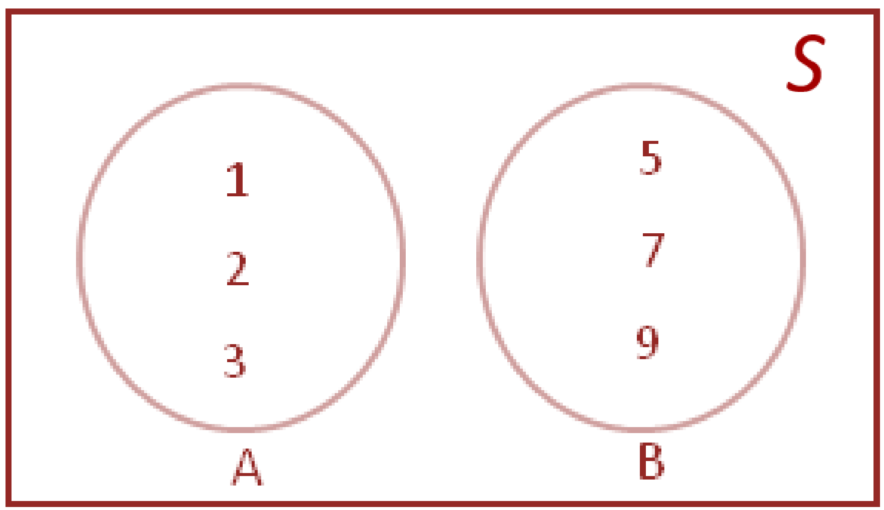


The intersection of A and B is also called the logical product.

Review of Set Theory

Definition 5. [**Exclusiveness**]

If A and B have no common basic outcomes, they are called mutually exclusive and their intersection is empty set ∅, i.e., $A \cap \mathrm{B} = \emptyset$, where ∅ denotes an empty set that contains nothing.

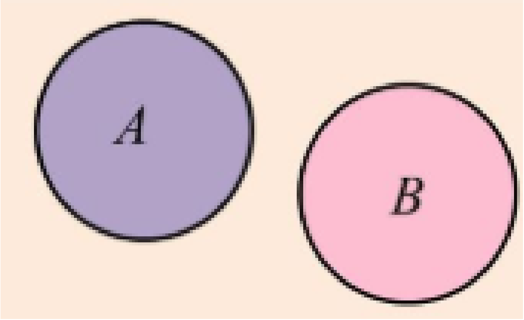


Disjoint Sets

Review of Set Theory

Remarks:

- Mutually exclusive events are also called *disjoint* because they do not overlap when represented in the Venn diagram.
- Any mutually exclusive events cannot occur simultaneously. As an example, any pair of the basic outcomes in sample space S are mutually exclusive.

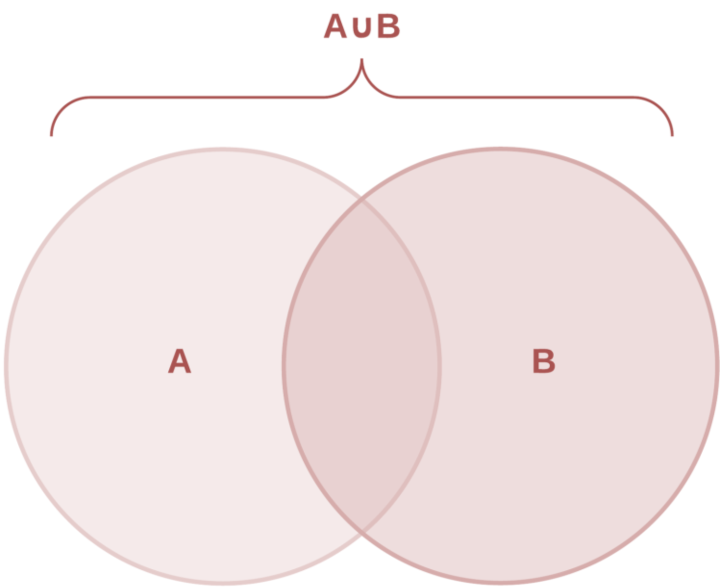


Review of Set Theory

Definition 6. [**Union**]

The union of A and B, $A \cup B$, is the set of all basic outcomes in S that belong to either A or B.

The union of A and B occurs if and only if either A or B (or both) occurs.



The union of A and B is also called the logical sum.

Review of Set Theory

Definition 7. [**Collective Exhaustiveness**]

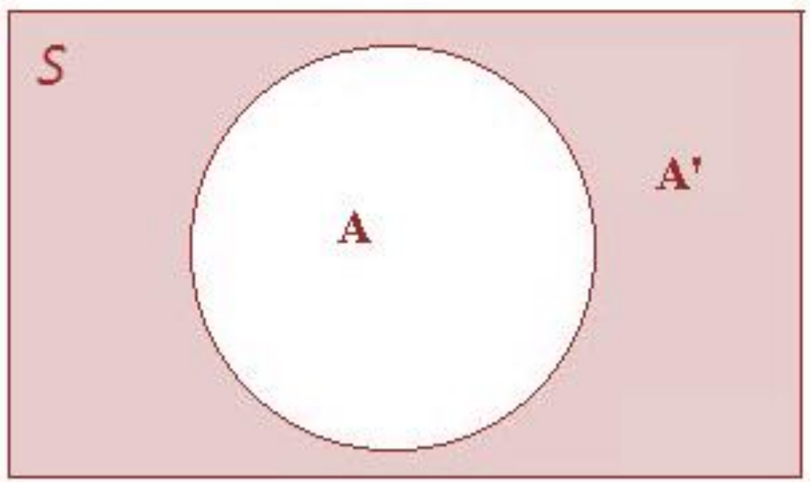
Suppose $A_1, A_2, \dots, A_n$ are n events in the sample space S, where n is any positive integer.

If $\bigcup_{i=1}^{n} A_i = S$, then these n events are said to be collectively exhaustive.

Review of Set Theory

Definition 8. [**Complement**]

The complement of A is the set of basic outcomes of a random experiment belonging to S but not to A, denoted as A^c.



Complement: A^C

Review of Set Theory

Remarks:

- The complement of event *A* is also called the negation of *A*.

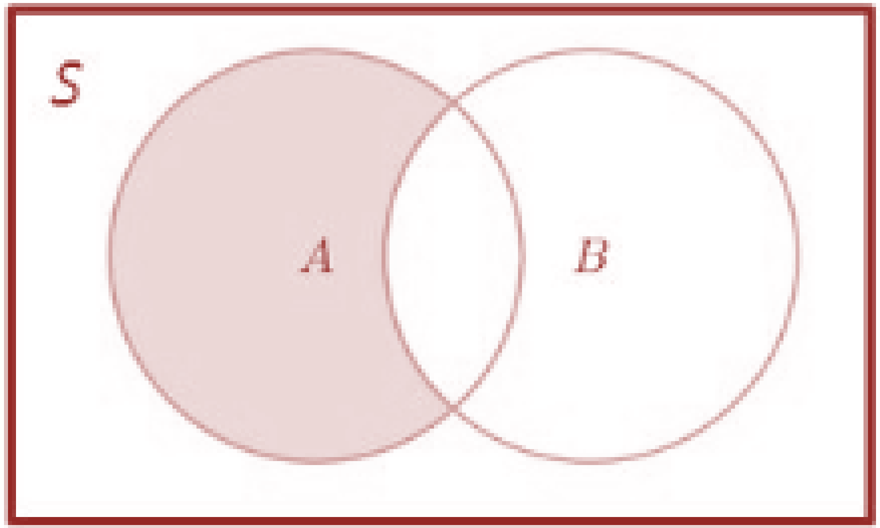
- Any event *A* and its complement A^c are mutually exclusive and collectively exhaustive. That is,

$$A \cap A^c = \emptyset \quad \text{and} \quad A \cup A^c = S.$$

Review of Set Theory

Definition 9. [**Difference**]

The difference of A and B, denoted as $A - B = A \cap B^c$, is the set of basic outcomes in S that belong to A but not to B.



Review of Set Theory

Example 7. [**Rolling a Die**]

The sample space $S = \{1,2,3,4,5,6\}$.

event A: the resulting number is even

event B: the resulting number is at least 4

Review of Set Theory

Then it follows that

$A = \{2,4,6\},$ $\quad B = \{4,5,6\},$

$A^c = \{1,3,5\},$ $\quad B = \{1,2,3\},$

$A - B = \{2\},$ $\quad B - A = \{5\},$

$A \cap B = \{4,6\},$ $\quad A \cup B = \{2,4,5,6\},$

$A \cap A^c = \emptyset\ ,$ $\quad A \cup A^c = \{1,2,3,4,5,6\} = S.$

Review of Set Theory

Theorem 1. [**Laws of Sets Operations**]

For any three events *A, B, C* defined on a sample space *S*;

Complementation

$$
\begin{aligned}
(A^c)^c &= A \\
\emptyset^c &= S \\
S^c &= \emptyset
\end{aligned}
$$

Review of Set Theory

Commutativity of union and intersection

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

Associativity of union and intersection

$$(A \cup B) \cup C = A \cup (B \cup C)$$
$$(A \cap B) \cap C = A \cap (B \cap C)$$

Review of Set Theory

Distributivity laws

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

More generally, for any $n \geq 1$,

$$B \bigcap \left(\bigcup_{i=1}^{n} A_i \right) = \bigcup_{i=1}^{n} \left(B \bigcap A_i \right)$$
$$B \bigcup \left(\bigcap_{i=1}^{n} A_i \right) = \bigcap_{i=1}^{n} \left(B \bigcup A_i \right)$$

Review of Set Theory

De Morgan's laws

$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

More generally, for any $n \geq 1$,

$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c$$
$$\left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c x$$

Review of Set Theory

Example 8

Suppose the events A and B are disjoint.

Under what condition are A^c and B^c also disjoint?

Solution

A^c and B^c are disjoint if and only if $A \cup B = S$. This can be shown by De Morgan's law (please try it).

Review of Set Theory

Example 9

Answer the following questions:

- Are $A \cap B$ and $A^c \cap B$ mutually exclusive?
- Is $(A \cap B) \cup (A^c \cap B) = B$?
- Are A and $A^c \cap B$ mutually exclusive?
- Is $A \cup (A^c \cap B) = A \cap B$?

Review of Set Theory

Example 10

Let the set of events $\{A_i = 1, \ldots, n\}$ be mutually exclusive and collectively exhaustive, and let *A* be an event in *S*.

- Are $A_1 \cap A, \ldots, A_n \cap A$ mutually exclusive?
- Is the union of $A_i \cap A, i = 1, \ldots, n$, equal to *A*?

That is, do we have

$$\bigcup_{i=1}^{n} (A_i \cap A) = A?$$

Review of Set Theory

Remarks:

- A sequence of collectively exhaustive and mutually exclusive events forms a partition of sample space S.
- A set of collectively exhaustive and mutually exclusive events can be viewed as a complete set of orthogonal bases.
- A complete set of orthogonal bases can represent any event A in the sample space S, and $A_i \cap A$ could be viewed as the projection of event A on the base A_i.

CONTENTS

Fundamental Probability Laws

Motivations:

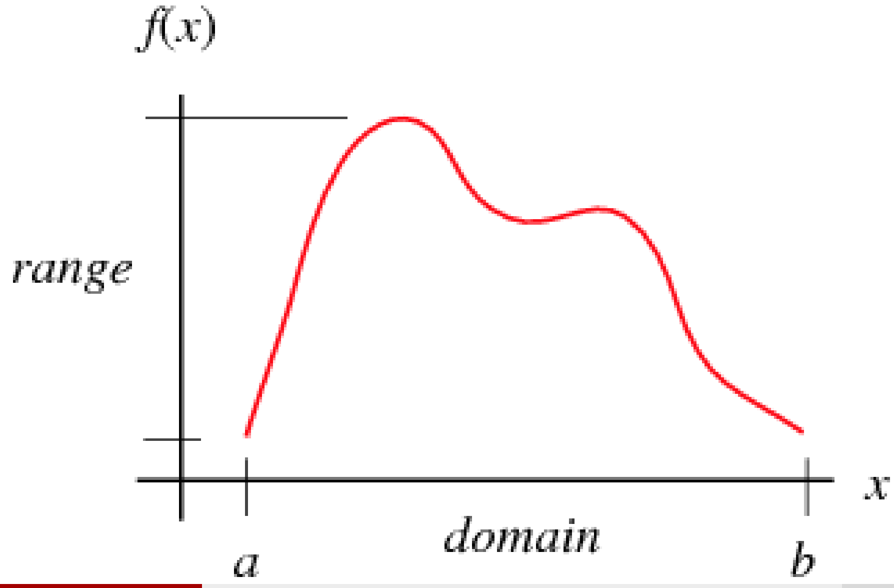
To assign a probability to an event *A* in *S*, we shall introduce a probability function, which is a function or a mapping from an event to a real number.

To assign probabilities to events, complements of events, unions and intersections of events, we want our collection of events to include all these combinations of events.

Fundamental Probability Laws

Motivations:

Such a collection of events is called a σ-field of subsets of the sample space S, which will constitute the domain of the probability function.



Fundamental Probability Laws

Definition 10. [**Sigma Algebra**]

A $sigma(\sigma)\ algebra$, denoted by $\mathbb{B}$, is a collection of subsets (events) of S that satisfies the following properties:

$\emptyset \in \mathbb{B}$ (i.e., the empty set is contained in $\mathbb{B}$)

If A $\in \mathbb{B}$, then $A^c \in \mathbb{B}$ (i.e., $\mathbb{B}$ is closed under countable complement)

If $A_1, A_2, \ldots \in \mathbb{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathbb{B}$ (i.e.,$\mathbb{B}$ is closed under countable unions)

Fundamental Probability Laws

Remarks:

- A σ-algebra is also called a σ-field.
- It is a collection of events in S that satisfies certain properties and constitutes the domain of a probability function.
- A σ-field is a collection of subsets in S, but itself is not a subset of S. In contrast, the sample space S is only an element of a σ-field .
- The pair, $(S, \mathbb{B})$, is called a measurable space.

Fundamental Probability Laws

Example 11

Show that for any sample space S, then set $\mathbb{B} = \{\emptyset, S\}$ is always a σ-field.

Solution

We verify the three properties of a σ-field :

- $\emptyset \in \{\emptyset, S\}$. Thus $\emptyset \in \mathbb{B}$.
- $\emptyset^c = S \in \mathbb{B}$ and $S^c = \emptyset \in \mathbb{B}$.
- $\emptyset \cup S = S \in \mathbb{B}$.

Fundamental Probability Laws

Example 12

Suppose the sample space $S = \{1,2,3\}$. Show that a set containing the following eight subsets {1}, {2}, {3}, {1,2}, {1,3}, {2,3}, {1,2,3}, and ∅ is a σ-field.

Fundamental Probability Laws

Example 13

Define $\mathbb{B}$ as the collection of all possible subsets (including an empty set $\emptyset$) in sample space S. Is $\mathbb{B}$ a σ-field ?

Fundamental Probability Laws

Definition 11. [**Probability Function**]

Suppose a random experiment has a sample space S and an associated σ-field $\mathbb{B}$. A probability function

$$P : \mathbb{B} \to [0,1]$$

is defined as a mapping that satisfies the following three properties:

(1) $0 \leq P(A) \leq 1$ for any event A in $\mathbb{B}$;

Condition (1) means that "everything is possible " or "any event is possible to happen".

