STA237 Notes

Contents

1 Introduction

1.1 Basic Definitions

1. Scientific Question - A question created by an experimenter.
2. Experiment - A task to collect information in order to answer a scientific question.
3. Sample Space (Ω) - The set of all possible outcomes or results of an experiment.
   For example, $\Omega = \{H, T\}$ is the sample space of tossing a coin.
4. Subsets of the sample space are called events.
   Events all use typical set operations (complements, union, intersection, etc.).

1.2 Properties of Events

1. We call events A, B mutually exclusive if A, B have no outcomes in common. That is, $A \cap B = \emptyset$
2. **Demorgan's Law** - For any two events A, B, we have $(A \cup B)^c = A^c \cap B^c$, and $(A \cap B)^c = A^c \cup B^c$.
3. A **Probability Function** (P) on a finite sample space Ω assigns to each event in A in Ω a number $P(A)$ in $[0, 1]$ such that:

   (a) $P(\Omega) = 1$, and
   (b) $P(A \cup B) = P(A) + P(B)$, if A, B are disjoint.
       The number $P(A)$ is the probability for which A occurs.

   Suppose we had two events A, B, and $P(A) \cap P(B) \neq \emptyset$. We have:

   (a) Elements of ONLY A: $A \cap B^c$
   (b) Elements of A AND B: $A \cap B$
   (c) Elements of ONLY B: $B \cap A^c$

   Then:

   (a) $P(A) = P(A \cap B^c) + P(A \cap B)$
   (b) $P(B) = P(B \cap A^C) + P(A \cap B)$
   (c) $P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(B \cap A^c)$
       Then: $P(A \cup B) = P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B)$
       $= P(A) + P(B) - P(A \cap B)$

   Therefore, we have $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

   We know that $P(A) \subseteq P(\Omega)$, and the complement A^c is mutually exclusive. $P(\Omega) = 1$, and thus:

   $$P(\Omega) = 1 = P(A^c) + P(A)$$

   Therefore: $P(A^c) = 1 - P(A)$.

4. A and B are **independent** if $P(A \cap B) = P(A) \cdot P(B)$.

1.2.1 Axioms

Suppose Ω is a sample space associated with an experiment. To every event A in Ω, we assign a number $P(A)$ (called the probability of A), so that the following axioms hold:

1. Axiom 1: $P(A) \geq 0$
2. Axiom 2: $P(S) = 1$
3. Axiom 3: If $A_1, A_2, ..., A_n$ form a sequence of pairwise mutually exclusive events in Ω (that is, $A_i \cap A_j = \emptyset$ if $i \neq j$), then
$$P(A_1 \cup A_2 \cup ... \cup A_n) = \sum_{i=1}^{n} P(A_i)$$

1.3 Tools for Counting Sample Points

With m elements $a_1, a_2, ..., a_m$, and $b_1, b_2, ..., b_n$, it is possible to form $mn = m \times n$ pairs containing one element from each group.

An ordered arrangement of r distinct objects is called a **permutation**. The number of ways of ordering n distinct objects taken r at a time will be designated by the symbol P_r^n. That is:

$$P_r^n = n(n-1)(n-2)...(n-(r+1)) = \frac{n!}{(n-r)!}$$

The number of unordered subsets of size r chosen (without replacement from n available objects is:

$$\binom{n}{r} = \frac{P_r^n}{r!} = \frac{n!}{r!(n-r)!}$$

Sometimes it is denoted as C_r^n.

2 Conditional Probability

Conditional probability is the likelihood of an event occurring based on the occurrence of a previous event. That is, for two events R, L, the conditional probability of R given L is $P(R|L)$.
It is denoted by:

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

provided $P(C) > 0$.

