# Lecture 2. Probability (Chapter 3)

Ping Yu

HKU Business School
The University of Hong Kong

# Plan of This Lecture

- Random Experiment, Outcomes, and Events
- Probability and Its Postulates
- Probability Rules
- Bivariate Probabilities
- Bayes' Theorem

# Random Experiment, Outcomes, and Events

# Random Experiment and Sample Space

- A random experiment is a process leading to two or more possible outcomes, without knowing exactly which outcome will occur.
- Examples:
  - A coin is tossed and the outcome is either a head or a tail.
  - A company has the possibility of receiving 0-5 contract awards.
- The possible outcomes from a random experiment are called the basic outcomes, and the set of all basic outcomes is called the sample space, denoted as $S$.
  - In the two examples above, $S = \{\text{head, tail}\}$ and $\{0, 1, 2, \cdots, 5\}$, respectively.
  - No two basic outcomes can occur simultaneously.
  - The random experiment must necessarily lead to the occurrence of one of the basic outcomes.
  - So after a random experiment is conducted, one and only one basic outcome will occur.

# Event, Intersection and Mutually Exclusive

- An event, $E$, is any subset of basic outcomes from the sample space. An event occurs if the random experiment results in one of its constituent basic outcomes.
  - The null event represents the absence of a basic outcome and is denoted as $\varnothing$.
  - e.g., {contract rewards are odd} and {contract rewards are less than 3} are both events.
  - This definition of "event" is different from our everyday notion, which requires that some changes occur (e.g., we would not refer to the contract reward being odd as an event, but we would refer to that the reward increases as such.).
  - Another way of thinking of an event is this: any declarative statement (a statement that can be true or false) is an event.
- The intersection of two events, $A$ and $B$, denoted as $A \cap B$, is the set of all basic outcomes that belong to both $A$ and $B$, i.e., $A \cap B$ occurs iff both $A$ and $B$ occur. [figure here]
  - We can similarly define $E_1 \cap E_2 \cap \cdots \cap E_K$.
- If the events $A$ and $B$ have no common basic outcomes (i.e., cannot co-occur), they are called mutually exclusive, i.e., $A \cap B = \varnothing$. [figure here]
  - We can similarly define $E_1, E_2, \cdots, E_K$ to be mutually exclusive as pairwisely mutually exclusive.

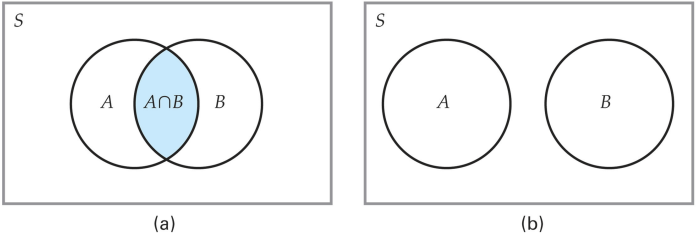





Figure: Venn Diagrams for $A\cap B$ and $A$ and $B$ are Mutually Exclusive

# Complement

- The complement of $A$, denoted as $\bar{A}$, is the set of basic outcomes belonging to $S$ but not to $A$. [figure here]

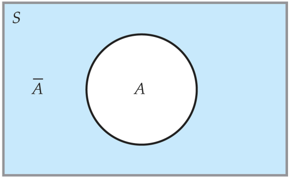





**Table 3.2** Intersection of and Mutually Exclusive Events

(a) Intersection of Events

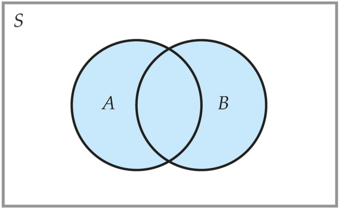
| | $B$ | $\bar{B}$ |
|---|---|---|
| $A$ | $A \cap B$ | $A - (A \cap B)$ |
| $\bar{A}$ | $B - (A \cap B)$ | $\bar{A} \cap \bar{B}$ |

(b) Mutually Exclusive Events

| | $B$ | $\bar{B}$ |
|---|---|---|
| $A$ | $\varnothing$ | $A$ |
| $\bar{A}$ | $B$ | $\bar{A} \cap \bar{B}$ |

