Nov 3, 2022

Engineering Statistics

Week 4: Random variable, probability distribution and expected value

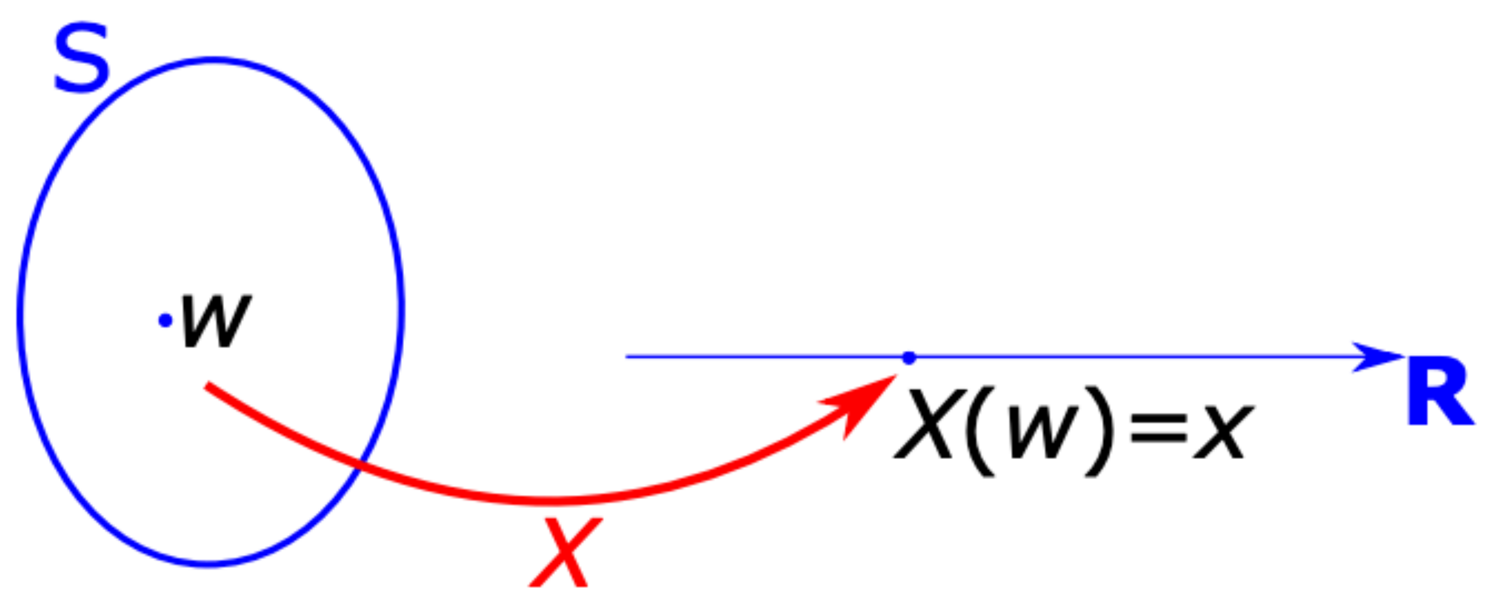
Random variable

Random variable

- For a given sample space S of some experiment, a random variable is any rule that associates a number with each outcome in S.
- Random variables are customarily denoted by uppercase letters, such as X and Y, near the end of our alphabet.
- The values that a random variable X assumes is denoted by lower x.

Random variable

In mathematical language, a random variable is a function whose domain is the sample space and whose range is the set of real numbers.



$$\begin{aligned} X : S &\longrightarrow \mathbb{R} \\ w &\longmapsto X(w) = x \end{aligned}$$

Types of random variables

Discrete

Assume a countable number of values are called discrete

Continuous

Assume values corresponding to any of the points contained in an interval are called continuous.

Examples

<u>For each random variable defined here, state whether the variable is discrete.</u>

(a) the number of unbroken eggs in a randomly chosen standard egg carton

(b) the number of students on a class list for a particular course who are absent on the first day of classes

(c) the number of times a duffer has to swing at a golf ball before hitting it.

