


INNOMATICS RESEARCH LABS



Topics:

- Introduction to Probability
- Various Probability and Bayes Theorem
- Random Variable
- Probability Distribution – Bernoulli Binomial and Poisson
- Normal Distribution

Introduction to Probability

Intuitive understanding…..

- You and a friend are at a cricket match, and out of the blue he offers you a bet that neither player will hit a century in that game. Should you take the bet?
- Your company is launching personalized marketing campaign to millions of potential customers. To which customer should you offer what type of product.
- A widget maker in your factory that normally breaks 4 widgets for every 100 it produces has recently started breaking 5 widgets for every 100. When is it time to buy a new widget maker?
- You are conducting poll on national election for a big media house. How many people do you have to poll? How do you ensure that your poll is free of bias? How do you interpret your results?

What is Probability?

- Measure of likeliness of something happening
  - Strength of belief that something is true
  - Mathematical way of expressing uncertainty

- Given n observations of an event, it denotes the proportion of observations where a given event occurs
  - $$\boldsymbol{P(E) = \frac{\#\ of\ outcomes\ in\ which\ the\ even\ occurs}{total\ possible\ \#\ of\ outcomes}}$$

  - Probability of a single event is always between 0 and 1
  - Probability of all possible outcomes always sums to 1

Examples:

- In a coin toss, probability of a head or tails appearing?
- In a roll of dice, probability of 3 appearing?
  - Six possible outcomes: {1,2,3,4,5,6}
  - Each outcome equally likely, therefore probability of an outcome:1/6
  - Probability of an odd number appearing?
- Probability of amount of rain in Mumbai in August?
- Probability of RCB winning IPL 2021?

Introduction to Probability Theory

- Analytics applications involve tasks such as prediction of probability of occurrence of an event, testing a hypothesis, building models to explain variation of importance to the business such as Profitability, market share, demand etc…
- Many important tasks in analytics deal with uncertain events & it is essential to understand probability theory that can be used to predict & measure them
- We don’t know the outcomes of a particular situation until it happens. Will it rain today? Will I pass the next math test? Will my favorite team win the toss? Will I get a promotion in next 6 months? All these questions are examples of uncertain situations we live in.

Probability Theory - Terminology

Experiment – It can be either deterministic or random

- Deterministic: Outcome always same and determined
- Random: Many possible outcomes from a range of value.

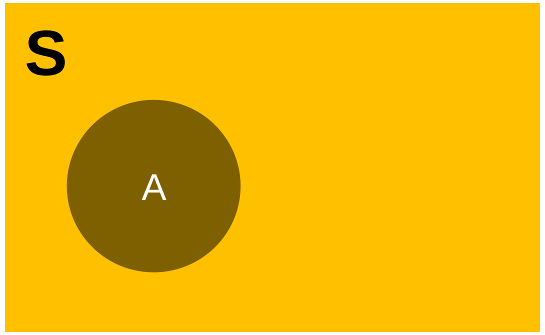
An experiment in which the outcome is not known with certainty .i.e. the output of this experiment cannot be predicted with certainty. Whether it rains on a daily basis is an experiment.

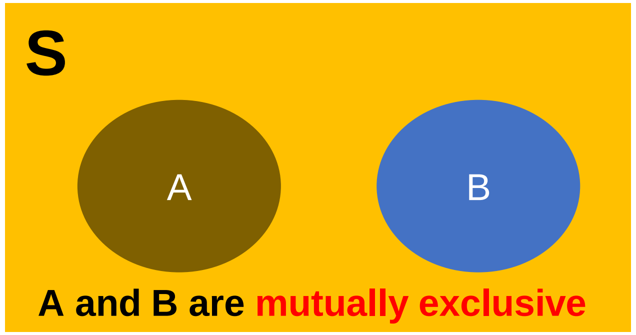
Outcome is the result of a single trial. So, if it rains today, the outcome of today's trial from the experiment is "It rained"

Event is one or more outcome from an experiment. "It rained" is one of the possible event for this experiment. Subset of sample space.