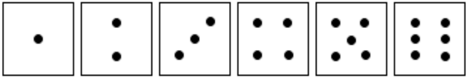
# Union, Collectively Exhaustive and Partition

- The union of two events, $A$ and $B$, denoted as $A \cup B$, is the set of all basic outcomes that belong to at least one of $A$ and $B$, i.e., $A \cup B$ occurs iff either $A$ or $B$ or both occur. [figure here]
  - We can similarly define $E_1 \cup E_2 \cup \cdots \cup E_K$.
- If $E_1 \cup E_2 \cup \cdots \cup E_K = S$, then these $K$ events are said to be collectively exhaustive.
- A mutually exclusive and collectively exhaustive set of events $\{B_i\}_{i=1}^K$ is called a partition of the sample space $S$.
  - Exactly one of the events $\{B_i\}_{i=1}^K$ must be true.
  - The set of all basic outcomes is a partition of $S$, and so are $\{A, \bar{A}\}$ and $\{A \cap B, A - (A \cap B), B - (A \cap B), \bar{A} \cap \bar{B}\}$ in Table 3.2.
  - We can also partition any event $A$ in the same way.



# [Example] Rolling a Die

- Let the Sample Space be the collection of all possible outcomes of rolling one die:



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

- Let $A$ be the event "Number rolled is even", and $B$ be the event "Number rolled is at least 4"; then

$$A = \{2, 4, 6\} \text{ and } B = \{4, 5, 6\}.$$

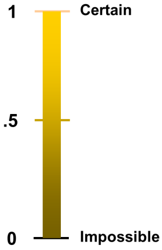
- Complements: $\bar{A} = \{1, 3, 5\}$, and $\bar{B} = \{1, 2, 3\}$.
- Intersections: $A \cap B = \{4, 6\}$, and $\bar{A} \cap B = \{5\}$.
- Unions: $A \cup B = \{2, 4, 5, 6\}$, and $A \cup \bar{A} = S$.
- Mutually exclusive: $A$ and $B$ are not mutually exclusive because the outcomes 4 and 6 are common to both.
- Collectively exhaustive: A and B are not collectively exhaustive because $A \cup B$ does not contain 1 or 3.

# Probability and Its Postulates

# Probability

- Probability is the chance that an uncertain event will occur (always between 0 and 1).
- Mathematically,

$$0 \le P(A) \le 1 \text{ for any event } A.$$




- We consider three definitions of probability: classical probability, relative frequency probability and subjective probability.

# Classical Probability

- The classical probability is the proportion of times that an event will occur, assuming that all outcomes in a sample space are equally likely to occur, specifically,

$$P(A) = \frac{N_A}{N} = \frac{\text{number of outcomes that satisfy the event } A}{\text{total number of outcomes in the sample space}}.$$

  - The basic idea is that the probability can be developed from fundamental reasoning about the process.
  - e.g., tossing a coin 10 times, what is the probability with 5 successive heads?

- The definition requires a count of the outcomes in the sample space as illustrated in the next slide.

# Formula for Counting the Number of Combinations

- The total number of possible ways of arranging $n$ objects in order is

$$n! := n \cdot (n-1) \cdot \cdots \cdot 2 \cdot 1,$$

read as "$n$ factorial", where ":=" is read as "is defined as".

- The number of permutations of $x$ objects chosen from $n$ (i.e., the number of possible arrangements when $x$ objects are to be selected from a total of $n$ objects and arranged in order, with $(n-x)$ objects left over) is

$$P_x^n := n \cdot (n-1) \cdot \cdots \cdot (n-x+1) = \frac{n!}{(n-x)!}.$$

- $P_n^n = P_{n-1}^n = n!$ with $0! = 1$, $P_1^n = n$, and $P_0^n = 1$.

- The number of combinations of $x$ objects chosen from $n$:

$$C_x^n = \frac{P_x^n}{x!} = \frac{n!}{x!(n-x)!}.$$

- $C_n^n = 1$, $C_{n-1}^n = C_1^n = n$, and $C_0^n = 1$.
- Sometimes, the notation $\begin{pmatrix} n \\ x \end{pmatrix}$ is used for $C_x^n$.