Fundamental Probability Laws

Definition 11. [**Probability Function**]

(2) $P(S) = 1$;

Condition (2) means that " something always occurs whenever a random experiment is performed".

(3) If $A_1, A_2, \ldots \in \mathbb{B}$ are mutually exclusive, then

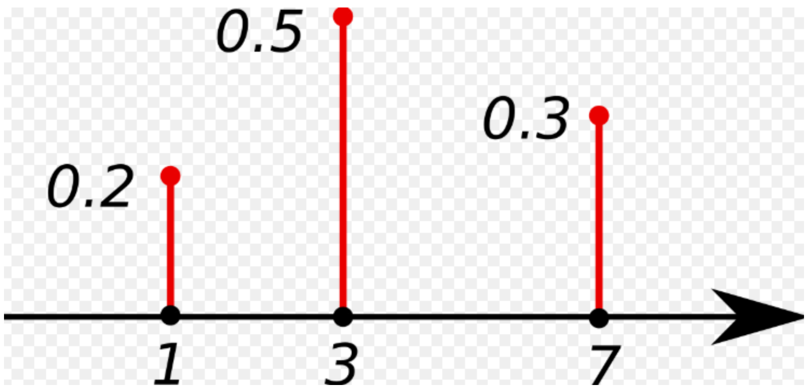
$$P(\cup_{i=1}^{\infty} A_i) = \Sigma_{i=1}^{\infty} P(A_i).$$

Condition (3) means that the probability of the "sum (i.e., union)" of exclusive events is equal to the sum of their individual probabilities.

Fundamental Probability Laws

Remarks:

- A probability function tells how the probability of occurrence is distributed over the set of events $\mathbb{B}$. In this sense we speak of a distribution of probabilities.



Fundamental Probability Laws

Remarks:

- For a given measurable space $(S, \mathbb{B})$, many different probability functions can be defined. The goal of econometrics and statistics is to find a probability function that most accurately describes the underlying DGP. This probability function is usually called the true probability function or true probability distribution model.

Interpretation of Probability

Question:

How to interpret the probability of an event?

Approach 1: Relative Frequency Interpretation

The probability of an event can be viewed as the limit of the " relative frequency" of occurrence of the event in a large number of repeated independent experiments under essentially the same conditions.

Colour	Frequency	Relative Frequency
blue	5	5/20 = 0.25
green	4	4/20 = 0.2
pink	6	6/20 = 0.3
red	3	3/20 = 0.15
orange	2	2/20 = 0.1
Totals	20	1

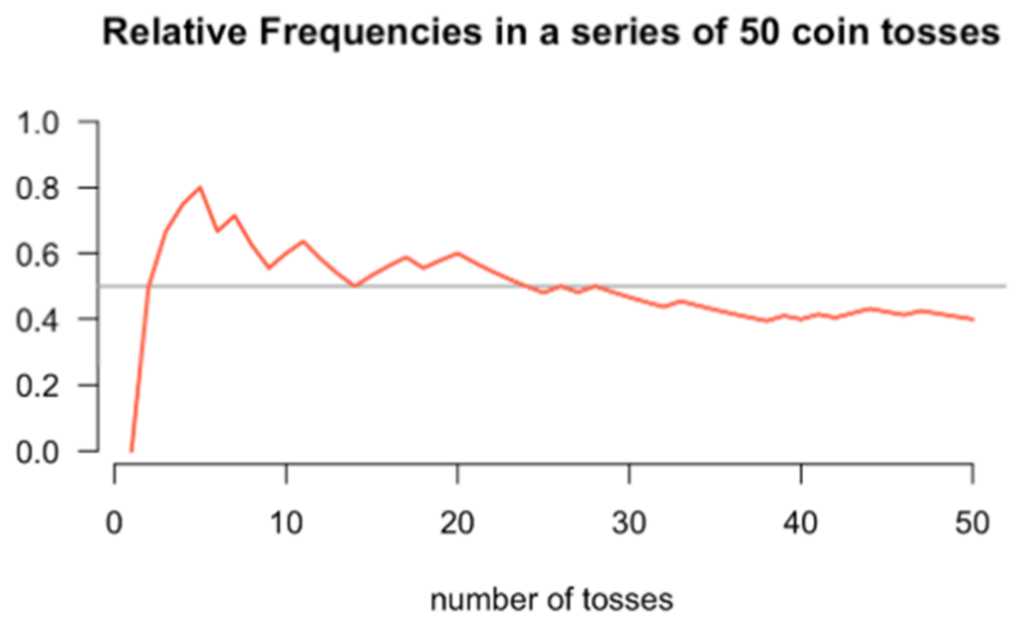
Approach 1: Relative Frequency Interpretation

For example, suppose we throw a coin for a total of N times. Each time, either "Head" or "Tail" occurs.

- Suppose that among the N trials, "Head" shows up N_h times. Then the proportion of occurrences of "Head" in the N trials is N_h/N.
- When $N \to \infty$, there will be little variation for the ratio N_h/N. This relative frequency will approach the probability of the event that "Head" occurs.

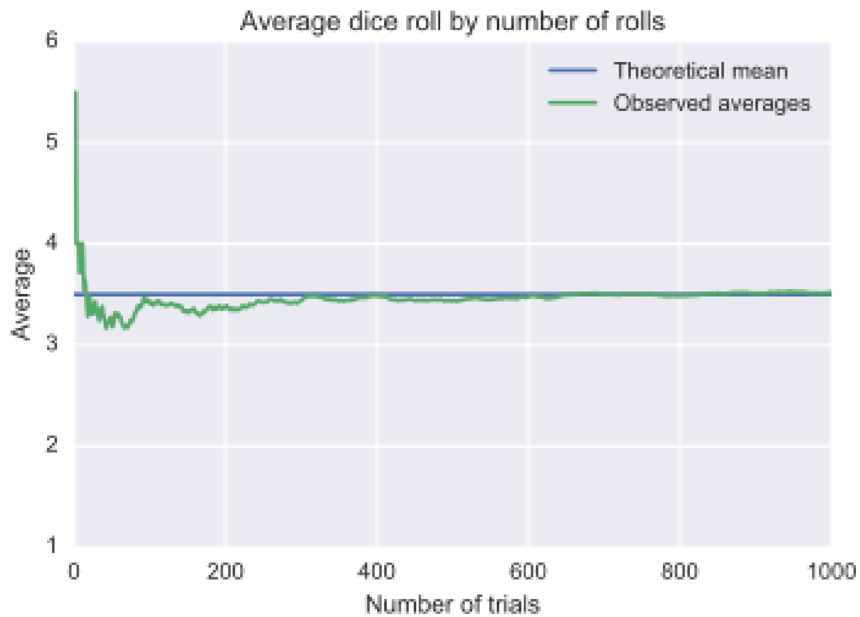
Approach 1: Relative Frequency Interpretation

The frequency interpretation views that the probability of an event is the proportion of the times that independent events of the same kind occur in the long run.



Approach 1: Relative Frequency Interpretation

The relative frequency interpretation is valid under the assumption of a large number of repeated experiments under the same condition.
In statistics, such an assumption is formally termed as "independence and identical distribution (IID)".



Approach 1: Relative Frequency Interpretation

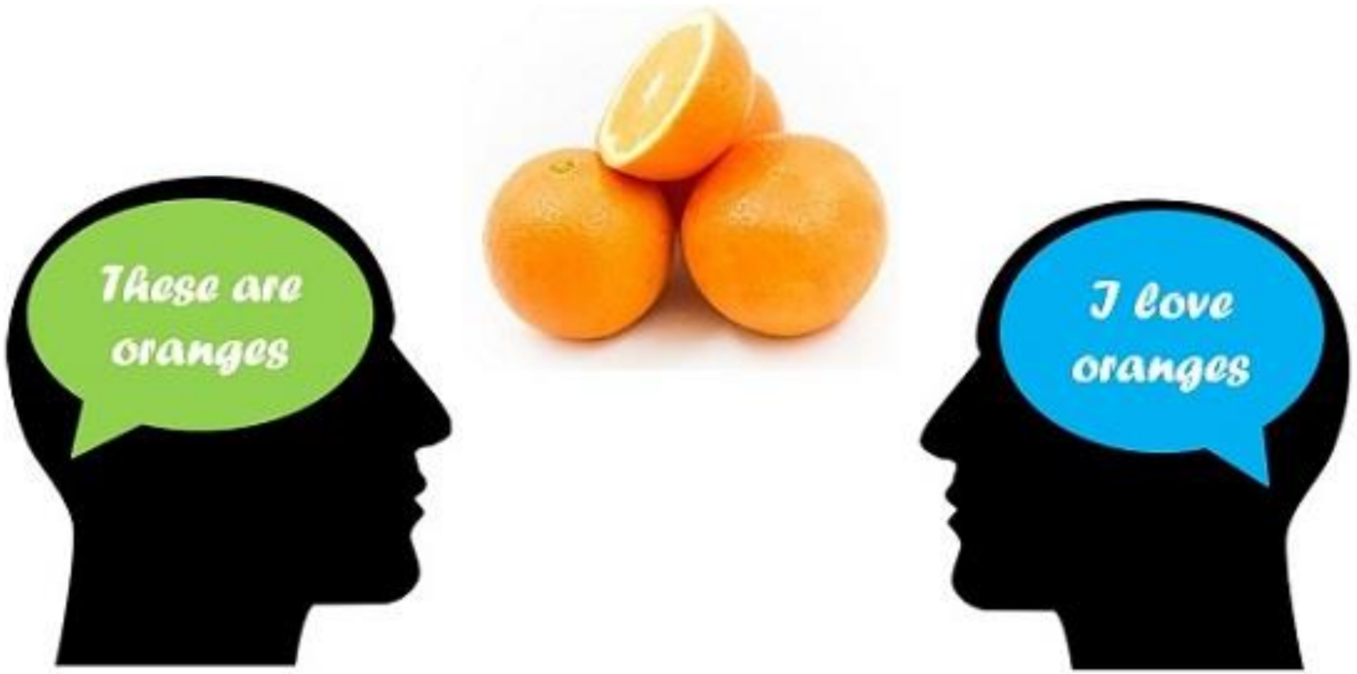
Example 14

When the weather forecast bureau predicts that there is a 30% chance for raining, it means that under the same weather conditions it will rain 30% of the times. We cannot guarantee what will happen on any particular occasion, but if we keep records over a long period of time, we should find that the proportion of "raining" is very close to 0.30 for the days with the same weather condition.

	M	T	W	TH	F	S	S
Chance of rainfall	70%	80%	90%	80%	60%	20%	0%

Approach 2: Subjective Probability Interpretation

The subjective method views probability as a belief in the chance of an event occurring.



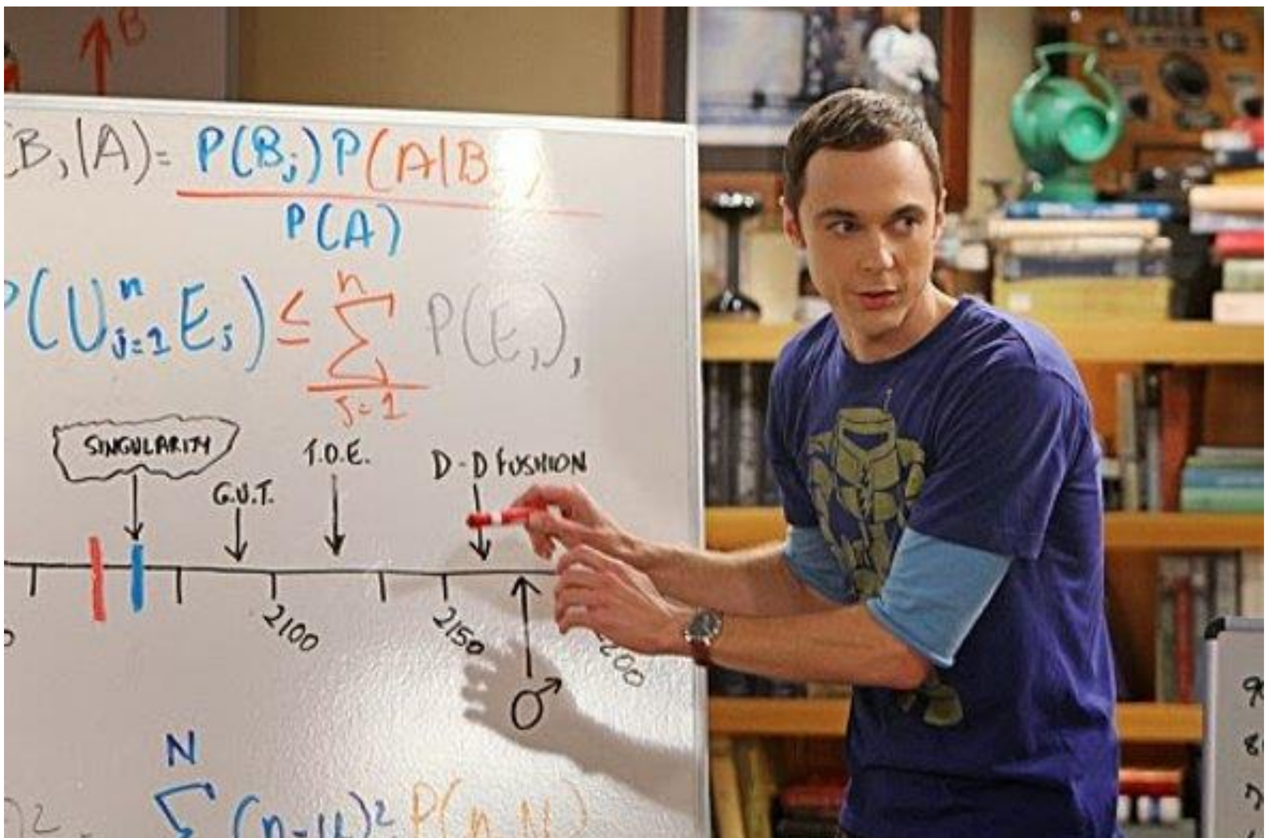
Approach 2: Subjective Probability Interpretation

A personal or subjective assessment is made of the probability of an event which is difficult or impossible to estimate in any way.

For example, the probability that S&P500 price index will go up in a given future period of time cannot be estimated very well by using the frequency interpretation, because economic and world conditions rarely replicate themselves closely.

Approach 2: Subjective Probability Interpretation

Subjective probability is the foundation of Bayesian statistics, which is a rival to classical statistics.



Approach 2: Subjective Probability Interpretation

Example 15. [Rational Expectations]

Rational expectations (Muth 1961) hypothesizes that the subjective expectation of an economic agent (i.e., the expectation under the subjective probability belief of the economic agent) coincides with the mathematical expectation (i.e., the expectation under the objective probability distribution).

Approach 2: Subjective Probability Interpretation

Example 16. [**Professional Forecasts**]

The U.S. central bank—Fed issues professional forecasts for important macroeconomic indicators such as GDP growth rate, inflation rate and unemployment rate.