Probability is a measure of how likely an event is. So, if it is 60% chance that it will rain tomorrow, the probability of Outcome "it rained" for tomorrow is 0.6

- **Sample Space-** Given a random experiment K, set of all possible outcomes for K is denoted by S as sample space. For example: Rain experiment K, has a sample space S = {“It rained”, “It did not rain”}





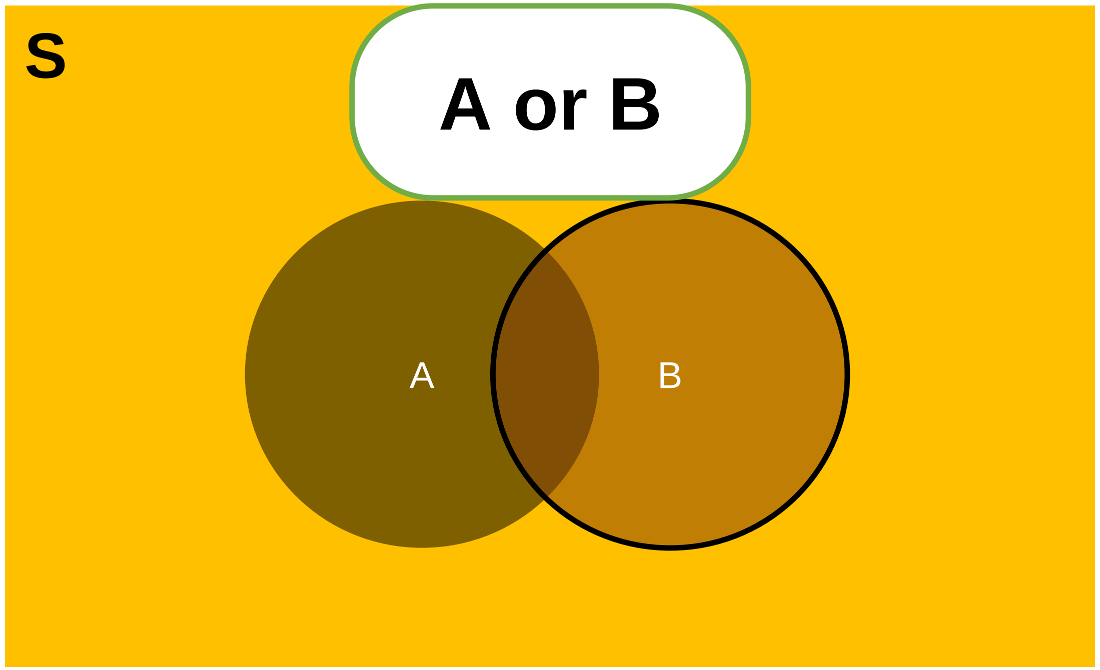
Area of the rectangle denotes sample space, and since probability is associated with area, it cannot be negative.

Mutually Exclusive – If event A happens, event B cannot.

Algebra of events:

Consider two events A and B

- **Intersection (A∩B):** Set of all elements common to A and B common
- **Union (A U B):** Set of all elements belonging to either A or B All
- **Difference (A-B):** Elements belonging to A but not to B
- **A and B are mutually exclusive** or disjoint if A∩B = Ø