# With/Without Replacement and Ordered/Unordered Counting

- When counting the number of objects in a set, there are two important distinctions. Counting may be with replacement or without replacement. Counting may be ordered or unordered.

| | Without Replacement | With Replacement |
|---|---|---|
| Ordered | $P_x^n$ | $n^x$ |
| Unordered | $C_x^n$ | $C_x^{n+x-1}$ |

- (**) This following part is optional.
- The unordered counting with replacement is more challenging. It is the number of distinct solutions to the equation

$$z_1 + z_2 + \cdots + z_n = x, \text{ where } z_i \in \{0, 1, 2, \cdots, x\}.$$

  - thinking of $z_i$ as $z_i$ bars, then we are arranging $x$ bars and $(n-1)$ plus signs $(+)$ to get different patterns.

# [Example] Choosing Two Letters From Four

- Suppose that two letters are to be selected from A, B, C, D and arranged in order. How many permutations are possible?
- Solution: The number of permutations, with $n = 4$ and $x = 2$, is

$$P_2^4 = \frac{4!}{(4-2)!} = 12.$$

  The permutations are $\{$AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, DC$\}$.
- Suppose that two letters are to be selected from A, B, C, D. How many combinations are possible (i.e., order is not important)?
- Solution: The number of combinations is

$$C_2^4 = \frac{4!}{2!(4-2)!} = 6.$$

  The combinations are {AB (same as BA), AC (same as CA), AD (same as DA), BC (same as CB), BD (same as DB), CD (same as DC)}.

# Relative Frequency Probability and Subjective Probability

- The relative frequency probability is the limit of the proportion of times that an event will occur in a large number of trials, specifically,

$$P(A) = \frac{n_A}{n} = \frac{\text{number of } A \text{ outcomes}}{\text{total number of trials or outcomes}}.$$

  The probability is the limit as $n$ becomes large (or approaches infinity).
  - e.g., the probability of family income above $\$75,000$ – we filter out the families with income above $\$75,000$, and then divide this number by the total number of families.
  - This probability can be obtained from more than one data sources to cross-validate each other.
- The subjective probability expresses an individual's degree of belief about the chance that an event will occur.
  - This probability is personal, so different individuals (with different information or different views) may have different probabilities.

# Probability Postulates

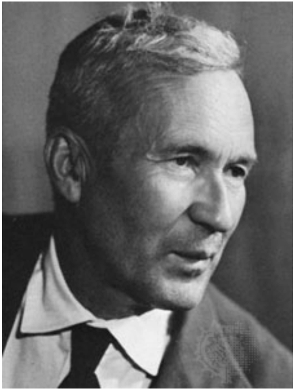
- We postulate the following properties of probability to assess and manipulate it.

1. If $A$ is an event in $S$, then
$$0 \le P(A) \le 1.$$
- An event with probability 0 is impossible; an event with probability 1 is certain.

2. Let $A$ be an event in $S$ and $O_i$ be the basic outcomes. Then
$$P(A) = \sum_{O_i \in A} P(O_i) =: \sum_A P(O_i),$$
where "$=:$" is read as "defines".
- Why? $P(A) = \lim_{n\to\infty} \frac{n_A}{n} = \lim_{n\to\infty} \frac{\sum_A n_i}{n} = \sum_A \lim_{n\to\infty} \frac{n_i}{n} = \sum_A P(O_i)$.

3. $P(S) = 1$.
- When a random experiment is carried out, something has to happen.

# History of Probability Theory



Andrey N. Kolmogorov (1903-1987), Russian[1]

[1] Vladimir Arnold, a student of Kolmogorov, once said: "Kolmogorov – Poincaré – Gauss – Euler – Newton, are only five lives separating us from the source of our science".

# Consequences of the Postulates

- If $S$ consists of $n$ equally likely basic outcomes, $O_1, O_2, \cdots, O_n$, then (from postulates 2 and 3)

$$P(O_i) = \frac{1}{n}.$$

  - tossing a coin, $P(\text{head}) = 1/2$.

- If $S$ consists of $n$ equally likely basic outcomes and event $A$ consists of $n_A$ of these outcomes, then (from consequence 1 and postulate 2)

$$P(A) = \frac{n_A}{n}.$$

- If $A$ and $B$ are mutually exclusive, then (from postulate 2)

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

  - A similar result applies to mutually exclusive events $E_1, E_2, \cdots, E_K$.

