Lecture 7: Bayes′ Rule, Revision

Kateřina Staňková

Statistics (MAT1003)

April 26, 2012

Outline

1. **Preliminaries**
   - Partition of a sample space
2. **Bayes′ rule**
   - Theory
   - Examples
3. **A lot of computing . . .**

And now . . .

Partition of a sample space

Partition of a sample space

Let S be a sample space. A set of subsets $B_1, B_2, \ldots, B_k$ of S is called a partition of S if

Partition of a sample space

Let S be a sample space. A set of subsets $B_1, B_2, \ldots, B_k$ of S is called a partition of S if

- $\bigcup_{i=1}^{k} B_i = B_1 \cup B_2 \cup \ldots B_k = S$

Partition of a sample space

Let S be a sample space. A set of subsets $B_1, B_2, \ldots, B_k$ of S is called a partition of S if

- $\bigcup_{i=1}^{k} B_i = B_1 \cup B_2 \cup \ldots B_k = S$
- $B_i \cap B_j = \emptyset$ for all $i, j \in \{1, \ldots, k\}$, $i \neq j$

Partition of a sample space

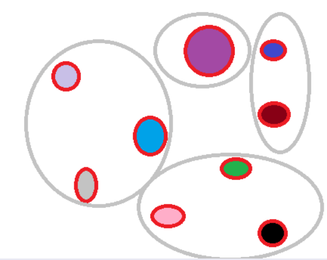
Let S be a sample space. A set of subsets $B_1, B_2, \ldots, B_k$ of S is called a partition of S if

- $\bigcup_{i=1}^{k} B_i = B_1 \cup B_2 \cup \ldots B_k = S$
- $B_i \cap B_j = \emptyset$ for all $i, j \in \{1, \ldots, k\}, i \neq j$

Partition of a sample space

Let S be a sample space. A set of subsets $B_1, B_2, \ldots, B_k$ of S is called a partition of S if

- $\bigcup_{i=1}^{k} B_i = B_1 \cup B_2 \cup \ldots B_k = S$
- $B_i \cap B_j = \emptyset$ for all $i, j \in \{1, \ldots, k\}$, $i \neq j$



Partition of a sample space

Example 1

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

Example 1

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

Partition $\{B_1, B_2, B_3\}$ with $B_1 = \{1, 3\}$, $B_2 = \{5\}$, $B_3 = \{2, 4, 6\}$

Example 1

$S = \{1, 2, 3, 4, 5, 6\}$
Partition $\{B_1, B_2, B_3\}$ with $B_1 = \{1, 3\}$, $B_2 = \{5\}$, $B_3 = \{2, 4, 6\}$

Example 2

Example 1

$S = \{1, 2, 3, 4, 5, 6\}$
Partition $\{B_1, B_2, B_3\}$ with $B_1 = \{1, 3\}$, $B_2 = \{5\}$, $B_3 = \{2, 4, 6\}$

Example 2

$S = [0, 1]$

Partition of a sample space

Example 1

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

Partition $\{B_1, B_2, B_3\}$ with $B_1 = \{1, 3\}$, $B_2 = \{5\}$, $B_3 = \{2, 4, 6\}$

Example 2

$S = [0, 1]$

Partition $\{I_1, I_2\}$ with $I_1 = \left[0, \frac{1}{2}\right)$, $I_2 = \left[\frac{1}{2}, 1\right]$

Example 1

$S = \{1, 2, 3, 4, 5, 6\}$
Partition $\{B_1, B_2, B_3\}$ with $B_1 = \{1, 3\}$, $B_2 = \{5\}$, $B_3 = \{2, 4, 6\}$

Example 2

$S = [0, 1]$
Partition $\{I_1, I_2\}$ with $I_1 = \left[0, \frac{1}{2}\right)$, $I_2 = \left[\frac{1}{2}, 1\right]$

Example 3

Example 1

$S = \{1, 2, 3, 4, 5, 6\}$
Partition $\{B_1, B_2, B_3\}$ with $B_1 = \{1, 3\}$, $B_2 = \{5\}$, $B_3 = \{2, 4, 6\}$

Example 2

$S = [0, 1]$
Partition $\{I_1, I_2\}$ with $I_1 = \left[0, \frac{1}{2}\right)$, $I_2 = \left[\frac{1}{2}, 1\right]$

Example 3

Any sample space S

Example 1

$S = \{1, 2, 3, 4, 5, 6\}$
Partition $\{B_1, B_2, B_3\}$ with $B_1 = \{1, 3\}$, $B_2 = \{5\}$, $B_3 = \{2, 4, 6\}$

Example 2

$S = [0, 1]$
Partition $\{I_1, I_2\}$ with $I_1 = \left[0, \frac{1}{2}\right)$, $I_2 = \left[\frac{1}{2}, 1\right]$