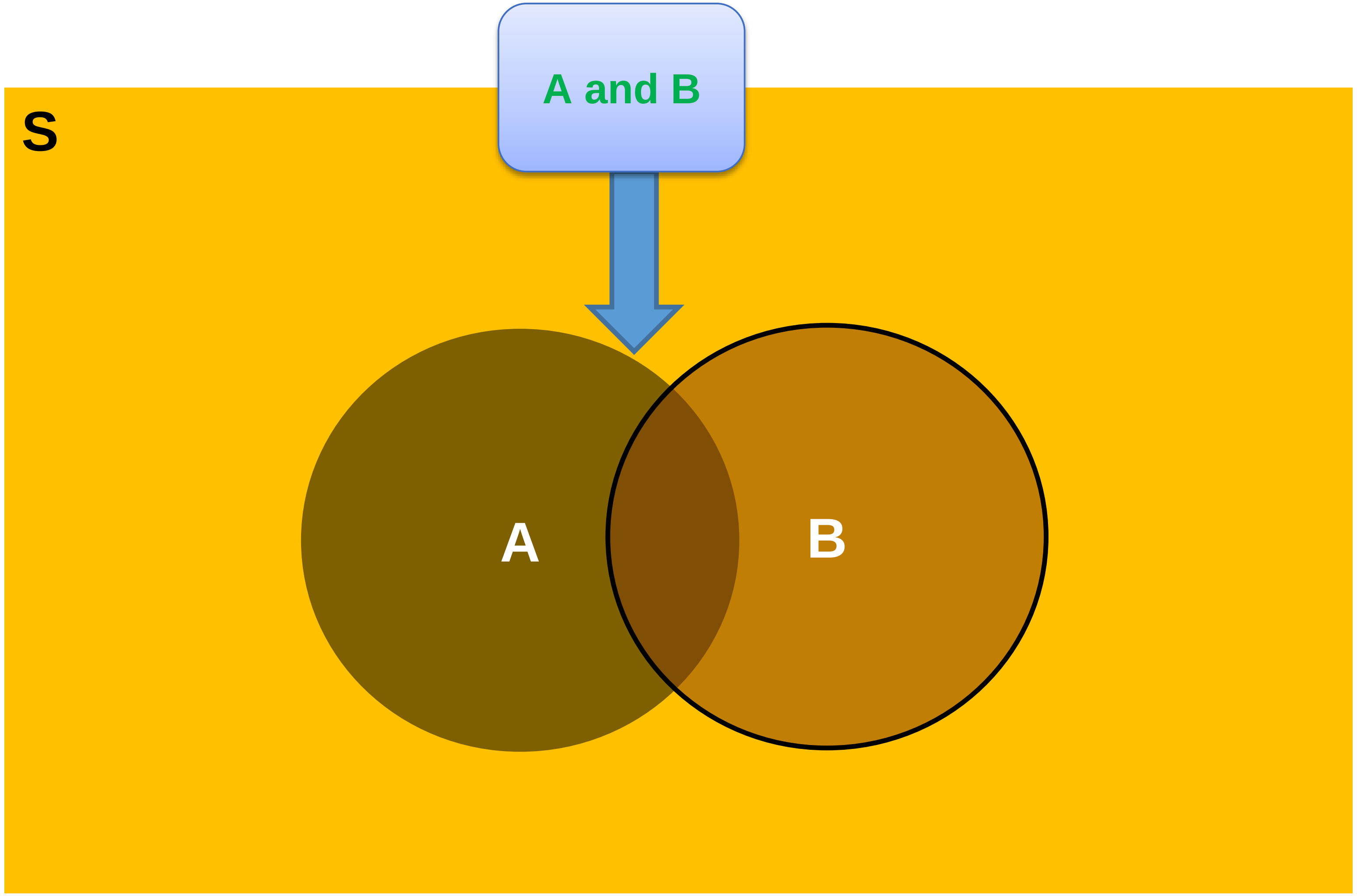
$$P(A\ or\ B) = P(A) + P(B)\ - P(A\ and\ B)$$