- If $E_1, E_2, \cdots, E_K$ are collectively exhaustive, then (from postulate 3)

$$P(E_1 \cup E_2 \cup \cdots \cup E_K) = 1.$$

# Probability Rules

# Complement Rule and Addition Rule

- We develop some rules for computing probabilities for compound events.
- The complement rule: for an event $A$ and its complement $\bar{A}$,

$$P\left(\bar{A}\right) = 1 - P\left(A\right).$$

  - This is because $1 = P\left(S\right) = P\left(A \cup \bar{A}\right) = P\left(A\right) + P\left(\bar{A}\right)$.
- The addition rule: for two events $A$ and $B$,

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

  - This is because $P\left(A \cup B\right) = P\left(A\right) + P\left(\bar{A} \cap B\right)$ and $P\left(B\right) = P\left(A \cap B\right) + P\left(\bar{A} \cap B\right)$. [figure here]
  - $P\left(A \cap B\right)$ is called the joint probability of $A$ and $B$.

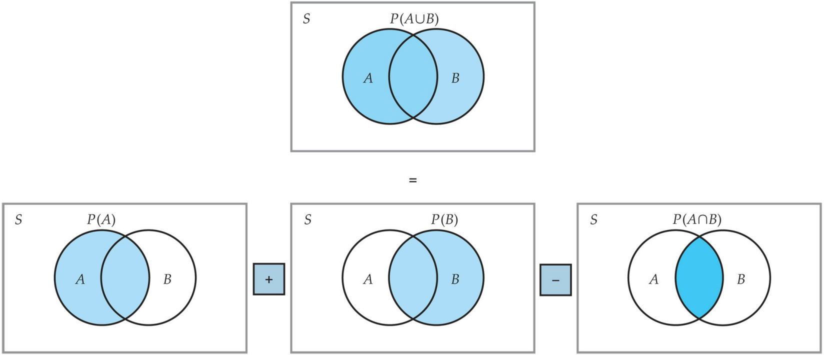




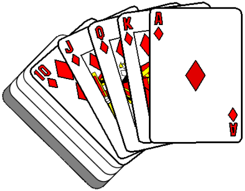

Figure: Venn Diagram for Addition Rule

# [Example] Cards

- Consider a standard deck of 52 cards, with four suits:

  ♥ ♣ ♦ ♠

- Let event $A$ = {card is an Ace}, and event $B$ ={card is from a red suit}.



- $P(\text{Red} \cup \text{Ace}) = P(\text{Red}) + P(\text{Ace}) - P(\text{Red} \cap \text{Ace})$

$$= \frac{26}{52} + \frac{4}{52} - \boxed{\frac{2}{52}} = \frac{28}{52}$$

Don't count the two red aces twice!

| Type | Color: Red | Color: Black | Total |
|---|---|---|---|
| Ace | 2 | 2 | 4 |
| Non-Ace | 24 | 24 | 48 |
| Total | 26 | 26 | 52 |

# Conditional Probability

- The conditional probability of event $A$ given that event $B$ occurred, denoted as $P(A|B)$, is
$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ provided that } P(B) > 0;$$
similarly,
$$P(B|A) = \frac{P(A \cap B)}{P(A)} \text{ provided that } P(A) > 0.$$
- refer to the figure in the previous slide.
- $P(A|B)$ can be thought of filtering or stratifying the data when calculating the "relative frequency" probability; it cannot be smaller than $P(A \cap B)$.

**Table 3.3**
Joint Probability of $A$ and $B$

| | $A$ | $\bar{A}$ | |
|---|---|---|---|
| $B$ | $P(A \cap B)$ | $P(\bar{A} \cap B)$ | $P(B)$ |
| $\bar{B}$ | $P(A \cap \bar{B})$ | $P(\bar{A} \cap \bar{B})$ | $P(\bar{B})$ |
| | $P(A)$ | $P(\bar{A})$ | 1.0 |



- $P(A)$ can be dramatically different from $P(A|B)$ (i.e., the probability we assign to an event depends on the knowledge we condition on; e.g., $P(\text{COVID-19})$ and $P(\text{COVID-19}|\text{temperature} = 40^\circ C)$).

