Introduction to Probability

Edgar Acuna

(based on Prof. A. Moore's notes)

Departamento de Ciencias Matematicas

Universidad de Puerto Rico-Mayaguez

academic.uprm.edu/eacuna

Probability

- The world is a very uncertain place
- Probability is a measurement of the uncertainty that some event A happens.
- Statistical methods and AI algorithms predict with uncertainty. They are not 100% sure.

Example of events

- A = The US president in 2023 will be male
- A = You wake up tomorrow with a headache
- A = You have diabetes

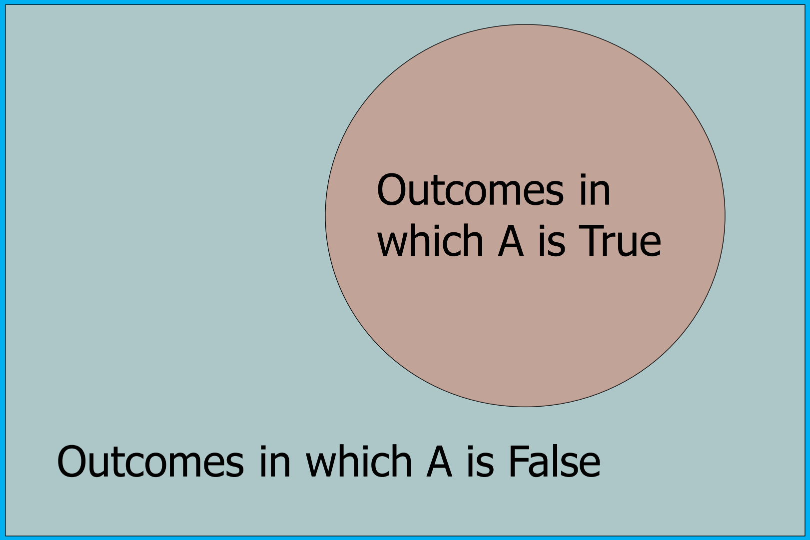
Probability

- The probability of an event A is denoted by P(A) and it represents “the fraction of all outcomes where A occurs” .
- There are four ways to assign probability to an event: Axiomatic approach, Classical approach, Frequential approach and the subjective approach.

Venn Diagram to visualize A

Sample space of all possible outcomes.

The area is 1



P(A) = Area of the oval region

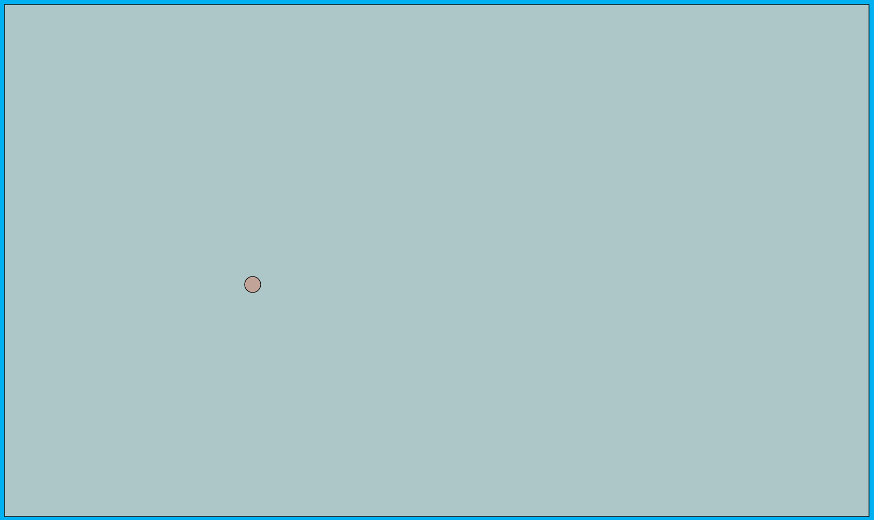
The Axioms of Probability

- 0 <= P(A) <= 1
- P(A is True in all the outcomes) = 1
- P(A is False in all outcomes) = 0
- P(A or B) = P(A) + P(B) if A and B are disjoints.