Note that $P(R|L) + P(R^c|L) = 1$:

$$\begin{aligned} P(R|L) + P(R^c|L) &= \frac{P(A \cap C}{P(C)} + \frac{P(A^c \cap C}{P(C)} \\ &= \frac{P(C)}{P(C)} && \text{Since } P(A), P(A^c) \text{ are mutually exclusive, the union of the intersections is } P(\\ &= 1 \end{aligned}$$

For example, suppose we had the following events:

1. L: Born in a long month (31 days)
$L = \{Jan, Mar, May, Jul, Aug, Oct, Dec\};$

2. R: Born in a month with letter r
   $R = \{Jan, Feb, Mar, Apr, Sep, Oct, Nov, Dec\}$

This means that the conditional probability of R given L is:

$$\begin{aligned} P(R|L) &= \frac{1/3}{7/12} \\ &= \frac{4}{7} \end{aligned}$$

2.0.1 Multiplication Rule

For any events A, C:

$$\begin{aligned} P(A|C) &= \frac{P(A \cap C)}{P(C)} \\ P(A \cap C) &= P(A|C) \cdot P(C) \end{aligned}$$

2.1 Independent Events

Events A, C are **independent** if and only if the probability of A is the same when we know that C has occurred. That is:

$$P(A|C) = P(A)$$

Then:

$$\begin{aligned} \frac{P(A \cap C)}{P(C)} &= P(A) \\ P(A \cap C) &= P(A) \cdot P(C) \end{aligned}$$

2.2 Partitions

For some positive integer k, let the sets $B_1, B_2, ..., B_k$ be such that:

1. $\Omega = B_1 \cup B_2 \cup ... \cup B_k$
2. $B_i \cap B_j = \emptyset$, for $\imath \neq j$.

Then, the collection of sets $\{B_1, B_2, ..., B_k\}$ is said to be a partition of Ω.

2.2.1 The Law of Total Probability

Suppose that $\{B_1, B_2, ..., B_k\}$ is a partitions of Ω such that $P(B_i) > 0$ for $i = 1, 2, ..., k$. Then, for any event A:

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + ... + P(A|B_k)P(B_k) \\ &= \sum_{i=1}^{k} P(A|B_i)P(B_i) \end{aligned}$$

2.3 Bayes' Theorem

Suppose that $\{B_1, B_2, ..., B_k\}$ is a partition of Ω such that $P(B_i) > 0$, for $i = 1, 2, ..., k$. Then, for any event A:

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

3 Random Variables

Discrete Variables are variables whose values can be measured by counting.
For example, a course mark: 0, 1, 2, ..., 100

Continuous Variables are impossible to count and can never properly be counted.
For example, time or weighs: 25 years, 10, months, ...

Categorical Variables take on a finite number of possible values, assigning units of observation to particular groups on the basis of qualitative properties.

For some event with sample space Ω taking multiple parameters ($e.g., \Omega = \{\sigma_1, \sigma_2\} : \sigma \in \{1, 2\}$), we can calculate the total outcome, i.e., the value of the function $X : \Omega \to \mathbb{R}$, given by:

$$X(\sigma_1, \sigma_2) = \sigma_1 + \sigma_2 \text{ for } (\sigma_1, \sigma_2) \in \Omega$$

We denote the event that the function S attains the value k by:

$$\{X = k\} = \{(\sigma_1, \sigma_2) \in \Omega : X(\sigma_1, \sigma_2) = k\}$$

We call X the **random variable**.
$X : \Omega \to \mathbb{R}$ is a **discrete random variable** if it takes on a finite number of values $a_1, a_2, ..., a_n$, **or** an infinite number of values $a_1, a_2, ...$
The probability that X takes on the value x, $P(X = x)$ is the sum of probabilities of all sample points in Ω that are assigned to the value x (i.e., $P(x) = P(X = x)$). We sometimes denote this as $p(x)$.
Then, the probability distribution of a discrete variable X can be represented by a formula, a table, or a graph that provides $P(X = x)$ for all x.

3.0.1 Result

For any discrete probability distribution, the following must be true:

1. $0 \leq p(x) \leq 1$ for all x
2. $\sum_x p(x) = 1$, where the summation is over all values of x with non-zero probability.

3.1 Expected Values of Random Variables

Let X be a discrete random variable with the probability function $p(x)$. Then, the expected value of X, $E(X)$, is defined as:

$$E(X) = \sum_x xP(x),$$

where $P(x) = P(X - x)$. Note that $E(x) = \mu = \sum_x xP(x)$.