# [Example] Cars with AC and CD

- Of the cars on a used car lot, 70% have air conditioning (AC) and 40% have a CD player (CD). 20% of the cars have both.
- What is the probability that a car has a CD player, given that it has AC? i.e., we want to find $P\left(\text{CD}|\text{AC}\right)$.
- Given AC, we only consider the top row (70% of the cars). Of these, 20% have a CD player. 20% of 70% is 28.57%:

| | CD | No CD | Total |
|---|---|---|---|
| AC | .2 | .5 | .7 |
| No AC | .2 | .1 | .3 |
| Total | .4 | .6 | 1.0 |

$$P(CD\,|\,AC) = \frac{P(CD \cap AC)}{P(AC)} = \frac{.2}{.7} = .2857$$

# Multiplication Rule

- The multiplication rule: for two events $A$ and $B$,

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A).$$

- In the cards example,

$$\begin{aligned} P(\text{Red} \cap \text{Ace}) &= P(\text{Red}|\text{Ace})P(\text{Ace}) \\ &= \frac{2}{4}\frac{4}{52} = \frac{2}{52} \\ &= \frac{\text{number of cards that are red and ace}}{\text{total number of cards}} = \frac{2}{52} \end{aligned}$$

| Type | Color: Red | Color: Black | Total |
|---|---|---|---|
| Ace | 2 | 2 | 4 |
| Non-Ace | 24 | 24 | 48 |
| Total | 26 | 26 | 52 |

# Statistical Independence

- Two events $A$ and $B$ are (statistically) independent iff

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

- Events $A$ and $B$ are independent when the probability of one event is not affected by the other event, i.e.,

$$\begin{aligned} P(A|B) &= P(A) \text{ (if } P(B) > 0\text{)}, \\ P(B|A) &= P(B) \text{ (if } P(A) > 0\text{)}. \end{aligned}$$

  - This can be also used as the definition of independence.

- In the cars example, are events AC and CD independent?

$$P(CD \cap AC) = 0.2 \neq 0.7 \times 0.4 = P(AC)P(CD),$$

  so the two events are not statistically independent.

- Generally, the events $E_1, E_2, \cdots, E_K$ are mutually independent iff

$$P(E_1 \cap E_2 \cap \cdots \cap E_K) = P(E_1) \cdots P(E_K) =: \prod_{i=1}^{K} P(E_i).$$

# continue

- Intuitively, independence between $A$ and $B$ means that knowing $B$ occurred will not change the assessment of $A$'s probability.
  - It is hard in practice for two events to be "strictly" independent, but we can "approximately" assume it for simplicity; e.g., $P\left(\text{you have the COVID-19}\right) \approx P\left(\text{you have the COVID-19} | \text{your friend Joe is 42 years old}\right)$.
- If two events $A$ and $B$ are not independent, then they are dependent.
  - Dependence and independence are symmetric relations – if $A$ is dependent on $B$, then $B$ is dependent on $A$, and if $A$ is independent of $B$, then $B$ is independent of $A$ (Formally, $P(A|B) = P(A) \Longrightarrow P(B|A) = P(B)$). This makes intuitive sense: if "smoke" tells us something about "fire", then "fire" must tell us something about "smoke".
- Independence is different from "mutually exclusive": the latter implies $P(A \cap B) = 0$ and the former means $P(A \cap B) = P(A)P(B)$.
  - $A \cap B = \varnothing$ implies "if $A$ occurs, then $B$ cannot", so they are not be independent (unless $P(A)$ or $P(B)$ is zero).

# (*) Examples 3.14 and 3.15: Choice of Cell Phone Features

- 75% customers use text messaging ($A$), 80% use photo capability ($B$), and 65% use both $(A \cap B)$.
- Then $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.75 + 0.80 - 0.65 = 0.90$.
- $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.65}{0.80} = 0.8125$ is the probability that a person who wants photo capability also wants texting messaging.
- $P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.65}{0.75} = 0.8667$ is the probability that a person who wants text messaging also wants photo capability.