(d) the length of a randomly selected rattlesnake

(e) the amount of royalties earned from the sale of a first edition of 10.000 textbooks.

(f) the pH of a randomly chosen soil sample

Probability distribution

Probability distribution of a discrete random variable

The probability distribution function of a discrete random variable X represents the probability that X takes the value x, as a function of x. That is,

$$f(x) = P(X = x), \text{ for all } x$$

Probability distribution of a discrete random variable

The probability distribution of a discrete random variable is a **graph**, **table** or **formula** that specifies the probability associated with each possible value that the random variable can assume.

Probability distribution of a discrete random variable

Required properties of probability distribution for discrete random variables:

1. $0 \leq P(X = x) \leq 1$, for all x.

2. $\sum_{x} P(X = x) = 1$

Example

Six lots of components are ready to be shipped by a certain supplier. The number of defective components in each lot is as follows:

Lot	1	2	3	4	5	6
Number of defectives	0	2	0	1	2	0

One of these lots is to be a randomly selected for shipment to a particular customer. Let X be the number of defectives in the selected lot. **Find the probability distribution of X.**

Example

The following table is a partial probability distribution for the MRA Company's projected profits (X: profit in $\$1000$) for the first year of operation (the negative value denotes a loss).

$X = x$	-100	0	50	100	150	200
$P(X = x)$	0.10	0.20	0.30	0.25	0.10	?

What is the proper value for $P(X = 200)$?

What is the probability that MRA will be profitable?

What is the probability that MRA will make at least $\$100.000$?

Example

A drought is a period of abnormal dry weather that causes serious problems in the farming industry of the region. University of Arizona researchers used historical annual data to study the severity of droughts in Texas. The researchers showed that the distribution of X, the number of consecutive years that must be sampled until a dry year is observed can be modeled using the following formula:

$$P(X = x) = 0.3(0.7)^{x-1},\ x = 1,2,3,...$$

Find the probability that exactly 3 years must be sampled before a drought year occurs.

A parameter of a probability distribution

Suppose $P(X = x)$ depends on a quantity that can be assigned any one of a number of possible values, with each different value determining a different probability distribution. Such a quantity is called **a parameter of the distribution**. The collection of all probability distributions for different values of the parameter is called **a family of probability distributions**.

Expected value

Expected value

Mean, or expected value of a discrete random variable X:

$$\mu = E(X) = \sum_{x} xP(X = x)$$

Expected value

- The average or mean value of X is a weighted average of the possible values of X, where the weights are the probability of those values.
- The expected value is the mean of the probability distribution, or a measure of its central tendency.
- It can be thought as the mean value of X in a very large (actually, infinite) number of repetitions of the experiment in which the values of X occur in proportions equivalent to the probabilities of X.
- We will often refer to μ as the population mean.

Variance

Variance

Variance of a discrete random variable X:

$$\sigma^2 = Var(X) = E(x-\mu)^2 = \sum_x (x-\mu)^2 P(X=x)$$

where μ is the expected value of X.

Variance

- The quantity $(x-\mu)^2$ is the squared deviation of X from its mean, and σ^2 is the expected squared deviation.

- The weighted average of squared deviations, where the weights are probabilities from the distribution.

- If most of the probability distribution is close to μ, then σ^2 will be relatively small. However, if there are x values far from μ that have large $P(X=x)$, then σ^2 will be quite large. Very roughly, σ^2 can be interpreted as the size of a representative deviation from the mean value μ.

- The variance of a random variable is a measure for the variability.

- We will often refer to σ^2 as the population variance.

Variance

Short-cut formula for the variance:

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

where $E(X^2) = \sum_x x^2 P(X = x)$.

Example

A library has an upper limit of 6 on the number of videos that can be checked out to an individual at one time. Consider only those who check out videos, and let denote the number of videos checked out to a randomly selected individual. The probability distribution of X as follows:

$X = x$	1	2	3	4	5	6
$P(X = x)$	0.30	0.25	0.15	0.05	0.10	0.15

- Find the mean of random variable X.
- Find the variance of random variable X.