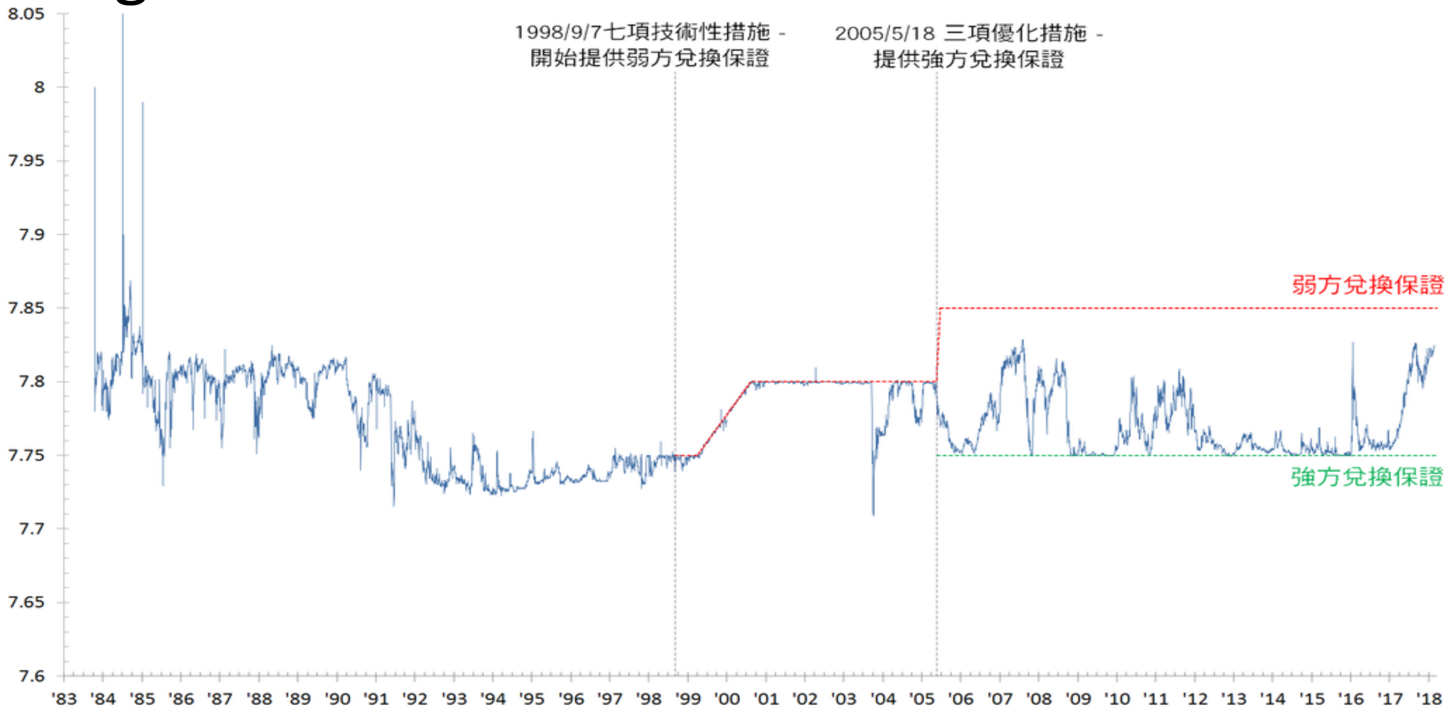
In each quarter, they send surveys to professional forecasters, asking their views on probability distributions of these important macroeconomic indicators.

Specifically, each forecaster will be asked what is his/her forecast of the probability that the inflation rate lies in various intervals.

Approach 2: Subjective Probability Interpretation

Example 17. [**Risk Neutral Probability**]

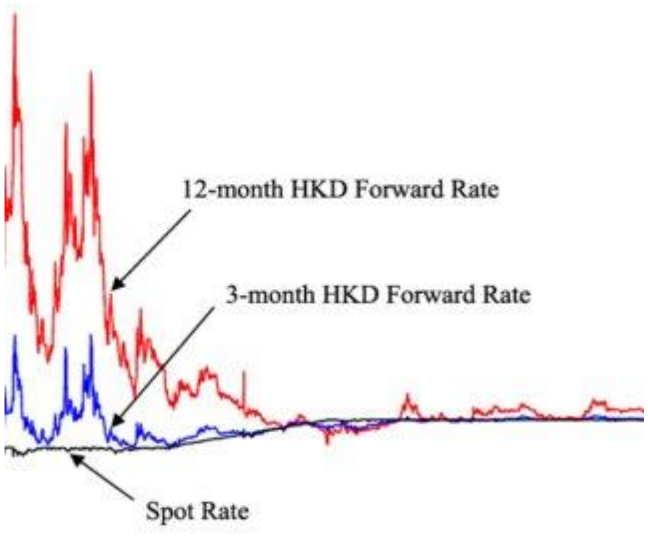
During the 1997-1998 Asian financial crisis, many investors were very concerned with the collapse of the Hong Kong peg exchange rate system with U.S. dollars and devaluations of Hong Kong dollars.



Approach 2: Subjective Probability Interpretation

Example 17. [**Risk Neutral Probability**]

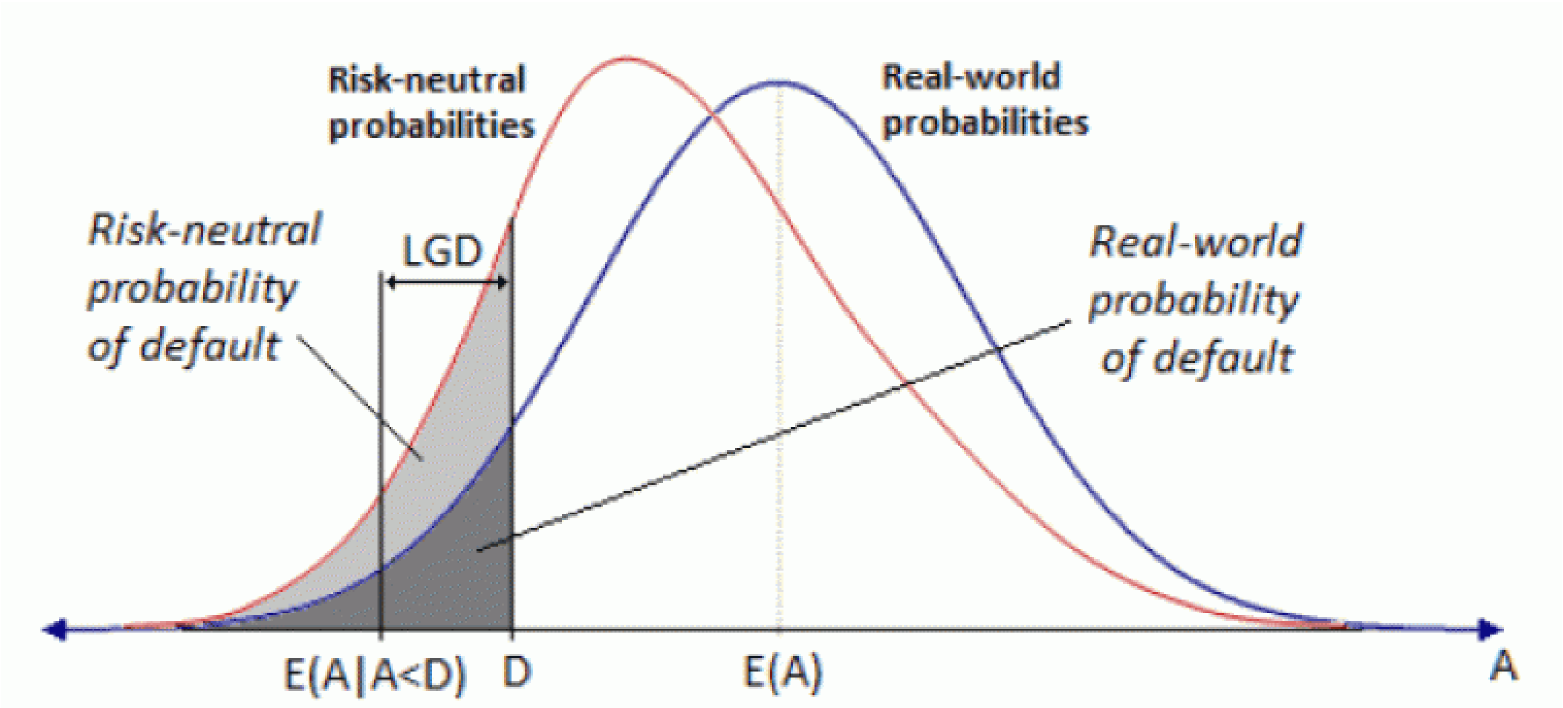
In other words, their subjective probabilities of Hong Kong dollar devaluation were higher than the objective probabilities of the Hong Kong dollar movements. The former are called risk-neutral probability distributions and the latter are called objective or physical probability distributions in finance.



Approach 2: Subjective Probability Interpretation

Example 17. [**Risk Neutral Probability**]

The gap between these two distributions reflects the risk attitude of market investors. The risk-neutral probability distribution is a financial instrument in derivative pricing.



Approach 2: Subjective Probability Interpretation

Example 18. [**Allais' Paradox**]

In experimental economics, suppose a set of prizes is $X = \{\$0,\ \$1{,}000{,}000,\ \$5{,}000{,}000\}$.

(1) Which probability distribution do you prefer: $P_1 =$ (0.00, 1.00, 0.00) or $P_2 = (0.01, 0.89, 0.10)$?

(2) Which probability do you prefer: $P_3 =$ (0.90, 0.00, 0.10) or $P_4 =$ (0.89, 0.11, 0.00)?

Approach 2: Subjective Probability Interpretation

Example 18. [**Allais' Paradox**]

Many subjects in the experiment report that they prefer P_1 over P_2, and P_3 over P_4. This is inconsistent with the well-known expected utility theory in microeconomics. Obviously, individuals tend to overweight low-probability events and underweight high-probability events.

Formally, suppose there is a prospect of payoff $\{(x_1, p_1), (x_2, p_2), \ldots (x_n, p_n)$ with $x_1 > \cdots > x_n$, where x_i is the payoff in state i and p_i is the probability of state i.

Approach 2: Subjective Probability Interpretation

Example 18. [**Allais' Paradox**]

We define a rank-dependent weighting:

$$\pi_i = w\left(\sum_{j=1}^{i} p_j\right) - w\left(\sum_{j=1}^{i-1} p_j\right),$$

where
w: $[0,1] \rightarrow [1,0]$ *is a strictly increasing and continuous weighting function, with* $w(0) = 0$ *and* $w(1) = 1$. *Then the value of the prospect is characterized as* $\sum_{i=1}^{n} \pi_i x_i$.

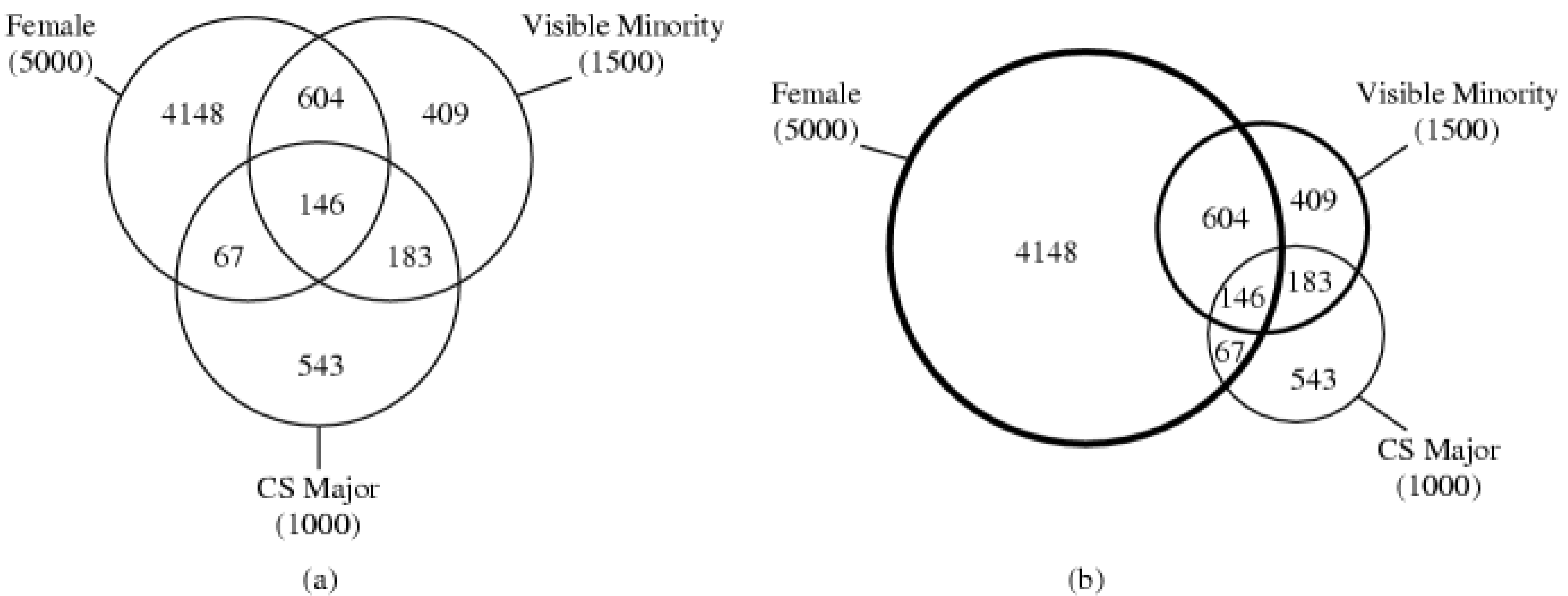
Approach 2: Subjective Probability Interpretation

Example 18. [**Allais' Paradox**]

The rank-dependent weightings $\{\pi_i\}_{i=1}^{n}$ can reasonably be interpreted as subjective probabilities. This is pretty much like the way we interpret the prices of the Arrow securities as subjective probabilities.

Approach 2: Subjective Probability Interpretation

Geometrically, based on the Venn diagram, the probability of any event *A* in sample space *S* can be viewed as equal to the area of event *A* in *S*, with the normalization that the total area of *S* is equal to unity.



Basic Probability Laws

Definition 12. [Probability Space]

A probability space is a triple $(S, \mathbb{B}, P)$ where:

- S is the sample space corresponding to the outcomes of the underlying random experiment.
- $\mathbb{B}$ is the σ-field of subsets of S. These subsets are called events.
- $P: \mathbb{B} \to [0,1]$ is a probability measure.

Basic Probability Laws

Remark:

A probability space $(S, \mathbb{B}, P)$ completely describes a random experiment associated with sample space S.

Because the probability function $P(\cdot)$ is defined on $\mathbb{B}$, the collection of sets (i.e., events), it is also called a set function.

Basic Probability Laws

Theorem 2

If $\emptyset$ denotes the empty set, then $P(\emptyset) = 0$.

Proof:

Given that $S = S \cup \emptyset$, and S and $\emptyset$ are mutually exclusive, we have

$P(S) = P(S \cup \emptyset) = P(S) + P(\emptyset)$.

It follows that $P(\emptyset) = 0$.

Basic Probability Laws

Remarks:

- Theorem 2 means that it is unlikely that nothing occurs when a random experiment is implemented.
- In other words, something always occurs when a random experiment is implemented.

