Chapter 3

Conditional Probability and Independence

Outline of Chapter 3

- 3.1 Conditional probability
- 3.2 The multiplicative rule
- 3.3 Law of total probability
- 3.4 Bayes' formula
- 3.5 Independence
- 3.6 Applications of probability to genetics

Section 3.1

Conditional Probability

Definition of Conditional Probability

- **Definition.** If $P(B) > 0$, the **conditional probability** of A given B, denoted by $P(A \mid B)$, is

$$P(A \mid B) = \frac{P(AB)}{P(B)}.$$

- Do the following two problems have the same answer?

Example 3.2 From the set of all families with two children, *a family* is selected at random and is found to have a girl. What is the probability that the other child of the family is a girl? Assume that in a two-child family all sex distributions are equally probable.

Example 3.3 From the set of all families with two children, *a child* is selected at random and is found to be a girl. What is the probability that the second child of this girl's family is also a girl? Assume that in a two-child family all sex distributions are equally probable.

- Let's see the analysis. G G

B B G B

Example 3.2

- From the set of all families with two children, a family is selected at random and is found to have a girl.
- What is the probability that the other child of the family is a girl?
  - Assume that in a two-child family all sex distributions are equally probable.
- **Solution.** Let B and A be the events that the family has a girl and the family has two girls, respectively.
- We are interested in $P(A \mid B)$.
- Now, in a family with two children there are four equally likely possibilities: (boy, boy), (girl, girl), (boy, girl), (girl, boy),
  - (girl, boy), we mean that the older child is a girl and the younger is a boy

Analysis of Example 3.2 - Continue

- $A = \{(\text{girl}, \text{girl})\}$
- $B = \{(\text{girl}, \text{girl}), (\text{boy}, \text{girl}), (\text{girl}, \text{boy})\}$
- $P(B) = \frac{3}{4}, P(AB) = \frac{1}{4}$
- $P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{1/4}{3/4} = 1/3$

Example 3.3

- From the set of all families with two children, a child is selected at random and is found to be a girl.
- What is the probability that the second child of this girl's family is also a girl?
  - Assume that in a two-child family all sex distributions are equally probable.
- **Solution.**
- Let B be the event that the child selected at random is a girl.
- Let A be the event that the second child of her family is also a girl.
- We want to calculate $P(A \mid B)$.

Analysis of Example 3.3 - Continue

- Now the set of all possibilities is as follows:
  - The child is a girl with a sister,
  - The child is a girl with a brother,
  - The child is a boy with a sister, and
  - The child is a boy with a brother.
- Thus, $P(B) = 2/4$ and $P(AB) = 1/4$.
- Hence, $P(A \mid B) = \frac{P(AB)}{P(B)} = \frac{1/4}{2/4} = \frac{1}{2}$

What happened?

- In Example 3.2, a family is chosen and found to have a girl.
  - Therefore, we have 3 possibilities: (girl, girl), (boy, girl), and (girl, boy)
    - In the 3 possibilities, only one is desirable: (girl, girl).
    - So, the probability is 1/3.
- In Example 3.3, a child is selected randomly.
  - Families with (girl, girl), (boy, girl), and (girl, boy) are not equally likely.
  - The families with (girl, girl) are twice likely to be selected than the families with only one girl.
  - That is, P(girl, girl) = 2P(boy, girl) = 2P(girl, boy).

$$P(\text{girl, girl}) + P(\text{boy, girl}) + P(\text{girl, boy}) = 1$$

$$P(\text{girl, girl}) + \frac{1}{2}P(\text{girl, girl}) + \frac{1}{2}P(\text{girl, girl}) = 1$$

  - This, and the fact that probabilities add up to 1, give P(girl, girl)=1/2.

Theorem 3.1 (Conditional probabilities are probability)

- Conditional probabilities are probability.
  - So they satisfy the same axioms that probabilities do.

- **Theorem 3.1** Let S be the sample space of an experiment, and let B be an event of S with $P(B) > 0$. Then

- $P(A \mid B) \geq 0$ for any event A of S
- $P(S \mid B) = 1$
- If $A_1, A_2, \ldots$ is a sequence of mutually exclusive events, then

$$P\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) = \sum_{i=1}^{\infty} P(A_i|B)$$

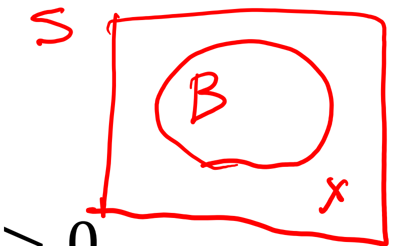
① $P(A|B) = \dfrac{P(AB)}{P(B)} \geq 0$ since $P(AB) \geq 0$ and $P(B) > 0$

② $P(S|B) = \dfrac{P(SB)}{P(B)} = \dfrac{P(B)}{P(B)} = 1$

③ $P\left(\bigcup_{1}^{\infty} A_n \middle| B\right) = \dfrac{P\left(\left(\bigcup_{1}^{\infty} A_n\right) B\right)}{P(B)} = \dfrac{P\left(\bigcup_{1}^{\infty} (A_n B)\right)}{P(B)}$

$$= \frac{\sum_{n=1}^{\infty} P(A_n B)}{P(B)} = \sum_{n=1}^{\infty} \frac{P(A_n B)}{P(B)} = \sum_{n=1}^{\infty} P(A_n|B).$$

Reduction of Sample Space



- Let B be an event of a sample space S with $P(B) > 0$.
- For a subset A of B, define $Q(A) = P(A \mid B)$.
- Then Q is a function from the set of subsets of B to [0, 1].
- Clearly, $Q(A) \geq 0, Q(B) = P(B \mid B) = 1$ and,
- By Theorem 3.1, if $A_1, A_2, \ldots$ is a sequence of mutually exclusive subsets of B, then

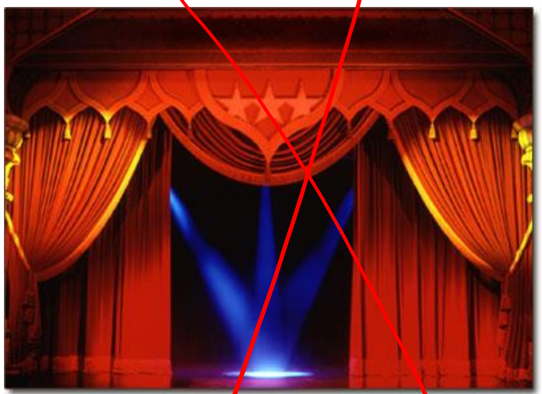
$$Q\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i \mid B\right) = \sum_{i=1}^{\infty} P(A_i \mid B) = \sum_{i=1}^{\infty} Q(A_i)$$

- Note that while P is defined for all subsets of S, the probability function Q is defined only for subsets of B.
- Therefore, for Q, the sample space is reduced from S to B.

Example 3.7

- A child mixes 10 good and 3 dead batteries.
- To find the dead batteries, his father tests them one by one and without replacement.
- If the first 4 batteries tested are all good, what is the probability that the fifth one is dead?
- Solution by reducing the sample space
  - Four batteries have been tested.
  - There are 9 remained.
  - The reduced sample space has 9 batteries consisting of 6 good batteries and 3 dead ones.
  - Since each battery is picked with equal probability, the probability that the next battery is a dead one is 3/9=1/3.