The axioms were introduced by the Russian school at the beginning of 1900's: Kolmogorov, Liapunov, Kinthchine, Chebychev

Interpreting the axioms

- 0 <= P(A) <= 1
- P(True) = 1
- P(False) = 0
- P(A or B) = P(A) + P(B)



The area of A can't get any smaller than 0

And a zero area would mean no world could ever have A true

Interpreting the axioms

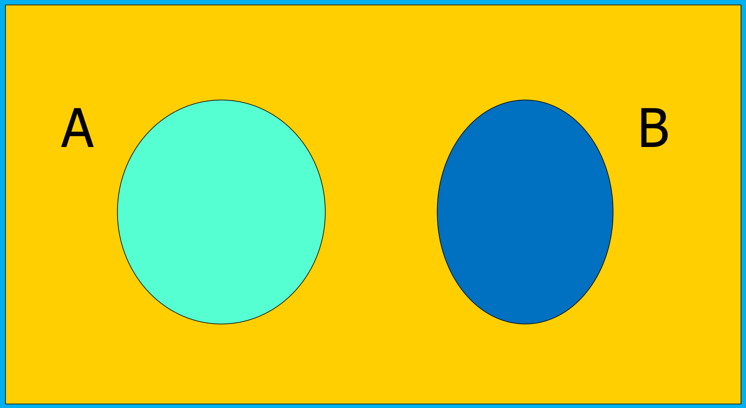
- 0 <= P(A) <= 1
- P(True) = 1
- P(False) = 0
- P(A or B) = P(A) + P(B)

The area of A can't get any bigger than 1

And an area of 1 would mean all worlds will have A true

Interpreting the axioms

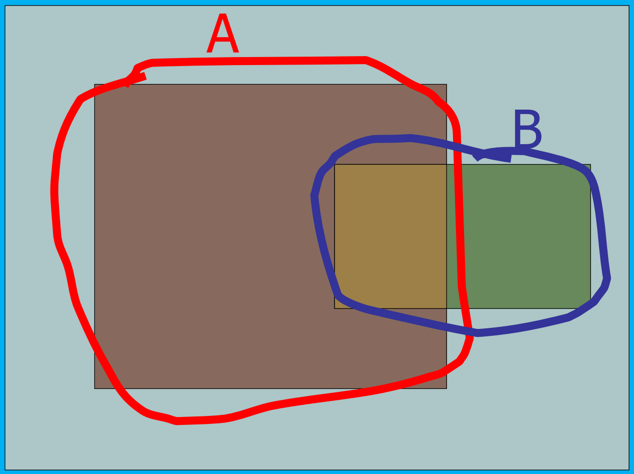
- $0 <= P(A) <= 1$
- $P(True) = 1$
- $P(False) = 0$
- $P(A \text{ or } B) = P(A) + P(B)$

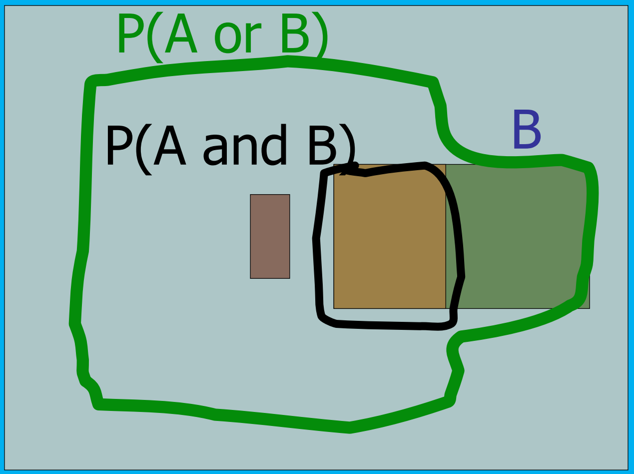


Properties obtained from the axioms

The additive Rule:

