UNIVERSITAS GADJAH MADA
DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING
BACHELOR IN CIVIL ENGINEERING

Statistics and Probability

Probability

Probability

- Why probability?
  - We cannot be sure on the result of an event (for example volcanic eruption, earthquake, tsunami) based on previous events (historical records)
  - Uncertainty or stochastic nature is an inherent characteristics of every process in nature

Delays of bus arrival

Delay (minutes)	Frequency	Relative frequency	Relative frequency
1	2	0.07	7%
2	3	0.10	10%
3	8	0.27	27%
4	4	0.13	13%
5	5	0.17	17%
6	3	0.10	10%
7	2	0.07	7%
8	0	0.00	0%
9	1	0.03	3%
10	2	0.07	7%
Σ	**30**	**1**	**100%**

How long will it be the delay of the 31st arrival?

Probability

- Definition #1
  - If a random event can occur in n equally likely and mutually exclusive ways, and if n_a of these ways have an attribute A, then the probability of the occurrence of the event having attribute A is n_a/n written as

$$\text{prob}(A) = n_a/n$$

    - In the above definition, n is the set of all possible results (events)
    - The above definition is an *a priori* definition because it assumes that one can determine before the fact all of the equally likely and mutually exclusive ways that an event can occur and all the ways that an event with attribute A can occur

Probability

- Definition #2
  - If a random event occurs a large number of times *n* and the event has attribute A in n_a of these occurrences, then the probability of the occurrence of the event having attribute A is

$$\text{prob}(A) = \lim_{n \to \infty} \frac{n_a}{n}$$

  - The above definition allows us to estimate probabilities based on observations and does not require that outcomes be equally likely or that they all be enumerated

Probability

- In definition #2
  - The estimates of probability, which are based on observations, are empirical and will only stochastically converge to the true probability as the number of observations becomes large
  - If two independent sets of observations are available (samples), an estimate of the probability of the event *A* could be determined from each set of observations
    - These two estimates of $\text{prob}(A)$ would not necessarily equal to each other nor
    - would either estimate necessarily equal the true probability (population) $\text{prob}(A)$ based on an infinitely large sample

How many observations are required to produce acceptable estimates of the probabilities of events?

Range of probability values

- From the two definitions, the probability ranges from 0 to 1
  - $0 \leq \text{prob}(A) \leq 1$
  - $\text{prob}(A) = 0$ nearly impossible
  - $\text{prob}(A) = 1$ almost certain

