Probability and Random Variables

ANA 500 – Foundations of Data Analytics

Module1_Week2A

Probability

- We assume the values taken on by the variables in our dataset are not know until they are observed.
- Our data is composed of observations of underlying random variables.
- Probability theory provides the mathematics of this randomness.
- Probability: fraction of occurrences an event occurs over many repeated trials.
- E.g. The probability of a coin flip coming up heads is the fraction of heads that occurs after many, many flips。

Probability

- Some terminology (couched in terms of the coin flipping example):
  - The **experiment** is the flipping of the coin. The **outcome** is not known until the coin is flipped, so there is **chance** involved. All possible outcomes that can occur represent the **sample space (S).** An **event** is one or more outcomes.
  - is repeated and conducted under controlled conditions.
  - S = {H, T} The experiment

Probability

- We use uppercase letters to denote events. Suppose A is the event of the coin coming up heads. We write P(A) to denote the probability of the event occurring.

- $0 < P(A) < 1$

- **Law of large numbers**: as the number of trials increases the empirical fraction of occurrences gets closer and closer to the theoretical probability of occurrence.

Probability

- Some rules:
  - “and”: an outcome is in event A and B if it is in A and B at the same time.

    - A = {1, 2, 3, 4, 5, 6}
    - B = { 4, 5, 6, 7, 8}
    - A and B = $A \cap B$ = {4, 5, 6}