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



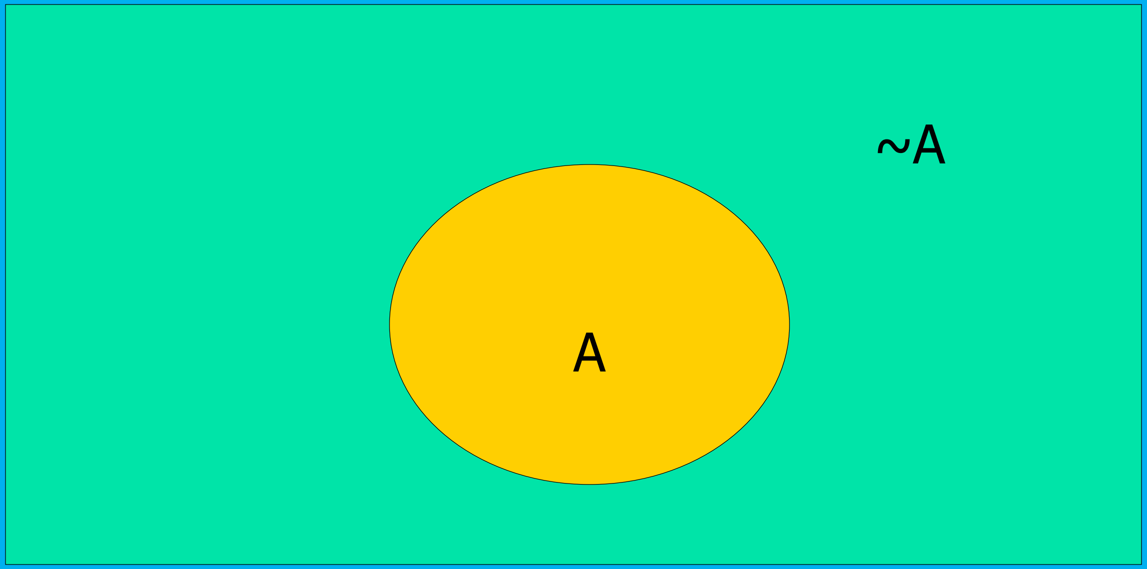


Simple addition and subtraction

Properties obtained from the Axioms

Probability of the complement:

$$P(\text{not } A) = P(\sim A) = 1 - P(A)$$

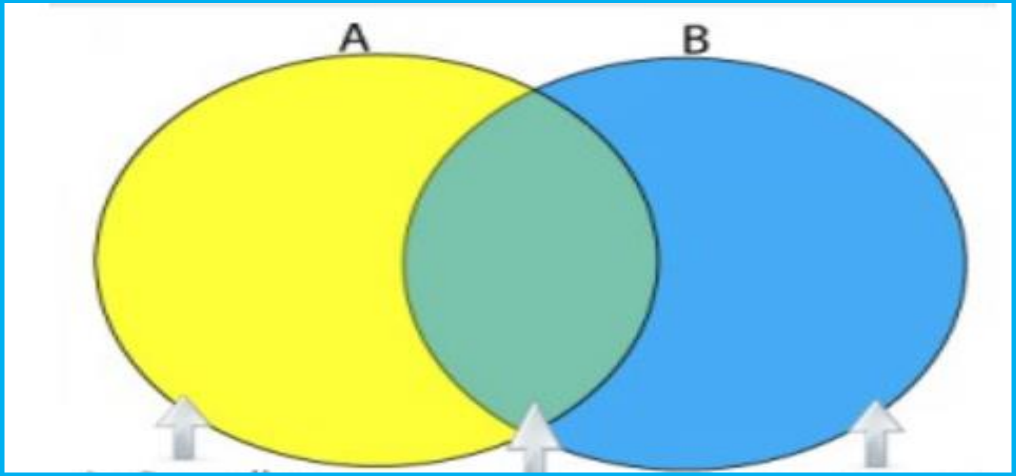


Properties obtained from the axioms

Total Probability:

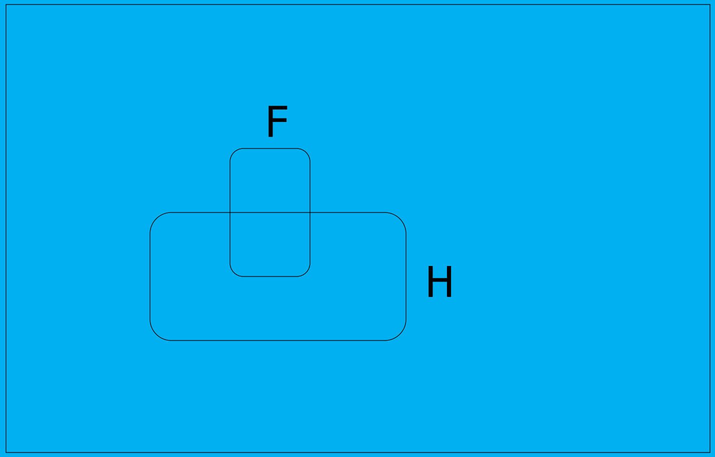
P(A) = P(A ^ B) + P(A ^ ~B)

