Probability Theory

upGrad
knowledge
hut

Learning Objectives

- Define probability theory
- Explain the probability basics
- Discuss simple calculating probabilities
- Describe the rule of addition
- Define the rule of multiplication, exercise and solution
- Explain the Bayes theorem

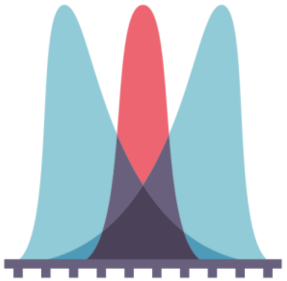
Learning Objectives

- Discuss the expected value and the examples
- Define the law of large numbers
- Discuss the central limit theorem through experiments
- Describe the central limit theorem intuition and challenge
- Explain the Binomial distribution
- Define Poisson distribution

Intro

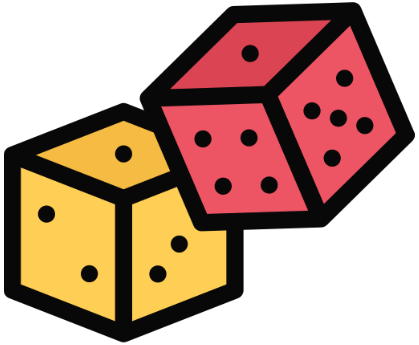
Probability Theory

- Calculates the probability
- Concepts of the expected value
- Important theories of the Law of large numbers
- Bayes theorem and the central limit theorem
- Works on the real-life problems



Probability Basics

Definition – Probability Theory



The transition from descriptive statistics to a mathematical concept is probability theory.

Probability Theory

Definition: Sample Space, S

A Sample Space is the set of all possible outcomes of an experiment. For example:

- Coin flip: $S = \{\text{Heads, Tails}\}$
- Flipping two coins: $S = \{(\text{H, H}), (\text{H, T}), (\text{T, H}), (\text{T, T})\}$
- Roll of 6-sided die: $S = \{1, 2, 3, 4, 5, 6\}$
- The number of emails you receive in a day: $S = \{x | x \in \mathbb{Z}, x \geq 0\}$ (non-neg. ints)
- YouTube hours in a day: $S = \{x | x \in \mathbb{R}, 0 \leq x \leq 24\}$

Definition: Event, E

An Event is some subset of S that we ascribe meaning to. In set notation ($E \subseteq S$).For example:

- Coin flip is heads: $E = \{\text{Heads}\}$
- At least 1 head on 2 coin flips = $\{(\text{H, H}), (\text{H, T}), (\text{T, H})\}$
- Roll of die is 3 or less: E = $\{1, 2, 3\}$
- You receive less than 20 emails in a day: $E = \{x | x \in Z, 0 \leq x < 20\}$ (non-neg. ints)
- Wasted day ($\geq$ 5 YouTube hours): $E = \{x | x \in R, 5 \leq x \leq 24\}$

Events can be represented as capital letters such as E or F.

Probability of Equally Likely Outcomes

Definition: Probability of Equally Likely Outcomes

If S is a sample space with equally likely outcomes, for an event E that is a subset of the outcomes in S:

$$\mathrm{P}(E) = \frac{\text{number of outcomes in } E}{\text{number of outcomes in } S} = \frac{|E|}{|S|}$$

Axioms of Probability

Axiom 1: $0 \leq \mathrm{P}(E) \leq 1$	All probabilities are numbers between 0 and 1.
Axiom 2: $\mathrm{P}(S) = 1$	All outcomes must be from the Sample Space.
Axiom 3: If E and F are mutually exclusive, then $\mathrm{P}(E \text{ or } F) = \mathrm{P}(E) + \mathrm{P}(F)$	The probability of "or" for mutually exclusive events

Identity 1: $\mathrm{P}(E^{\mathrm{C}}) = 1 - \mathrm{P}(E)$	The probability of event E not happening
Identity 2: If $E \subseteq F$, then $\mathrm{P}(E) \leq \mathrm{P}(F)$	Events which are subsets

Demonstration: Probability Basics

Practical Session

Calculating Simple Probability

Simple Probability - Exercise

Two normal dice with 6 sides are thrown.

The result of dice 1 and dice 2 is noted.

