DS203: Programming in Data Science
IE 605 Engineering Statistics

Introduction to Probability and Statistics
Lecture 01

Manjesh K. Hanawal

19th August 2020

Introduction

- In real world problems exhibit inherent randomness
- In modeling real-world problems, we need to take into account possible variations (randomness)
- This is done by allowing the models to be probabilistic

- Data observed from real world problems represents their behavior/property/nature
- We want to build models that describes the observed data
- Probability models helps us systematically capture variations in the data and gives rules for consistent reasoning

Outline

- Sample Space and Events
- Axioms of Probability
- Conditional Probability
- Independent Events
- Baye's formula

Sample Space and Events

Consider an experiment whose outcomes are not predictable in advance. Examples: Coin toss, throw of dice, stock prices, weather, demand for goods, arrival of customer

Definition (Sample Space)

Possible outcomes of an experiment is known as sample space. We denote it as Ω.

Definition (Event)

Any subset of the sample space is known as an event.

Analysis of an random experiment begins by defining its outcome.

Examples

1. Example 1: (Flipping a coin) $\Omega = \{H, T\}$
2. Example 2: (Rolling a dice) $\Omega = \{1, 2, 3, 4, 5, 6\}$
3. Example 3: (Flipping two coins)
   $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$
4. Example 4: (Rolling two dice)

$$\Omega = \begin{bmatrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{bmatrix}$$

5. Example 4: (Temperature of a room) $\Omega = [a, b]$ for some real values a, b.

Examples contd.

1. Example 1: (Flipping a coin) $E = \{H\}$ or $E = \{T\}$ or $E = \{H, T\}$
2. Example 2: (Rolling a dice) $E = \{2, 4, 6\}$ or $E = \{1, 3, 5\}$ or $E = \{3, 6\}$
3. Example 3: (Flipping two coins) $E = \{(H, H), (H, T)\}$ or $E = \{(T, H), (T, T)\}$ or, ...
4. Example 4: (Rolling two dice)
   $E = \{(1, 4), (4, 1), (2, 3), (3, 2)\}$ (sum of outcome is 5) ,
   $E = \{(1, 5), (5, 1), (2, 4), (4, 2), (3, 3)\}$ (sum of outcome is 6)
5. Example 4: (Temperature of a room) $E = [c, d]$ for $a \leq c, d \leq b$.

Operations on Events

Consider an experiment with sample space Ω and events E and F.

- We say event E occurs when outcome of the experiment lies in E. In rolling dice problem if $E = \{1, 4, 6\}$, event E occurs if face of the dice throws 1, 4 or 6.
- **Complement:** $E^c = \Omega \backslash E$. $E \cup E^c = \Omega$ and $E \cap E^c = \emptyset$.
- **Union:** $E \cup F$ consists of all elements in E and F. $G = E \cup F$ occurs if E or F occurs
- **Intersection:** $E \cap F$ consists of elements belonging to both E and F. $G = E \cap F$ occurs only if both E and F occur
- **Mutually Exclusive:** If there is no common element between E and F, $E \cap F = \emptyset$, i.e., then E and F are mutually exclusive.
- For any sequence of events $E_1, E_2, ...,$ $\cup_{i=1}^{\infty} E_i$ and $\cap_{i=1}^{\infty} E_i$, denote their union and intersection, respectively

Probability of Events

In a random experiment we want to know/assign 'likelihood' of each event. This is done by defining probabilities. Intuitively, probability should satisfy some basic properties given by following axioms:

- Non-negativity: $P(E) \geq 0$ for all $E \subset \Omega$
- Normalization: $P(\Omega) = 1$
- (Finite) additivity For mutually exclusive events E_1 and E_2, $P(E_1 \cup E_2) = P(E_1) + P(E_2)$. (to be extended)

Consequences of Axioms

- $P(E) \leq 1$ for all $E \subset \Omega$
  - Claim: $A \subset B \implies P(A) \leq P(B)$
    $B = A \cup (B \backslash A) \implies P(B) = P(A) + P(B \backslash A)$
    (as B and $B \backslash A$ are mutually exclusive Axiom 3 applied)
    As $P(B \backslash A) \geq 0$, Axiom 1 applies and the claim holds.
- $P(E \cup F \cup G) =$
  $P(E) + P(F) + P(G) - P(EF) - P(EG) - P(FG) + P(EFG)$
- For any $E, F \subset \Omega$, $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

$$
\begin{aligned}
& E \cup F = E \cup (F \backslash (E \cap F)) \\
\implies & P(E \cup F) = P(E) + P(F \backslash (E \cap F)) (\text{Axiom 3}) \\
& P(E \cup F) = P(E) + P(F) - P(E \cap F)
\end{aligned}
$$

Extending Finite Additivity Property

Example: $\Omega = \{1, 2, 3, \ldots\}$ and $P(i) = 1/2^i$ for all $i \in \Omega$.
What is the probability of finding an even number?