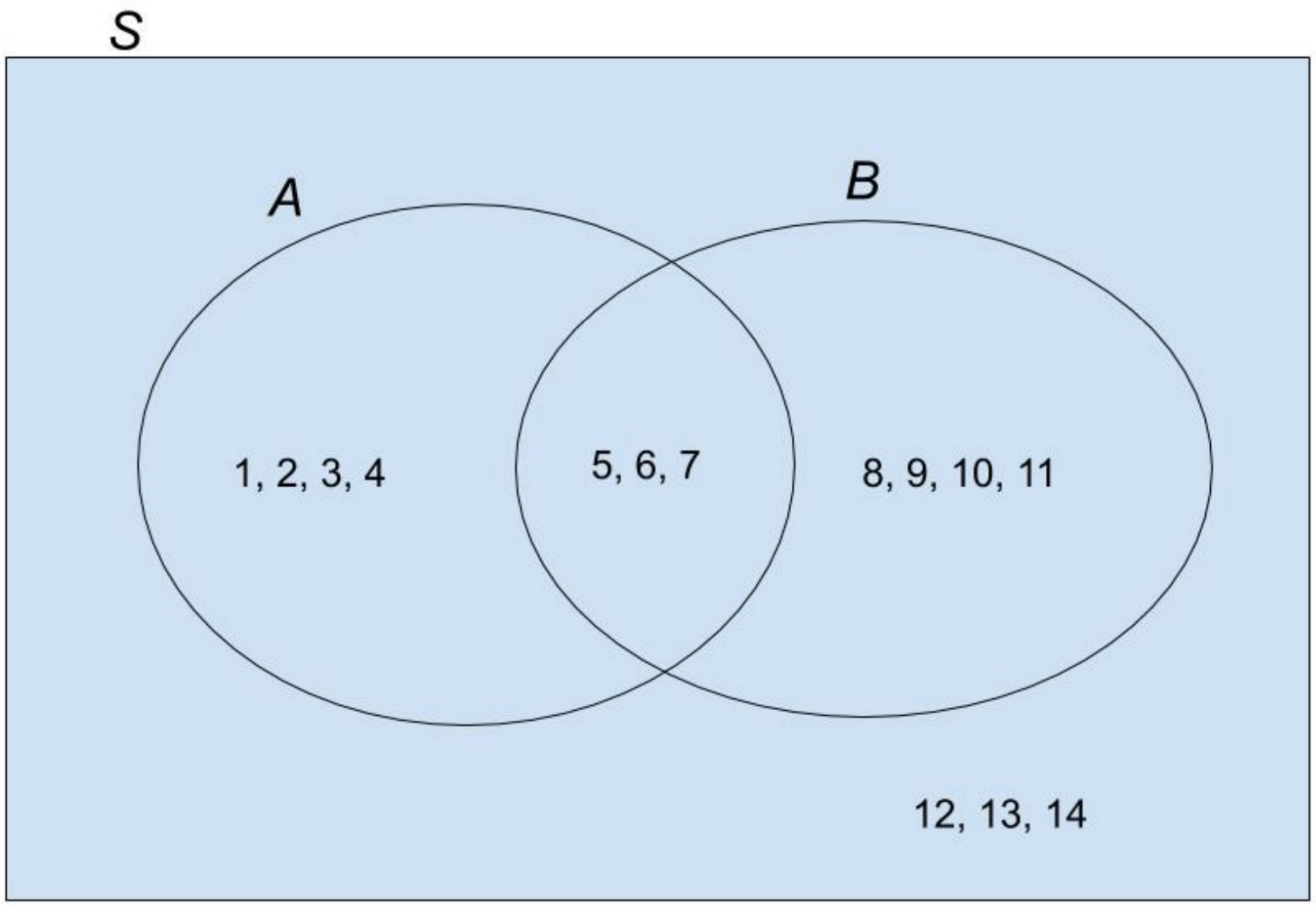
Probability

- Some rules:

  - “or”: an outcome is in event A or B if it is in A or B
    - A = {1, 2, 3, 4, 5, 6}
    - B = { 4, 5, 6, 7, 8}
    - A or B = $A \cup B$ = {1, 2, 3, 4, 5, 6, 7, 8}

  - The complement of event A is denoted A’.
    - P(A’) = 1 – P(A)

Probability



$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14\}$
$A = \{1, 2, 3, 4, 5, 6, 7\}$
$B = \{5, 6, 7, 8, 9, 10, 11\}$

Conditional probability

- Conditional probability:

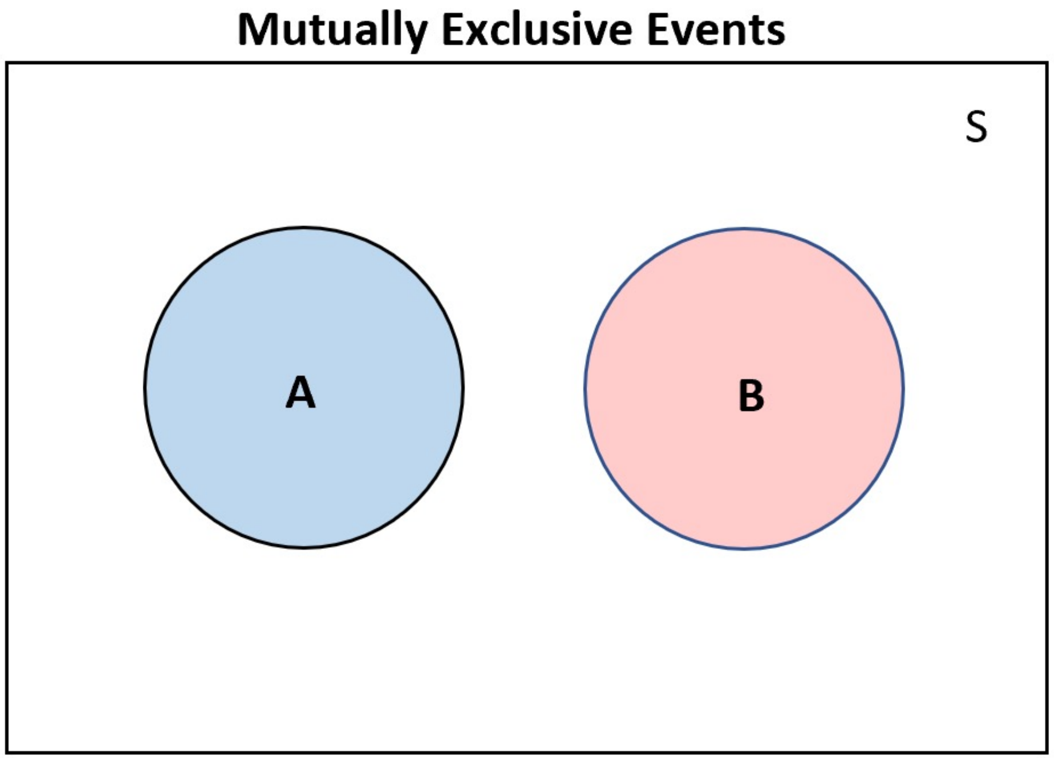
$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

- *A conditional probability reduces the sample size*
  - *E.g.*, For example, perhaps event B is someone having an income over \$100,000, and event A is buying a particular product. The conditional probability is the probability that someone with over \$100,000 buys the product.

Mutually Exclusive

- Two events are mutually exclusive if $P(A \cap \mathrm{B}) = 0$

Two events cannot occur together



Independent Events

The outcome of one event does not affect the outcome of the other.