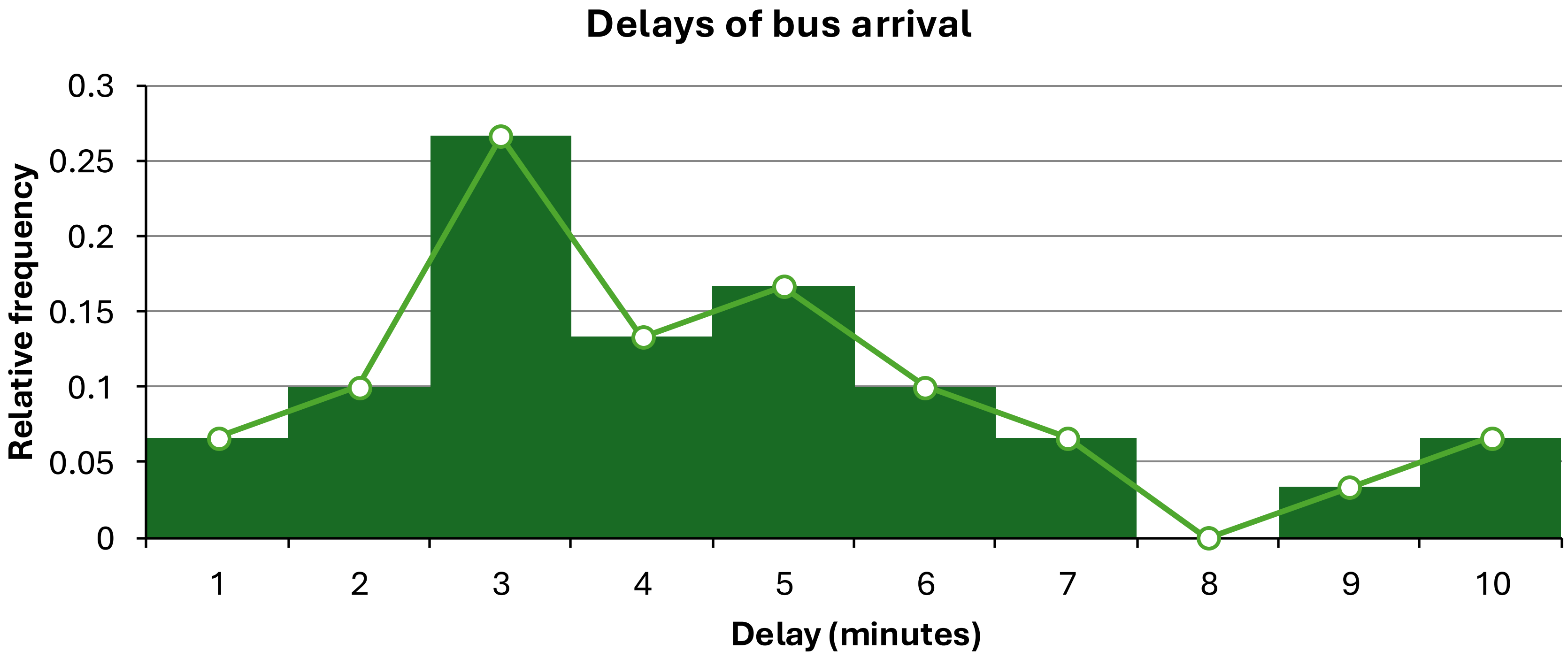
Probability

- Set and measure
  - Consider that an experiment is any process that generates values of random variables
    - All possible outcomes of an experiment constitutes the sample space
    - Any particular point in the sample is a sample point or element
    - A collection of elements, known as a set, is an event
  - To each element in the sample space, a non-negative weight is assigned such that the sum of the weights on all of the element is one
    - The magnitude of the weight is proportional to the likelihood that the experiment will result in a particular element
    - The weight assigned to the elements of the sample space are known as probabilities

Probability and frequency table

Delay (minutes)	Frequency	Relative frequency	Relative frequency
1	2	0.07	7%
2	3	0.10	10%
3	8	0.27	27%
4	4	0.13	13%
5	5	0.17	17%
6	3	0.10	10%
7	2	0.07	7%
8	0	0.00	0%
9	1	0.03	3%
10	2	0.07	7%
Σ	**30**	**1**	**100%**

Probability and histogram



Probability

Sample space and sample elements

Sample space and sample elements

- Example #1
  - A catchment area has three stations: Sta-1, Sta-2, Sta-3
  - An experiment was conducted to decide whether a particular station needs a recalibration
  - Outcomes of the experiment (*y*, *n*, *y*)
    - Sta-1 requires recalibration (*y* = yes)
    - Sta-2 does not require recalibration (*n* = no)
    - Sta-3 requires recalibration (*y* = yes)

Sample space and sample elements

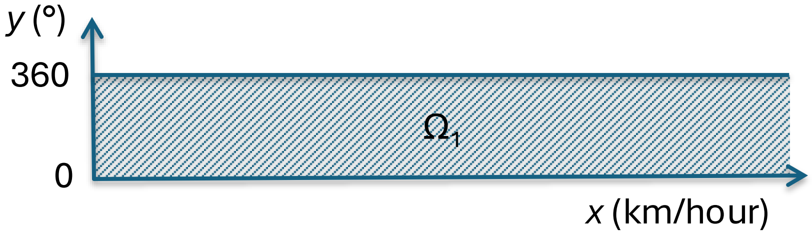
- Example #1
  - Sample space – Alternative 1
    - $S_1 = \{(y, y, y), (y, y, n), (y, n, y), (n, y, y), (y, n, n), (n, y, n), (n, n, y), (n, n, n)\}$
    - S_1 is a discrete sample space, i.e. it has a countable number of elements
    - If the experiment is carried out once, then the outcome is one of its element
  - Sample space – Alternative 2
    - $S_2 = \{0,1,2,3\}$
    - S_2 is a discrete sample space
    - We are interested in the number of stations that needs recalibration
    - We do not need information on which station that needs recalibration