1. What is the number of possible outcomes? 36
2. What is the probability that the sum of the two dice is greater than 9? 1/6
3. What is the probability that the least one of the dice is greater than 4? 5/9
4. What is the probability that both are less than 3? 1/9

Demonstration: Simple Probability Solution

Practical Session

Rule of Addition

Probability of OR

Definition: Probability of **or** for mututally exclusive events

If two events: E, F are mutually exclusive then the probability of E ***or*** F occurring is:

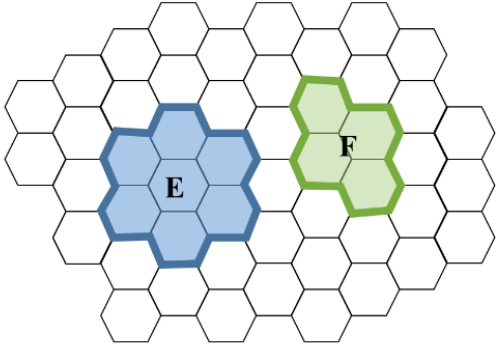
$$\mathrm{P}(E \text{ or } F) = \mathrm{P}(E) + \mathrm{P}(F)$$

This property applies regardless of how you calculate the probability of E or F. Moreover, the idea extends to more than two events. Lets say you have n events $E_1, E_2, \ldots E_n$ where each event is mutually exclusive of one another (in other words, no outcome is in more than one event). Then:

$$\mathrm{P}(E_1 \text{ or } E_2 \text{ or } \ldots \text{ or } E_n) = \mathrm{P}(E_1) + \mathrm{P}(E_2) + \cdots + \mathrm{P}(E_n) = \sum_{i=1}^{n} \mathrm{P}(E_i)$$

What is the probability getting 1 or 2 for rolling a dice ?

$$P(1 \; or \; 2) = P(1) + P(2) = ?$$



Example of two events: E, F, which are mutually exclusive.

Probability of OR

Buggy derivation: Incorrectly assuming mutual exclusion

Calculate the probability of E, getting an even number on a dice role (2, 4 or 6), or F, getting three or less (1, 2, 3) on the same dice role.

$$\begin{aligned} \mathrm{P}(E \text{ or } F) &= \mathrm{P}(E) + \mathrm{P}(F) && \text{Incorrectly assumes mutual exclusion} \\ &= 0.5 + 0.5 && \text{substitute the probabilities of } E \text{ and } S \\ &= 1.0 && \text{uh oh!} \end{aligned}$$

The probability can't be one since the outcome 5 is neither three or less nor even. The problem is that we double counted the probability of getting a 2, and the fix is to subtract out the probability of that doubly counted case.

Probability of OR

Definition: Inclusion Exclusion principle

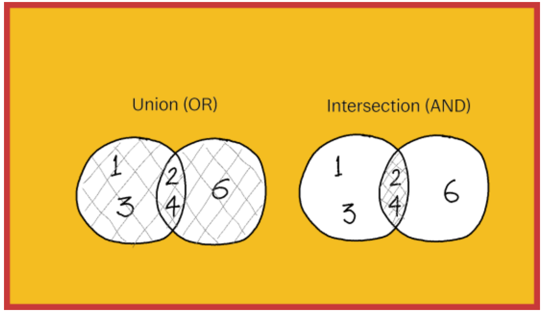
For any two events: E, F:

$$\mathrm{P}(E \text{ or } F) = \mathrm{P}(E) + \mathrm{P}(F) - \mathrm{P}(E \text{ and } F)$$

This formula does have a version for more than two events, but it gets rather complex. See the next two sections for more details.

$$P(E\ or\ F) = P(E\ \cup F)$$

$$P(E\ and\ F) = P(E \cap F)$$



Exercise: Addition Rule

1

You work as an analyst at a car dealer and evaluate the sales.

The probability that a customer will choose a light car color (= event A) is 20%.

The probability that the decision is made in favor of a station wagon is 40% (= event B).

And the probability that the decision is made for a light-colored station wagon is 15%.

What is the probability that the decision to buy is made in favor of a light-colored car or a station wagon?

2

You work at a retail chain as an analyst.

While analyzing the sales, you found out that 30% of the customers included clothes in their purchase and 50% of the customers included food in their purchase. In addition, you found that 60% of all purchases included clothing or food.

What is the probability of food and clothing being included in a purchase?

Rule of Multiplication

Rule of Multiplication

The Multiplication rule enables the calculation of the combination of two events.

Probability of AND

Definition: Probability of ***and*** for independent events.

If two events: E, F are independent then the probability of E ***and*** F occurring is:

$$\mathrm{P}(E\ \text{and}\ F) = \mathrm{P}(E) \cdot \mathrm{P}(F)$$

This property applies regardless of how the probabilities of E and F were calculated and whether or not the events are mutually exclusive.