Example 3.8



The host reveals

The player picks

Example 3.9

- On a TV game show, there are 3 curtains.
- Behind two of the curtains, there is nothing.
- Behind the third curtain, there is a prize that the player might win.
- The probability that the prize is behind a given curtain is 1/3.
- The rule of the game
  - The player randomly picks a curtain.
  - The host of the show, who knows behind which curtain the prize is, will pull back a curtain other than the one picked by the player.
  - The host will not pull back the curtain selected by the player, nor will he pulls back the curtain with the prize, if different from the player's choice.
  - At this point, the host offers the player an opportunity to change his choice.
- Should the player change his choice?

Solution of Example 3.9

- We will show that the conditional probability that the player wins, given that he always changes his original choice is 2/3.
  - Therefore, the player should always change his choice.
- Suppose that the prize is behind curtain 1 and the player always changes his choice.
- These two assumptions reduce the sample space.
  - The elements of the reduced sample space can be described by 3-tuples (x, y, z), where
    - x is the curtain the player guesses first,
    - y is the curtain the host pulls back, and
    - z is the curtain that the player switches to.
- Under these assumptions, the reduced sample space is

$$S=\{(1, 2, 3),(1, 3, 2),(2, 3, 1),(3, 2, 1)\}$$

(2, 1, 3)

- The reduced sample space is

$$S=\{(1,2,3),(1,3,2),(2,3,1),(3,2,1)\}$$

The player guesses curtain 1

The player guesses curtain 2

The player guesses curtain 3

- The event that the player wins is $\{(2,3,1),(3,2,1)\}$.
- Since the player guesses a curtain with probability 1/3,

$$P(\{(2,3,1)\})=P(\{(3,2,1)\})=P(\{(1,2,3),(1,3,2)\})=1/3$$

- This shows that no matter what values are assigned to $P(\{(1,2,3)\})$ and $P(\{(1,3,2)\})$, as long as their sum is 1/3, the conditional probability that the player will win is $P(\{(2,3,1)\})+P(\{(3,2,1)\})=2/3$.


日本
台湾
福山雅治
破案天才2013
伽利略
www.niotv.com


無実を証明できる確率、0.1%。
99.9
日曜劇場 99.9 刑事専門弁護士 4月17日スタート 日曜よる9時
初回25分拡大スペシャル


KKTV 全網獨播
你給我5000圓、我給你唯一真相
不在場證明
破解少女
アリバイ崩し承ります


密室パズルミステリー
鍵のかかった部屋
4.16〈月〉9:00pm Start
夜9時～10時9分放送
大野智 戸田恵梨香 佐藤浩市
月9初 全エピソード 密室事件
密室は、破れました。


貴族探偵
毎週月曜ヨル9時


雅芝V2
破案貴族帥執事
INTRODUCTION | CHARACTER | STORY | BACK | TOP

By Theorem 3.1, the function Q defined above by $Q(A) = P(A \mid B)$ is a probability function. Therefore, $Q(A) = P(A \mid B)$ satisfies the theorems stated for P. In particular, for all choices of B, $P(B) > 0$,

1. $P(\emptyset \mid B) = 0$.

2. $P(A^c \mid B) = 1 - P(A \mid B)$.

3. If $C \subseteq A$, then $P(AC^c \mid B) = P(A - C \mid B) = P(A \mid B) - P(C \mid B)$.

4. If $C \subseteq A$, then $P(C \mid B) \leq P(A \mid B)$.

5. $P(A \cup C \mid B) = P(A \mid B) + P(C \mid B) - P(AC \mid B)$.

6. $P(A \mid B) = P(AC \mid B) + P(AC^c \mid B)$.

7. To calculate $P(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n \mid B)$, we calculate conditional probabilities of all possible intersections of events from $\{A_1, A_2, \ldots, A_n\}$, given B, add the conditional probabilities obtained by intersecting an odd number of the events, and subtract the conditional probabilities obtained by intersecting an even number of events.

8. For any increasing or decreasing sequences of events $\{A_n, n \geq 1\}$,

$$\lim_{n \to \infty} P(A_n \mid B) = P(\lim_{n \to \infty} A_n \mid B).$$

Section 3.2

Law of Multiplication

- To compute $P(AB)$, one can compute $P(A \mid B)$ or $P(B \mid A)$ depending on which is easier to compute
- Then we use the definition of condition probability

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

$$P(AB) = P(A)P(B \mid A) = P(B)P(A \mid B) \qquad (3.5)$$

to compute $P(AB)$

Example 3.10

- Suppose that five good fuses and two defective fuses have been mixed up.
- To find the defective fuses, we test them one-by-one, at random and without replacement.
- What is the probability that we find both of the defective fuses in the first two tests?
- **Solution.** Let D_1 and D_2 be the events of finding a defective fuse in the first and second tests, respectively.
- Using (3.5)

$$P(D_1 D_2) = P(D_1)P(D_2|D_1) = \frac{2}{7} \times \frac{1}{6} = \frac{1}{21}$$

Generalization

- Assume that $P(AB) > 0$.
- It follows that $P(A) > 0$.

$$
\begin{aligned}
P(ABC) &= P(C|AB)P(AB) \\
&= P(C|AB)P(B|A)P(A)
\end{aligned}
$$

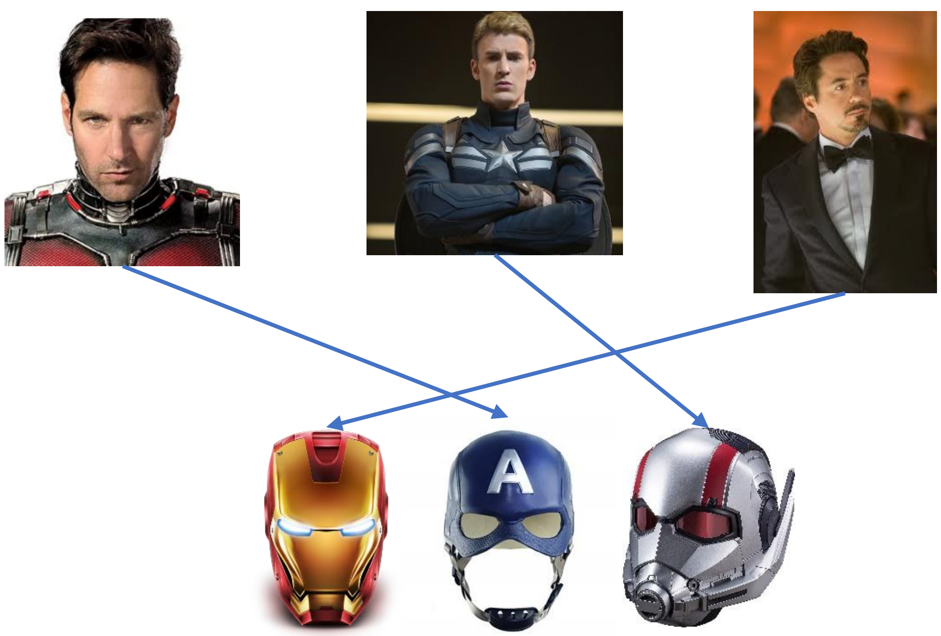
Theorem 3.2 *If* $P(A_1A_2A_3\cdots A_{n-1}) > 0$, *then*

$$
\begin{aligned}
&P(A_1A_2A_3\cdots A_{n-1}A_n) \\
&\quad = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1A_2)\cdots P(A_n \mid A_1A_2A_3\cdots A_{n-1}).
\end{aligned}
$$

Example – A Hat Matching Problem

- n men throw their (distinct) helmets on the floor
- Everyone picks up a helmet randomly
- What is the probability that exactly k out of n people have matched?

$$P(\text{no match}) = \sum_{i=0}^{n} (-1)^i / i!$$



Sheldon Ross, ``A first course in probability,” Example 2g, p. 63, 8-th ed.