Sample space and sample elements

- Example #2
  - Measurement on wind velocity
    - speed (km/hour)
    - direction (°)
  - Outcomes: (x, y)
    - x = speed (km/hour)
    - y = direction (°)

Sample space and sample elements

- Example #2
  - Sample space – Alternative 1
    $\Omega_1 = \{(x, y): x \geq 0, 0 \leq y \leq 360\}$ continuous sample space



- Example #2
  - Sample space – Alternative 2
    $\Omega_2 = \{+, -\}$ discrete sample space
    - $+$ = wind speed > 60 km/hour
    - $-$ = wind speed < 60 km/hour

Events

- An event is a subset of a sample space
- An event occurs if and only if outcomes of the experiment are members of the event
- Example, recalibration of Sta-1, Sta-2, Sta-3
  - Event *A*: at least two stations need recalibration
    $A = \{(y, y, y), (y, y, n), (y, n, y), (n, y, y)\}$
  - Event *B*: none of those stations requires recalibration
    $B = \{(n, n, n)\}$
  - Event *C*: two stations need recalibration
    $C = \{(y, y, n), (y, n, y), (n, y, y)\}$

Venn diagram

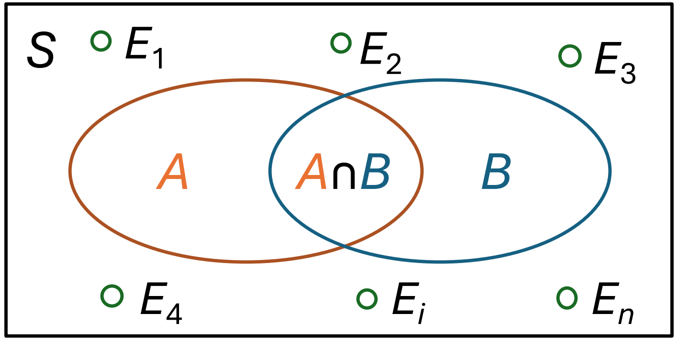
- Notation

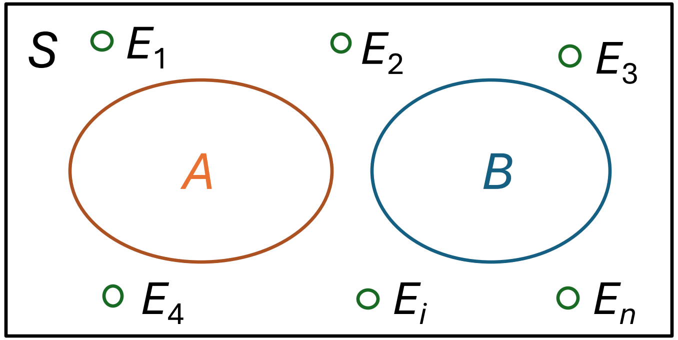
S	= sample space
E_i	= element in S
A,B	= *events in* S
prob(E_i)	= probability of element E_i

$$0 \leq \text{prob}(E_1) \leq 1$$

$$S = \bigcup_i E_i$$

$$\text{prob}(S) = \sum \text{prob}(E_i) = 1$$





Probability of an event

- Event A

$$A = \bigcup_{i=m}^{n} E_i \qquad 0 \leq \text{prob}(A) = \sum_{i=m}^{n} \text{prob}(E_i) \leq 1$$