Question:

Does $P(A) = 0$ implies $A = \emptyset$?

Basic Probability Laws

Theorem 3

$$P(A) = 1 - P(A^c).$$

Proof:

Observe $S = A \cup A^c$. Then

$$P(S) = P(A \cup A^c).$$

Because $P(S) = 1$

A and A^c are mutually exclusive,

we have

$$1 = P(A) + P(A^c).$$

Basic Probability Laws

Remarks:

The ratio of the probability of an event *A* to the probability of its complement,

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

is called the ratio of *odds*.

Basic Probability Laws

Example 19

Suppose X denotes the outcome of some random experiment. The following is the probability distribution for X, namely, the probability that X takes various values:

$$P(X = i) = \frac{1}{2^i}, \qquad i = 1, 2, \cdots.$$

Find the probability that X is larger than 3.

Basic Probability Laws

Solution

The sample space $S = \{1,2,\ldots\}$. Let A be the event that $X > 3$. Then $A = \{4,5,\cdots\}$. It follows that

$$\begin{aligned} P(A) &= P(X > 3) \\ &= P(X=4) + P(X=5) + P(X=6) + \cdots \\ &= \sum_{i=4}^{\infty} P(X=i) \\ &= \sum_{i=4}^{\infty} \frac{1}{2^i} X \end{aligned}$$

To be Continued

Basic Probability Laws

Direct calculation of this infinite sum may be a bit tedious. Instead, we can apply Theorem 3 and compute

$$
\begin{aligned}
P(A) &= 1 - P(A^c) \\
&= 1 - P(X \leq 3) \\
&= 1 - [P(X = 1) + P(X = 2) + P(X = 3)] \\
&= 1 - \left(\frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} \right) \\
&= \frac{1}{8}
\end{aligned}
$$

Basic Probability Laws

Theorem 4

If A and B are two events in $\mathbb{B}$, and $A \subseteq B$, then

$$P(A) \leq P(B).$$

Proof:

Using the fact that $S = A \cup A^c$ and the distributive law, we have

$$\begin{aligned} B &= S \cap B = (A \cup A^c) \cap B \\ &= (A \cap B) \cup (A^c \cap B) \\ &= A \cup (A^c \cap B), \end{aligned}$$

To be Continued

Basic Probability Laws

where the last equality follows from $A \subseteq B$ so that $A \cap B = A$. Because A and $A^c \cap B$ are mutually exclusive, we have

$$\begin{aligned} P(B) &= P(A) + P(A^c \cap B) \\ &\geq P(A) \end{aligned}$$

given that $P(A^c \cap B) \geq 0$.

Basic Probability Laws

Corollary 1

For any event $A \in \mathbb{B}$ such that

$$\emptyset \subseteq A \subseteq S,$$

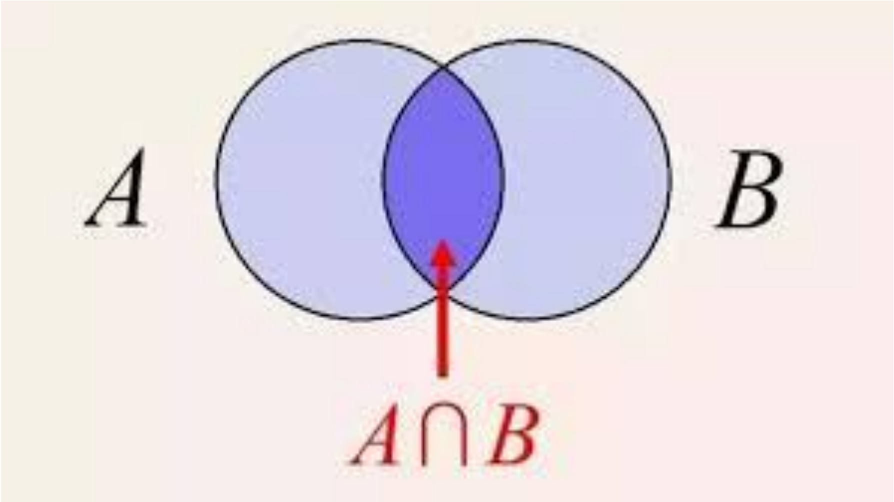
$$0 \leq P(A) \leq 1.$$

Basic Probability Laws

Theorem 5

For any two events A and B in $\mathbb{B}$,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$



Basic Probability Laws

Proof: Since $A \cup B = A \cup (A^c \cap B)$, and A and $A^c \cap B$ are mutually exclusive, we have

$$P(A \cup B) = P(A) + P(A^c \cap B). \tag{1}$$

On the other hand, because $B = S \cap B = (A \cap B) \cup (A^c \cap B)$, and both $A \cap B$ and $A^c \cap B$ are mutually exclusive, we have

$$P(B) = P(A \cap B) + P(A^c \cap B). \tag{2}$$

Adding both Equations (1) and (2) yields

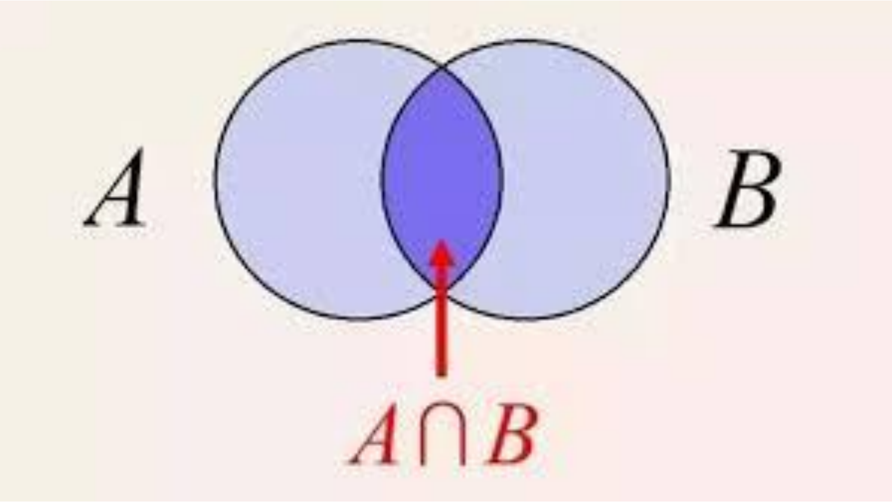
$$\begin{aligned} &P(A \cup B) + P(A \cap B) + P(A^c \cap B) \\ &= P(A) + P(A^c \cap B) + P(B). \end{aligned}$$

This delivers the desired result.

Basic Probability Laws

Remark:

Theorem 5 can be illustrated via a Venn diagram, keeping in mind that the probability of an event is equal to the area it occupies in the sample space S.



Basic Probability Laws

Example 20. [**Bonferroni's Inequality**]

Show $P(A \cup B) \geq P(A) + P(B) - 1$.

Solution

Since $A \cap B \subseteq S$, we have $P(A \cap B) \leq P(S) = 1$ by Theorem 4. It follows from Theorem 5 that

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &\geq P(A) + P(B) - 1. \end{aligned}$$

Basic Probability Laws

Example 21

Suppose there are two events A and B in S, with $P(A) = 0.20$, $P(B) = 0.30$ and $P(A \cap B) = 0.10$. Then

- Are A and B mutually exclusive?
- $P(A^c) =?$ $P(B^c) =?$
- $P(A \cup B) =?$
- $P(A^c \cup B^c) =?$
- $P(A^c \cap B^c) =?$

Basic Probability Laws

Theorem 6. [Rule of Total Probability]

If $A_1, A_2, \cdots \in \mathbb{B}$ are mutually exclusive and collectively exhaustive, and A is an event in S, then

$$P(A) = \sum_{i=1}^{\infty} P(A \cap A_i).$$

Basic Probability Laws

Proof:

Noting $S = \cup_{i=1}^{\infty} A_i$ and

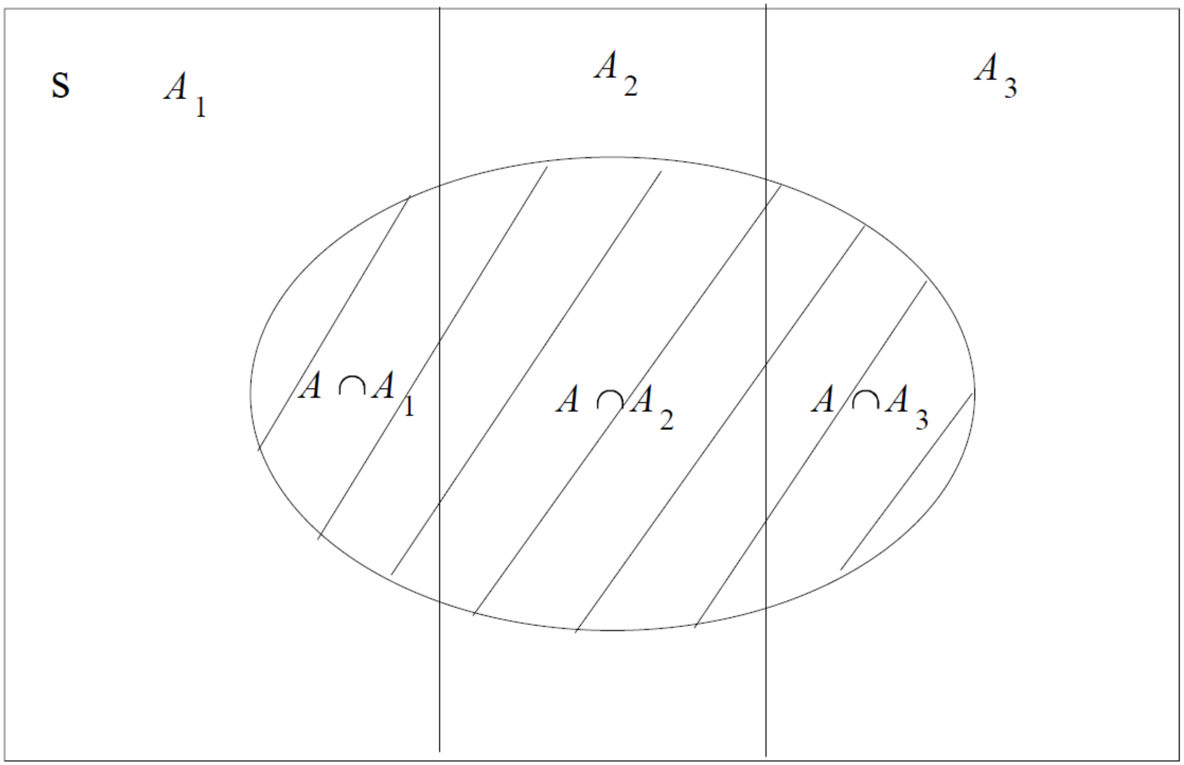
$A = A \cap S = A \cap \cup_{i=1}^{\infty} A_i) = \cup_{i=1}^{\infty} (A \cap A_i)$,

where the last equality follows by the distributive law. The result follows because $A \cap A_i$ and $A \cap A_j$ are disjoint for all $i \neq j$.

Basic Probability Laws

Remarks:

- The rule of total probability can be illustrated clearly in a Venn diagram (with $n = 3$, see Figure below):



Basic Probability Laws

Remarks:

- With a set of mutually exclusive and collectively exhaustive events $A_1, \dots, A_n$ any event A can be represented as the union of the mutually exclusive intersections $A \cap A_1, \cdots, A \cap A_n$. As a result, the probability of A is equal to the sum of the probabilities of these intersections.

Basic Probability Laws

Example 22

If

$$A = \{\text{students whose scores} > 90 \text{ points}\},$$
$$A_i = \{\text{students from country } i\},$$

then

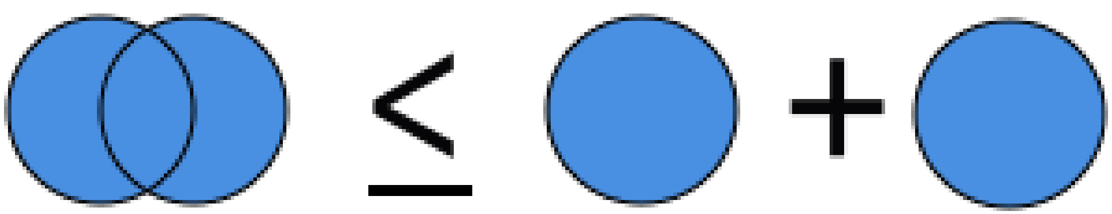
$$A \cap A_i = \{\text{students from country } i \text{ whose scores are} > 90 \text{ points}\}.$$

Basic Probability Laws

Theorem 7. [**Subadditivity: Boole' Inequality**]

For any sequence of events $\{A_i \in \mathbb{B}, i = 1,2,\cdots\}$,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i).$$



Basic Probability Laws

Proof:

Put $B = \cup_{i=2}^{\infty} A_i$. Then $\cup_{i=1}^{\infty} A_i = A_1 \cup B$. It follows that

$$
\begin{aligned}
P\left(\bigcup_{i=1}^{n} A_i\right) &= P(A_1 \cup B) \\
&= P(A_1) + P(B) - P(A_1 \cap B) \\
&\leq P(A_1) + P(B),
\end{aligned}
$$

To be Continued

Basic Probability Laws

where the inequality follows given $P(A_1 \cap B) \geq 0$. Again, put $C = \cup_{i=3}^{\infty} A_i$. Then

$$P(B) = P(A_2 \cup C) \leq P(A_2) + P(C).$$

It follows that

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq P(A_1) + P(A_2) + P(C).$$

Repeating this process, we have

$$P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i).$$

Basic Probability Laws

Remarks:

- The probability of the "sum (i.e., union)" of events is less than the sum of their individual probabilities.
- Equality occurs only when all events are mutually exclusive.
- Whenever there are overlapping events, the probability of total union will be strictly smaller than the sum of individual probabilities.

CONTENTS

Methods of Counting

How to calculate the probability of event *A*?

Suppose event A includes k basic outcomes $A_1, \cdots, A_k$ in sample space S. Then

$$P(A) = \sum_{i=1}^{k} P(A_i)$$

Methods of Counting

If in addition S consists of n equally likely basic outcomes $A_1, \cdots, A_n$, and event A consists of k basic outcomes, then

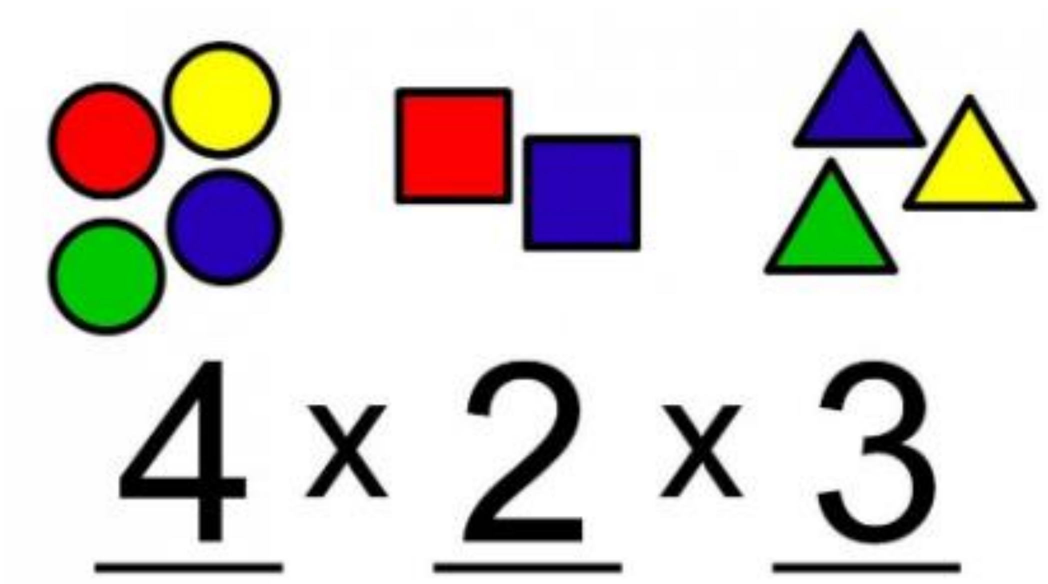
$$P(A) = \frac{k}{n}$$

Calculation of probability for event A boils down to the counting of the numbers of basic outcomes in event A and in sample space S.