Solution

k set n

- Focus our attention on a ***particular*** set of k people and determine the probability that these k individuals have matches and no one else does.
- Let E be the event that everyone in this set has a match.
- Let G be the event that none of the other $n - k$ people have a match.
- We are interested in $P(EG) = P(E)P(G \mid E)$.
- Let $F_i, i = 1, 2, \ldots, k$, be the event that the i-th member of the set of k individuals has a match. Then,

$$\begin{aligned} P(E) &= P(F_1 F_2 \cdots F_k) \\ &= P(F_1)P(F_2|F_1)P(F_3|F_2F_1)\cdots P(F_k|F_1\cdots F_{k-1}) \\ &= \frac{1}{n}\frac{1}{n-1}\frac{1}{n-2}\cdots\frac{1}{n-k+1} \\ &= \frac{(n-k)!}{n!} \end{aligned}$$

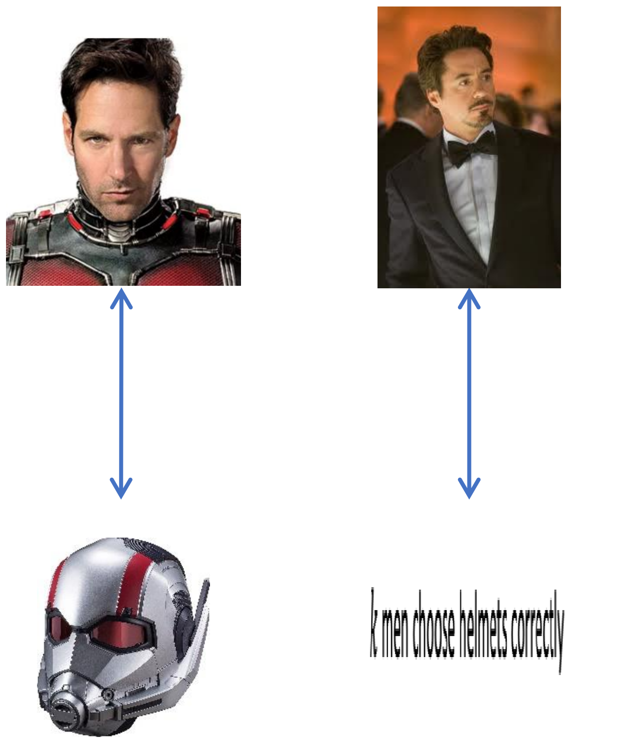
Solution - Continue

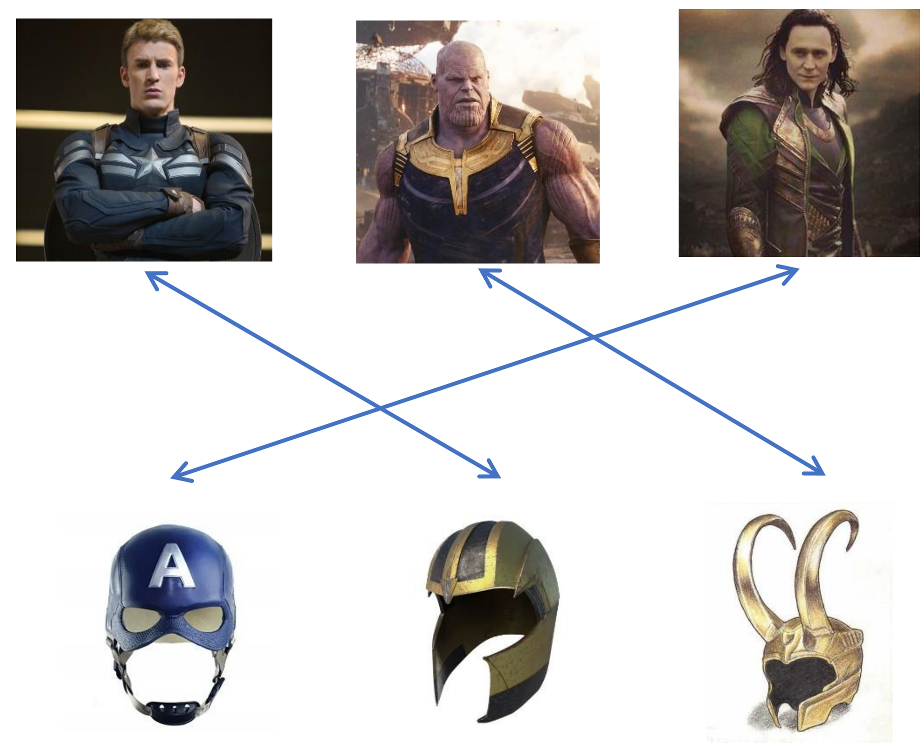
- Now we make a **crucial observation** that is useful in solving many probability problems as follows.
- $P(G \mid E)$ is the same as the probability of no match in a problem of $n - k$ men randomly picking helmets out of $n - k$ helmets.
  - This smaller problem of size $n - k$ has the same probability structure as the problem of size n.
  - Thus,

$$P(G|E) = P(\text{no match in a problem with } n - k \text{ individuals}) = \sum_{i=0}^{n-k} (-1)^i / i!$$

$n = 5, k = 2$

A smaller problem of size $n - k$





k men choose helmets correctly

$n - k$ men choose helmets incorrectly

$$P(E) = \frac{(n-k)!}{n!}$$

$$P(G|E) = P(\text{no match}) = \sum_{i=0}^{n-k} \frac{(-1)^i}{i!}$$

$$P(EG) = P(G|E)P(E) = \frac{(n-k)!}{n!} \sum_{i=0}^{n-k} \frac{(-1)^i}{i!}$$

$$P(\text{exactly } k \text{ matches}) = P(EG)\binom{n}{k}$$

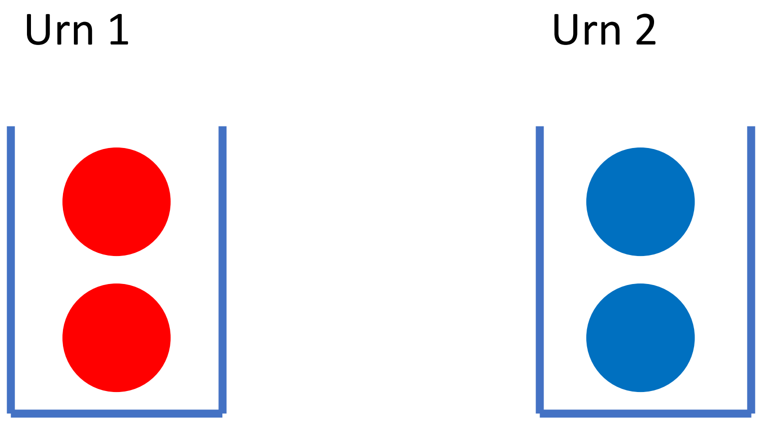
$$\approx \frac{e^{-1}}{k!}$$

Another Example

- Urn 1 contains n red balls and urn 2 contains n blue balls.
- Balls are randomly removed from urn 1 in the following manner:
  - After each removal from urn 1, a ball is taken randomly from urn 2 (if urn 2 is not empty yet) and placed in urn 1.
- This process continues until all the balls have been removed.
  - That is, there are $2n$ removals totally.
- Find $P(R)$, where R is the event that the final ball removed from urn 1 is red.