- Events A and B

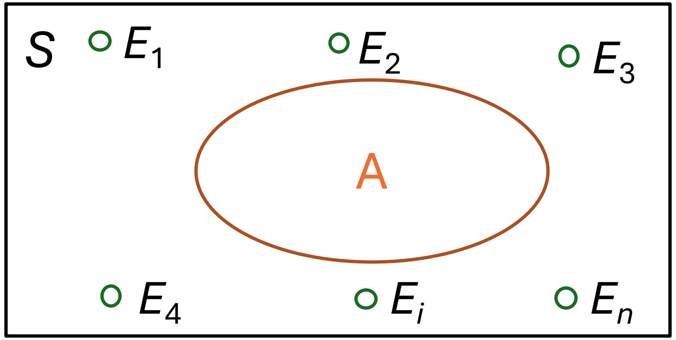
$$\text{prob}(A \cup B) = \text{prob}(A) + \text{prob}(B) - \text{prob}(A \cap B)$$

- Events A and B are independent

$$\text{prob}(A \cup B) = \text{prob}(A) + \text{prob}(B)$$

Probability of an event

- Event A^c = complement of event A
- A^c represents all elements in the sample space S are not in A



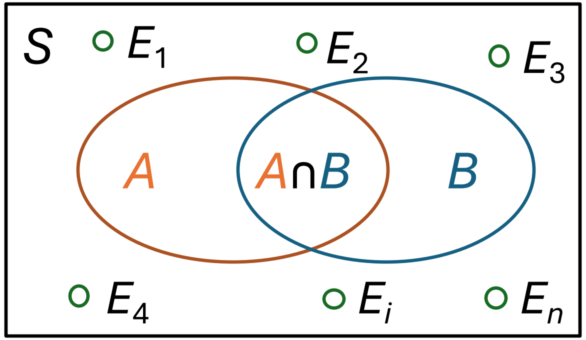
$$\text{prob}(A \cap A^c) = 0$$

$$\text{prob}(A \cap A^c) = \text{prob}(A) + \text{prob}(A^c) = 1$$

$$\text{prob}(A) = 1 - \text{prob}(A^c)$$

Conditional probability

- The probability of an event (event B) depends on the occurrence of another event (event A)



$\text{prob}(B|A) = \text{prob}(B)$ with the condition that event A has occurred

- the sample space S reduces to A
- the event is represented by

$$\text{prob}(B|A) = \frac{\text{prob}(A \cap B)}{\text{prob}(A)}, \qquad \text{prob}(A) \neq 0$$

$$\Rightarrow \text{prob}(A \cap B) = \text{prob}(A) \cdot \text{prob}(B|A)$$

Conditional probability

- If event B is independent on the occurrence of event A

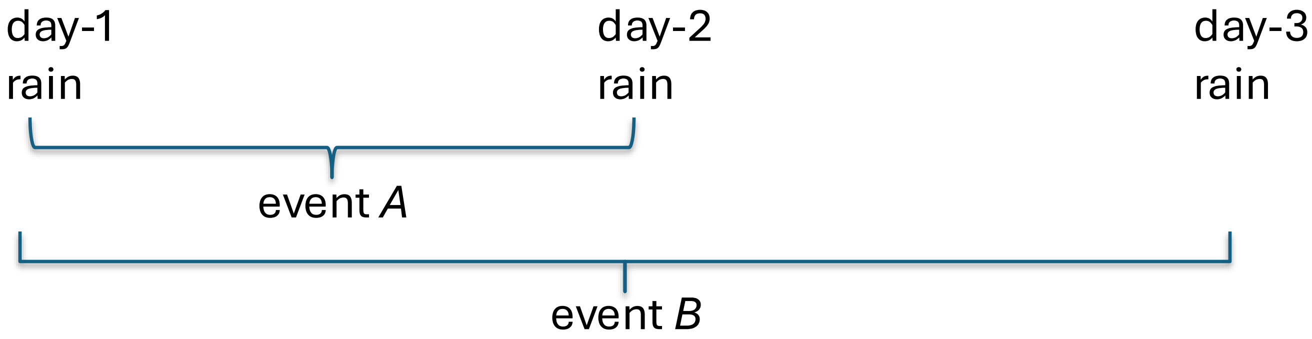
$$\text{prob}(B|A) = \text{prob}(B)$$

$$\text{prob}(A \cap B) = \text{prob}(A) \cdot \text{prob}(B)$$

Conditional Probability

- Example
  - Rainfall data at a station show the probability of rainy day as follow
    - probability of a rainy day following a rainy day is 0.444
    - probability of a dry day following a rainy day is 0.556
    - probability of a dry day following a dry day is 0.724
    - probability of a rainy day following a dry day is 0.276
  - If it rains one day, what is the probability of having rain during the following two days?