The independence principle extends to more than two events. For n events $E_1, E_2, \ldots E_n$ that are ***mutually*** independent of one another -- the independence equation also holds for all subsets of the events.

$$\mathrm{P}(E_1\ \text{and}\ E_2\ \text{and} \ldots \text{and}\ E_n) = \prod_{i=1}^{n} \mathrm{P}(E_i)$$

Probability of AND

Definition: The chain rule.

The formula in the definition of conditional probability can be re-arranged to derive a general way of calculating the probability of the ***and*** of any two events:

$$\mathrm{P}(E \text{ and } F) = \mathrm{P}(E|F) \cdot \mathrm{P}(F)$$

Of course there is nothing special about E that says it should go first. Equivalently:

$$\mathrm{P}(E \text{ and } F) = \mathrm{P}(F \text{ and } E) = \mathrm{P}(F|E) \cdot \mathrm{P}(E)$$

We call this formula the "chain rule." Intuitively it states that the probability of observing events E ***and*** F is the probability of observing F, multiplied by the probability of observing E, given that you have observed F. It generalizes to more than two events:

$$\begin{aligned}\mathrm{P}(E_1 \text{ and } E_2 \text{ and} \ldots \text{and } E_n) = &\mathrm{P}(E_1) \cdot \mathrm{P}(E_2|E_1) \cdot \mathrm{P}(E_3|E_1 \text{ and } E_2) \ldots \\ &\mathrm{P}(E_n|E_1 \ldots E_{n-1})\end{aligned}$$

Demonstration: Multiplication Rule

Practical Session

Bayes Theorem

Bayes Theorem

The Bayes theorem calculates the conditional probability.

likelihood

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

prior: general belief before seeing data

posterior : belief after seeing data.

evidence

It is used in the multiplication rule.

Demonstration: Bayes Theorem

$P(illness|pos)$ if the test result is not 100% reliable ?

$P(pos|illness)$: probability of a positive result given illness = 0.92

$P(pos|not\ illness)$: probability of a positive result given illness = 0.1

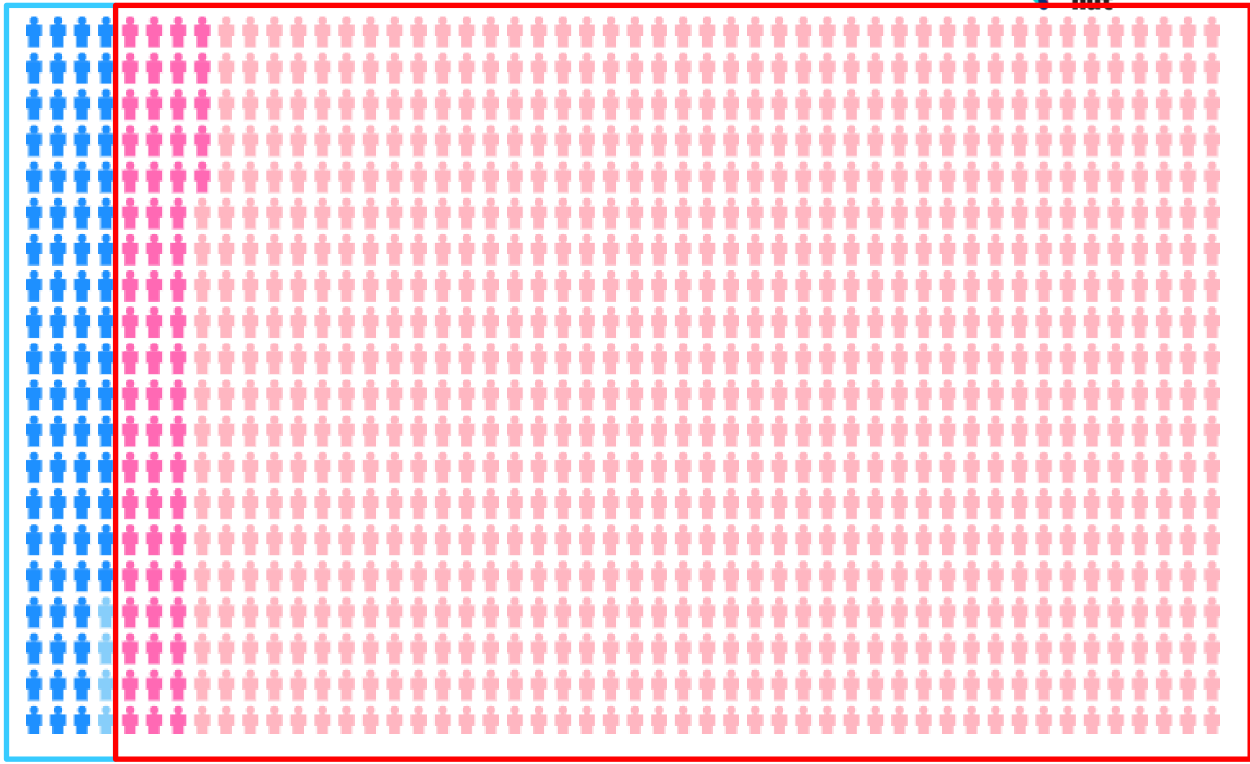
$P(illness)$: probability of illness = 0.13