Is P a valid probability function. Sanity check

- $0 \leq P(i) \leq 1$ for all $i \in \Omega$
- $P(\Omega) = \sum_{i=1}^{\infty} 1/2^i = \frac{1}{2}\left(\frac{1}{1-1/2}\right) = 1$

We are interested in event $E = \{2, 4, 8, 10, \ldots\} = \cup_{i=1}^{\infty}\{2i\}$
$P(E) = P(2) + P(4) + P(6) + \ldots = \sum_{i=1}^{\infty} P(\{2i\})$
We added infinitely many (countable) events!

Extended Axiom 3: For a sequence of mutually exclusive events $E_1, E_2, E_3, \ldots$ defined on the same sample space

$$P(E_1 \cup E_2 \cup E_3 \ldots) = P(E_1) + P(E_2) + P(E_3) + \ldots$$

Additive property for uncountable case?

Continuous case:: $\Omega = \{(x, y), 0 \leq x, y \leq 1\}$

- We know $P(\Omega) = 1$
- $P(\Omega) = \sum_{0 \leq x, y \leq 1} P(x, y)$. For any (x, y), $P(x, y) = 0$. Hence $P(\Omega) = 0$. A contradiction!

> Additivity axiom applies to finite and 'countable' number events not to uncountable number of events!

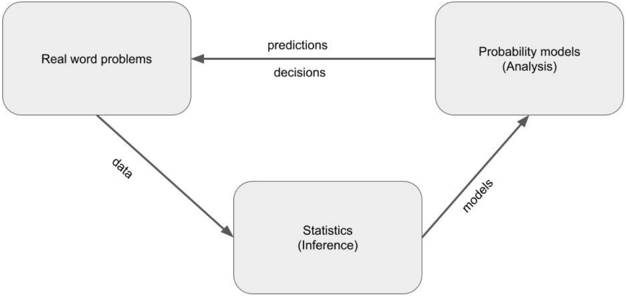
Interpretation of Probability

- Frequentist view
  - Probability of an event is the fraction of times it appears
  - In coin tossing: $P(H) = \frac{\text{number of times head appears}}{\text{total number of trials}}$ when number of trials is repeated indefinitely.
- Probabilities are interpreted as
  - Description of beliefs
  - Preference of events

Role of probability and Statistics In Data Science

Probability and Statistics provide framework for inferring and analyzing uncertain outcomes

- consistent inference
- consistent reasoning
- prediction and decision in uncertain environments



Conditional Probability

- Many time we would like to know probability of an event given that another event has occurred.
- For any pair of events E, F, probability of event E given that event F occurs is denoted as $P(E|F)$ and defined as

$$P(E|F) = \frac{P(E \cap E)}{P(F)}.$$

- Conditional probability is well defined if $P(F) > 0$.

Conditional Probability

- Many time we would like to know probability of an event given that another event has occurred.
- For any pair of events E, F, probability of event E given that event F occurs is denoted as $P(E|F)$ and defined as

$$P(E|F) = \frac{P(E \cap E)}{P(F)}.$$

- Conditional probability is well defined if $P(F) > 0$.

Example: In rolling a fair dice example, what is the probability that an observed outcome is even given that it is divisible by 3?
We have $E = \{2, 4, 6\}$ and $F = \{3, 6\}$.
$P(E|F) = P(EF)/P(F) = P(\{6\})/P(\{3, 6\}) = 1/2$.
Examples: If it rains, what is the chance it be be sunny? If enemy airfract intrudes, what is the chances that our radar will miss it.

Independence of Two Events

- Two event are "independent" if occurrence of one event does not provide any information about the other
- Example, $P(E|F) = P(E)$ and $P(F|E) = P(E)$. Uncertainity of one remains the same, even after observing the other.
- From conditional probability this implies that $P(E \cap F) = P(E)P(F)$. Formally,

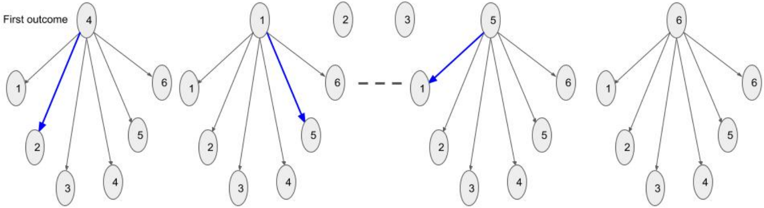
> Definition: Two event E and F are independent if
>
> $$P(E \cap F) = P(E)P(F).$$

- If two events are not independent, then they are dependent.