Binomial distribution

Binomial experiment

A binomial experiment exhibits the following four properties:

1. The experiment consists of n identical trials.

2. There are only two possible outcomes on each trial. We will denote one outcome by S (for **Success**) and the other by F (for **Failure**).

3. The probability of S remains the same from trial to trial. This probability is denoted by p, and the probability of F is denoted by $q = 1 - p$.

4. The trials are independent.

Example #1

The same coin is tossed successively and independently n times. We arbitrarily use S to denote the outcome H (heads) and F to denote the outcome T (tails).

Example #2

Before marketing a new product on a large scale, many companies conduct a consumer-preference survey to determine whether the product is likely to be successful. Suppose a company develops a new diet soda and then conducts a taste-preference survey in which 100 randomly chosen consumers state their preferences from among the new soda and the two leading sellers.

The binomial random variable and distribution

The binomial random variable X associated with a binomial experiment consisting of n trials is defined as:

X: the number of successes among the n trials

The binomial random variable and distribution

The binomial probability distribution

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

p : probability of a success on a single trial

n : number of trials

x : number of successes in n trials

The binomial random variable and distribution

The random variable X having the probability distribution given is said to have Binomial distribution with parameters n and p. This is shortly denoted by:

$$X \sim Binom(n, p)$$

The binomial random variable and distribution

The mean and variance of a binomial random variable:

Suppose that $X \sim Binom(n, p)$. Then,

$$\mu = E(X) = np \text{ and } \sigma^2 = np(1-p)$$

Example

A study of various brands of bottled water conducted by the Natural Resources Defense Council found that 25% of the bottled water is just tap water packaged in a bootle (Scientific American, July 2003).

Consider a sample of five bottled-water brands, and let X equals the number of these brands that use tap water.

- Explain why X is (approximately) a binomial random variable.
- Give the probability distribution for X as a formula.
- Find $P(X = 2)$.
- Find $P(X \geq 2)$.
- Find mean and variance of X.

Poisson distribution

Poisson distribution

- A type of probability distribution that is often useful in describing the number of rare events that will occur during a specific period or in a specific area or volume is the Poisson distribution.
- The name of the distribution comes from the 18th-century physicist and mathematician Simeon-Denis Poisson.

Examples

- The number of traffic accidents per month at a busy intersection
- The number of noticeable surface defects (scratches, dents, etc.) found by quality inspectors on a new automobile.
- The number of parts per million (ppm) of some toxin found in the water or air emissions from a manufacturing plant.
- The number of diseased trees per acre of a certain woodland.

Characteristics of Poisson distribution

- The experiment consists of counting the number of times a certain event occurs during a given unit of time or in a given area or volume (or weight, distance, or any other unit of measurement).

- The probability that an event occurs in a given unit of time, area, or volume is the same for all the units.

- The number of events that occur in one unit of time, area, or volume is independent of the number that occur in other units.

- The mean (or expected) number of events in each unit is denoted by the Greek letter λ (called lambda).

The Poisson random variable and distribution

Poisson probability distribution:

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \; x = 0,1,2,3,...$$

where

λ : mean number of events during a given unit of time, area, volume, etc.

e : Euler constant (~ 2.71828)

The Poisson random variable and distribution

The random variable X having the probability distribution given is said to have Poisson distribution with parameter λ. This is shortly denoted by:

$$X \sim Poisson(\lambda)$$

The Poisson random variable and distribution

Suppose that $X \sim Poisson(\lambda)$. Then,

$$\mu = E(X) = \lambda \text{ and } \sigma^2 = \lambda$$

Example

Airline passengers arrive randomly and independently at the passenger-screening facility at a major international airport. The mean arrival rate is 10 passengers per minute.

- Compute the probability that three or fewer passengers arrive in a one-minute period.
- Compute the probability of no arrivals in a 15-second period.
- Compute the probability of at least one arrival in a 15-second period.

Course materials

You can download the notes and codes from:

https://github.com/mcavs/ESTUMatse_2022Fall_EngineeringStatistics

SCAN ME

Contact

Do not hesitate to contact me on:

https://twitter.com/mustafa_cavus

https://www.linkedin.com/in/mustafacavusphd/

mustafacavus@eskisehir.edu.tr