$(A \cup B)$ $(A \cap B)$

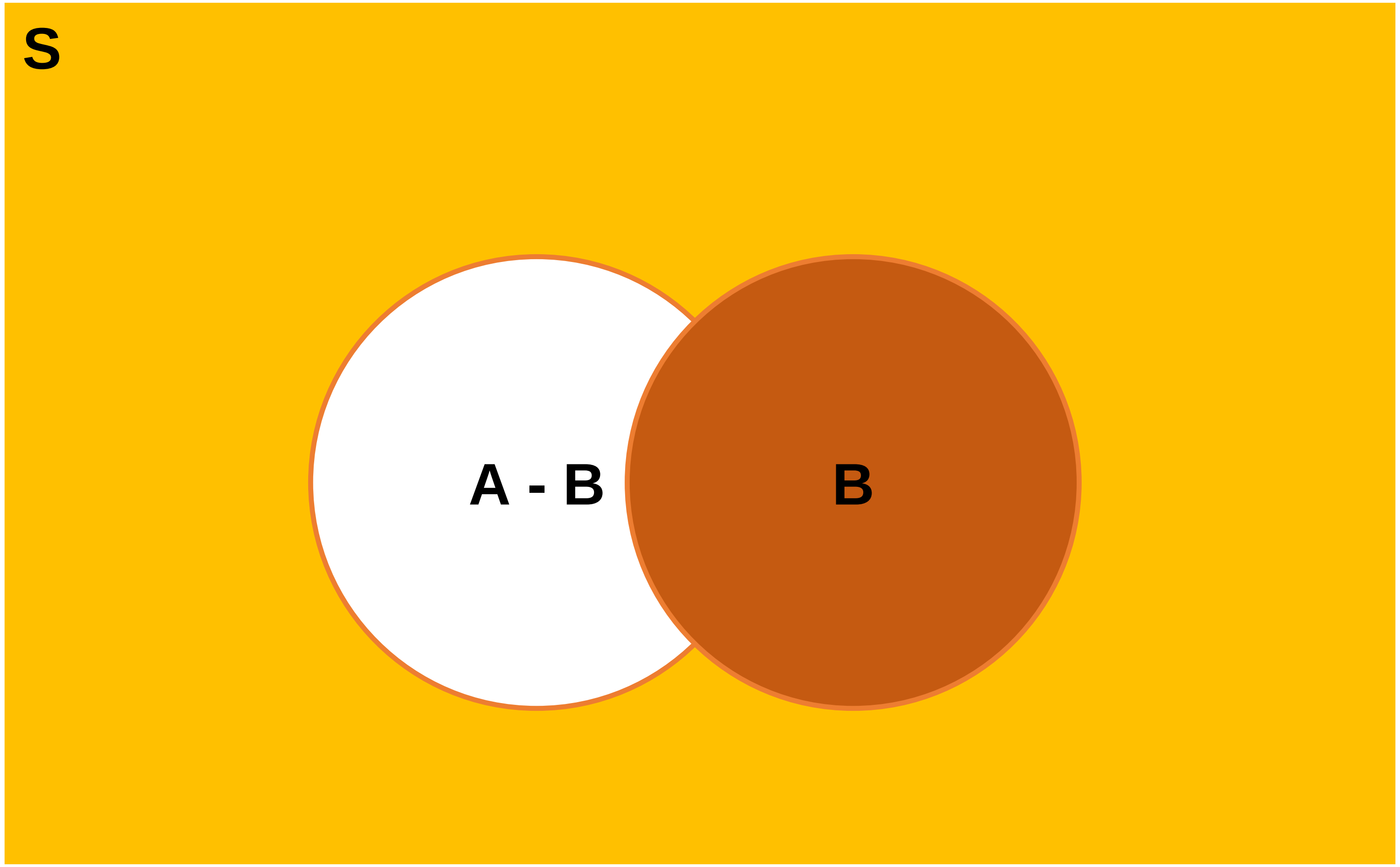
Example

- Event A – Customers who default on loans
- Event B – Customers who are High Net Worth Individuals

Intersection (A and B):



Difference (A – B) :



Axioms of Probability:

According to axiomatic theory of probability, the probability of an event E satisfies the following axioms:

For an experiment K, and event A:

- The probability of an event E always lies between 0 and 1
    - $0<=P(A)<=1$
- The probability of universal set = 1
    - $P(S) = 1$
- If A1, A2, A3…are mutually exclusive then $P(A1 \cup A2 \cup A3…) = P(A1)+P(A2)+P(A3)…$

Probability – Types:

Contingency table summarizing 2 variables, *Loan Default* and *Age:*

		Age			
		Young	Middle-aged	Old	Total
Loan Defaults	No	5252	13684	130	19066
	Yes	1793	2426	60	4279
	Total	7045	16110	190	23345

Probability – Types:

Convert it into probabilities

		Age			
		Young	Middle-aged	Old	Total
Loan Defaults	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

Probability - Types

Marginal Probability:

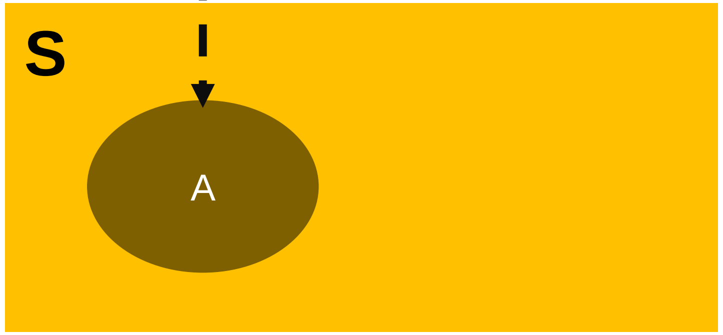
		Age			
		Young	Middle-aged	Old	Total
Loan Defaults	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