$P(not\ illness) = 1 - P(illness)$

Demonstration: Bayes Theorem

$P(illness|pos)$ if the test result is not 100% reliable ?

- We are going to color people who **have the illness** in blue and those **without the illness** in pink.
- A certain number of people **with the illness will test positive** (which we will draw in Dark Blue).
- A certain number of people **without the illness will test positive** (which we will draw in Dark Pink):



$N * P(illness)$

$N * P(not\ illness)$

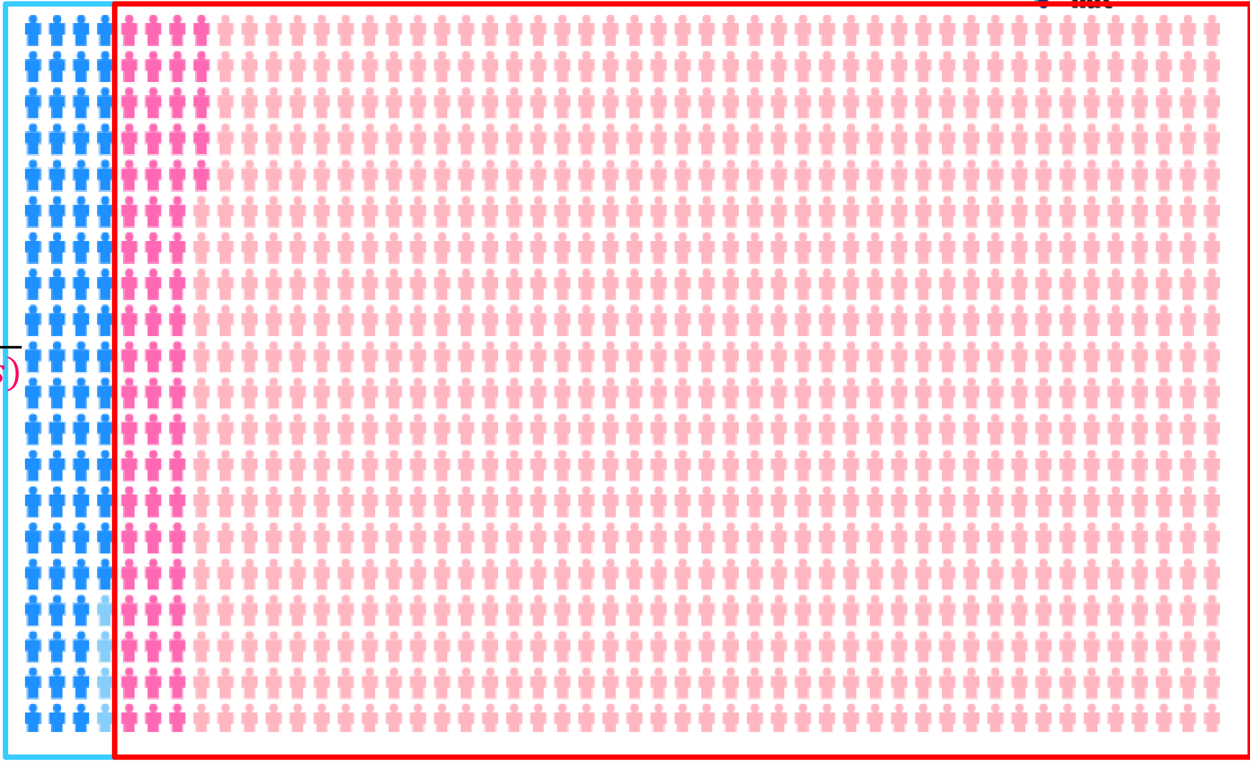
$N * P(illness) * P(pos|illness) = N * P(illness \,\&\, pos)$

$N * P(not\ illness) * P\ (pos|not\ illness) = N*P(not\ illness \,\&\, pos)$

Demonstration: Bayes Theorem

$P(illness|pos)$ if the test result is not 100% reliable ?