Methods of Counting

Theorem 8. [Fundamental Theorem of Counting]

If a random experiment consists of k separate tasks, the i^{th} of which can be done in n_i ways, $i = 1,2,\cdots,k$, then the entire job can be done in $n_1 \times n_2 \times\cdots\times n_k = \Pi_{i=1}^{k} n_i$ ways.



Permutations

Example 23

Suppose we will choose two letters from four letters $\{A, B, C, D\}$ in different orders, with each letter being used at most once each time. How many possible orders could we obtain?

Solution

There are 12 ways:
$AB, BA, AC, CA, AD, DA, BC, CB, BD, DB, CD, DC$

Permutations

Example 24

How many different ways to choose 20 letters from the 26 letters?

Permutations

Example 24

A general problem: Suppose there are x **boxes** arranged in row and there are n distinguishable **objects**, where $x \leq n$. We will choose x from the n objects to place them in the x boxes. Each object can be used at most once in each arrangement (i.e., no replacement is allowed).

How many possible different **sequences** could we obtain? That is, how many different ways can we fill the x boxes?

Permutations

Example 24

- First, how many ways can we fill Box 1? There are n objects available, so there are n ways.
- Second, suppose we have filled Box 1. How many different ways can we fill Box 2?
- Because there has been one object used to fill box 1, $n-1$ objects remain and each of these $n-1$ objects can be used to box 2. Therefore, there are $(n-1)$ ways to fill box 2.

Permutations

Example 24

- Third, suppose we have filled the first two boxes. Then there are $n-2$ ways to fill the Box 3.

- $\cdots$

- For the last box (i.e. Box x), given that $x-1$ objects have been used to fill the first $x-1$ boxes, there are $[n-(x-1)]$ objects left, so there are $[n-(x-1)]$ ways to fill the last box.

Permutations

Example 24

The total number of possible orderings of choosing x out of n objects is

$$P_n^x = \frac{n!}{(n-x)!}$$

where the notation $k!$ is called a "k factorial":

$$k! = k \times (k-1) \times \cdots \times 2 \times 1$$

By convention, $0! = 1$.

Permutations

Example 25

A company has six sales representatives and has the following incentive scheme. It decides that the most successful representative during the previous year will be awarded a January vacation in Hawaii, while the second most successful representative will win a vacation in Las Vegas. The other representatives will be required to take a course on probability and statistics.

How many different outcomes are possible?

Permutations

Solution

Ordering matters here because who goes to Hawaii and who goes to Las Vegas will be considered as different outcomes. Thus, we use permutation: with n = 6, x = 2, we have

$$P_n^x = \frac{6!}{(6-2)!} = 30$$

Permutations

Example 26. [**Birthday Problem**]

Suppose there are k students in a class, where $2 \leq k \leq 365$. What is the probability that at least two students have the same birthday? Here, by the same birthday, we mean the same day of the same month, but not necessarily of the same year. Moreover, we make the following assumptions:

- *No twins in the class;*
- *Each of the 365 days is equally likely to be the birthday of anyone in the class;*
- *Anyone born on Feb. 29 will be considered as on March 1.*

Permutations

Solution

First, how many possible ways in which the whole class could be born? This is a problem of ordering with replacement:

$$365^k = 365 \times 365 \times \cdots \times 365$$

where each student has 365 days to be born: This is the total number of basic outcomes in the sample space S.

To be Continued

Permutations

Second, the event *A* that at least 2 students have the same birthday is complement to the event A^c that all k students have different birthdays.

How many ways that k students can have different birthdays? This is a problem of choosing k different days out of 365 days to k students: By permutation, this number is

$$\frac{365!}{(365-k)!}$$

To be Continued

Permutations

Therefore,

$$
\begin{aligned}
P(A) &= 1 - P(A^c) \\
&= 1 - \frac{365!/(365-k)!}{365^k}
\end{aligned}
$$

Remark:

probabilities for various class sizes:

k	20	30	40	50
$P(A)$	0.411	0.706	0.891	0.970

Combinations

Example 27

Suppose now we will choose two letters from four letters $\{A, B, C, D\}$. Each letter is used at most once in each arrangement but now we are not concerned with their ordering. In other words, we are choosing a set that contains two different letters.

How many possible such sets could we obtain?

Solution

There are six sets that contain two different letters:

$$\{A, B\}, \{A, C\}, \{A, D\}, \{B, C\}, \{B, D\}, \{C, D\}$$

To be Continued

Combinations

A general problem: suppose we are interested in the number of different ways of choosing x objects from n objects but are not concerned about the order of the selected x objects. Here, each object can be used at most once in each arrangement. How many subsets containing x distinct objects can we obtain?

To be Continued

Combinations

We consider the following formula:

- # of choosing x from n objects with ordering = # of choosing x from n objects without ordering × # of ordering x objects.
- This implies
$$\frac{n!}{(n-x)!} = C_n^x \times x!$$
- It follows that the number of combinations of choosing x from n without ordering is
$$C_n^x = \binom{n}{x} = \frac{n!}{x!\,(n-x)!}.$$

Combinations

Lemma 1 (2.9). [Properties of Combinations]

(1) $\binom{n}{k} = \binom{n}{n-k}$;

(2) $\binom{n}{1} = n$;

(3) $\binom{n}{k} = \frac{P_n^k}{k!}$.

Combinations

Example 28

A personnel officer has 8 candidates to fill 4 positions. Five candidates are men, and three are women. If in fact, every combination of candidates is equally likely to be chosen, what is the probability that no women will be hired?

Combinations

Solution

Since we do not care who will fill a specific position, ordering does not matter.

(1) How many ways to select 4 out of the total of 8 candidates? The total number of possible combinations of choosing 4 out of 8 candidates is

$$C_8^4 = \frac{8!}{4!\,4!} = 70$$

To be Continued

Combinations

(2) When no women is hired, the four successful candidates must come from the available five men. The number of choosing 4 out of 5 male candidates is

$$C_5^4 C_3^0 = \frac{5!}{4!\,1!} = 5$$

To be Continued

Combinations

Therefore, the probability that no women will be hired is

$$P(A) = \frac{C_5^4 C_3^0}{C_8^4} = \frac{5}{70} = \frac{1}{14}$$

Combinations

Example 29

Suppose a class contains 15 boys and 30 girls, and 10 students will be selected randomly to form a team. Here, by "randomly" we mean that in each case, all possible selections are equally likely.

What is the probability that exactly 3 boys will be selected?

Combinations

Solution

(1) How many ways to form a term with 10 members: It is given by

$$C_{45}^{10} = \frac{45!}{10!\,35!}$$

To be Continued

Combinations

(2) How many ways that exactly 3 boys (and therefore 7 girls) will be selected? Here, we choosing 3 out of 15 boys and 7 out of 30 girls. The number is $C_{15}^{3}C_{30}^{7}$.

(3) It follows that the probability that exactly 3 boys will be selected is $P(A)=\frac{C_{15}^{3}C_{30}^{7}}{C_{45}^{10}}=0.2904$.

Combinations

Example 30

A manager has four assistants—John, George, Mary and Jean to assign to four tasks. Each one will be assigned to one task.

(1) How many different arrangements of assignments will be possible?

(2) What is the probability that Mary will be assigned to a specific task?

Combinations

Solution

We shall use the permutation method.

(1) There are $P_4^4 = 4! = 24$ different arrangements in total.

To be Continued

Combinations

(2) If Mary is assigned to a specific task, the manager has to arrange the other three candidates to the remaining 3 tasks. There are a total of $P_3^3 = 3!$ =6 different ways for the managers to make such arrangements. It follows that the probability that Mary will be assigned to a specific task is given by

$$P(A) = \frac{P_3^3}{P_4^4} = \frac{6}{24} = \frac{1}{4}$$

Combinations

Example 31

Suppose a team of 12 people is selected in a random manner from a group of 100 people. Determine the probability that two particular persons A and B will be selected.

Combinations

Solution

(1) How many ways to form the team?

There are C_{100}^{12}.

(2) Suppose two persons, say John and Tom, are included in the team, how many ways to select the other 10 members?

There C_{98}^{10}.

(3) Therefore, the probability that two particular persons, John and Tom, will be selected is given by

$$P(A) = \frac{C_{98}^{10}}{C_{100}^{12}}$$

Combinations

Example 32

The U.S. senate has 2 senators from each of the 50 states.

(1) If a committee of 8 senators is selected at random, what is the probability that it will contain at least one of the two senators from the New York state?

(2) What is the probability that a group of 50 senators selected at random will contain one senator from each state?

Combinations

Solution

(1) (a) How many ways to form a 8-member committee?

There are a total of C_{100}^{8} different ways.

To be Continued

Combinations

(b) Suppose A denotes the event that at least one of the 2 senators from the New York state will be selected. Then A^c is the event that no senator from the New York state will be selected. How many ways to select a 8-member committee if no senator from NY is considered?

There area total of C_{98}^{8} different ways.

It follows that

$$\begin{aligned} P(A) &= 1 - P(A^c) \\ &= 1 - \frac{C_{98}^{8}}{C_{100}^{8}} \end{aligned}$$

To be Continued

Combinations

(2) (a) How many ways to select a group of 50 senators?

There are a total of C_{100}^{50} different ways.

To be Continued

Combinations

(b) If the group contains one senator from each state, then there are 2 possible choices for each state. Thus, the total number of choosing one senator from each state is

$$2^{50} = 2 \cdot 2 \cdots 2$$

It follows that the probability that a group of 50 senators will contain one senator from each state is given by

$$P(A) = \frac{2^{50}}{C_{100}^{50}}$$

Combinations

Example 33

Choose an integer randomly from 1 to 2000. How many possible ways to choose an integer that can be divided exactly neither by 6 nor by 8?

Combinations

Solution

Define

A = {the integer that can be divided exactly by 6},

B = {the integer can be divided exactly by 8}.

Then by de Morgan's law

$$
\begin{aligned}
P(A^c \cap B^c) &= P[(A \cup B)^c] \\
&= 1 - P(A \cup B) \\
&= 1 - [P(A) + P(B) - P(A \cap B)]
\end{aligned}
$$

To be Continued

Combinations

Because

$$333 < \frac{2000}{6} < 334$$

then

$$P(A) = \frac{333}{2000}$$

Similarly,

$$P(B) = \frac{250}{2000}$$

To be Continued

Combinations

Moreover, an integer that can be divided exactly both by 6 and 8 is an integer that can be divided exactly by 24. Because

$$83 < \frac{2000}{24} < 84$$

we have

$$P(A \cap B) = \frac{83}{2000}$$

It follows that

$$P(A^c \cap B^c) = 1 - \left(\frac{333}{2000} + \frac{250}{2000} - \frac{83}{2000}\right) = \frac{3}{4}$$

Combinations

Example 34

Suppose we throw a fair coin 10 times independently.

(1) What is the probability of obtaining exactly three heads?

(2) What is the probability of obtaining three or fewer heads?

Combinations

Solution

(1) How many possible outcomes in the experiment of throwing a coin 10 times?

There are 2 possible outcomes each time, so there are a total of 2^{10} possible outcomes.

To be Continued

Combinations

Now, how many possible ways to obtain exactly three heads in the experiment?
Because the heads are indistinguishable, we have to use the combination method. Thus, we have $C_{10}^{3} =$ 120 different ways to obtain three heads in the total of 10 trials. It follows that

$$P(3\text{ heads obtained}) = \frac{120}{2^{10}} = 0.1172$$

To be Continued

Combinations

(2)

P(3 or fewer heads)

$$= P(0\ head) \ + P(1\ head) +$$

$$P(2\ heads) + P(3\ heads)$$

$$= \frac{176}{2^{10}}$$

$$= 0.1719$$

Combinations

Example 35

Suppose we throw a fair coin n times independently.
What is the probability that exactly x heads will show up?

Combinations

Solution

(1) How many possible outcomes in total?

There are $2 \times 2 \times \cdots 2 = 2^n$ outcomes.

To be Continued

Combinations

(2) How many possible ways to obtain exactly x heads in the experiment?

Since heads are indistinguishable, we have to use the combination method. There are a total of C_n^x different ways to obtain x heads.

To be Continued

Combinations

It follows that

$$P(\text{exactly } x \text{ heads}) = \frac{C_n^x}{2^n}$$

$$= C_n^x (\frac{1}{2})^x (1 - \frac{1}{2})^{n-x}$$

This is a special case of the so-called binomial distribution $B(n, p)$ with $P = \frac{1}{2}$.

Combinations

Example 36

We would like to choose r elements out of n elements. For the following cases, how many ways do we have?

(1) ordered, without replacement;

(2) unordered, without replacement;

(3) ordered, with replacement;

(4) unordered, with replacement.

Combinations

Solution

(1) $P_n^r = \dfrac{n!}{(n-r)!}$;

(2) $C_n^r = \dfrac{P_n^r}{r!} = \dfrac{n!}{r!(n-r)!}$;

(3) n^r;

(4) C_{r+n-1}^r.