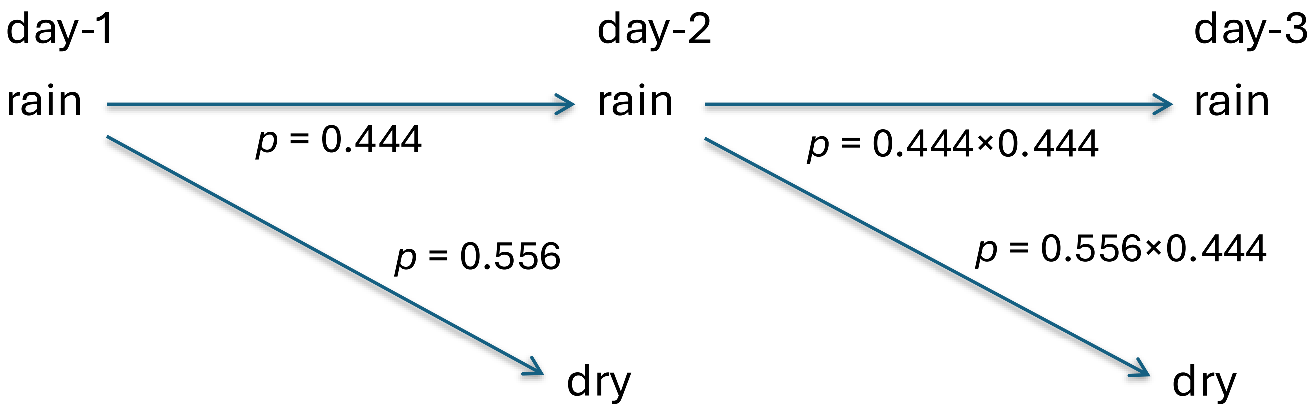
Conditional probability



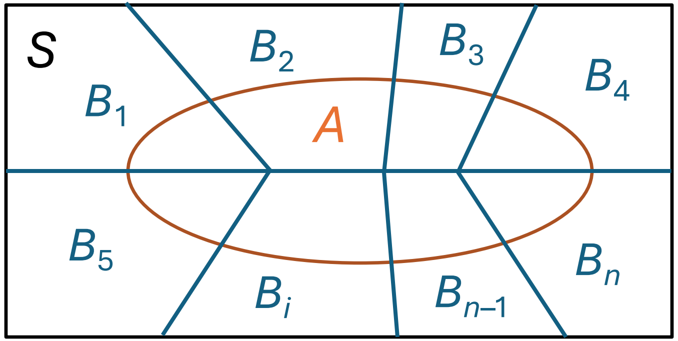
- Event A = rain in day-2 after rain in day-1
  Event B = rain in day-3 after rain in day-1
- Looking for rain in 3 consecutive days: $\text{prob}(A \cap B) = \text{prob}(A)\ \text{prob}(B|A)$
- It is known that $\text{prob}(A) = 0.444$
  $\text{prob}(B|A) = 0.444$ (rainy day following a rainy day)
  $\Longrightarrow \text{prob}(A \cap B) = 0.444 \times 0.444 = 0.197$

Conditional probability

- Another approach
  - Probability of a rainy day following a rainy day is $p = 0.444$
  - It rains one day