3.1.1 Variance of Random Variables

If X is a random variable with mean $E(X) = \mu$, then the variance of a random variable X is the expected value of $(X - \mu)^2$. That is:

$$\sigma^2 = V(X) = E[(-\mu)^2]$$

The standard deviation of X is the positive square root of $V(X)$, or σ.

3.1.2 Results

1. Let X be a discrete random variable with probability function $p(x)$, and let c be a constant. Then,

$$\begin{aligned} E(c) &= \sum_x c \sum P(x) \\ &= c \cdot 1 \\ &= c \end{aligned}$$

   Therefore, $E(c) = c$.

2. Note that for the variance:

   (a)

$$\begin{aligned} V(c) &= E((c-\mu)^2) \\ &= E((c-c)^2) \\ &= 0 \end{aligned}$$

   (b)

$$\begin{aligned} V(cX) &= c^2 V(X) \\ V(aX+b) &= a^2 V(X) \end{aligned}$$

3. Let X be a discrete random variable with probability function $p(x)$, $g(x)$ be a function of X, and let c be a constant. Then:

$$\begin{aligned} E(cx) &= cE(x) \\ &= E[ax+b] \\ &= aE(x)+b \end{aligned}$$

   Therefore, $E[cg(X)] = cE(g(X))$.

4. Let X be a discrete random variable with probability function $p(x)$, and $g_1(X), g_2(X), ..., g_k(X)$ be k functions of X. Then:

$$E[g_1(X) + g_2(X) + ...g_k(X)] = E[g_1(X)] + E[g_2(X)] + ... + E[g_k(X)]$$

3.2 Distribution Function

The distribution function F of a random variable X is the function $F : \mathbb{R} \to [0,1]$, defined by:

$$F(a) = P(X \leq a) \text{ for } -\infty < a < \infty$$

3.3 Bernoulli Distributions

The Bernoulli distribution is used to model an experiment with only two possible outcomes, referred to as a 'success' and 'failure', usually encoded as 1 and 0. A Bernoulli Trial is the term used to describe these experiments.
A discrete random variable X has a Bernoulli distribution with parameter p, where $0 \leq p \leq 1$, if its probability mass function is given by:

$$P(X=1) = p \text{ and } P(X=0) = 1-p$$

We denote this distribution by $Ber(p)$.
Also, we have:

1.

$$\begin{aligned}\mu = E(x) &= \sum_x xP(x)\\ &= 0\cdot(1-p) + 1\cdot p\\ E(x) &= p\end{aligned}$$

Similarly,

$$\begin{aligned}E(x^2) &= \sum_x x^2P(x)\\ &= 0^2\cdot(1-p) + 1^2\cdot p\\ &= p\end{aligned}$$

2.

$$\begin{aligned}\sigma^2 = V(X) &= E(x^2) - \mu^2\\ &= p - p^2\\ V(x) &= p(1-p)\end{aligned}$$

For example: Suppose we flip a coin. Heads is a success (S), and Tail is a failure (F).
We have $P(S) = p$, and $P(F) = 1 - p$. We denote X as the number of heads (i.e., $X = 0, 1$).
Then, $P(X = 0) = 1 - p$, and $P(X = 1) = p$.

3.3.1 Probability Mass Functions

A probability mass function (pmf) is a function over the sample space of a discrete random variable X that shows $P(X)$ is equal to a specific value. That is:

$$P(X = x) = p^x(1-p)^{1-x}, \text{ where } x = 0, 1$$

3.4 Binomial Distributions

A discrete random variable X has a binomial distribution with parameters n, p, where $n = 1, 2, ...,$ and $0 \le p \le 1$, if its probability mass function is given by:

$$P(X = x) = \binom{n}{x}p^x(1-p)^{n-x} \text{ for } x = 0, 1, 2, ..., n,$$

where $\binom{n}{k} = \frac{n!}{(n-x)!x!}$.
We denote this distribution by $B(n, p)$. We also have:

1. $E(X) = np$
2. $V(X) = np(1-p)$

3.4.1 Properties of Binomial Distribution

1. The experiments consist of a fixed number, n, identical trials.
2. Each trial results in one of two outcomes (S, F).
3. $P(S) = p$ for every trial, and $P(F) = 1 - p$.
4. The trials are independent.