P(Smoker)=P(Smoker and Drinker)+P(Smoker and No Drinker)



Conditional Probability

- P(A|B) = fraction among the outcomes where B is True and where also A is True

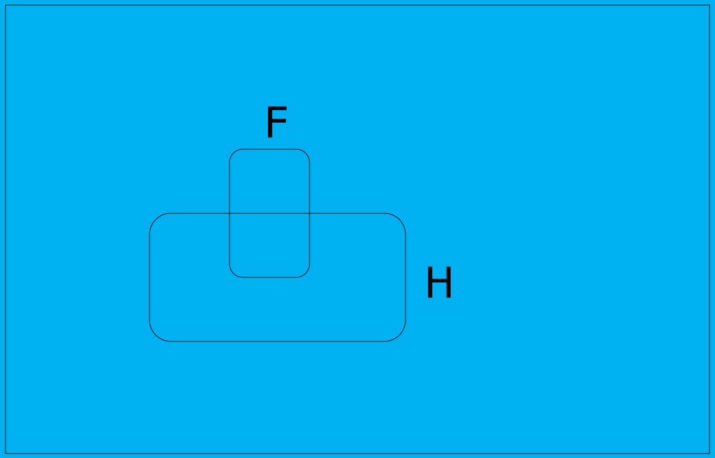


H = "Have a headache"
F = "Coming down with Flu"

P(H) = 1/10
P(F) = 1/40
P(H|F) = 1/2

"Headaches are rare and flu is rarer, but if you're coming down with 'flu there's a 50-50 chance you'll have a headache."

Conditional Probability



H = "Have a headache"
F = "Coming down with Flu"

P(H) = 1/10
P(F) = 1/40
P(H|F) = 1/2

P(H|F) = fraction among people with flu that have headache

= #outcomes with flu and headache

#outcomes with flu

= Area of "H and F" region

Area of "F" region

= P(H ^ F)

P(F)

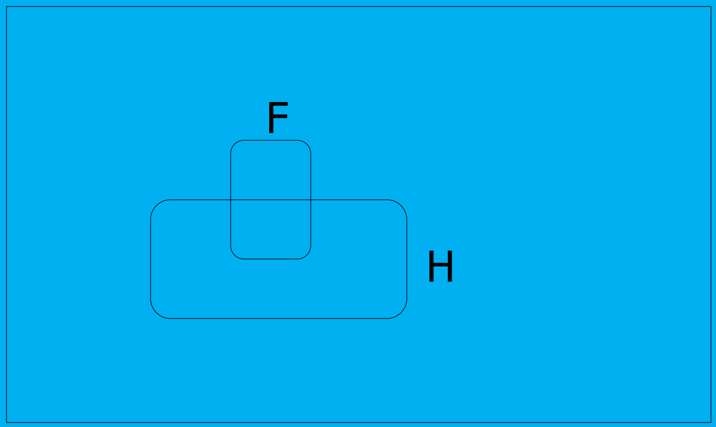
Definition of Conditional Probability

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

Corollary: The Chain Rule

$$P(A \wedge B) = P(A|B)\, P(B)$$

Probabilistic Inference



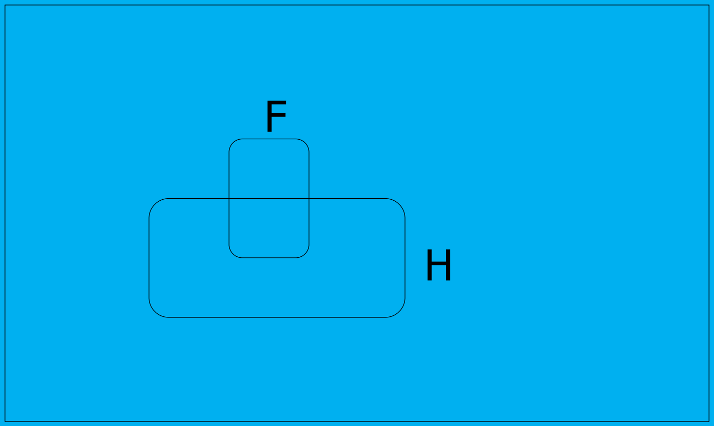
H = "Have a headache"
F = "Coming down with Flu"

P(H) = 1/10
P(F) = 1/40
P(H|F) = 1/2

One day you wake up with a headache. You think: "50% of flus are associated with headaches so I must have a 50-50 chance of coming down with flu"

Is this reasoning good?