CONTENTS

Conditional Probability

Motivations:

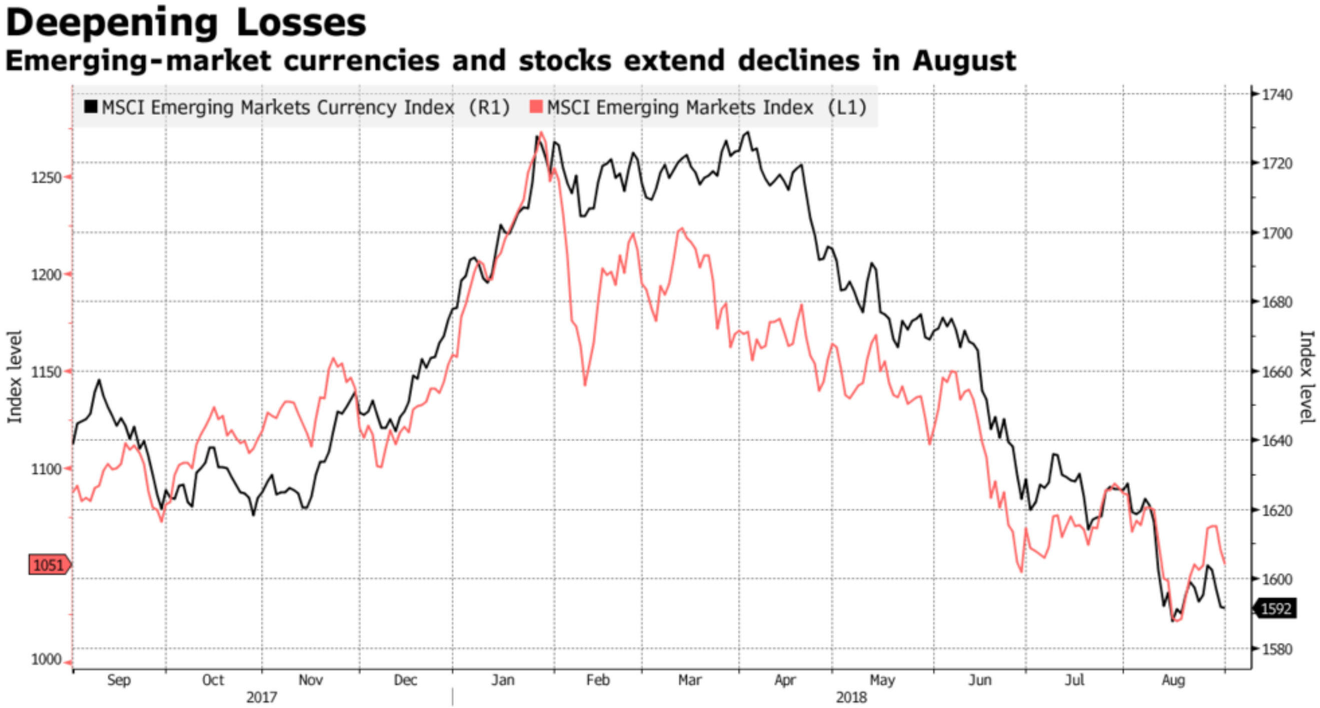
- Economic events are generally related to each other. For example, there may exist causal relationships between economic events. Because of the connectedness, the occurrence of event B may affect or contain the information about the probability that event A will occur.

- Thus, if we have information about event B, then we can know better about the occurrence of event A. This can be described by the concept of conditional probability.

Conditional Probability

Example 37. [**Financial Contagion**]

A large drop of the price in one market can cause a large drop of the price in another market, given the speculations and reactions of market participants. This can occur regardless of market fundamentals.



Source: MSCI Inc.

Conditional Probability

Definition 13. [**Conditional Probability**]

Let A and B be two events in probability space $(S, \mathbb{B}, P)$. Then the conditional probability of event A given event B, denoted as $P(A|B)$, is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

provided that $P(B) > 0$.

Similarly, the conditional probability of event B given A is defined as

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

provided $P(A) > 0$.

Conditional Probability

Remarks:

- In defining the conditional probability $P(A|B)$, we assume $P(B) > 0$. This is because $P(B) = 0$ implies that B is unlikely to happen, and conditioning on an unlikely event is practically meaningless.

Conditional Probability

Remarks:

- In a Venn diagram, $P(A|B)$ can be represented. It is the area occupied by event *A* within the area occupied by *B* relative to the area occupied by B. Intuitively, when event *B* has occurred, the complement B^c will never occur. The uncertainty has been reduced from *S* to *B*. Thus, we will treat *B* as a new sample space when we consider $P(A|B)$. All further occurrences are then calibrated with respect to their relationships to *B*.

Conditional Probability

Remarks:

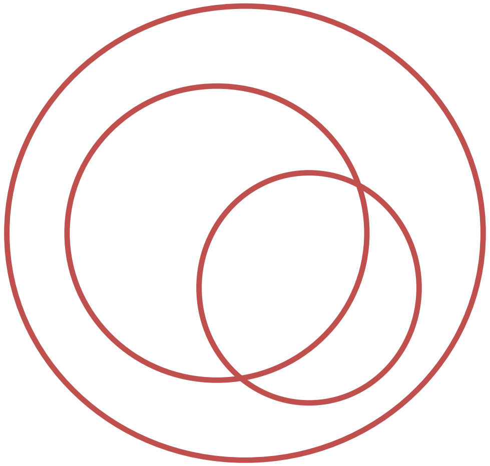
- The triple $(S \cap B, \mathbb{B} \cap B, P(\cdot \mid B))$ is a probability space associated with $P(A \mid B)$. In particular, $P(A \mid B)$ satisfies all probability laws defined on the sample space B. For example, we have (please show it!)

$$P(A^C|B) = 1 - P(A|B)$$

Conditional Probability

Example 38

Suppose the sample space S contains 25 sample points, which are chosen equally. Moreover, event A contains 15 points, event B contains 7 points, while $A \cap B$ contains 5 points,



$$
\begin{aligned}
P(A) &= \frac{15}{25}, \\
P(B) &= \frac{7}{25}, \\
P(A \cap B) &= \frac{5}{25}.
\end{aligned}
$$

Conditional Probability

Then we have

$$
\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{5}{7}, \\
P(B|A) &= \frac{P(B \cap A)}{P(A)} = \frac{1}{3}.
\end{aligned}
$$

Conditional Probability

Example 39

Let A and B be disjoint and $P(B) > 0$. What is $P(A|B)$?

Solution

Given $P(A \cap B) = 0$, we have

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = 0$$

Intuitively, mutually exclusive events cannot occur simultaneously. If event B has occurred, then event A will never occur.

Conditional Probability

Remarks:

- $P(A|B)$ describes how to use the information on event B to predict the probability of event A. This is a predictive relationship between A and B.
- A predictive relationship is not necessarily a causal relationship from B to A, even if the information of event B can be used to predict event A.

To identify a causal relationship, we have to use economic theory outside probability and statistics.

Conditional Probability

Lemma 2. [**Multiplication Rules**]

(1) If $P(B) > 0$, then $P(A \cap B) = P(A|B)P(B)$;

(2) If $P(A) > 0$, then $P(A \cap B) = P(B|A)P(A)$.

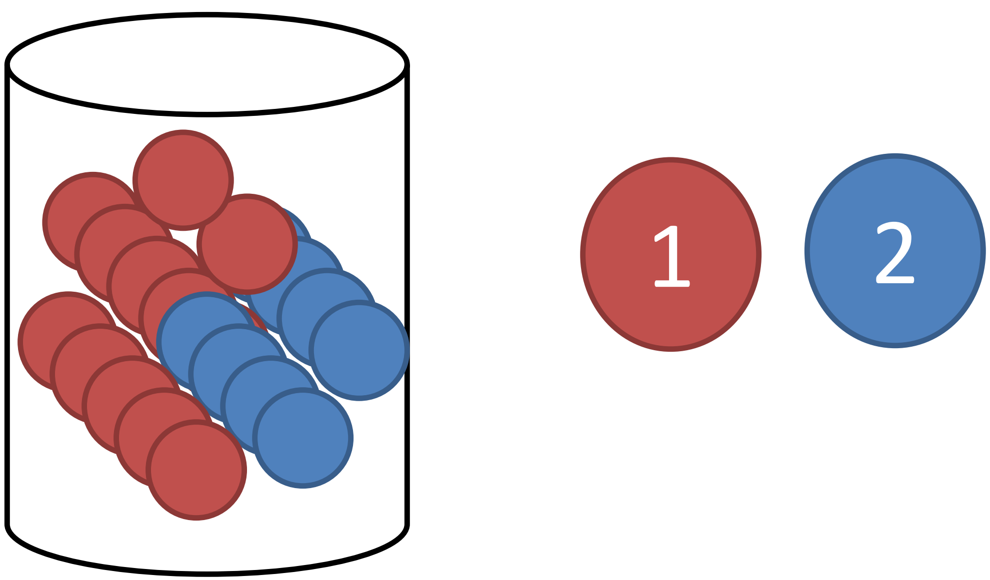
Remark:

These formula can be used to compute the joint probability of events A and B, that is, $P(A \cap B)$.

Conditional Probability

Example 40. [**Selecting Two Balls**]

Suppose two balls are to be selected, without replacement, from a box containing r red balls and b blue balls. What is the probability that the first is red and the second is blue?



Conditional Probability

Solution

Define

A = {the first ball is red}, B = {the second ball is blue}.

$$\begin{aligned} P(A) &= \frac{r}{r+b}, \\ P(B|A) &= \frac{b}{r+b-1}. \\ P(A \cap B) &= P(B|A)P(A) \\ &= \frac{rb}{(r+b)(r+b-1)}. \end{aligned}$$

Conditional Probability

Theorem 9

Suppose $\{A_i \in \mathbb{B}, i = 1, \ldots, n\}$ is a sequence of n events. Then the joint probability of these n events

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P\left(A_i \middle| \cap_{j=1}^{i-1} A_j\right)$$

with the convention that $P(A_1 | \cap_{j=1}^{0} A_j) = P(A_1)$.

Conditional Probability

Example 41. [**Computation of Joint Probabilities**]

For $n = 3$, we have

$$P(A_1 \cap A_2 \cap A_3) = P(A_3|A_2 \cap A_1)P(A_2|A_1)P(A_1)$$

Conditional Probability

Remarks:

- The multiplication rule can be repeatedly used to obtain the joint probability of multiple events.
- In fact, in expressing the joint probability $P(\cap_{i=1}^{n} A_i)$, there are $n!$ different ways of conditioning sequences.

Conditional Probability

Remarks:

- In time series analysis, where i is an index for time, the partition that the event A_i is conditional on $\cap_{j=1}^{i-1} A_j$ has a nice interpretation: A_i is conditional on the past information available at time $i-1$.
- Joint probability calculations are important for maximum likelihood estimation (MLE, Chapter 8).

Conditional Probability

Theorem 10. [**Rule of Total Probability**]

Let $\{A_i\}_{i=1}^{\infty}$ be a partition (i.e., mutually exclusive and collectively exhaustive) of sample space *S*, with $P(A_i) > 0$ for $i \geq 1$. Then for any event *A* in $\mathbb{B}$,

$$P(A) = \sum_{i=1}^{\infty} P(A|A_i)P(A_i)$$

Conditional Probability

Proof:

We have shown in Theorem 6 that

$$P(A) = \sum_{i=1}^{\infty} P(A \cap A_i)$$

The desired result follows immediately from the multiplication rule that $P(A \cap A_i) = P(A|A_i)P(A_i)$.

Conditional Probability

Remarks:

- This is called the rule of total probability because it says that if event A can be partitioned as a set of mutually exclusive subevents, then the probability of event A is equal to the sum of probabilities of this set of mutually exclusive subevents contained in A.
- It is also called the rule of elimination.

Conditional Probability

Example 42

Let $B_1, \cdots, B_k$ be mutually exclusive, and let $B = \cup_{i=1}^{k} B_i$. Suppose $P(B_i) > 0$ and $P(A|B_i) = p$ for $i = 1, \cdots, k$. Find $P(A|B)$.

Conditional Probability

Solution

By the definition of conditional probability, we have:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{P\left[\bigcup_{i=1}^{k} (A \cap B_i)\right]}{\sum_{i=1}^{k} P(B_i)}$$

$$= \frac{\sum_{i=1}^{k} P(A \cap B_i)}{\sum_{i=1}^{k} P(B_i)}$$

$$= \frac{\sum_{i=1}^{k} P(A|B_i)P(B_i)}{\sum_{i=1}^{k} P(B_i)}$$

$$= \frac{p\sum_{i=1}^{k} P(B_i)}{\sum_{i=1}^{k} P(B_i)}$$

$$= p.$$

Conditional Probability

Example 43

Suppose B_1, B_2, and B_3 are mutually exclusive.

If $P(B_i) = \frac{1}{3}$ and $P(A|B_i) = \frac{1}{6}$ for $i = 1,2,3$,

what is $P(A)$?

Conditional Probability

Solution

Noting that B_1, B_2, B_3 are collectively exhaustive (why?), we have

$$A = S \cap A = (B_1 \cup B_2 \cup B_3) \cap A = \bigcup_{i=1}^{3} (A \cap B_i).$$

It follows that

$$\begin{aligned} P(A) &= P\left[\bigcup_{i=1}^{3} (A \cap B_i)\right] \\ &= \sum_{i=1}^{3} P(A \cap B_i) \\ &= \sum_{i=1}^{3} P(A|B_i)P(B_i) \\ &= \frac{1}{3}\sum_{i=1}^{3} P(A|B_i) \\ &= \frac{1}{3}\left(\frac{1}{6} + \frac{2}{6} + \frac{3}{6}\right) \\ &= \frac{1}{3}. \end{aligned}$$

CONTENTS

Bayes' Theorem

Motivations:

- The knowledge that an event B has occurred can be used to revise or update the prior probability that an event A will occur.
- Bayes' theorem describes the mechanism of revising or updating the prior probability.
- This theorem leads to the Bayesian school of statistics, a rival to the school of classical statistics.

Bayes' Theorem

Theorem 11. [Bayes' Theorem]

Suppose A and B are two events with $P(A) > 0$ and $P(B) > 0$. Then

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

Bayes' Theorem

Remarks:

- $P(A)$ is called a " prior" probability (i.e., before the fact or evidence) about event *A* since it is the probability of *A* before new information *B* arrives.
- The conditional probability $P(A|B)$ is called a "posterior" probability (i.e., after the fact or evidence) since it represents the revised assignment of probability of *A* after the new information that *B* has occurred is obtained.

Bayes' Theorem

Remarks:

- Bayes' theorem can be verbally stated as the posterior probability of event A is proportional to the probability of the sample evidence B after A has occurred times the prior probability of A.
- Bayes' theorem expresses how a subjective degree of belief should rationally change to account for availability of related evidence.

Bayes' Theorem

Remarks:

- Bayes' theorem has been the subject of extensive controversy. There is no question about the validity of Bayes' theorem, but considerable arguments have been raised about the assignment of the prior probabilities.

Bayes' Theorem

Remarks:

- A good deal of mysticism surrounds Bayes' theorem because it entails a "backward" or "inverse" sort of reasoning, that is, reasoning "from effect to cause". In fact, this is a rather useful approach in economics and finance given no irreversibility or the nonexperimental nature of an economic process.

Bayes' Theorem

Theorem 12. [**Alternative Statement of Bayes' Theorem**]

Suppose $A_1, \cdots, A_n$ are n mutually exclusive and collectively exhaustive events in the sample space S, and A is an event with P(A) > 0. Then the conditional probability of A_i given A is

$$P(A_i|A) = \frac{P(A|A_i)P(A_i)}{\sum_{j=1}^{n} P(A|A_j)P(A_j)}, \qquad i = 1, \cdots, n$$

Bayes' Theorem

Proof:

By the conditional probability definition and multiplication rule, we have

$$
\begin{aligned}
P(A_i|A) &= \frac{P(A_i \cap A)}{P(A)} \\
&= \frac{P(A|A_i)P(A_i)}{P(A)}
\end{aligned}
$$

To be Continued

Bayes' Theorem

Because $\{A_i\}_{i=1}^{n}$ are collectively exhaustive and mutually exclusive, from the rule of total probability in Theorem 10, we have

$$
\begin{aligned}
P(A) &= \sum_{j=1}^{n} P(A \cap A_j) \\
&= \sum_{j=1}^{n} P(A|A_j)P(A_j)
\end{aligned}
$$

The desired result then follows immediately.