Total probability



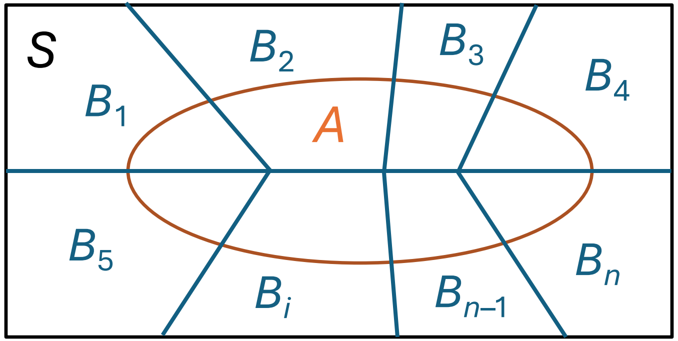
- $B_1, B_2, \ldots, B_n$ are mutually exclusive events, each has non-zero probability of occurrence, $\text{prob}(B_i) \neq 0$, for all i

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

$$B_i \cap B_j = 0, \forall i, j\ (i \neq j)$$

$$\text{prob}(B_i) > 0, \forall i$$

Total probability



- Probability of event A can be written as follows

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n)$$

$$\text{prob}(A) = \text{prob}[(A \cap B_1) \cup (A \cap B_2) \cup \cdots \cup (A \cap B_n)]$$

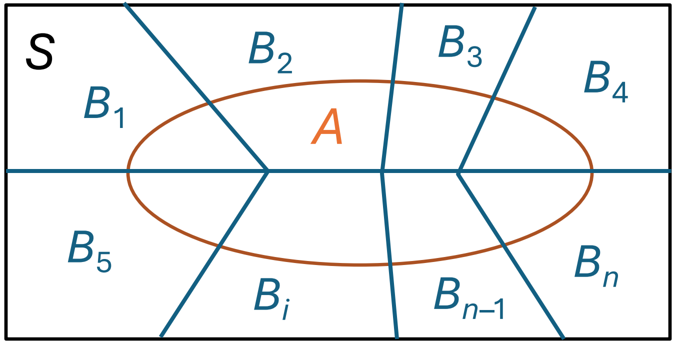
$$\text{prob}(A) = \text{prob}(A \cap B_1) + \text{prob}(A \cap B_2) + \cdots + \text{prob}(A \cap B_n)$$

$$\text{prob}(A) = \text{prob}(B_1 \cap A) + \text{prob}(B_2 \cap A) + \cdots + \text{prob}(B_n \cap A)$$

Total probability

- From the conditional probability

$$\left.\begin{aligned} \text{prob}(A \cap B_1) &= \text{prob}(A)\text{prob}(B_1|A) \\ \text{prob}(B_1 \cap A) &= \text{prob}(B_1)\text{prob}(A|B_1) \end{aligned}\right\} \quad \begin{aligned} &\text{prob}(A \cap B_1) = \text{prob}(B_1 \cap A) \Longrightarrow \\ &\text{prob}(A)\text{prob}(B_1|A) = \text{prob}(B_1)\text{prob}(A|B_1) \end{aligned}$$



$$\text{prob}(A) = \text{prob}(A \cap B_1) + \text{prob}(A \cap B_2) + \cdots + \text{prob}(A \cap B_n)$$

$$\text{prob}(A) = \text{prob}(B_1)\,\text{prob}(A|B_1) + \cdots + \text{prob}(B_n)\,\text{prob}(A|B_n)$$

$$\text{prob}(A) = \sum_{i=1}^{n} \text{prob}(B_i)\,\text{prob}(A|B_i)$$

Total probability

- Example
  - Records in a residential area show that probability of inundation is 0.80 for rainy days and 0.25 for non-rainy days
  - It is known that the probability of a rainy day in that area is 0.36
  - What is the probability of inundation in that residential area?