- Two events are independent if $P(A|B) = P(A)$
- If two events are independent, then $P(A \cap B) = P(A) * P(B)$
- Conditional probability:

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

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

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

Probability

- Multiplication rule $P(A \cap B) = P(B) * P(A|B)$
- Additional rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Random variables

- A random variable provides a numerical description of the outcomes of an experiment.

- A discrete random variable has outcomes that are countable.
  - E.g. Number of children, number of computer crashes

- Uppercase letters denote random variables (X and Y). These are described with words.

- Lowercase letters denote the outcome (x and y) . These are described with numbers.
  - E.g. we roll a six-sided die and it comes up 4
    - X = rolling a six-sided die
    - x = 4

Random variables

- A **probability distribution function** describes how the probabilities are assigned over all outcomes of a discrete random variable.

  - We use $f(x)$ to denote a probability distribution function
  - E.g. if the probability of rolling a 4 with a six-sided die is 1/6
    - $f(4) = 1/6$

Uniform distribution

- A probability distribution that describes a random variable for which all outcomes are equally likely.

- E.g. $n=6$; $f(x) = 1/n = 1/6$

Random variables

Rules:

- $0 \leq f(x) \leq 1$

- $\sum f(x) = 1$
  - E.g. 1/6 + 1/6 + 1/6 + 1/6 + 1/6 + 1/6 =1/6 * 6=1

Random variables

- The center and spread are important characteristics of a probability distribution.

- Center: mean = Expected Value: $E[x] = \sum f(x) * x = \mu$

- Note that the expected value does not have to be a value that can actually occur.

    e.g., the expected value of rolling a six-sided die is 3.5.

Random variables

- Law of large numbers: as the number of observations increases the difference between the empirical mean $\bar{x}$. and the theoretical mean (μ) gets infinitely smaller.

Random variables

- Spread: variance = $E[(x - \mu)^2] = \sum f(x) * (x - \mu)^2 = \sigma^2$

- A variance can be difficult to interpret because of the squaring. The units of measurement are no longer the original units of measurement after the squaring. For this reason, the standard deviation is often a preferred measure of spread.

- The standard deviation is simply equal to the square root of the variance.

Binomial distribution

- A binomial probability distribution is a well-known **discrete** probability distribution
- n independent trials
- Each trial has the outcome of either ”success” or “failure”
- The probability of “success” is p and the probability of “failure” is q
- The number of ”successes” is denoted k
- Each trial is repeated under identical conditions
- The outcomes of a binomial experiment fit a binomial distribution
- $X \sim B(n, p)$

Binomial distribution

$$f(x) = \binom{n}{k} p^k q^{n-k} \qquad \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

- Where:
  - the k is the number of Successes;
  - n is the number of total trials;
  - p and q are the probabilities of a success and failure, respectively.

- n factorial: $n! = [1*2*3*4* \ldots *n]$

Binomial distribution

- Mean = np

- Variance = npq

- A binomial experiment with a single trial is known as a **Bernoulli trial.**

  - One example of a Bernoulli trial is the coin-tossing experiment, which results in heads or tails.

Poisson distribution

- An important **discrete** probability distribution.
- Often used to model the number of occurrences of an event over a specified time interval or space.
- Instead of the probability of an event, the Poisson Distribution requires knowing how often it occurs for a specific period of time or distance.
- It is assumed that the rate of occurrence is the same for any two intervals of equal length, and the occurrence of an event in any interval is independent of the occurrence in any other.

Poisson distribution

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

- the “e” is known as Euler’s number or Napier’s constant. It is a fixed value approximately equal to 2.72. $e^{-\lambda} = \frac{1}{e^{\lambda}}$
- x: number of occurrences in an interval
- λ: average mean rate of occurrence

- $X \sim Pois(\lambda)$
- Mean = λ
- Variance = λ

Poisson distribution

Example:

- Imagine you created an online course on probability. Usually, your students ask you around 4 questions per day, but yesterday they asked 7. Surprised by this sudden spike in interest from your students, you wonder how likely it was that they asked exactly 7 questions.

Poisson distribution

Example:

- In this example, the average questions you anticipate is 4, so lambda equals 4. (λ=4)

- The time interval is one entire workday and the singular instance you are interested in is 7. (x=7)

- $f(x) = \frac{\lambda^x e^{-\lambda}}{x!}$: $f(7) = \frac{4^7 e^{-4}}{7!} \approx \frac{16384 * 0.0183}{5040} \approx 0.06$

- Therefore, there was only a 6% chance of receiving exactly 7 questions.