Bayes' Theorem

Remarks:

- Our interest is to update the probability about A_i given that event A has occurred.
- When event A has occurred, we may have better knowledge about the occurrence of A_i. Event A provides useful information for our updating knowledge on A_i.

Bayes' Theorem

Example 44.[**How to Determine Auto-insurance Premium?**]

Suppose an insurance company has three types of customers—high risk, medium risk and low risk. From the company's historical consumer database, it is known that 25% of its customers are high risk, 25% are medium risk, and 50% are low risk. Also, the database shows that the probability that a customer has at least one speeding ticket in one year is 0.25 for high risk, 0.16 for medium risk, and 0.10 for low risk.

Now suppose a new customer wants to be insured and reports that he has had one speeding ticket this year. What is the probability that he is a high risk customer, given that he has had one speeding ticket this year?

Bayes' Theorem

Solution

It is important for the auto-insurance company to determine whether the new customer belongs to the category of high risk customers, because it will affect the insurance premium to be charged.

To be Continued

Bayes' Theorem

We denote events

H = {the customer is of high risk},

M = {the customer is of medium risk},

L = {the customer is of low risk},

A = {the customer has received a speeding ticket this year}.

To be Continued

Bayes' Theorem

Then

$$P(H) = 0.25, \quad P(M) = 0.25, \quad P(L) = 0.50.$$
$$P(A|H) = 0.25, \quad P(A|M) = 0.16, \quad P(A|L) = 0.10.$$

It follows that

$$\begin{aligned} P(H|A) &= \frac{P(A|H)P(H)}{P(A)} \\ &= \frac{P(A|H)P(H)}{P(A|H)P(H) + P(A|M)P(M) + P(A|L)P(L)} \\ &= 0.410 \end{aligned}$$

To be Continued

Bayes' Theorem

Without the speeding ticket information reported by the new customer, the auto-insurance company, based on its customer database, only has a prior probability $P(H) = 0.25$ for the new customer. With the new information (A); the auto-insurance company has an updated probability $P(H|A) = 0.41$ for the new customer.

Bayes' Theorem

Example 45. [Is It Useful for Publishers to Send Free Sample Textbooks to Professors?]

A publisher sends a sample statistics textbook to 80% of all statistics professors in the U.S. schools. 30% of the professors who receive this sample textbook adopt the book, as do 10% of the professors who do not receive the sample book.

What is the probability that a professor who adopts the book has received a sample book?

Bayes' Theorem

Solution

Define event A = {A professor has received a sample copy}. Then

$$P(A) = 0.80, \qquad P(A^c) = 1 - 0.8 = 0.2$$

Also define B = {the professor adopts the textbook}. Then

$$P(B|A) = 0.3, \qquad P(B|A^c) = 0.1$$

To be Continued

Bayes' Theorem

It follows from Bayes' theorem that

$$
\begin{aligned}
P(A|B) &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \\
&= \frac{0.3 \cdot 0.8}{0.3 \cdot 0.8 + 0.1 \cdot 0.2} \\
&= 0.923
\end{aligned}
$$

Bayes' Theorem

Example 46.[**Are Stock Analysts Helpful?**]

Data evidence shows that last year 25% of the stocks in a stock exchange performed well, 25% poorly, and the remaining 50% performed on average. Moreover, 40% of those that performed well were rated "good buy" by a stock analyst at the beginning of last year, as were 20% of those that performed on average, and 10% of those that performed poorly. What is the probability that a stock rated a "good buy" by the stock analyst will perform well this year?

Bayes' Theorem

Solution

Define events

A = {the stock is rated as "good buy" by the stock analyst},

A_1 = {the stock performs better than the market average},

A_2 = {the stock performs as the market average},

A_3 = {the stock performs worse than the market average}.

To be Continued

Bayes' Theorem

Then

$$P(A_1) = 0.25, \quad P(A_2) = 0.5, \quad P(A_3) = 0.25.$$
$$P(A|A_1) = 0.4, \quad P(A|A_2) = 0.2, \quad P(A|A_3) = 0.1.$$

By Bayes' theorem, we have

$$\begin{aligned} P(A_1|A) &= \frac{P(A|A_1)P(A_1)}{\sum_{i=1}^{3} P(A|A_i)P(A_i)} \\ &= \frac{0.4 \cdot 0.25}{0.4 \cdot 0.25 + 0.2 \cdot 0.50 + 0.1 \cdot 0.25} \\ &= 0.444. \end{aligned}$$

To be Continued

Bayes' Theorem

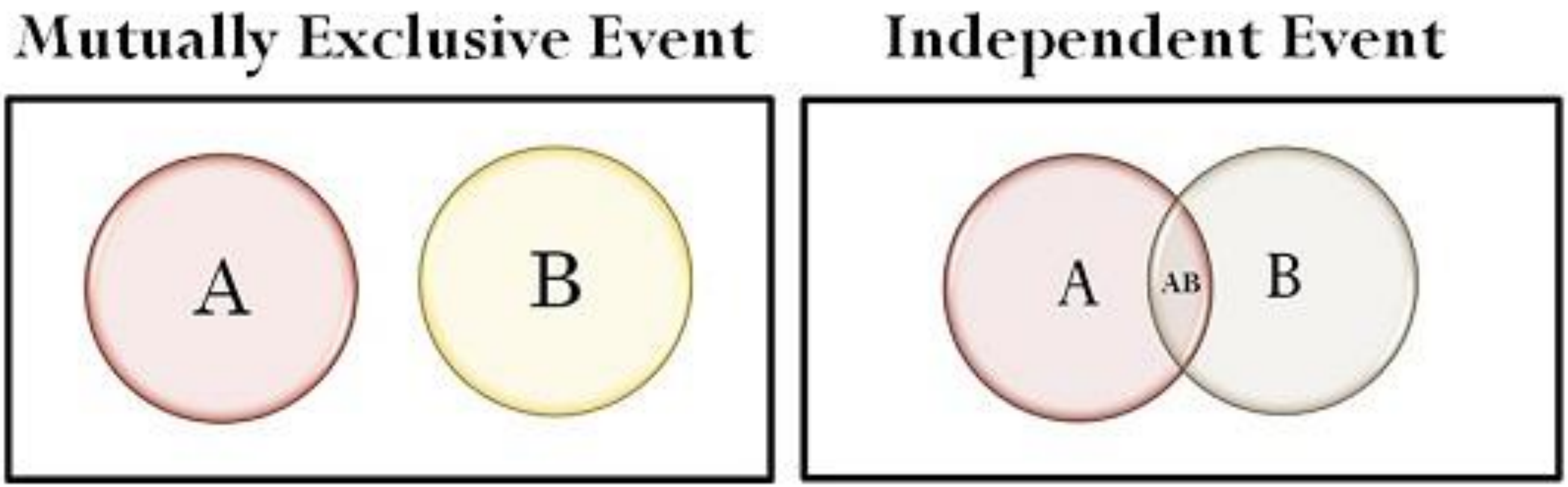
Without the recommendation by the stock analyst, an investor, based on the historical data of the stock market, will only have the prior probability $P(A_1) = 0.25$. With the recommendation by the stock analyst (A), the investor will update his belief to $P(A_1|A) = 0.444$.

CONTENTS

Independence

Definition 14 (2.14). [Independence]

Events A and B are said to be statistically independent if $P(A \cap B) = P(A)P(B)$.



Independence

Remarks:

- Events that are independent are called statistically independent, stochastically independent, or independent in a probability sense. In most instances, we use the word "independent" without a modifier if there is no possibility of misunderstanding.

- Independence is a probability notion to describe nonexistence of any kind of relationship between two events. It plays a fundamental role in probability theory and statistics.

Independence

Question:

What is the implication of independence?

Independence

Suppose $P(B) > 0$. Then by definition of independence,

$$
\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
&= \frac{P(A)P(B)}{P(B)} \\
&= P(A)
\end{aligned}
$$

Independence

The knowledge of *B* does not help in predicting *A*.

Similarly, we have $P(B|A) = P(B)$, i.e. the occurrence of *A* has no effect on the occurrence or probability of *B*.

Intuitively, independence implies that *A* and *B* are "irrelevant", or there exists no relationship between them.

Independence

Example 47

Let

A = {Raining in Ithaca}

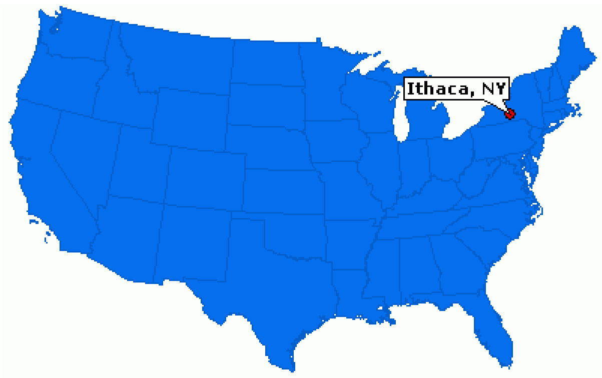
B = {Standard & Poor 500 price index going up}

These two events are likely to be independent.

Independence

Remarks:

While whether in Ithaca, which is a small town in the upstate New York State, is likely to be independent of S&P 500 price changes, there have been some empirical evidence that weather is associated with stock returns given that weather may affect the mood or sentiment of investors (see, e.g., Goetzman, Kim, Kumar and Wang (2015), and Hirshleifer and Shumway (2003)).



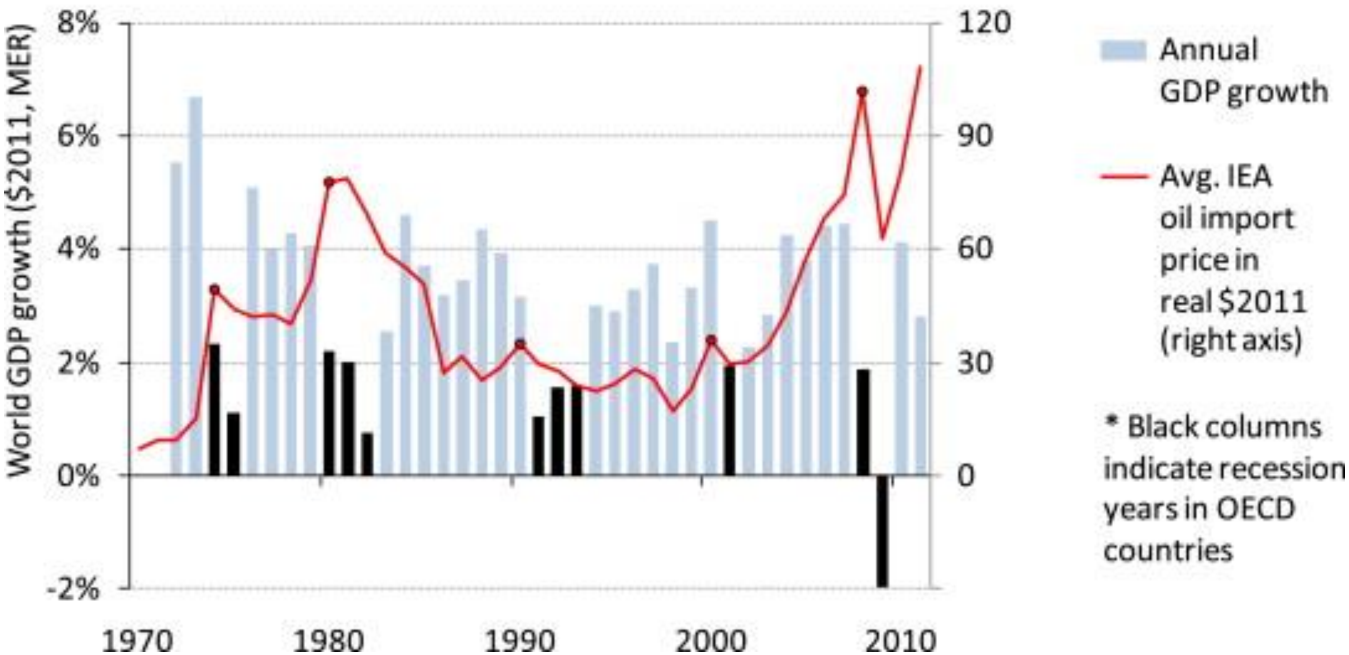
Independence

Example 48

Let A = {Oil price goes up}
B = {Output growth slows down}

These two events are likely to be dependent of each other.

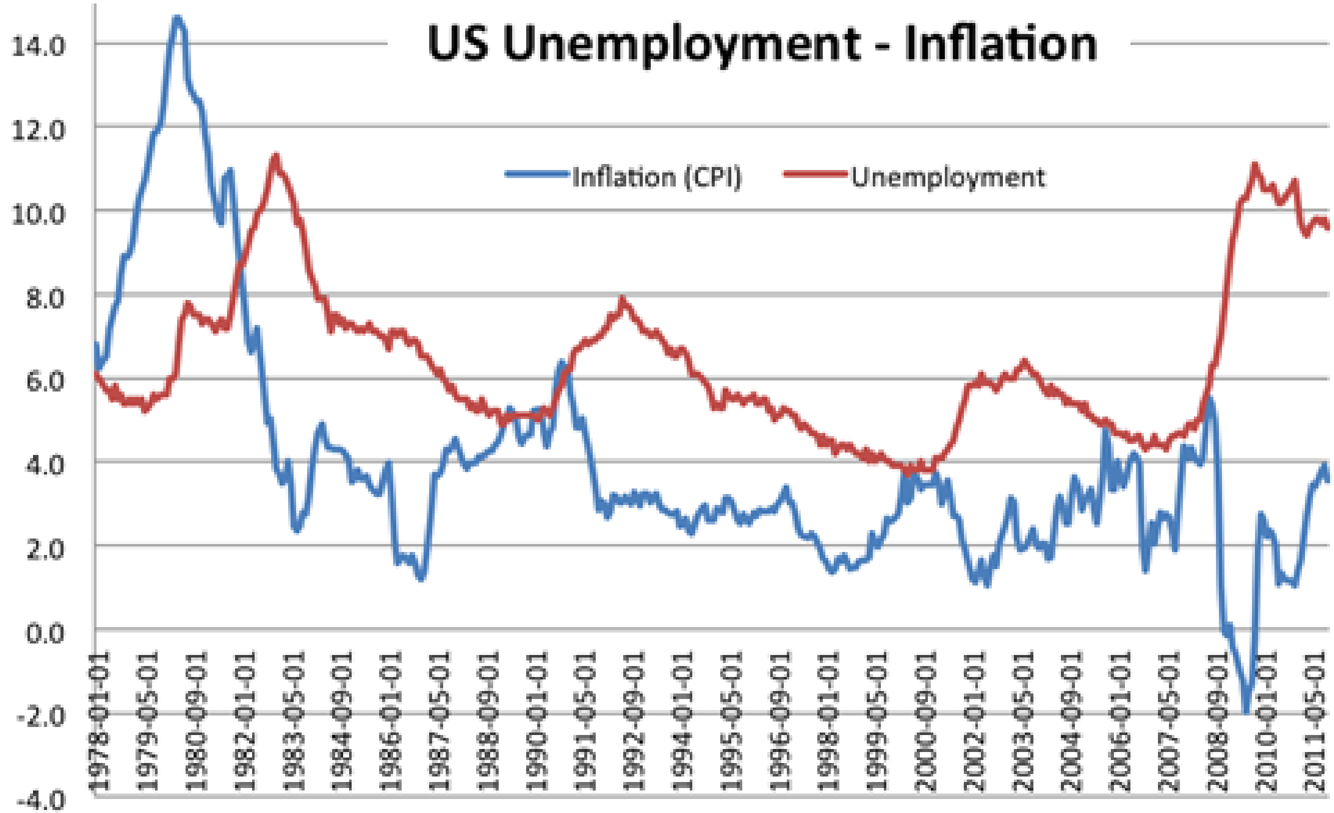
International crude oil prices and global GDP growth



Independence

Example 49

Let A = {Inflation rate increases}
B = {Unemployment decreases}
A and B are most likely to be dependent of each other.