**Table 3.4** Joint Probability for Example 3.15

| | Text Messaging | No Text Messaging | |
|---|---|---|---|
| Photo | 0.65 | 0.15 | 0.80 |
| No Photo | 0.10 | 0.10 | 0.20 |
| | 0.75 | 0.25 | 1.0 |



**Table 3.5** Joint Probability for Photo and Messaging When They Are Independent

| | Messaging | No Messaging | |
|---|---|---|---|
| Photo | 0.60 | 0.20 | 0.80 |
| No photo | 0.15 | 0.05 | 0.20 |
| | 0.75 | 0.25 | 1.0 |

# Example 3.20: The Birthday Problem

- What is the probability that at least 2 people in a party have the same birthday (month and day, neglecting Feb. 29)?
- It is easier to calculate $\bar{A}$, i.e., the probability that "all $M$ people have different birthdays".
- After some thinking, you can figure out that

$$P\left(\bar{A}\right)=\frac{P_M^{365}}{365^M},$$

so $P\left(A\right)=1-P\left(\bar{A}\right)$:

| $M$ | 10 | 20 | 22 | 23 | 30 | 40 | 60 |
|---|---|---|---|---|---|---|---|
| $P\left(A\right)$ | 0.117 | 0.411 | 0.476 | 0.507 | 0.706 | 0.891 | 0.994 |

- The probability that any given pair of people will have the same birthday is $1/365$, but as $M$ increases, the number of possible matches increases, until $P\left(A\right)$ becomes quite large.

# Bivariate Probabilities

# Bivariate Probabilities

**Table 3.6**
Outcomes for Bivariate Events

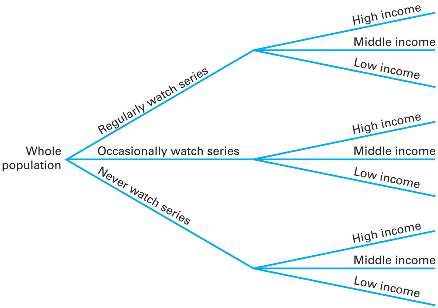
| | $B_1$ | $B_2$ | $\dots$ | $B_K$ |
|---|---|---|---|---|
| $A_1$ | $P(A_1 \cap B_1)$ | $P(A_1 \cap B_2)$ | $\dots$ | $P(A_1 \cap B_K)$ |
| $A_2$ | $P(A_2 \cap B_1)$ | $P(A_2 \cap B_2)$ | $\dots$ | $P(A_2 \cap B_K)$ |
| . | . | . | . | . |
| . | . | . | . | . |
| . | . | . | . | . |
| $A_H$ | $P(A_H \cap B_1)$ | $P(A_H \cap B_2)$ | $\dots$ | $P(A_H \cap B_K)$ |

- The events $A_i$ and $B_j$ are mutually exclusive and collectively exhaustive within their sets; all intersections $A_i \cap B_j$ can be regarded as basic outcomes of a random experiment (because they are the finest events that will be used).
- The probabilities $P\left(A_i \cap B_j\right)$ are called bivariate probabilities.
- The following slide provides an empirical example.

**Table 3.7** Probabilities for Television Viewing and Income Example

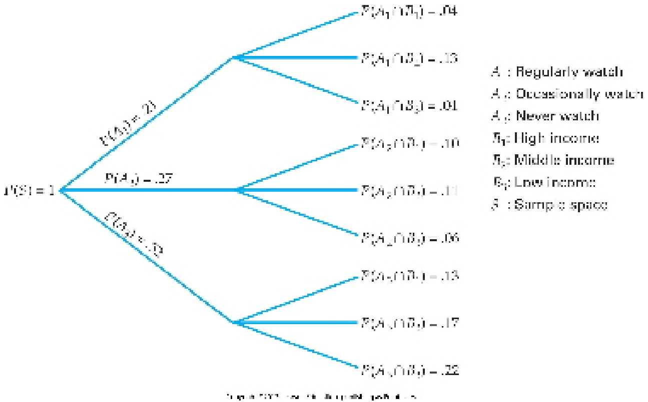
| Viewing Frequency | High Income | Middle Income | Low Income | Totals |
|---|---|---|---|---|
| Regular | 0.04 | 0.13 | 0.04 | 0.21 |
| Occasional | 0.10 | 0.11 | 0.06 | 0.27 |
| Never | 0.13 | 0.17 | 0.22 | 0.52 |
| Totals | 0.27 | 0.41 | 0.32 | 1.00 |





# Joint and Marginal Probabilities

- $P(A_i \cap B_j)$ are joint probabilities, and $P(A_i)$ or $P(B_j)$ are called marginal probabilities and put at the margin of a table as above.
- The marginal probabilities $P(A_i)$ ($P(B_j)$) are obtained by summing the probabilities for a particular row (column) or from tree diagrams as below. [why?]