Total probability

- Solution
  - Suppose
    - Event A = inundation
    - Event B_1 = rainy day
    - Event B_2 = dry day
  - Inundation might occur in a rainy or dry day
    - $\text{prob}(A) = \text{prob}(B_1)\ \text{prob}(A|B_1) + \text{prob}(B_2)\ \text{prob}(A|B_2)$

$$= 0.36 \times 0.80 + (1 - 0.36) \times 0.25$$

$$= 0.448$$

Bayes theorem

- From the conditional probability
  - $\text{prob}(A \cap B) = \text{prob}(A)\ \text{prob}(B|A)$ (1)
  - $\text{prob}(B \cap A) = \text{prob}(B)\ \text{prob}(A|B)$ (2)
  - Since $\text{prob}(A \cap B) = \text{prob}(B \cap A) \Longrightarrow$
    $\text{prob}(A)\ \text{prob}(B|A) = \text{prob}(B)\ \text{prob}(A|B)$ (3)
- For events A and B_j, Eq. (3) becomes
  - $\text{prob}(A)\ \text{prob}(B_j|A) = \text{prob}(B_j)\ \text{prob}(A|B_j)$ (4)

Bayes theorem

- From the total probability
  - $\text{prob}(A) = \sum_{j=1}^{n} \text{prob}(B_j)\, \text{prob}(A|B_j)$ (5)
- Substituting (5) to (4)

$$\text{prob}(B_j|A) = \frac{\text{prob}(B_j)\, \text{prob}(A|B_j)}{\sum_{i=1}^{n} \text{prob}(B_i)\, \text{prob}(A|B_i)} \qquad (6)$$

Bayes theorem

- Use of Bayes theorem
  - Finding probabilities of one event (B_j) provided that another event has occurred (A)
  - Estimating probabilities of one event (B_j) by observing a second event (A)

Bayes theorem

- Example
  - Information from an early warning system are transmitted by four relays.
  - R_i (i = 1, 2, 3, 4) is event where information are transmitted by relay i.
  - Probabilities of event R_i are 0.1, 0.2, 0.3, and 0.4.
  - It is known from previous experiences that probabilities of transmittal errors of each relay are 0.05, 0.10, 0.15, and 0.20.
  - An error has just been noticed recently.
  - What is the probability that the information were sent through relay R_2?

Bayes theorem

- Solution
  - Given $\text{prob}(R_1) = 0.1$ $\quad\text{prob}(E|R_1) = 0.10$
    $\text{prob}(R_2) = 0.2$ $\quad\text{prob}(E|R_2) = 0.15$
    $\text{prob}(R_3) = 0.3$ $\quad\text{prob}(E|R_3) = 0.20$
    $\text{prob}(R_4) = 0.4$ $\quad\text{prob}(E|R_4) = 0.25$

  - Probability that the transmission was done through relay R_2, given that an error had occurred is

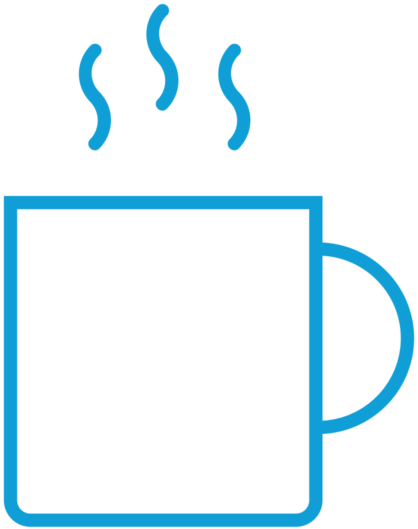
$$\text{prob}(R_2|E) = \frac{\text{prob}(R_2) \cdot \text{prob}(E|R_2)}{\sum_{i=1}^{n} \text{prob}(R_i) \cdot \text{prob}(E|R_i)}$$

$$\text{prob}(R_2|E) = \frac{0.2 \times 0.15}{0.1 \times 0.10 + 0.2 \times 0.15 + 0.3 \times 0.20 + 0.4 \times 0.25} = \frac{0.03}{0.20} = 0.15$$

Bayes theorem

i	prob(R_i)	prob($E|R_i$)	prob(R_i). prob($E|R_i$)	prob($R_i|E$)
1	0.1	0.10	0.01	0.05
2	0.2	0.15	0.03	0.15
3	0.3	0.20	0.06	0.30
4	0.4	0.25	0.10	0.50
			0.20	**1.00**

prob(E)



Statistics and Probability

Probability