Sheldon Ross, ``A first course in probability," Example 3h, p. 70, 8-th ed.

Animation of the Problem

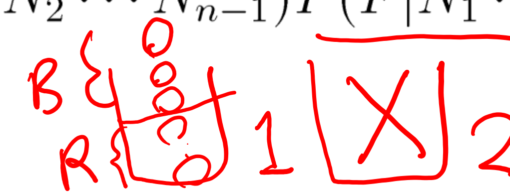


Solution

- Focus attention on ***any particular*** red ball, and let F be the event that this ball is the final one selected.
- In order for F to occur, the ball in question must still be in the urn after the first n balls have been removed (at which time urn 2 is empty).
- Let N_i be the event that this ball is **not** the i-th ball to be removed. Then,

$$\boxed{F \subset N_n \subset N_{n-1} \subset \cdots \subset N_1}$$

$$\begin{aligned} P(F) &= P(N_1 N_2 \cdots N_n F) \\ &= P(N_1)P(N_2|N_1)\cdots P(N_n|N_1 N_2 \cdots N_{n-1})P(F|N_1 \cdots N_n) \\ &= \left(1-\frac{1}{n}\right)\cdots\left(1-\frac{1}{n}\right)\frac{1}{n} \end{aligned}$$



  - This formula uses the fact that the conditional probability that the ball under consideration is the final ball to be removed, given that it is still in urn 1 when only n balls remains, is, by **symmetry**, $1/n$.

- We number the n red balls and let R_j be the event that red ball j is the final ball removed.
- Then,

$$P(R_j) = \left(1 - \frac{1}{n}\right)^n \frac{1}{n}$$

- Because events $\{R_j\}$ are mutually exclusive, we have

$$P(R) = P\left(\cup_{j=1}^{n} R_j\right) = \sum_{j=1}^{n} P(R_j) = \left(1 - \frac{1}{n}\right)^n \approx e^{-1}.$$

for large n.

$\frac{f}{g} \to ?$ $\quad f \to 0$, $g \to 0$

$\approx \frac{f'}{g'}$ L'Hospital's rule

Homework 3

- Section 3.1: A.18, B.22, B.23, B.25
- Section 3.2: A.10, B.13, B.14

- Due date: 5 pm, Friday, March 17, 2022

- Hint for problem B.25 of Section 3.1
  - A family with two girls is **twice** likely to have a girl named Mary than a family with one girl.

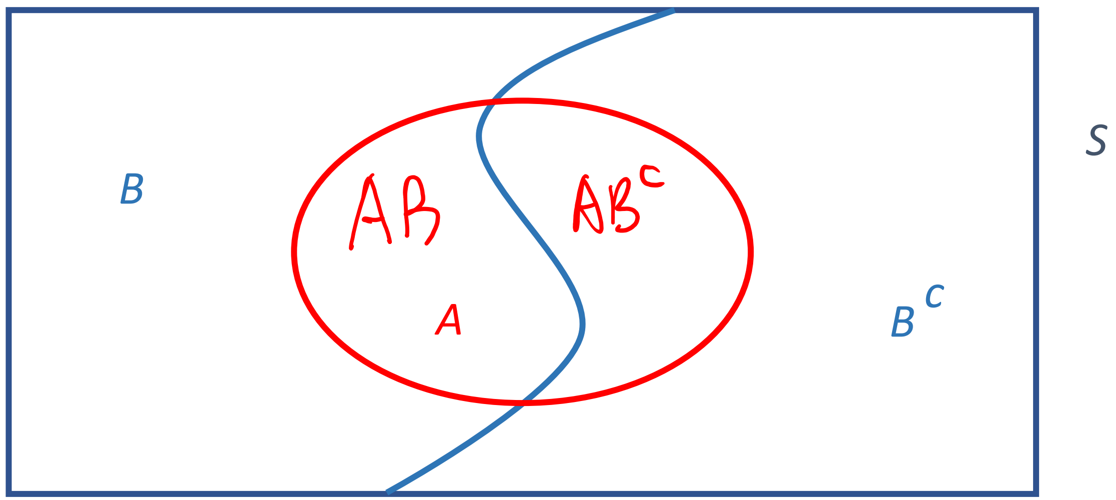
Section 3.3

Law of Total Probability

Theorem 3.3 Law of Total Probability

- **Theorem 3.3** Let B be an event with $P(B) > 0$ and $P(B^c) > 0$. Then for any event A,

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



Theorem 3.3 and its Proof

- **Theorem 3.3** Let B be an event with $P(B) > 0$ and $P(B^c) > 0$. Then for any event A,

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

$$P(A|C) = P(A|BC)P(B|C) + P(A|B^cC)P(B^c|C)$$

AB, AB^c mutally exclusive

- **Proof.**
- By Theorem 1.7,

$$P(A) = P(AB) + P(AB^c) \tag{3.7}$$

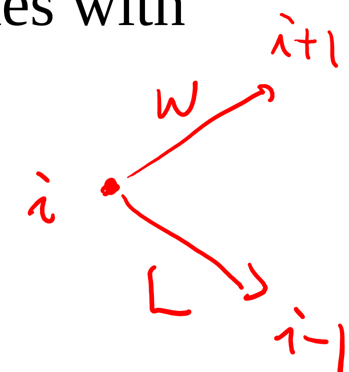
- Now $P(B) > 0$ and $P(B^c) > 0$.
- These imply that $P(AB) = P(A \mid B)P(B)$ and $P(AB^c) = P(A \mid B^c)P(B^c)$
- Putting these in (3.7) proves theorem

Example 3.15 (Gambler's Ruin Problem)

- Gamblers A and B play a series of gambling games.
- In each game, A or B wins with an equal probability of 1/2.
  - In each game, the loser gives the winner one dollar.
- Assume that A begins with a dollars and B begins with b dollars.
- The series of games is over if anyone loses all his/her money.
  - The gambler who loses is said to be ruined.
- What is the probability that
  - A is eventually ruined,
  - B is eventually ruined, or
  - the games go on forever.

Solution of Example 3.15

- Let E be the event that A is ruined eventually.
- Let D_i be the event that A begins a series of games with i dollars initially.
- We are interested in $P(E|D_i)$.
- Denote $p_i = P(E|D_i)$ and derive an equation.