Example 3

Any sample space S
Partition $\{B, B'\}$ for any $B \subseteq S$

Observation

Observation

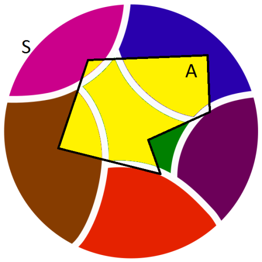
Let $B_1, \ldots, B_k$ be a partition of S. Then for any $A \subseteq S$:

$$\begin{aligned} A = A \cap S &= A \cap (B_1 \cup B_2 \cup \ldots \cup B_k) \\ &= \underbrace{(A \cap B_1)}_{A_1} \cup \underbrace{(A \cap B_2)}_{A_2} \cup \ldots \cup \underbrace{(A \cap B_k)}_{A_k} \end{aligned}$$

Observation

Let $B_1, \ldots, B_k$ be a partition of S. Then for any $A \subseteq S$:

$$\begin{aligned} A = A \cap S &= A \cap (B_1 \cup B_2 \cup \ldots \cup B_k) \\ &= \underbrace{(A \cap B_1)}_{A_1} \cup \underbrace{(A \cap B_2)}_{A_2} \cup \ldots \cup \underbrace{(A \cap B_k)}_{A_k} \end{aligned}$$



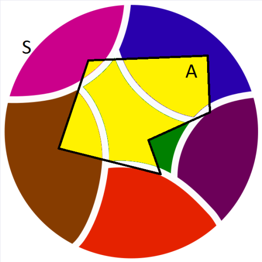
Observation

Let $B_1, \ldots, B_k$ be a partition of S. Then for any $A \subseteq S$:

$$\begin{aligned} A = A \cap S &= A \cap (B_1 \cup B_2 \cup \ldots \cup B_k) \\ &= \underbrace{(A \cap B_1)}_{A_1} \cup \underbrace{(A \cap B_2)}_{A_2} \cup \ldots \cup \underbrace{(A \cap B_k)}_{A_k} \end{aligned}$$

$P(A_i) = P(A \cap B_i) = P(B_i) \cdot P(A|B_i)$

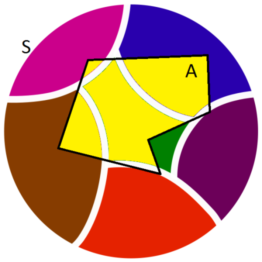
$A_1, \ldots, A_k$ disjoint sets



Observation

Let $B_1, \ldots, B_k$ be a partition of S. Then for any $A \subseteq S$:

$$A = A \cap S = A \cap (B_1 \cup B_2 \cup \ldots \cup B_k)$$
$$= \underbrace{(A \cap B_1)}_{A_1} \cup \underbrace{(A \cap B_2)}_{A_2} \cup \ldots \cup \underbrace{(A \cap B_k)}_{A_k}$$



$P(A_i) = P(A \cap B_i) = P(B_i) \cdot P(A|B_i)$

$A_1, \ldots, A_k$ disjoint sets

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

And now ...

Bayes′ rule

Bayes′ rule

If $P(B_i)$ and $P(A|B_i)$ are given for all i, we can calculate $P(B_i|A)$ as follows:

Bayes′ rule

If $P(B_i)$ and $P(A|B_i)$ are given for all i, we can calculate $P(B_i|A)$ as follows:

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

Bayes′ rule

If $P(B_i)$ and $P(A|B_i)$ are given for all i, we can calculate $P(B_i|A)$ as follows:

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

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

Exercise 2.95 + forget about 40

Exercise 2.95 + forget about 40

- Events:

Exercise 2.95 + forget about 40

- Events:

  C : A randomly chosen person has cancer

Exercise 2.95 + forget about 40

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

Exercise 2.95 + forget about 40

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'

Exercise 2.95 + forget about 40

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:

Exercise 2.95 + forget about 40

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$

Exercise 2.95 + forget about 40

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$

Exercise 2.95 + forget about 40

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$

Exercise 2.95 + forget about 40

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
- Question 1: $P(D) = ?$

Exercise 2.95 + forget about 40

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
- Question 1:$P(D) =?$
- Solution:

$$P(D) =$$

Exercise 2.95 + forget about 40

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
- Question 1:$P(D) =?$
- Solution:

$$P(D) = P(D \cap C) + P(D \cap C')$$

Exercise 2.95 + forget about 40

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
- Question 1: $P(D) = ?$
- Solution:

$$\begin{aligned} P(D) &= P(D \cap C) + P(D \cap C') \\ &= P(C) \cdot P(D|C) + P(C') \cdot P(D|C') \end{aligned}$$

Exercise 2.95 + forget about 40

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
- Question 1: $P(D) =?$
- Solution:

$$\begin{aligned} P(D) &= P(D \cap C) + P(D \cap C') \\ &= P(C) \cdot P(D|C) + P(C') \cdot P(D|C') \\ &= 0.05 \cdot 0.78 + 0.95 \cdot 0.06 = 0.039 + 0.054 = 0.093 \end{aligned}$$

Exercise 2.95

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
  - $P(D) = 0.093 \quad (P(D') = 1 - 0.093 = 0.907)$

Exercise 2.95

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
  - $P(D) = 0.093 \quad (P(D') = 1 - 0.093 = 0.907)$
- Question 2: What is the probability that someone diagnosed with cancer actually has the disease?