Marginal Probability

Probability describing a single attribute

P(Middle) = 0.690

P(old) = 0.008



Probability - Types

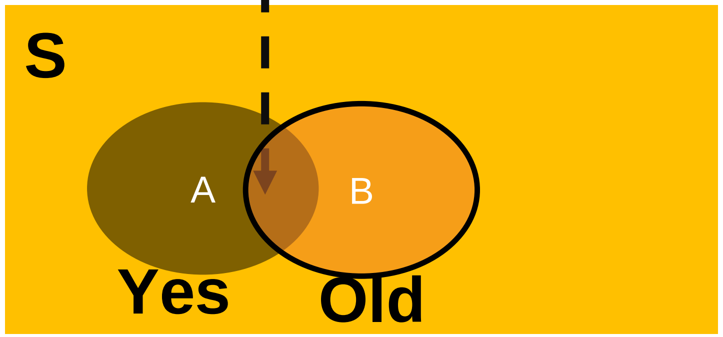
Joint Probability:

		Age			
		Young	Middle-aged	Old	Total
Loan Defaults	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

Joint Probability

Probability describing a combination of attribute

P(Yes and old) = 0.003



Probability - Types

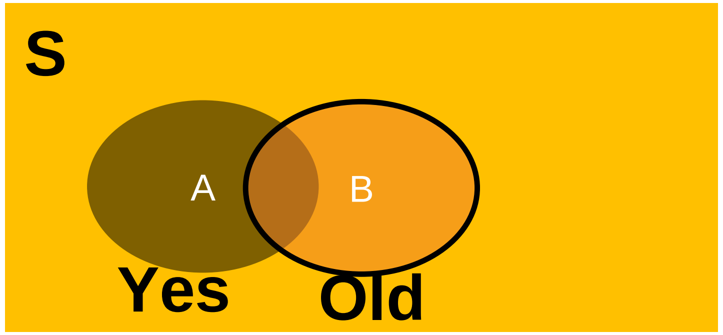
Union Probability

		Age			
		Young	Middle-aged	Old	Total
Loan Defaults	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

P(Yes or old) = P(Yes) + P(old) – P(Yes and old)

= 0.184 + 0.008 – 0.003

= 0.189

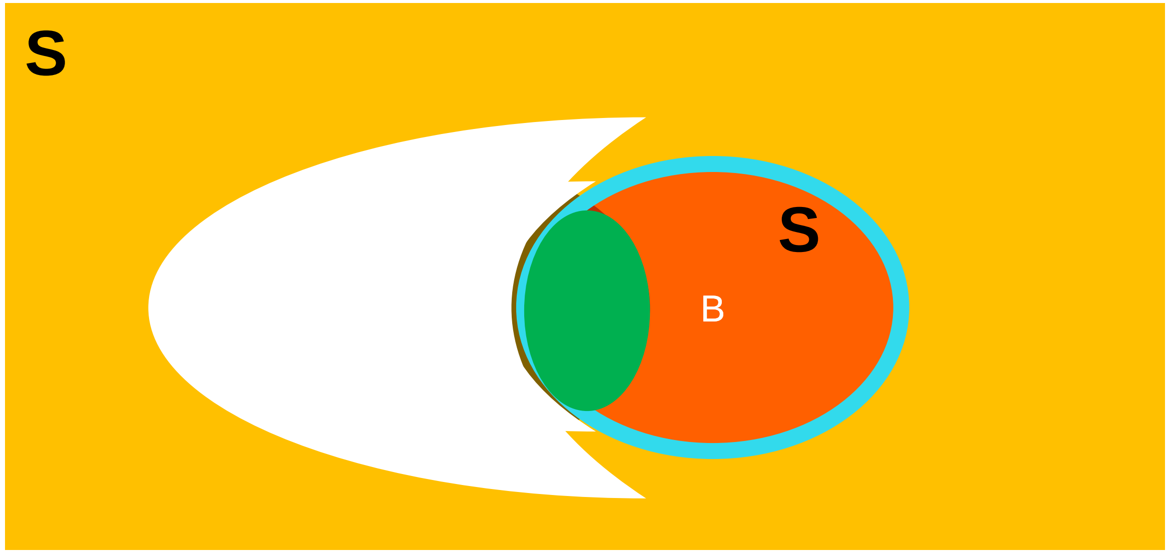


Probability - Types

Conditional Probability

- Probability of *A occurring* ***given that*** B has occurred.
- The sample space is restricted to a single row or column.
- This makes rest of the sample space irrelevant.