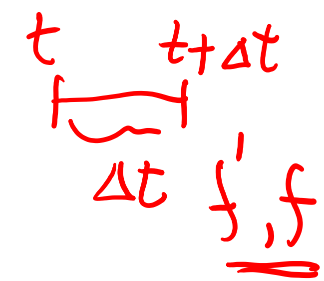


$$
\begin{aligned}
P(E|D_i) &= P(E|D_i, \{\text{A wins a game}\})P(\text{A wins a game}) \\
&\quad + P(E|D_i, \{\text{A loses a game}\})P(\text{A loses a game}) \\
&= P(E|D_{i+1})P(\text{A wins a game}) + P(E|D_{i-1})P(\text{A loses a game})
\end{aligned}
$$

$$
p_i = p_{i+1} \cdot \frac{1}{2} + p_{i-1} \cdot \frac{1}{2}
$$

Difference equation

differential



$$
p_0 = 1 \text{ and } p_{a+b} = 0
$$

- $p_i = p_{i+1}/2 + p_{i-1}/2$.
- $p_0 = 1$ because if A starts with 0 dollars, he/she is already ruined.
- $p_{a+b} = 0$ because the capital of A reaches $a + b$, then B is ruined; thus $p_{a+b} = 0$.
- Rewrite the equation above as $p_{i+1} - p_i = p_i - p_{i-1}$.
- Letting $p_1 - p_0 = \alpha$, we get

$$\begin{aligned} p_i - p_{i-1} &= p_{i-1} - p_{i-2} = p_{i-2} - p_{i-3} = \ldots \\ &= p_2 - p_1 = p_1 - p_0 = \alpha. \end{aligned}$$

- Thus,

$$\begin{aligned} p_1 &= p_0 + \alpha \\ p_2 &= p_1 + \alpha = p_0 + \alpha + \alpha = p_0 + 2\alpha \\ &\vdots \\ p_i &= p_0 + i\alpha \\ &\vdots \end{aligned}$$

- Now $p_0 = 1$ gives $p_i = 1 + i\alpha$.
- But $p_{a+b} = 0$; thus $0 = 1 + (a + b)\alpha$.
- This gives

$$\alpha = -1/(a + b)$$

- Therefore,

$$p_i = 1 - \frac{i}{a+b} = \frac{a+b-i}{a+b}$$

- In particularly, $p_a = b/(a + b)$.
- The probability that A will be ruined is $b/(a + b)$.

P(A eventually ruins) + P(B ruins)
+ P(goes on forever) = 1

- Similarly, the probability that B will be ruined if he/she starts with i dollars is

$$q_i = \frac{a+b-i}{a+b}.$$

- With $i = b$, the probability that B will be ruined is $\frac{a}{a+b}$.
- Since these two probabilities add up to 1, it follows that the probability that the games go on for ever is zero.

A Remark

- **Conditional probabilities satisfy the law of total probability**.
- According to Section 3.1, page 81, $Q(A) = P(A|B)$ is a probability function and possesses all the properties of probability functions.
- Thus, by the law of total probability,

$$Q(A) = Q(A \mid C)Q(C) + Q(A \mid C^c)Q(C^c)$$

- Therefore, replacing $Q(A)$ by its definition, we have

$$P(A \mid B) = P(A \mid BC)P(C \mid B) + P(A \mid BC^c)P(C^c \mid B)$$

Definition of Partition

- **Definition.** Let $\{B_1, B_2, \cdots, B_n\}$ be a set of nonempty subsets of the sample space S of an experiment. If the events $B_1, B_2, \cdots, B_n$ are mutually exclusive and $\bigcup_{i=1}^{n} B_i = S$, the set $\{B_1, B_2, \cdots, B_n\}$ is called a partition of S.

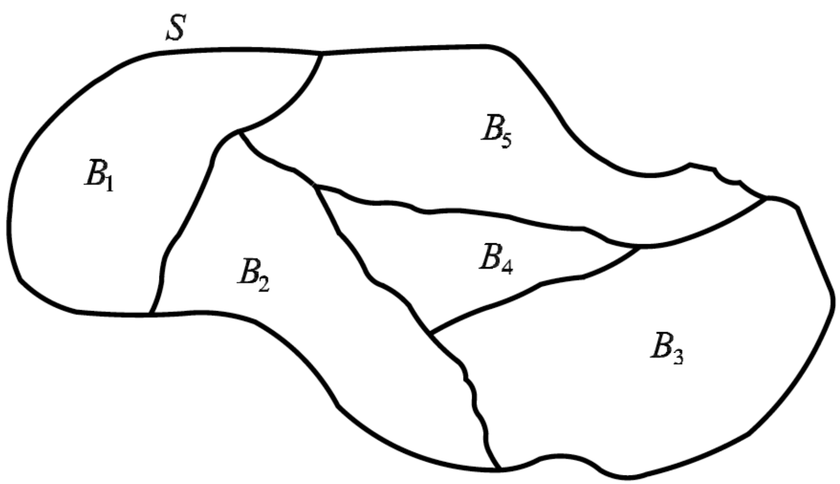


Figure 3.3 Partition of the given sample space S.

Theorem 3.4 (Law of Total Probability)

- **Theorem** 3.4 (Law of Total Probability) If $\{B_1, B_2, \cdots, B_n\}$ is a partition of the sample space of an experiment and $P(B_i) > 0$ for $i = 1, 2, \cdots, n$, then for any event A of S,

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_n)P(B_n)$$
$$= \sum_{i=1}^{n} P(A \mid B_i)P(B_i)\,.$$

- More generally, let $\{B_1, B_2, \cdots\}$ be a sequence of mutually exclusive events of S such that $\bigcup_{i=1}^{\infty} B_i = S$.
- Suppose that, for all $i \geq 1, P(B_i) > 0$. Then for any event A of S,

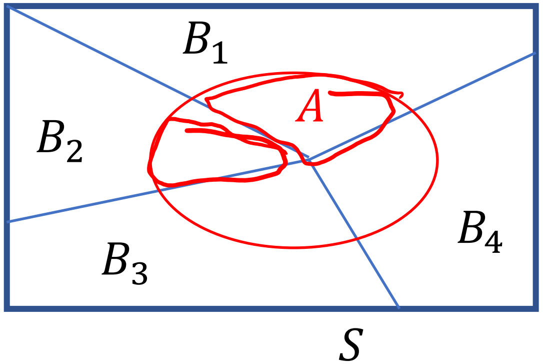
$$P(A) = \sum_{i=1}^{\infty} P(A \mid B_i)P(B_i).$$

Proof of Theorem 3.4

$$\boxed{P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_n)P(B_n)}$$

Proof: Since $B_1, B_2, \ldots, B_n$ are mutually exclusive, $B_iB_j = \emptyset$ for $i \neq j$. Thus $(AB_i)(AB_j) = \emptyset$ for $i \neq j$. Hence $\{AB_1, AB_2, \ldots, AB_n\}$ is a set of mutually exclusive events. Now

$$S = B_1 \cup B_2 \cup \cdots \cup B_n$$



gives

therefore,

$$P(A) = P(AB_1) + P(AB_2) + \cdots + P(AB_n).$$

But $P(AB_i) = P(A \mid B_i)P(B_i)$ for $i = 1, 2, \ldots, n$, so

$$P(A) = P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_n)P(B_n).$$

The proof of the more general case is similar. ♦

Section 3.4

Bayes' Formula

$P(A|B)$ $P(B|A)$

$$P(B_3 \mid A) = \frac{P(B_3 A)}{P(A)}$$

multiplication rule

$$P(B_3 A) = P(A \mid B_3) P(B_3)$$

law of total prob

$$P(A) = P(A \mid B_1) P(B_1) + P(A \mid B_2) P(B_2) + P(A \mid B_3) P(B_3)$$

$$P(B_3 \mid A) = \frac{P(B_3 A)}{P(A)}$$

$$= \frac{P(A \mid B_3) P(B_3)}{P(A \mid B_1) P(B_1) + P(A \mid B_2) P(B_2) + P(A \mid B_3) P(B_3)}$$

Theorem 3.5 (Bayes' Theorem)

Theorem 3.5 (Bayes' Theorem) *Let* $\{B_1, B_2, \ldots, B_n\}$ *be a partition of the sample space* S *of an experiment. If for* $i = 1, 2, \ldots, n$, $P(B_i) > 0$, *then for any event* A *of* S *with* $P(A) > 0$,

$$P(B_k \mid A) = \frac{P(A \mid B_k)P(B_k)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2) + \cdots + P(A \mid B_n)P(B_n)}.$$

In statistical applications of Bayes' theorem, $B_1, B_2, \ldots, B_n$ are called **hypotheses**, $P(B_i)$ is called the **prior** probability of B_i, and the conditional probability $P(B_i \mid A)$ is called the **posterior** probability of B_i after the occurrence of A.