Exercise 2.95

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
  - $P(D) = 0.093 \quad (P(D') = 1 - 0.093 = 0.907)$
- Question 2: What is the probability that someone diagnosed with cancer actually has the disease?
  $P(C|D) = ?$

Exercise 2.95

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
  - $P(D) = 0.093 \quad (P(D') = 1 - 0.093 = 0.907)$
- Question 2: What is the probability that someone diagnosed with cancer actually has the disease? $P(C|D) =?$
- Solution:

$$P(C|D) =$$

Exercise 2.95

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
  - $P(D) = 0.093 \quad (P(D') = 1 - 0.093 = 0.907)$
- Question 2: What is the probability that someone diagnosed with cancer actually has the disease? $P(C|D) =?$
- Solution:

$$P(C|D) = \frac{P(C \cap D)}{P(D)}$$

Exercise 2.95

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
  - $P(D) = 0.093 \quad (P(D') = 1 - 0.093 = 0.907)$
- Question 2: What is the probability that someone diagnosed with cancer actually has the disease? $P(C|D) = ?$
- Solution:

$$P(C|D) = \frac{P(C \cap D)}{P(D)} = \frac{P(D|C) \cdot P(C)}{P(D)}$$

Exercise 2.95

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

  Complementary events: C' and D'
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
  - $P(D) = 0.093 \quad (P(D') = 1 - 0.093 = 0.907)$
- Question 2: What is the probability that someone diagnosed with cancer actually has the disease? $P(C|D) =?$
- Solution:

$$P(C|D) = \frac{P(C \cap D)}{P(D)} = \frac{P(D|C) \cdot P(C)}{P(D)} = \frac{0.78 \cdot 0.05}{0.093} = \frac{0.039}{0.093} = \frac{13}{31}$$

Exercise 2.95 - Question 3

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
  - $P(D) = 0.093 \quad (P(D') = 1 - 0.093 = 0.907)$

Exercise 2.95 - Question 3

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
  - $P(D) = 0.093 \quad (P(D') = 1 - 0.093 = 0.907)$

- Question 3: What is the probability that someone who is diagnosed as not having cancer actually has the disease?

Exercise 2.95 - Question 3

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer
- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
  - $P(D) = 0.093 \quad (P(D') = 1 - 0.093 = 0.907)$
- Question 3: What is the probability that someone who is diagnosed as not having cancer actually has the disease? $P(C|D') = ?$

Exercise 2.95 - Question 3

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
  - $P(D) = 0.093 \quad (P(D') = 1 - 0.093 = 0.907)$
- Question 3: What is the probability that someone who is diagnosed as not having cancer actually has the disease? $P(C|D') = ?$
- Solution:

$$P(C|D') =$$

Exercise 2.95 - Question 3

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
  - $P(D) = 0.093 \quad (P(D') = 1 - 0.093 = 0.907)$
- Question 3: What is the probability that someone who is diagnosed as not having cancer actually has the disease? $P(C|D') = ?$
- Solution:

$$P(C|D') = \frac{P(C \cap D')}{P(D')}$$

Exercise 2.95 - Question 3

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
  - $P(D) = 0.093 \quad (P(D') = 1 - 0.093 = 0.907)$
- Question 3: What is the probability that someone who is diagnosed as not having cancer actually has the disease? $P(C|D') =?$
- Solution:

$$P(C|D') = \frac{P(C \cap D')}{P(D')} = \frac{P(D'|C) \cdot P(C)}{P(D')}$$

Exercise 2.95 - Question 3

- Events:

  C : A randomly chosen person has cancer
  D : A randomly chosen person is diagnosed as having cancer

- Data:
  - $P(C) = 0.05 \quad (P(C') = 0.95)$
  - $P(D|C) = 0.78 \quad (P(D'|C) = 0.22)$
  - $P(D|C') = 0.06 \quad (P(D'|C') = 0.94)$
  - $P(D) = 0.093 \quad (P(D') = 1 - 0.093 = 0.907)$
- Question 3: What is the probability that someone who is diagnosed as not having cancer actually has the disease? $P(C|D') = ?$
- Solution:

$$P(C|D') = \frac{P(C \cap D')}{P(D')} = \frac{P(D'|C) \cdot P(C)}{P(D')}$$
$$= \frac{0.05 \cdot 0.22}{0.907} = \frac{0.0011}{0.907} \approx 0.012$$

And now . . .

1. Preliminaries
   - Partition of a sample space
2. Bayes′ rule
   - Theory
   - Examples
3. A lot of computing . . .

- Try all odd exercises from Section 2.7 (pp. 76–77), Exercise 2.99 was done in the class
- Also, you should be able to compute all Review exercises from pp. 77-79, check it!