Probability, i.e., $P(\boldsymbol{A}|\boldsymbol{B}) = \frac{\boldsymbol{P(A \text{ and } B)}}{\boldsymbol{P(B)}}$



Probability – Types:

Conditional Probability

		Age			
		Young	Middle-aged	Old	Total
Loan Defaults	No	0.225	0.586	0.005	0.816
	Yes	0.077	0.104	0.003	0.184
	Total	0.302	0.690	0.008	1.000

What is the probability that a person will not default on the loan payment **given she is middle-aged?**

Probability, i.e., $P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$

P(No | Middle-Aged) = 0.586/0.690 = 0.85

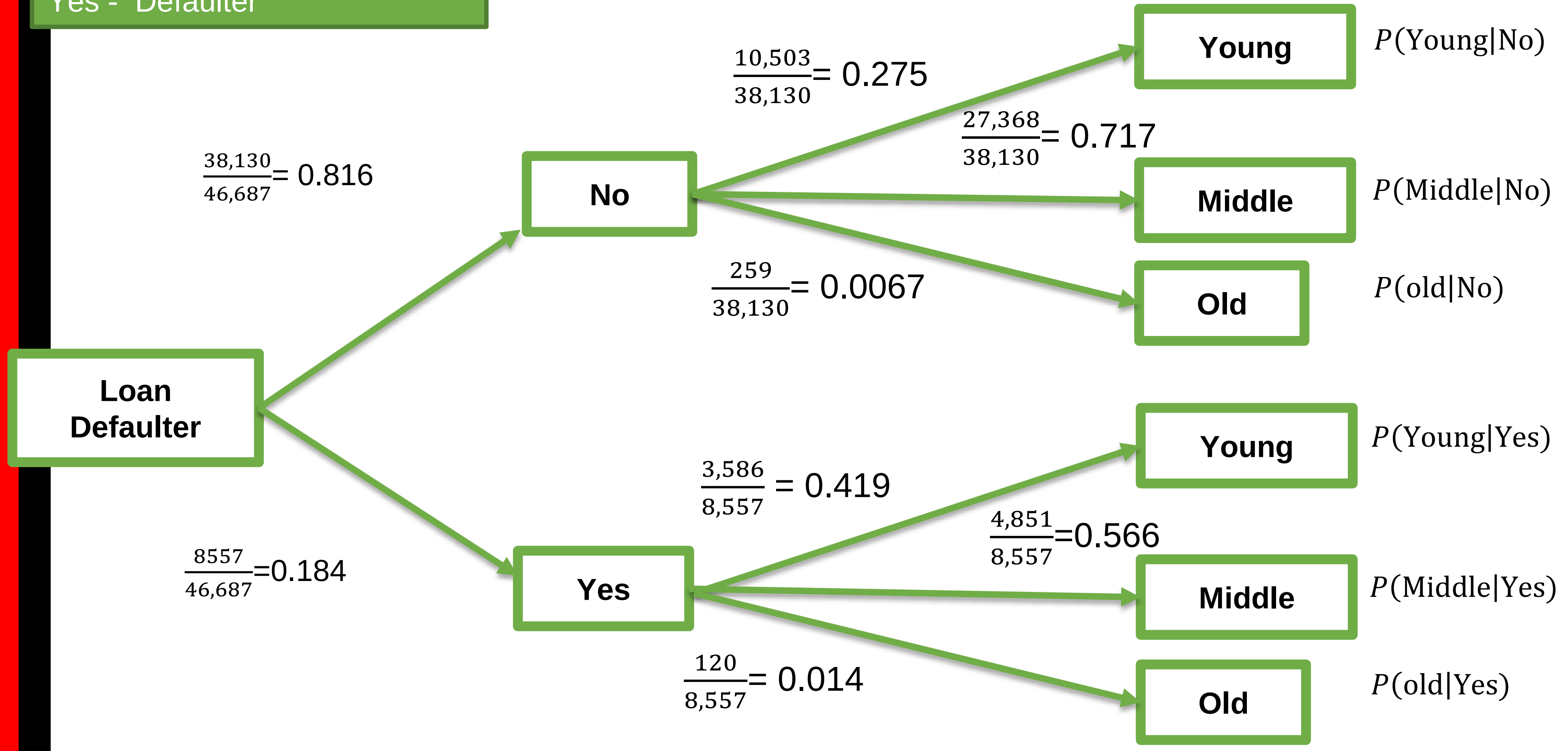
Note that this is the ratio of **Joint Probability to Marginal**