Probabilistic Inference



H = "Have a headache"
F = "Coming down with Flu"

P(H) = 1/10
P(F) = 1/40
P(H|F) = 1/2

P(F ^ H) =P(F)P(H/F)=(1/40)(1/2)=1/80

P(F|H) =P(F^H)/P(H)=(1/80)/(1/10)=1/8=.125

What we just did…

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

This is Bayes Rule. P(B) es llamada la probabilidad apriori, P(A/B) es llamada la veosimiltud, y P(B/A) es la probabilidad posterior

Bayes, Thomas (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London,* **53:370-418**

More General Forms of Bayes Rule

$$P(A|B)=\frac{P(B\mid A)P(A)}{P(B\mid A)P(A)+P(B|\sim A)P(\sim A)}$$

$$P(A=v_i\,|B)=\frac{P(B\mid A=v_i)P(A=v_i)}{\sum_{k=1}^{n_A}P(B\mid A=v_k)P(A=v_k)}$$

The Joint Distribution

Example: Binary variables A, B, C

Recipe for making a joint distribution of M variables:

1. Make a truth table listing all combinations of values of your variables (if there are M Boolean variables then the table will have 2^M rows).

A	B	C
0	0	0
0	0	1
0	1	0
0	1	1
1	0	0
1	0	1
1	1	0
1	1	1

A=1 Subject is Male, 0 subject s Female, B=1 the subject is married, 0 the subject is not married and C =1 subject is sick ,0 subject is not sick

The Joint Distribution

Example: Binary variables A, B, C

Recipe for making a joint distribution of M variables:

1. Make a truth table listing all combinations of values of your variables (if there are M Boolean variables then the table will have 2^M rows).
2. For each combination of values, say how probable it is.

A	**B**	**C**	**Prob**
0	0	0	0.30
0	0	1	0.05
0	1	0	0.10
0	1	1	0.05
1	0	0	0.05
1	0	1	0.10
1	1	0	0.25
1	1	1	0.10

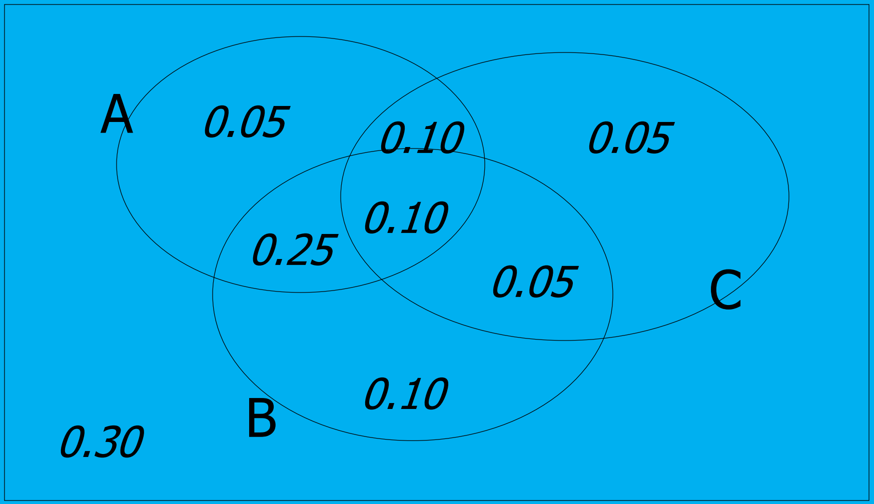
The Joint Distribution

Example: Binary variables A, B, C

Recipe for making a joint distribution of M variables:

1. Make a truth table listing all combinations of values of your variables (if there are M Boolean variables then the table will have 2^M rows).
2. For each combination of values, say how probable it is.
3. If you subscribe to the axioms of probability, those numbers must sum to 1.

A	**B**	**C**	**Prob**
0	0	0	0.30
0	0	1	0.05
0	1	0	0.10
0	1	1	0.05
1	0	0	0.05
1	0	1	0.10
1	1	0	0.25
1	1	1	0.10



A note about independence

- Assume A and B are Boolean Random Variables. Then

"A and B are independent"

if and only if

$$P(A|B) = P(A)$$

- "A and B are independent" is often notated as

$$A \perp B$$

Independence Theorem

Si A and B are independent then

$$P(A \wedge B)=P(A)P(B)$$

,