Example of dependent and Independent set

Example 1: Rolling of two fair dice. Event E denotes the sum of outcomes is 6 and event F denotes the outcome of first dice is 4.
$E = \{(1,5),(5,1),(2,4),(4,2),(6,6)\}$,

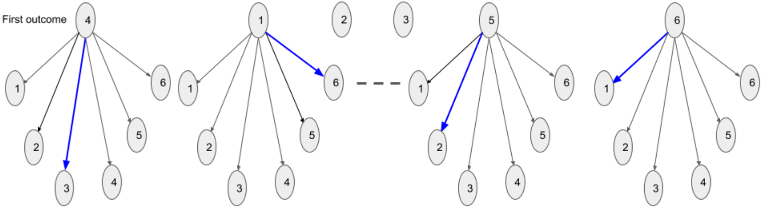
- $F = \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\}$
- $P(E \cap F) = P(\{4,2\}) = 1/36, P(E) = 5/36, P(F) = 1/6$.
- $P(E \cap F) \neq P(E)P(F)$. Hence E and F are dependent.
- If first outcome is 4, we have some hope of getting the sum 6
- if the first outcome is not 4, say 6, we do not have any hope.



Example of dependent and Independent set

Example 1: Rolling of two fair dice. Event E denotes the sum of outcomes is 7 and event F denotes the outcome of first dice is 4. $E = \{(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)\}$,

- $F = \{(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)\}$
- $P(E \cap F) = P(\{4,3\}) = 1/36, P(E) = 1/6, P(F) = 1/6.$
- $P(E \cap F) = P(E)P(F)$. Hence E and F are independent.
- If first outcome is 4, we have some hope of getting the sum 7
- if first outcome is not 4, the amount of hope is the same.



Independence of Collection of Events

- Intuitively, a collection of events are independent if occurrence of any of them have no effect on the probability of occurrence of other events. Formally,

> **Definition (Independence of Events)**
>
> A finite set of events $E_1, E_2, E_3, \ldots, E_n$ are independent if any subset $E_{1'}, E_{2'}, \ldots E_{r'}$, where $r' \leq n$,
>
> $$P(E_{1'} \cap E_{2'} \cap \ldots \cap E_{r'}) = P(E_{1'})P(E_{2'}) \ldots, P(E_{r'})$$

- Number of conditions to check Independence of $n > 2$ events is $\binom{n}{2} + \binom{n}{3} \ldots, \binom{n}{n} = 2^n - n$. Exponential in n!

Pairwise Independence

- A weaker notion independence of collection of events is pairwise independence

> **Definition (Pairwise independence)**
>
> A finite set of events $E_1, E_2, E_3, \ldots, E_n$ are pairwise independent if for any pair (i, j) such that $1 \leq i, j \leq n$ and $i \neq j$
> $$P(E_i \cap E_j) = P(E_i)P(E_j)$$

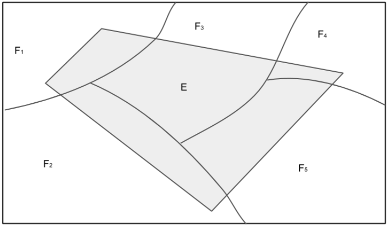
- For pairwise independence, only need check $\binom{n}{2}$ conditions. Quadratic in n!

Total Probability Law

For any sets E and F

$$
\begin{aligned}
E &= (E \cap F) \cup (E \cap F^c) \\
\implies P(E) &= P(E \cap F) + P(E \cap F^c) \quad \text{(Axiom 3)} \\
P(E) &= P(E|F)P(F) + P(E|F^c)P(F^c)
\end{aligned}
$$

Total Probability Law: For any mutually exclusive sets $F_1, F_2, \ldots, F_n$, $P(E) = \sum_{i=1}^{n} P(E|F_i)P(F_i)$



Baye's Formula

Suppose E has occured and we are interested in determining which one of the F_j has occurred

$$P(F_j|E) = \frac{P(E \cap F_j)}{P(E)} = \frac{P(E|F_j)P(F_j)}{\sum_{i=1}^{n} P(E|F_i)P(F_i)}$$

Example: Assume that the symptoms mild fever (F_1), body ache (F_2), high fever (F_3), cold and cough (F_4) occur with probabilities $P(F_1) = .2, P(F_2) = 0.1, P(F_3) = 0.5$ and $P(F_4) = 0.2$. Conditional probabilities of these causing Corona infection (E) are given as $P(E|F_1) = .5, P(E|F_2) = .2, P(E|F_3) = .7, P(E|F_4) = .3$. If a person is tested positive for Corona, what is the probability that the patient had mild fever (asymptomatic).