Statistics for Data Science - 2

Week 0 Part 2 Important formulas

Basic Probability

1. **Experiment:** Process or phenomenon that we wish to study statistically.
   Example: Tossing a fair coin.

2. **Outcome:** Result of the experiment.
   Example: head is an outcome on tossing a fair coin.

3. **Sample space:** A sample space is a set that contains all outcomes of an experiment.
   • Sample space is a set, typically denoted S of an experiment.
   • example: Toss a coin: S = { heads, tails }

4. **Event:** An event is a subset of the sample space.

   - Toss a coin: S = { heads, tails }
     - Events: empty set, {heads}, {tails}, { heads, tails }
     - 4 events
   - An event is said to have "occurred" if the actual outcome of the experiment belongs to the event.
   - One event can be contained in another, i.e. $A \subseteq B$
   - Complement of an event A, denoted A^C= { outcomes in S not in A } = $(S \setminus A)$.
   - Since events are subsets, one can do complements, unions, intersections.

5. **Disjoint events:** Two events with an empty intersection are said to be disjoint events.

   - Throw a die: even number, odd number are disjoint.
   - Multiple events: $E_1, E_2, E_3, ...$ are disjoint if, for any $i \neq j$, $E_i \cap E_j$ = empty set.

6. **De Morgan's laws:** For any two events A and B,
   $(A \cup B)^C = A^C \cap B^C$ and $(A \cap B)^C = A^C \cup B^C$.

7. **Probability:** "Probability" is a unction P that assigns to each event a real number between 0 and 1 and satisfies the following two axioms:

   (i) $P(S) = 1$ (probability of the entire sample space equals 1).

   (ii) If $E_1, E_2, E_3, ...$ are disjoint events (Could be infinitely many),

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

   - Probability function Assigns a value that represents chance of occurrence of the event.

- Higher value of the probability of an event means higher chance of occurring that event.
- 0 means event cannot occur and 1 means event always occurs.

8. Probability of the empty set (denoted ϕ) equals 0. that is

$$P(\phi) = 0$$

9. Let E^C be the complement of Event E. Then,

$$P(E^C) = 1 - P(E)$$

10. If event E is the subset of event F, that is $E \subseteq F$, then

$$P(F) = P(E) + P(F \setminus E)$$

$$\Rightarrow P(E) \leq P(F)$$

11. If E and F are events, then

$$P(E) = P(E \cap F) + P(E \setminus F)$$

$$P(F) = P(E \cap F) + P(F \setminus E)$$

12. If E and F are events, then

$$P(E \cup F) = P(E) + P(F) - P(E \cap F)$$

13. **Equally likely events:** assign the same probability to each outcome.
14. If sample space S contains the equally likely outcomes, then
    - $P(\text{one outcome}) = \dfrac{1}{\text{Number of outcomes in } S}$
    - $P(\text{event}) = \dfrac{\text{Number of outcomes in event}}{\text{Number of outcomes in } S}$

15. **Conditional probability space:** Consider a probability space (S, E, P), where S represents the sample space, E represents the collection of events, and P represents the probability function.

   - Let B be an event in S with $P(B) > 0$. Now, conditional probability space given B is defined as
     For any event A in the original probability space (P, S, E), the conditional probability of A given B is $\dfrac{P(A \cap B)}{P(B)}$.
   - It is denoted by $P(A \mid B)$. And

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

16. **Law of total probability:**

   - If the events B and B^c partitioned the sample space S such that $P(B_1), P(B_2) \neq 0$, then for any event A of S,

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

   - In general, if we have k events $B_1, B_2, \cdots, B_k$ that partition S, then for any event A in S,

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

17. **Bayes' theorem:** Let A and B are two events such that $P(A) > 0, P(B) > 0$.

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

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

   In general, if the events $B_1, B_2, \cdots, B_k$ partition S such that $P(B_i) \neq 0$ for $i = 1, 2, \cdots, k$, then for any event A in S such that $P(A) \neq 0$,

$$P(B_r \mid A) = \frac{P(B_r)P(A \mid B_r)}{\sum_{i=1}^{k} P(B_i)P(A \mid B_i)}$$

   for $r = 1, 2, \cdots, k$.

18. **Independence of two events:** Two events A and B are independent iff

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

   - A and B independent $\Rightarrow P(A \mid B) = P(A)$ and $(B \mid A) = P(B)$ for $P(A), P(B) > 0$.

- Disjoint events are never independent.
- A and B independent $\Rightarrow$ A and B^c are independent.
- A and B independent $\Rightarrow$ A^c and B^c are independent.

19. **Mutual independence of three events:** Events A, B, and C are mutually independent if

    (a) $P(A \cap B) = P(A)P(B)$
    (b) $P(A \cap C) = P(A)P(C)$
    (c) $P(A \cap B) = P(A)P(B)$
    (d) $P(A \cap B \cap C) = P(A)P(B)P(C)$

20. **Mutual independence of multiple events:** Events $A_1, A_2, \cdots,_n$ are mutually independent if, $\forall i_1, i_2, \cdots, i_k$,

    $$P(A_{i_1} \cap A_{i_2} \cap \cdots A_{i_k} \cap) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$$

    n events are mutually independent $\Rightarrow$ any subset with or without complementing are independent as well.

21. Occurrence of event A in a sample space is considered as *success.*

22. Non - occurrence of event A in a sample space is considered as *failure.*

23. **Repeated independent trials:**

    (a) **Bernoulli trials**
    - Single Bernoulli trial:
      - Sample space is {success, failure} with P(success) $= p$.
      - We can also write the sample space S as $\{0, 1\}$, where 0 denotes the failure and 1 denotes the success with $P(1) = p, P(0) = 1 - p$.
        This kind of distribution is denoted by $Bernoulli(p)$.
    - Repeated Bernoulli trials:
      - Repeat a Bernoulli trial multiple times independently.
      - For each of the trial, the outcome will be either 0 or 1.

    (b) **Binomial distribution:** Perform n independent $Bernoulli(p)$ trials.
    - It models the number of success in n independent Bernoulli trials.
    - Denoted by $B(n, p)$.
    - Sample space is $\{0, 1, \cdots, n\}$.
    - Probability distribution is given by

      $$P(B(n, p) = k) = nC_k p^k (1 - p)^{n-k}$$

      where n represents the total number trials and k represent the number of success in n trials.

- $P(B = 0) + P(B = 1) + \cdots + P(B = n) = 1$
  $\Rightarrow (1 - p)^n + nC_2 p^2 (1 - p)^{n-2} + \cdots + p^n = 1.$

(c) **Geometric distribution:** It models the number of failures the first success.

- Outcomes: Number of trials needed for first success and is denoted by $G(p)$.
- Sample space: $\{1, 2, 3, 4, \cdots\}$
- $P(G = k) = P(\text{first } k - 1 \text{ trials result in 0 and } kth \text{ trial result in 1.}) = (1 - p)^{k-1} p.$
- Identity: $P(G \leq k) = 1 - (1 - p)^k$.