Independence

Question:

Why is the concept of independence useful in economics and finance?

Independence

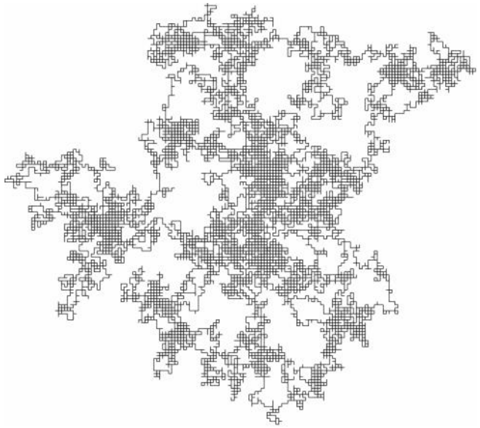
Example 50. [Random Walk Hypothesis (Fama 1970)]

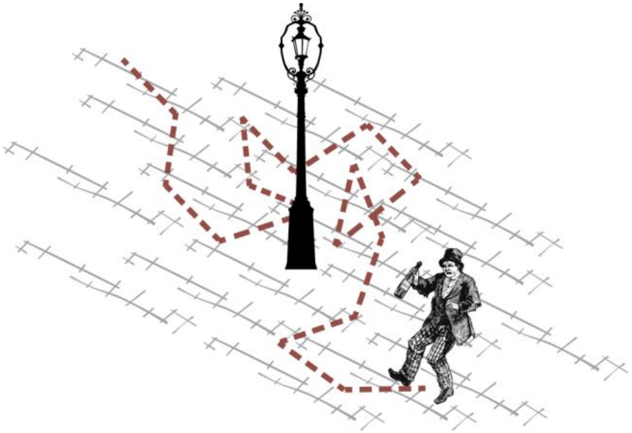
A stock price P_t will follow a random walk if

$$P_t = P_{t-1} + X_t$$

where $\{X_t\}$ is independent across different time periods.

Note that here $X_t = P_t - P_{t-1}$ is the stock price change from time $t-1$ to time t.





Independence

Remarks:

A closely related concept is the **geometric random walk hypothesis**. The stock price $\{P_t\}$ is called a geometric random walk if

$$\ln P_t = lnP_{t-1} + X_t$$

Where

$\{X_t\}$ *is independent across different time periods*.

Independence

Remarks:

The increment

$$
\begin{aligned}
X_t &= \ln\left(\frac{P_t}{P_{t-1}}\right) \\
&= \ln(1 + \frac{P_t - P_{t-1}}{P_{t-1}}) \\
&\simeq \frac{P_t - P_{t-1}}{P_{t-1}}
\end{aligned}
$$

can be interpreted as the relative stock price change.

Independence

Remarks:

The most important implication of the random walk hypothesis is:

if $\{X_t\}$ is serially independent across different time periods, then a future stock price change X_t is not predictable using the historical stock price information. In such a case, we call the stock market is **informationally efficient.**

Independence

Example 51

Can two independent events A and B be mutually exclusive?

Can two mutually exclusive events A and B be independent?

Independence

Solution

We first consider a regular case where $P(A) > 0$ and $P(B) > 0$:

To be Continued

Independence

Case (1):

If A and B are independent with $P(A) > 0$ and $P(B) > 0$, then

$$P(A \cap B) = P(A)P(B) > 0$$

Therefore, if A and B are independent, then they cannot be mutually exclusive.

To be Continued

Independence

On the other hand, if *A* and *B* are mutually exclusive (so $P(A \cap B) = 0$), then they cannot be independent.

However, there exists a pathological case where independent events can be mutually exclusive as well. This happens when $P(A) = 0$ or $P(B) = 0$.

To be Continued

Independence

Case (2):

Suppose $P(A) = 0$ or $P(B) = 0$. if *A* and *B* are independent, then

$$P(A \cap B) = P(A)P(B) = 0$$

This implies that *A* and *B* could be mutually exclusive. On the other hand, if *A* and *B* are mutually exclusive, they are independent.

Independence

Remarks:

- When $P(A) > 0$ and $P(B) > 0$, independent events cannot be mutually exclusive. This implies that independent events contain common basic outcomes and so can occur simultaneously.
- For example, Standard & Poor 500 price index can increase when there is raining in Ithaca. Intuitively, two independent events can occur simultaneously, so they are not mutually exclusive.
- On the other hand, two mutually exclusive events cannot occur simultaneously, so they are not independent.

Independence

Example 52

There are four cards, numbered 1, 2, 3, 4. The experiment is to select one card randomly. Define events A_1 = {1 or 2}, A_2 = {1 or 3}. Then

$$P(A_1) = \frac{1}{2} = P(A_2)$$

Since $A_1 \cap A_2 = \{1\}$, we have $P(A_1 \cap A_2) = 1/4$. Therefore, events A_1 and A_2 are independent, although they have a common element.

Independence

Theorem 13

Let A and B are two independent events. Then

- A and B^c are independent
- A^c and B are independent
- A^c and B^c are independent

Independence

Proof:

(1) If $P(A \cap B^c) = P(A)P(B^c)$, then A and B^c are independent. Because $(A \cap B) \cup (A \cap B^c) = A$, we have

$$P(A \cap B) + P(A \cap B^c) = P(A).$$

It follows from the multiplication rule that

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) \\ &= P(A) - P(A)P(B) \\ &= P(A)[1 - P(B)] \\ &= P(A)P(B^c) \end{aligned}$$

To be Continued

Independence

(2) By symmetry.

(3) Because $(A \cap B^c) \cup (A^c \cap B^c) = B^c$, we have

$$P(A \cap B^c) + P(A^c \cap B^c) = P(B^c)$$

It follows that

$$\begin{aligned} P(A^c \cap B^c) &= P(B^c) - P(A \cap B^c) \\ &= P(B^c) - P(A)P(B^c) \\ &= P(A^c)P(B^c) \end{aligned}$$

Independence

Remark:

Theorem 13 could be understood intuitively:

Suppose A and B are independent.
Then A and B^c should be independent as well because if not, one would be able to predict the probability of B^c using A, and thus predict the probability of B using A via the complement probability formula
$P(B|A) = 1 - P(B^c|A)$.

Independence

Definition 15. [Independence Among Several Events]

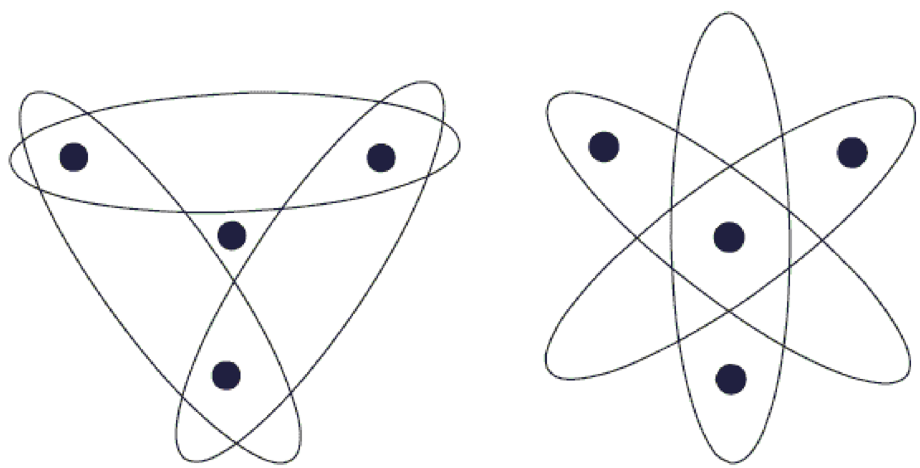
k events $A_1, A_2, \cdots, A_k$ are mutually independent if, for every possible subset $A_{i_1}, \cdots, A_{i_j}$ of j of those events ($j = 2,3,\cdots,k$),

$$P(A_{i_1} \cap \cdots \cap A_{i_j}) = P(A_{i_1}) \times \cdots \times P(A_{i_j}).$$

Independence

Remarks:

- For three or more events, independence is called mutual independence or joint independence. If there is no possibility of misunderstanding, independence is often used without the modifier "mutual" or "joint" when considering several events.



Example 1 Example 2

Independence

Remarks:

- A collection of events are mutually independent if the joint probability of any sub-collection of the events is equal to the product of the individual probabilities. There are $(2^k - 1 - k)$ conditions to characterize independence among k events (because $\sum_{j=0}^{k} \binom{k}{j} = 2^k, \binom{k}{0} = 1, \binom{k}{1} = k$).

Independence

Remarks:

- Three events A, B and C are independent, if the following $2^3 - (1+3) = 4$ conditions are satisfied:

$$
\begin{aligned}
P(A \cap B) &= P(A)P(B), \\
P(A \cap C) &= P(A)P(C), \\
P(B \cap C) &= P(B)P(C), \\
P(A \cap B \cap C) &= P(A)P(B)P(C)
\end{aligned}
$$

Independence

Example 53. [**Mooncake Betting**]

In the city of Xiamen (also called Amoy), Southeast China, which is a main origin port for Oversea Chinese, there is a traditional activity called "Mooncake Betting" to celebrate the Mid-Autumn Festival, a traditional Chinese holiday.

Independence

Example 53. [**Mooncake Betting**]

This is essentially a game of rolling 6 dices. The final winner, called "Zhuang-Yuan", will result if she rolls 6 dices with at least 4 dices showing number 4 simultaneously. If at least two candidates have 4 dices with number 4, then the one who gets a larger number on the fifth dice will win.

科举	常用名	图标
状元	状元插金花	⚃ ⚃ ⚃ ⚃ ⚀ ⚀
	六抔红	⚃ ⚃ ⚃ ⚃ ⚃ ⚃
	六抔黑	⚅ ⚅ ⚅ ⚅ ⚅ ⚅
	五王	⚃ ⚃ ⚃ ⚃ ⚃
	五子带一秀	⚂ ⚂ ⚂ ⚂ ⚂ ⚃
	五子登科	⚂ ⚂ ⚂ ⚂ ⚂
	状元	⚃ ⚃ ⚃ ⚃

Independence

Example 53. [**Mooncake Betting**]

Suppose 2 friends play such a game. On average, how many rounds of rolling that both of them have to complete to order to produce a winner?

Independence

Example 54. [**Reliability**]

A project, such as launching a satellite, consists of k independent subprojects, denoted as $A_1, A_2, \cdots, A_k$. Suppose subproject i has a failure rate f_i, where $i = 1, \cdots, k$.

What is the probability that the project will be successfully implemented?

Independence

Solution

The success of the project requires that all subprojects be successful. Thus, the probability of a successful project is given by

$$
\begin{aligned}
P\Big(\bigcap_{i=1}^{k} A_i\Big) &= \prod_{i=1}^{n} P(A_i) \\
&= \prod_{i=1}^{n} [1 - P(A_i^c)] \\
&= \prod_{i=1}^{n} (1 - f_i)
\end{aligned}
$$

Independence

Remarks:

- Joint independence implies pairwise independence. However, the converse is not true.
- It is possible to find that three events are pairwise independent but not jointly independent.

Independence

Example 55

Suppose

$$S = \{aaa, bbb, ccc, abc, bca, cba, acb, bac, cab\}$$

and each basic outcome is equally likely to occur. For $i = 1,2,3$, define A_i={i-th place in the triple is occupied by letter a}.

For example

$$A_1 = \{aaa, abc, acb\}$$

Independence

Example 55

It is then easy to see that

$$P(A_1) = P(A_2) = P(A_3) = \frac{3}{9} = \frac{1}{3}$$

and

$$P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{9},$$

so that A_1, A_2 and A_3 are pairwise independent. However,

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{9} > P(A_1)P(A_2)P(A_3) = \frac{1}{27}$$

Therefore, A_1, A_2, A_3 are not mutually independent.

Independence

Remarks:

- The implication of Example 53 is that if one uses A_2 or A_3 to predict A_1; then A_2 or A_3 is not helpful. However, if one uses both A_2 and A_3 jointly to predict A_1, then A_1 is predictable.

Independence

Remarks:

- The difference between pairwise independent events and joint independent events: in the former, knowledge about the joint occurrence of any two of them may be useful in predicting the chance of the occurrence of the remaining one, but in the latter it would not.

Independence

Example 56.[**Complementarity Between Economic Reforms**]

In the fields of economic growth and development, many studies find that one economic policy usually necessities another policy to stimulate the economic growth, which is called policy complementarities. In traditional economics, individual reforms or sequential reforms may not be effective or fully effective, or even back-ring. Reforms must be packed together in order to be effective.

For example, in order to improve firm productivity (A_1), changing a manager (A_2) should be together with granting autonomy to the firm (A_3).

Independence

Remarks:

- There are many other examples of economic complementarities. Harrison (1996), Ro-driguez and Rodrik (2000), Loayza *et al*. (2005), Chang *et al*. (2005) document that international trade openness, only when combined with other policies that improve a country's educational investment, financial depth, inflation stabilization, public infrastructure, governance, labor market flexibility, and ease of firm entry and exit, can promote economic growth.

Independence

Remarks:

- In some special cases, it is possible that pairwise independence implies joint dependence.
- For multiple joint independence, it is important to check the joint probability of every possible subset of events should be equal to the product of the probabilities of all individual events in the subset.

Independence

Remarks:

- For example, for independence among three events A, B, C, it is not sufficient to only check the condition that $P(A \cap B \cap C) = P(A)P(B)P(C)$. Conditions on all possible pairs of events should be considered as well.

Independence

Question:

When can we have pairwise independence imply joint independence? If so, please give an example.

Independence

Question:

Suppose A, B, C are three events. Does $P(A \cap B \cap C) = P(A)P(B)P(C)$ imply independence among the three events A, B, C? If yes, prove it. If not, give an example.

CONTENTS

Conclusion

Conclusion:

- This chapter is a foundation of probability theory.
- We first characterize a random experiment by a probability space $(S, \mathbb{B}, P)$. Interpretations for probabilities are provided.

Conclusion

Conclusion:

- Given a measurable space $(S, \mathbb{B})$, one can dene many probability functions. The main objective of econometrics is to use the observed economic data to infer a suitable probability function which truly represents the true probability distribution for the data generating process.

Conclusion

Conclusion:

- For random experiments with equally likely outcomes, methods of counting are very useful in calculating probabilities of interested events.
- The conditional probability function characterizes predictive relationships between or among economic events. An application is Bayes' theorem.

Conclusion

Conclusion:

- Finally, we introduce the concept of independence and its implications in economics and finance.

Thank You !