# Law of Total Probability

- Given a partition of $S$, $\{B_i\}_{i=1}^K$, it is not hard to see that $\{A \cap B_i\}_{i=1}^K$ is a partition of $A$. So we have the law of total probability:

$$P(A) = \sum_{i=1}^{K} P(A \cap B_i). \tag{1}$$

  - Calculating $P(A)$ in this way is called marginalizing over $\{B_i\}_{i=1}^K$, and the resulting probability $P(A)$ is of course the marginal probability of $A$.

- Because $P(A \cap B_i) = P(A|B_i)P(B_i)$, (1) can be rewritten as

$$P(A) = \sum_{i=1}^{K} P(A|B_i)P(B_i). \tag{2}$$

  - The decomposition in (2) is often referred to as conditionalizing on $\{B_i\}_{i=1}^K$.
  - This conditionalizing is useful because it is often hard to assess $P(A)$ directly, but easier to assess conditional probabilities such as $P(A|B_i)$, which are tied to specific context [we will give an example when discussing Bayes' theorem below].

# Conditional Probabilities and Independent Events

**Table 3.8** Conditional Probabilities of Viewing Frequencies, Given Income Levels

| Viewing Frequency | High Income | Middle Income | Low Income |
|---|---|---|---|
| Regular | 0.15 | 0.32 | 0.12 |
| Occasional | 0.37 | 0.27 | 0.19 |
| Never | 0.48 | 0.41 | 0.69 |

- $\sum_{i=1}^{H} P\left(A_i|B_j\right) = \sum_{i=1}^{H} \frac{P(A_i \cap B_j)}{P(B_j)} = \frac{\sum_{i=1}^{H} P(A_i \cap B_j)}{P(B_j)} = \frac{P(B_j)}{P(B_j)} = 1$, i.e., $P\left(\cdot|B_j\right)$ is like a "probability".
- The joint and marginal probabilities can also be used to check whether paired events are statistically independent: $P\left(A_i \cap B_j\right) = P\left(A_i\right) P\left(B_j\right)$?
  - e.g., $P\left(A_2 \cap B_1\right) = 0.1 \neq 0.27 \times 0.27 = P\left(A_2\right) P\left(B_1\right)$.
- For a pair of events, $A$ and $B$, $A$ is partitioned into $A_i, i = 1, \cdots, H$, and $B$ is partitioned into $B_j, j = 1, \cdots, K$. If every $A_i$ is independent of every $B_j$, then $A$ and $B$ are independent events.
  - "viewing frequency" and "income" are not independent since $A_2$ and $B_1$ are not.

# Odds

- The odds in favor of a particular event are given by the ratio of the probability of the event divided by the probability of its complement. The odds in favor of $A$ are

$$\text{Odds} = \frac{P(A)}{1 - P(A)} = \frac{P(A)}{P(\bar{A})}.$$

- Conversely, we can convert the odds in favor of $A$ to the probability of $A$, e.g., the odds in favor of $A$, 2 to 1, implies

$$\frac{2}{1} = \frac{P(A)}{1 - P(A)},$$

i.e., $P(A) = 0.67$.

# (*) Overinvolvement Ratios [Next Tutorial]

- Sometimes, the desired conditional probabilities are hard to obtain due to high enumeration costs or some critical, ethical, or legal restrictions, but alternative conditional probabilities are available.
- Given an event $A_1$, and two mutually exclusive and collectively exhaustive events $B_1$ and $B_2$, the overinvolvement ratio is defined as

$$\frac{P(A_1|B_1)}{P(A_1|B_2)}.$$

  - e.g., $A_1$ is "seeing our advertisement", $B_1$ is "purchasing our products" and $B_2 = \overline{B_1}$. We want to know whether advertising influences purchase behavior, but we only observes $P(A_1|B_1)$ and $P(A_1|B_2)$.
- An overinvolvement ratio greater than 1 implies that event $A_1$ increases the conditional odds ratio in favor or $B_1$:

$$\frac{P(B_1|A_1)}{P(B_2|A_1)} > \frac{P(B_1)}{P(B_2)}.$$

  - e.g., in the above example, the overinvolvement ratio greater than 1 implies that advertising influences purchase behavior.
- Why? $\frac{P(B_1|A_1)}{P(B_2|A_1)} = \frac{P(A_1 \cap B_1)/P(A_1)}{P(A_1 \cap B_2)/P(A_1)} = \frac{P(A_1|B_1)P(B_1)/P(A_1)}{P(A_1|B_2)P(B_2)/P(A_1)} = \frac{P(A_1|B_1)}{P(A_1|B_2)} \cdot \frac{P(B_1)}{P(B_2)} > \frac{P(B_1)}{P(B_2)}$ if $\frac{P(A_1|B_1)}{P(A_1|B_2)} > 1$.