Example 3.23

- A box contains 7 red and 13 blue balls.
- Two balls are removed randomly without their colors being seen. without replacement
- Now a third ball is drawn randomly and observed to be red, what is the probability that both of the discarded balls were blue?

Solution of Example 3.23

- Let BB, BR, and RR be the events that the discarded balls are blue and blue, blue and red, red and red, respectively.
- Let R be the event that the third ball is red.
- The problem is to find $P(BB|R)$.

$\frac{6}{18}$

$$P(BB \mid R) = \frac{P(R \mid BB)P(BB)}{P(R \mid BB)P(BB) + \underline{P(R \mid BR)}P(BR) + P(R \mid RR)P(RR)}.$$

Now

$$P(BB) = \frac{13}{20} \times \frac{12}{19} = \frac{39}{95}, \qquad P(RR) = \frac{7}{20} \times \frac{6}{19} = \frac{21}{190},$$

and

$$P(BR) = \frac{13}{20} \times \frac{7}{19} + \frac{7}{20} \times \frac{13}{19} = \frac{91}{190},$$

Solution of Example 3.23

$$P(BB \mid R) = \frac{\dfrac{7}{18} \times \dfrac{39}{95}}{\dfrac{7}{18} \times \dfrac{39}{95} + \dfrac{6}{18} \times \dfrac{91}{190} + \dfrac{5}{18} \times \dfrac{21}{190}} \approx 0.46.$$

Example

- In many real-life problems, people may be psychologically confuse $P(A|B)$ with $P(B|A)$. It is demonstrated by the following example.

- A lab blood test is 95% effective in detecting a certain disease when it is, in fact, present.
- However, the test also yields a "false positive" result for 1% of the healthy persons tested.
  - That is, if a healthy person is tested, then, with probability 0.01, the test result will imply that he/she has the disease.
- If 0.5% of the population actually has the disease, what is the probability that a person has the disease given that his/her test result is positive?
  - Many people have high faith in medical tests and believe the conclusion drawn by the tests. However, this can be wrong!

Sheldon Ross, "A first course in probability", Example 3d, p. 67, 8-th ed.

Solution

- Randomly select a person from the population and test him/her.
- Let D be the event that the tested person had the disease.
- Let E be the event that the test result is positive.
- The desired probability is $P(D|E)$.
  - We are given $P(E|D) = 0.95$.
- One would psychically expect that both $P(E|D)$ and $P(D|E)$ are large. Is it true?

$$\begin{aligned} P(D|E) &= \frac{P(DE)}{P(E)} \\ &= \frac{P(E|D)P(D)}{P(E|D)P(D) + P(E|D^c)P(D^c)} \\ &= \frac{0.95 \times 0.005}{0.95 \times 0.005 + 0.01 \times 0.995} \\ &= 0.323 \end{aligned}$$

Section 3.5

Independence

Definition of Independence

- If $P(A|B) \neq P(A)$, we say that A is dependent of B.
- $P(AB) = P(A|B)P(B) = P(A)P(B)$ if A is independent of B.

Definition *Two events A and B are called* ***independent*** *if*

$$P(AB) = P(A)P(B).$$

If two events are not independent, they are called ***dependent****. If A and B are independent, we say that* $\{A, B\}$ *is an independent set of events.*

- Note that with this definition, we did not require that $P(A)$ or $P(B)$ to be strictly positive
  - Thus, any event A with $P(A) = 0$ or $P(A) = 1$ is independent of every event B.

Theorem 3.6

- **Theorem 3.6** *If A and B are independent, then A and B^c are independent as well.*
- **Proof.**
- By Theorem 1.7,

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

$A = (AB) \cup (AB^c)$

disjoint

- Therefore,

$$P(AB^c) = P(A) - P(AB) = P(A) - P(A)P(B)$$
$$= P(A)[1 - P(B)] = P(A)P(B^c)$$

- **Corollary.** *If A and B are independent, then A^c and B^c are independent as well.*
- **Proof.** Apply Theorem 3.6 twice.

Generalization

- **Definition.** *The set of events* $\{A_1, A_2, \ldots, A_n\}$ *is called independent if for every subset* $\{A_{i_1}, A_{i_2}, \ldots, A_{i_k}\}$, $k \geq 2$, *of* $\{A_1, A_2, \ldots, A_n\}$,

$$P(A_{i_1} A_{i_2} \cdots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_k}). \tag{3.15}$$

- This definition can be generalized to infinite sets of events.
- Sequence of events $\{A_i\}_{i=1}^{\infty}$ is called independent if for any finite subset $\{A_{i_1}, A_{i_2}, \ldots, A_{i_k}\}$, $k \geq 2$, (3.15) is valid.

Example 3.31

- The following example shows that $P(ABC) = P(A)P(B)P(C)$ does not imply that $\{A, B, C\}$ is independent
- **Example 3.31** Let an experiment consist of throwing a dice twice.
- Let A be the event that in the second throw the die lands 1, 2, or 5;
- B the event that in the second throw it lands 4, 5, or 6; and
- C the event that the sum of the two outcomes is 9.
- $P(A) = P(B) = \frac{1}{2}, P(C) = \frac{4}{36} = \frac{1}{9}$.
- $P(AB) = \frac{1}{6} \neq \frac{1}{4} = P(A)P(B)$
- $P(AC) = \frac{1}{36} \neq \frac{1}{18} = P(A)P(C)$
- $P(BC) = \frac{1}{12} \neq \frac{1}{18} = P(B)P(C)$
- $P(ABC) = \frac{1}{36} = P(A)P(B)P(C)$

$x_1 = 1, 2, \cdots, 6$

$x_2 = 1, 2, \cdots 6$

$x_1 + x_2 = 9$

$(3,6), (6,3), (5,4)$

$(4,5)$

$\frac{3}{36} = \frac{1}{12}$

Example 3.28

- A spinner is mounted on a wheel, on which there are three equal length arcs A, B, and C.

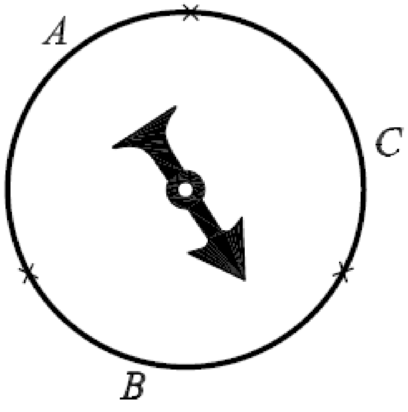


Figure 3.7 Spinner of Example 3.26.

