MIKA Group Meeting – חוג מיק"א

Lecture # 4 – November 28th 2012

Bayesian statistics & conditional probability

Lecturer: Prof. David Brandon

Department of Materials Science & Engineering

Technion – Israel Institute of Technology, Haifa, Israel

Bayesian Logic

Thomas Bayes: 1702-1761

Based on probability, P(X) NOT frequency distributions, BUT all common distributions can be incorporated into a Bayesian analysis : Poisson, Normal (Gaussian) etc.

Employs LOGIC to analyze related events and derive event probabilities based on incomplete knowledge.

EXAMPLE 1:

1. *In families with two children, if one child is a girl, what is the probability that the other will be a boy?*
2. *Does this probability change if the girl was the first-born?*

What arte the options?

1^{st} born	**2^{nd} born**
Boy	Boy
Boy	Girl
Girl	Boy
Girl	Girl

Note:

1. *'One child is a girl'*- eliminates the 1^{st} option
2. *'The eldest is a girl'* – eliminates both the 1^{st} and 2^{nd} options
3. *NO probability distribution is involved, only logic*

EXAMPLE 2:
If we select one box out of three in the expectation trhat it may contain a prize, then how should we react when one of the remaining boxes is shown to be empty?

1. *Stick to the box that we selected previously OR*
2. *Change our choice to the remaining box.*

WHY?

Bayesian logic uses PARTIAL INFORMATION to improve the accuracy of predictions!

(elections, card games, dice, even research)

EXAMPLE 3: Life Expectancy (USA, 2000)

Population 275million: Deaths 2.4 million: $P(M)=2.4/275=0.00873$

Elderly (74+) 16.6million, of whom 1.36million died.

What is the probability that a specific death was ‘elderly’? That is $P_E(M)$

The probability of death conditional on being elderly

Divide the probability of being elderly & dead:
$P(M\&E)=1.36/275=0.00495$, by the probability of being elderly:
$P(E)=16.6/275=0.06036$:

$P_E(M)= P(M\&E)/ P(E)=0.00495/0.06036=0.082$

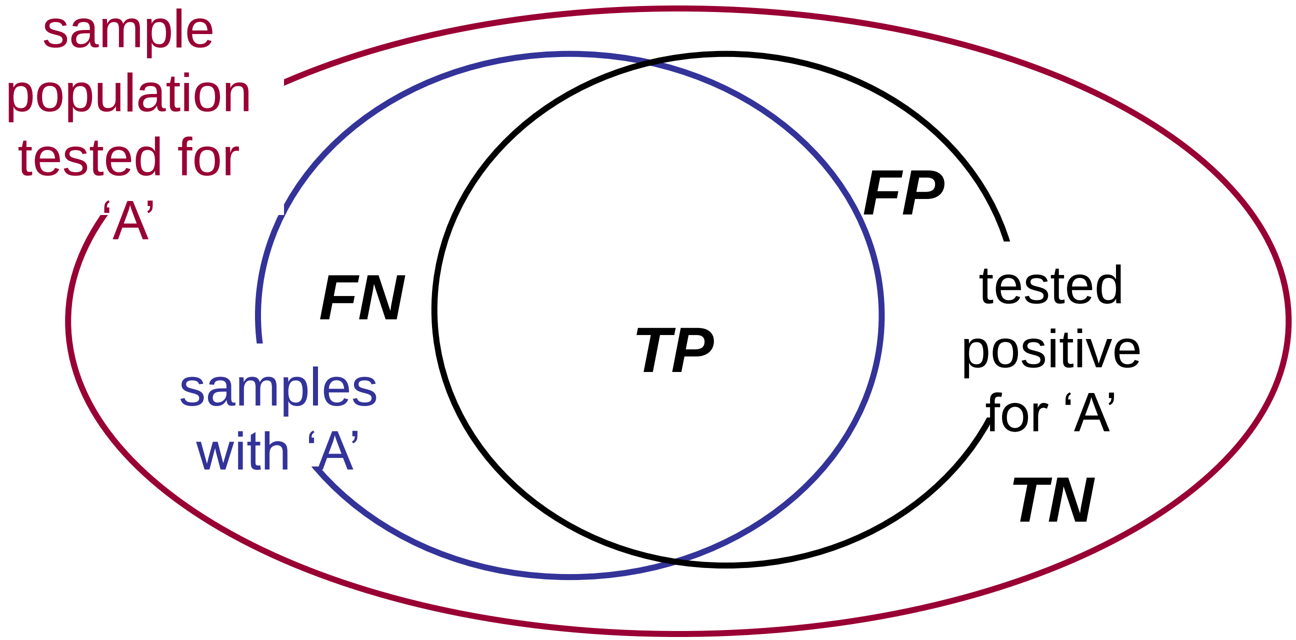
What is the probability that those who died were elderly?