# Bayes' Theorem

# Bayes' Theorem

- Bayes' Theorem: For two events $E$ and $H$,

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}. \text{ [figure here]} \tag{3}$$

- Alternative Form: Let $H_1, \cdots, H_K$ be a partition of $S$ and $E$ be some other event. Then

$$P(H_i|E) = \frac{P(E|H_i)P(H_i)}{\sum_{j=1}^K P(E|H_j)P(H_j)},$$

  where from (2), $P(E) = \sum_{i=1}^K P(E|H_i)P(H_i)$.
  - The advantage of this form of Bayes' theorem is that the probabilities it involves are often those that are available as argued in (2) [see the example below].

# History of Bayes' Theorem

Thomas Bayes (1701-1761), English Reverend[2]

[2]He never published what would eventually become his most famous accomplishment; his notes were edited and published after his death by Richard Price.

# Subjective Probabilities Interpretation of Bayes' Theorem

- We can loosely refer to the event $H$ as "hypothesis" and $E$ as "evidence". In many cases, we can easily determine $P(E|H)$ (the probability that a piece of evidence will occur, given that our hypothesis is correct), but it is much harder to figure out $P(H|E)$ (the probability of the hypothesis being correct, given that we obain a piece of evidence). Yet the latter is the question that we most often want to answer in real world.
  - deduction $P(E|H)$ vs. induction $P(H|E)$; the latter is much more difficult than the former to human beings.
- Bayes' theorem provides a mechanism for updating a prior probability of $H$ to a posterior probability when some additional evidence $E$ is available.
- Subjective Probabilities Interpretation of Bayes' Theorem: In (3), we are interested in the probability of hypothesis $H$. $P(H)$ is its prior probability, $E$ is the additional evidence, and $P(H|E)$ is the updated probability of $H$ after observing $E$, termed as the posterior probability of $H$. The updating is through multiplying $P(H)$ by the likelihood ratio $\frac{P(E|H)}{P(E)}$ – the relative improvement on the assessment of evidence $E$'s probability given $H$.
  - The more surprising the evidence $E$, the more convinced one should become of the hypothesis $H$; e.g., $E = \{$Christ rose from the dead$\}$, and $H = \{$Christ is the son of God$\}$.

# Example 3.23: Drug Screening

- In practice, we should first well define $H_i$ and $E$, and then obtain the required probabilities and conditional probabilities in Bayes' theorem, and finally apply Bayes' theorem to get the desired conditional probability.
- Let $D_1$ be the event of actually using performance-enhancing drugs, $D_2 = \overline{D_1}$, $T_1$ be the event that a screening test indicates using drugs. From experiences, $P(D_1) = 0.1$, $P(T_1|D_1) = 0.9$ and $P(T_1|D_2) = 0.1$. How effective is the test?
- Solution: From Bayes' theorem,

$$\begin{aligned} P(D_1|T_1) &= \frac{P(T_1|D_1)P(D_1)}{P(T_1|D_1)P(D_1)+P(T_1|D_2)P(D_2)} = \frac{0.9\times 0.1}{0.9\times 0.1+0.1\times(1-0.1)} \\ &= 0.5 > 0.1 = P(D_1), \\ P(D_2|T_2) &= \frac{P(T_2|D_2)P(D_2)}{P(T_2|D_1)P(D_1)+P(T_2|D_2)P(D_2)} = \frac{(1-0.1)\times(1-0.1)}{(1-0.9)\times 0.1+(1-0.1)\times(1-0.1)} \\ &= 0.988 > 0.9 = P(D_2). \end{aligned}$$

So a negative test result is reliable, but a positive one is not although it enhances the unconditional probability from 0.1 to 0.5.