- The spinner is flicked and the player wins 1, 2, and 3 points if the spinner stops on A, B, and C respectively.

- Suppose that the player spins the spinner twice.
- Let E be the event that the player wins 1 point in the first spin and any number of points in the second.
- Let F be the event that he wins a total of 3 points in the two spins.
- Let G be the event that he wins a total of 4 points in two spins.
- The sample space of this experiment has 9 elements which are equally likely.
- $E = \{(1,1),(1,2),(1,3)\}$
- $F = \{(1,2),(2,1)\}$
- $G = \{(1,3),(2,2),(3,1)\}$
- $P(E) = \frac{1}{3}, P(G) = \frac{1}{3}, P(F) = 2/9$
- $P(GE) = P(\{(1,3)\}) = \frac{1}{9} = P(G)P(E)$
- $P(FE) = P(\{(1,2)\}) = \frac{1}{9} \neq P(E)P(F) = \frac{2}{9}$
- E and G are independent and E and F are not

Example – Coupon Collecting

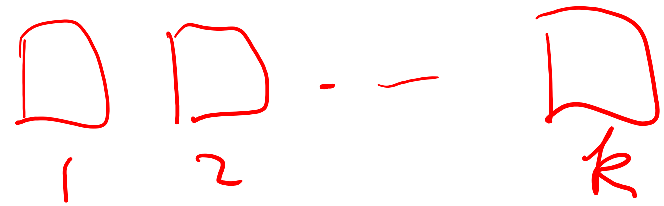
- There are n types of coupons, and each new one collected is independently of type i with probability p_i, where

$$\sum_{i=1}^{n} p_i = 1$$

- Suppose that k coupons are to be collected.
- Let A_i be the event that there is at least one type i coupon among those collected.
- For $i \neq j$, find $P(A_i), P\left(A_i \cup A_j\right)$, and $P(A_i \mid A_j)$

Sheldon Ross, ``A first course in probability,'' Example 4i, p. 85, 8-th ed.

Solution



$$
\begin{aligned}
P(A_i) &= 1 - P(A_i^c) \\
&= 1 - P(\text{no coupon is of type } i) \\
&= 1 - P(\cap_{j=1}^{k}\{\text{coupon } j \text{ is not of type } i\}) \\
&= 1 - \prod_{j=1}^{k} P(\{\text{coupon } j \text{ is not of type } i\}) \\
&= 1 - (1 - p_i)^k
\end{aligned}
$$

$$
\begin{aligned}
P(A_i \cup A_j) &= 1 - P((A_i \cup A_j)^c) \\
&= 1 - P(\text{no coupon is either type } i \text{ or type } j) \\
&= 1 - (1 - p_i - p_j)^k
\end{aligned}
$$

$P(A_i \cup A_j) = P(A_i) + P(A_j)$
$- P(A_i A_j)$

$$P(A_i) = 1 - (1 - p_i)^k$$

$$P(A_i \cup A_j) = 1 - (1 - p_i - p_j)^k$$

$$\begin{aligned} P(A_i A_j) &= P(A_i) + P(A_j) - P(A_i \cup A_j) \qquad \text{(inclusion-exclusion)} \\ &= 1 - (1 - p_i)^k + 1 - (1 - p_i)^k - [1 - (1 - p_i - p_j)^k] \\ &= 1 - (1 - p_i)^k - (1 - p_j)^k + (1 - p_i - p_j)^k \end{aligned}$$

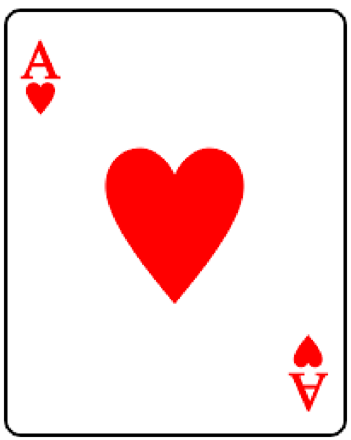
$$\begin{aligned} P(A_i \mid A_j) &= \frac{P(A_i A_j)}{P(A_j)} \\ &= \frac{1 - (1 - p_i)^k - (1 - p_j)^k + (1 - p_i - p_j)^k}{1 - (1 - p_j)^k} \end{aligned}$$

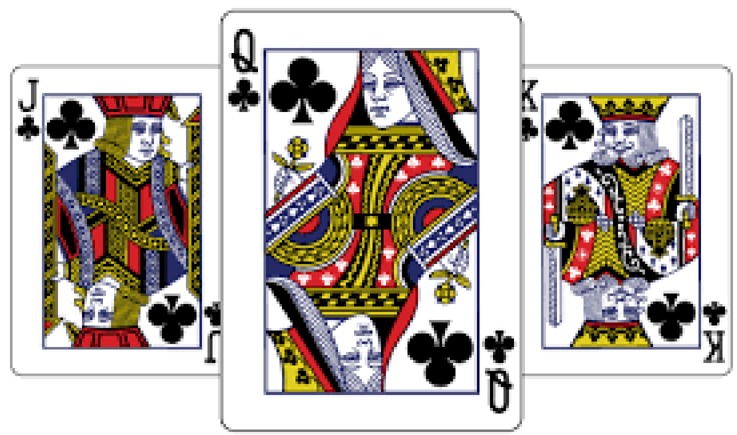
Example 3.34

- Draw cards, one at a time, at random and successfully from an ordinary deck of 52 cards **with replacement**
- What is the probability that an ace card appears before a face card?

$x\ x\ x \ldots \boxed{A} \ldots \boxed{F}$

2, 3





- We present two techniques to solve this problem.
- **Technique 1.** Let E be the event of an ace appearing before a face card.
- Let A, F, and B be the events of ace, face card, and neither in the first experiment, respectively.
- Then, by the law of total probability,

$$P(E) = P(E \mid A)P(A) + P(E \mid F)P(F) + P(E \mid B)P(B)$$

$$P(E) = 1 \times \frac{4}{52} + 0 \times \frac{12}{52} + P(E|B) \times \frac{36}{52}. \tag{3.16}$$

- Now note that since the outcomes of successive experiments are all independent of each other, when the second experiment begins, the whole probability process starts all over again.
- That is, $P(E \mid B) = P(E)$.
- Eq. (3.16) gives

$$P(E) = \frac{4}{52} + P(E) \times \frac{36}{52}.$$

- Thus, $P(E) = 1/4$.