$$P(illness|pos) = \frac{P(pos|illness) * P(illness)}{P(pos)} = \frac{P(illness \,\&\, pos)}{P(illness \,\&\, pos) + P(not\ illness \,\&\, pos)}$$



$$P(illness \,\&\, pos) = P(illness) \ * P(pos|illness) = 0.92 * 0.13$$

$$P(not\ illness \,\&\, pos) = P(not\ illness) * P(pos|not\ illness) = (1 - 0.13) * 0.1$$

$$P(illness|pos) = 0.5789$$

Expected Value

Expected Value

Definition: Expectation

The expectation of a random variable X, written $\mathrm{E}[X]$ is the average of all the values the random variable can take on, each weighted by the probability that the random variable will take on that value.

$$\mathrm{E}[X] = \sum_x x \cdot \mathrm{P}(X = x)$$

Recall the definition of mean or average ?

Law of Large Numbers

Law of Large Numbers

Law of large numbers

X $E(X)$ $\frac{X_1 + X_2 + \dots + X_n}{n} = \bar{X}_n$

$\bar{X}_n \rightarrow E(X)$ for $n \rightarrow \infty$

Central Limit Theorem - Theory

Central Limit Theorem - Intuition

Central Limit Theorem

X $\quad E(X)$

$$\frac{X_1 + X_2 + \dots + X_n}{n} = \bar{X}_n \qquad n = 5, 10, \dots$$

The distribution of sample means approaches a normal distribution even if the original experiment is not normally distributed.

Binomial Distribution

Binomial Distribution

Our random variable

Num trials

Probability of success on each trial

$$X \sim \mathrm{Bin}(n, p)$$

Is distributed as a

Binomial

With these parameters

Binomial Distribution

$$P(X = k) = \binom{n}{k} p^k (1-p)^{(n-k)}$$

$$k = 0, 1, 2, 3, \ldots, n$$

X represents the number of successes in n independent trials.

Binomial Distribution

Examples:

(1) If 20% of customers are happy with the new product, what is a probability of having exactly 20 out of 100 customers will buy the product ?

n= 100; p=0.2; k=20

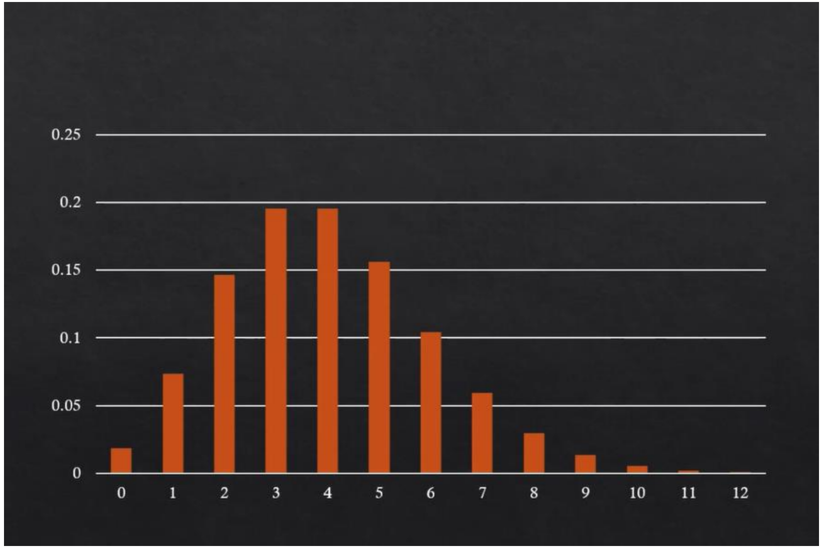
(1) What is a probability of having 99 positive reviews of 100 reviews if a chance of having a positive review is 50%?

n= 100; p=0.5; ; k=99

Poisson Distribution

Poisson Distribution

Poisson distribution is a discrete distribution taking only a discrete set of values, and the values have to be whole numbers.



Discrete Distribution

- Calculates the probability of certain events.
- Poisson distribution is used when the given period is an hour.
- Lambda is the only parameter.
- The distribution is different if the value is set to another number.

Probabilities Functions

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

e is the "Euler number"

e ≈ 2,72

Density Function

$$P(X \leq n) = e^{-\lambda} \sum_{k=0}^{n} \frac{\lambda^k}{k!}$$

Cumulative Density Function

Expected value and Variance

Expected value:

$$E(X) = \lambda$$

Variance:

$$V(X) = \lambda$$

Lambda is the parameter for the expected value and the variance.

Demonstration: Real Life Problems

Practical Session

Thank you!