C&EE 110

Contents

1 Descriptive Statistics **2**
1.1 Sample/Population Mean . . . 2
1.2 Sample Variance/Standard Deviation . . . 2
1.3 Empirical Rule . . . 2
1.4 Skew . . . 2
1.4.1 Negative/Left Skew . . . 2
1.4.2 Symmetric/No Skew . . . 2
1.4.3 Positive/Right Skew . . . 2

2 Probability and Bayes Theorem **2**
2.1 Overview . . . 2
2.1.1 Experiment/Aleatory Experiment . . . 2
2.1.2 Sample Space . . . 3
2.1.3 Event . . . 3
2.2 Set Notation and Boolean Algebra . . . 3
2.2.1 Sub/Sets . . . 3
2.2.2 Set Operations . . . 3

3 Frequentist Probability **3**
3.1 Axioms . . . 4

4 Discrete Random Variables **4**
4.1 Probability Mass Function . . . 4
4.2 Cumulative Distribution Function . . . 4
4.3 Expectation . . . 4
4.4 Binomial Distribution . . . 4
4.4.1 Mean . . . 4
4.4.2 Variance/Standard Deviation . . . 5
4.4.3 Probability . . . 5

1 Descriptive Statistics

1.1 Sample/Population Mean

Sample Mean:

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Popoulation Mean:

$$\mu = \frac{\sum_{i=1}^{N} y_i}{N}$$

1.2 Sample Variance/Standard Deviation

Sample Variance:

$$s^2 = \frac{\sum_{y_i-\overline{y}}^{n} y_i}{n-1}$$

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \overline{y})}{n-1}}$$

1.3 Empirical Rule

For approximately symmetrical (*Gaussian*) distributions:

(i) $\mu \pm \sigma$: contains ∼68% of the samples

(ii) $\mu \pm 2\sigma$: contains ∼95% of the samples

(iii) $\mu \pm 3\sigma$: contains ∼99% of the samples

1.4 Skew

1.4.1 Negative/Left Skew

Mass of the distribution is concentrated on the right. Consequently, the **tail** of the distribution points left. Note that in a negative/left skew: mean $\leq$ median $\leq$ mode.

1.4.2 Symmetric/No Skew

Statistician's wet dream: mean = median = mode.

1.4.3 Positive/Right Skew

Mass of the distribution is concentrated on the left. Consequently, the **tail** of the distribution points right. Note that in a positive/right skew: mean $\leq$ median $\leq$ mode.

2 Probability and Bayes Theorem

2.1 Overview

2.1.1 Experiment/Aleatory Experiment

Experiment: Process by which an observation is made
Aleatory Experiment: When replicated under the same conditions, the experiment may not yield the same results; that is, conditions of the experiment determine the probabilistic behavior of the results.

2.1.2 Sample Space

Set consisting of all possible outcomes (samples) of an aleatory experiment. Sample spaces can be:

- continuous or discrete
- finite, non/enumerable infinity

2.1.3 Event

A subset of the sampling space ($\subseteq$)

2.2 Set Notation and Boolean Algebra

2.2.1 Sub/Sets

A set is a collection of items/elements, each with a specified characteristic. A set that includes **all** items of interest is called the **universal set** and is denoted by Ω. Sets can be discrete or continuous.

A subset is a set derived from the universal set Ω. Set relationships can be illustrated via a Venn Diagram. If A is a subset of B and B is a subset of Ω, we denote that as $A \subseteq B \subseteq \Omega$.

2.2.2 Set Operations

Given sets A and B,

Intersection ($\cap$)

The intersection of two sets is defined as $A \cap B := \{x : x \in A \text{ and } x \in B\}$.
Note: We call two sub/sets A, B **disjoint mutually exclusive** if $A \cap B = \emptyset$

Union ($\cup$)

The union of two sets is defined as $A \cup B := \{x : x \in A \text{ (inclusive) or } x \in B\}$.

Complement ($\overline{A}$)

The complement of a set is defined as $\overline{A} := \{x : x \notin A\}$

Boolean Algebra

JUST TAKE THE PICS FROM SLIDES

3 Frequentist Probability

Based on previous relative frequences, the following conditions must hold:

(i) The relative frequence of occurerence of an event must be ≥ 0

(ii) The relative frequency of the whole sample space must be 1

(iii) If two events are mutually exclusive, the relative frequency of their union is the sum of their respective relative frequencies

Each random event E is associated with a probability $P(E)$ of the occurrence of the event:

$$P(E) = \frac{N_E}{N}$$

where N_E is the number of elements in the set E (i.e., the number of outcomes favorable to the event E) and N is the number of elements in S (i.e., all possible outcomes)

3.1 Axioms

Supposed S is a sample space associated with an experiment and E is a random event in S. Then,

(i) $0 \leq P(E) \leq 1$

(ii) $P(S) = 1$

(iii) If $\bigcap_{i=1}^{\infty} E_i = \emptyset$, then $P\big(\bigcup_{i=1}^{\infty} E_i\big) = \sum_{i=1}^{\infty} P(E_i)$

4 Discrete Random Variables

A random variable X is a numerical representation of every outcome. They can be discrete or continuous.

4.1 Probability Mass Function

Probability Mass Function: Assigns a probability to each possible domain value of X

$$P(X = x_i) = p_i,\ i = 1, 2, \ldots$$

where $0 \leq P(X = x_i) \leq 1$ and $\sum_i P(X = x_i) = 1$.

4.2 Cumulative Distribution Function

$$F(X) = P(X \leq x) = \sum_{i:x_i \leq x} P(X = x_i)$$

such that

(i) $F(x_1) \leq F(x_2)\ \forall x_1 \leq x_2$

(ii) $\lim_{x \to -\infty} F(x) = 0$

(iii) $\lim_{x \to +\infty} F(x) = 1$

4.3 Expectation

If X is a discrete random variable with a probability distribution $p(x)$, then the expectation of X is:

$$E(X) = \sum x \cdot p(x)$$

Composite expectations: $p(x), g(X)$:

$$E\big[g(X)\big] = \sum g(x) \cdot p(x)$$

4.4 Binomial Distribution

Has the following properties:

(i) Binary outcomes

(ii) Independent trials

(iii) n number of trials

(iv) Same probability p per trial

4.4.1 Mean

$$\mu = n \cdot p$$

4.4.2 Variance/Standard Deviation

Variance:

$$\sigma^2 = n \cdot p \cdot (1 - p)$$

or

$$\sigma^2 = n \cdot p \cdot q \text{ where } q = (1 - p)$$

Standard Deviation:

$$\sigma = \sqrt{\sigma^2} = \sqrt{n \cdot p \cdot (1 - p)}$$

or

$$\sigma = \sqrt{\sigma^2} = \sqrt{n \cdot p \cdot q}$$

4.4.3 Probability

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

or

$$P(X = x) = \frac{n!}{(n-x)!x!} \cdot p^x q^{n-x}$$

where $n =$ trials, $p =$ success, $q =$ failure