~~A~~ 3 5 · · ~~A~~ $\boxed{A}$

1 2 · · · n-1 n

- **Technique 2.**
- Let A_n be the event that no face card or ace appears on the first $(n-1)$ drawings, and the nth draw is an ace.
- Then, $E =$ {an ace before a face card} $= \bigcup_{n=1}^{\infty} A_n$.
- Now $\{A_n, n \geq 1\}$ forms a sequence of mutually exclusive events because, if $n \neq m$, simultaneous occurrence of A_n and A_m is the impossible event that an ace appears for the first time in the nth and mth draws.
- Hence

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

- To compute $P(A_n)$, note that

$$P(\text{an ace on any draw}) = 1/13$$

$$P(\text{no face card and no ace in any trial}) = (52-16)/52 = 9/13.$$

$\boxed{A} \times$
$\boxed{F} \times$ $\quad$ A
1 2 - - - $n-1$ n
$1/13$

- By the independence of trials we obtain

$$P(A_n) = \left(\frac{9}{13}\right)^{n-1} \frac{1}{13}.$$

- Therefore,

$$P(E) = P\left(\bigcup_{i=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \left(\frac{9}{13}\right)^{n-1} \frac{1}{13}$$

$$= \frac{1}{13} \sum_{n=1}^{\infty} \left(\frac{9}{13}\right)^{n-1}$$

$$= \frac{1}{13} \cdot \frac{1}{1 - 9/13} = 1/4$$

Example 3.37 – Coin Tossing

- Adam tosses a **fair** coin $n + 1$ times.
- Andrew tosses the same coin n times.
- What is the probability that Adam get more heads than Andrew?

$$\boxed{P(H_1 > H_2) = ?}$$

Solution

- Let H_1 and H_2 be the number of heads obtained by Adam and Andrew respectively.
- Also let T_1 and T_2 be the number of tails obtained by Adam and Andrew, respectively.

- Since the coin is fair, we have

$$P(H_1 > H_2) = P(T_1 > T_2)$$

- But,

$$P(T_1 > T_2) = P(n + 1 - H_1 > n - H_2) = P(H_1 \leq H_2).$$

Therefore, $P(H_1 > H_2) = P(H_1 \leq H_2)$. So

$$P(H_1 > H_2) + P(H_1 \leq H_2) = 1$$

implies that

$$P(H_1 > H_2) = P(H_1 \leq H_2) = \frac{1}{2}.$$

$$
\begin{aligned}
P(H_1 > H_2) &= \sum_{i=0}^{n} P(H_1 > H_2 \mid H_2 = i)P(H_2 = i) \text{ (law of total probability)} \\
&= \sum_{i=0}^{n} P(H_1 > i \mid H_2 = i)P(H_2 = i) \\
&= \sum_{i=0}^{n} P(H_1 > i)P(H_2 = i) \qquad \text{(independence)} \\
&= \sum_{i=0}^{n} \sum_{j=i+1}^{n+1} P(H_1 = j)P(H_2 = i) \\
&= \sum_{i=0}^{n} \sum_{j=i+1}^{n+1} \frac{\frac{(n+1)!}{j!(n+1-j)!}}{2^{n+1}} \frac{\frac{n!}{i!(n-i)!}}{2^{n}} \\
&= \frac{1}{2^{2n+1}} \sum_{i=0}^{n} \sum_{j=i+1}^{n+1} \binom{n+1}{j} \binom{n}{i}.
\end{aligned}
$$

A Combinatorial Solution

Another Coin Tossing Problem

- Toss a coin repeatedly.
  - The probability of obtaining a head is p and the probability of getting a tail is $q = 1 - p$.
  - Results of successive tossing are independent.
- What is the probability that a run of n consecutive heads occurs before a run of m consecutive tails?
  - Another two related problems.
  - What is the probability that a run of m consecutive tails occurs before a run of n consecutive heads?
  - What is the probability that the game goes on for ever?

n=3, m=2

Sheldon Ross, ``A first course in probability," Example 5c, p. 95, 8-th ed.

Solution

- Let E be the event that a run of n consecutive heads occurs before a run of m consecutive tails.
- To obtain $P(E)$, we condition on the outcome of the first toss.
  - Some authors call this technique "first-step decomposition".
- Let H be the event that the first toss results in a head and apply the law of total probability

$$P(E) = p\,P(E \mid H) + q\,P(E \mid H^c)$$

- Now we derive
  - $P(E|H)$
  - $P(E|H^c)$
- Given that the first toss was a head, one way to get a run of n heads before a run of m tails would be to have the next $n-1$ tosses result in heads.
  - Let F be the event that the second toss up to the n-th toss are all heads

$$P(E|H) = \underbrace{P(E|FH)}_{=1}P(F|H) + \underbrace{P(E|F^cH)}_{P(E|H^c)}P(F^c|H)$$

  - $P(E|FH) = 1$, since n heads already appear
  - $P(E|F^cH) = P(E|H^c)$, since the first tail wipes out all the previous heads

- For example, let $n = 5$ and consider the following realization.

2 - - - —— n

F^c

This situation would be exactly the same as if we have started with a tail initially

This tail wipes out all the previous heads

- $P(E|F^cH) = P(E|H^c)$

$$
\begin{aligned}
P(E|H) &= P(E|FH)P(F|H) + P(E|F^cH)P(F^c|H) \\
&= P(F|H) + P(E|H^c)P(F^c|H) \\
&= P(F) + P(E|H^c)P(F^c) \\
&= p^{n-1} + P(E|H^c)(1-p^{n-1})
\end{aligned}
$$

Events F and H are independent

- $P(F) = p^{n-1}$
- $P(F^c) = 1 - p^{n-1}$

- We now consider $P(E|H^c)$
- Let G be the event that the second toss up to the m-th toss are all tails
- By the same analysis, we have

$$
\begin{aligned}
P(E|H^c) &= P(E|GH^c)P(G|H^c) + P(E|G^cH^c)P(G^c|H^c) \\
&= 0 + P(E|H)P(G^c|H^c) \\
&= P(E|H)P(G^c|H^c) \\
&= P(E|H)P(G^c) \\
&= P(E|H)(1 - q^{m-1})
\end{aligned}
$$

  - $P(E|GH^c) = 0$, since m tails have appeared
  - $P(E|G^cH^c) = P(E|H)$), since a head wipes out all the accumulation of tails
  - Events G^c and H^c are independent
  - $P(G^c) = 1 - q^{m-1}$

- We have

$$\boxed{P(E|H^c) = (1 - q^{m-1})P(E|H)}$$

$$\boxed{P(E|H) = p^{n-1} + (1 - p^{n-1})P(E|H^c)}$$

- Solving the two equations in red boxes, we obtain

$$\begin{aligned} P(E|H) &= \frac{p^{n-1}}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}} \\ P(E|H^c) &= \frac{(1 - q^{m-1})p^{n-1}}{p^{n-1} + q^{m-1} - p^{n-1}q^{m-1}} \end{aligned}$$

$$\begin{aligned} P(E) &= pP(E|H) + qP(E|H^c) \\ &= \frac{p^{n-1}(1-q^m)}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}} \end{aligned}$$

- Due to the symmetry of the problem, the probability of obtaining a run of m consecutive tails before a run of n consecutive heads would be given by the above equation with p and q interchanged and n and m interchanged.

$$\begin{aligned} &P(\text{run of } m \text{ tails before a run of } n \text{ heads}) \\ &= \frac{q^{m-1}(1-p^n)}{q^{m-1}+p^{n-1}-q^{m-1}p^{n-1}} \end{aligned}$$

- Since these two probabilities add up to 1, it follows that with probability 1, <u>either a run of n heads or a run of m tails will eventually occur</u>.

Homework 4

- Section 3.3: A.14, B.21, B.24
- Section 3.4: A.18, A.24
- Section 3.5: A.24, B.44, B.49, B.50

- You don’t need to submit solution to this homework set of problems.
- Solution will be distributed in the weekend.