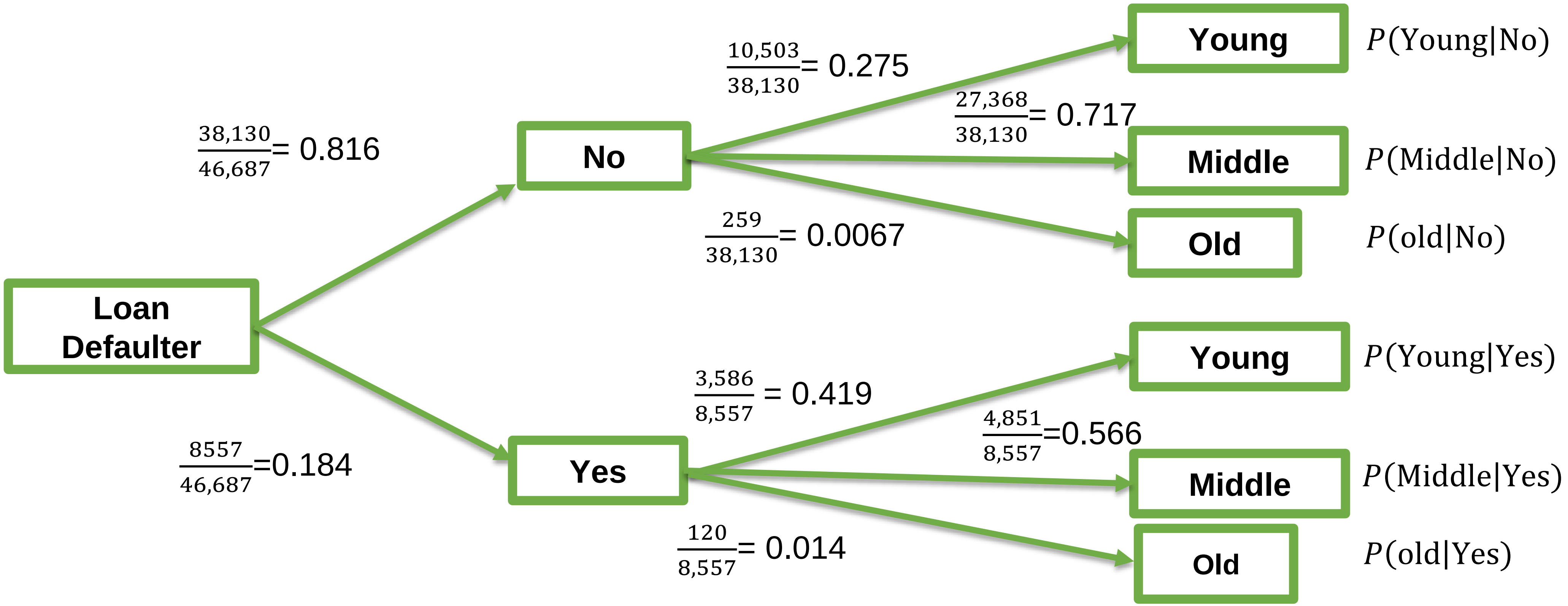
P(Middle-Aged | No) = $\frac{0.586}{0.816}$ = 0.72 (Order Matters)

		Age			Total
		Young	Middle-aged	Old	Total
Loan Defaults	No	10,503	27,368	259	38,130
	Yes	3,586	4,851	120	8,557
	Total	14,089	32,219	379	46,687

Age			Total
Young	Middle-aged	Old	Total
0.225	0.586	0.005	0.816
0.077	0.104	0.003	0.184
0.302	0.690	0.008	1.000

No – Non-defaulter
Yes - Defaulter




Loan Defaulter
$\frac{38,130}{46,687}= 0.816$
No
$\frac{8557}{46,687}=0.184$
Yes
$\frac{10,503}{38,130}= 0.275$
$\frac{27,368}{38,130}= 0.717$
$\frac{259}{38,130}= 0.0067$
Young
Middle
Old
$P(\text{Young}|\text{No})$
$P(\text{Middle}|\text{No})$
$P(\text{old}|\text{No})$
$\frac{3,586}{8,557} = 0.419$
$\frac{4,851}{8,557}=0.566$
$\frac{120}{8,557}= 0.014$
Young
Middle
Old
$P(\text{Young}|\text{Yes})$
$P(\text{Middle}|\text{Yes})$
$P(\text{old}|\text{Yes})$