$P_M(E)= P(M\&E)/ P(M)=0.00495/00873=0.57$:

The proportion of senior deaths in the population

Venn Diagrams

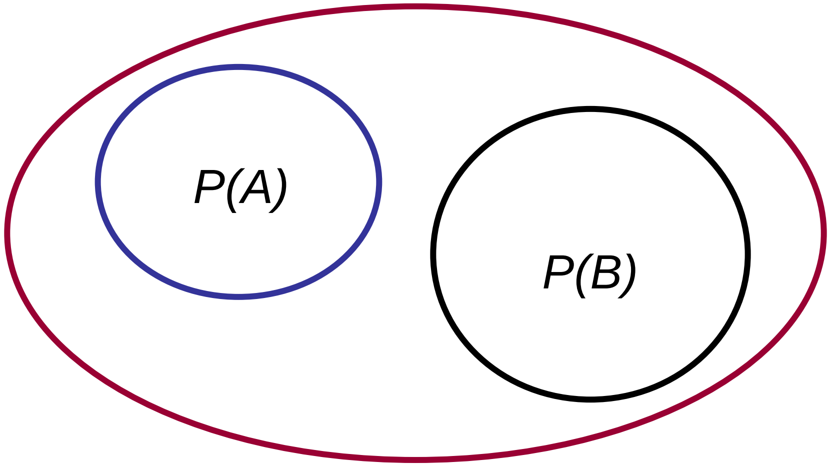
John Venn – 1834-1923



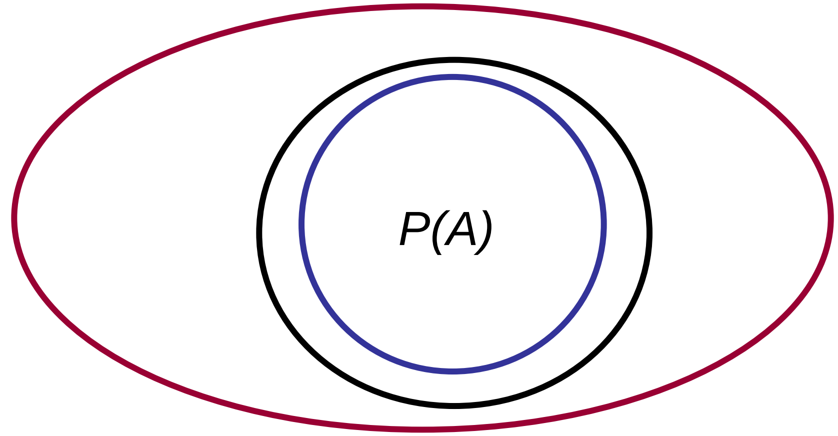
		Property 'A'	
		Yes	*No*
Test Results	*Positive*	TP	FP
	Negative	FN	TN

What are the options?

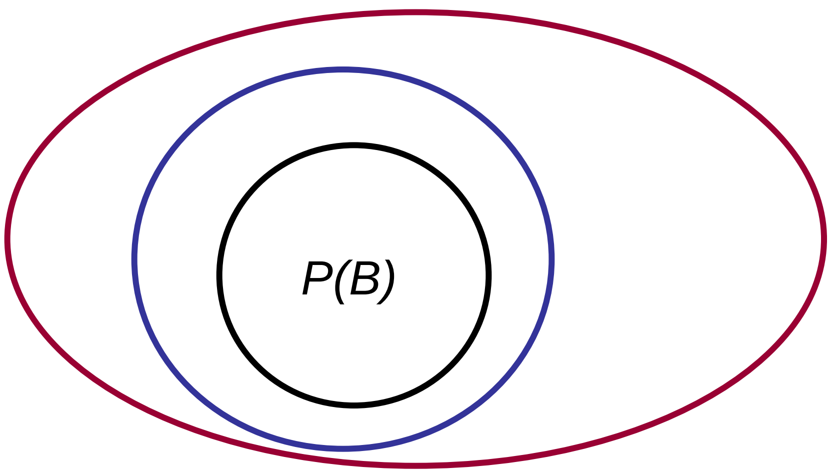
Mutually exclusive events



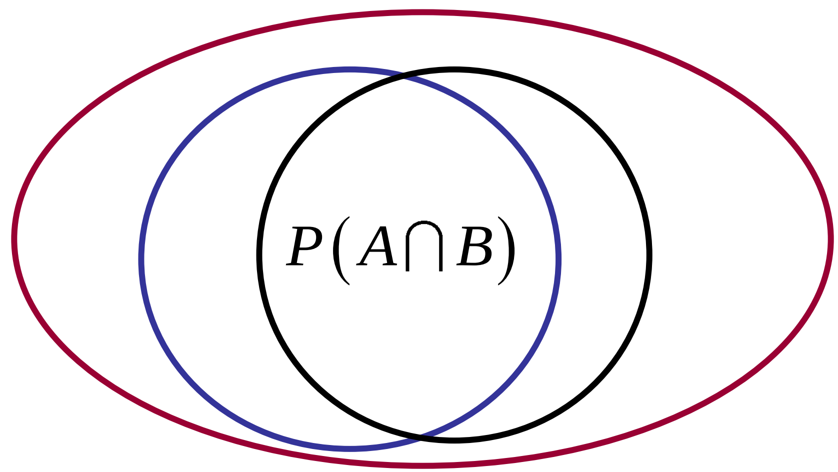
'A' is a subset of' B'



'B' is a subset of' A'



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



Basic Bayesian logic:

either A or NOT A=A must be true:*

$$P((A) + P(A^*) = 1$$

when ALL event probabilities are summed:

$$P((A) + P(B) + P(C) + \ ... = 1$$

An OR statement has events A and B <u>mutually exclusive</u>:

$$P(A \cup B) = P(A) + P(B)$$

An AND statement has A and B <u>independent events</u>:

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

Exclusivity & Independence

If two events are NOT exclusive, an AND statement is modified:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

That is, there is an area of overlap, as for true-positive tests in the previous Venn diagram

If two events are NOT independent:

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

This is an AND statement with B 'conditional' on A

Conditional statements sow confusion, so consider these options:

1. *The probability of two independent events coinciding is ALWAYS less than a 'conditional' probability*
2. *The 'conditional' probability is ALWAYS less than the probability of either event by itself*

Bayes Law

In Bayesian theory the *order* of events is not important, so:

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

From which it follows that:

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

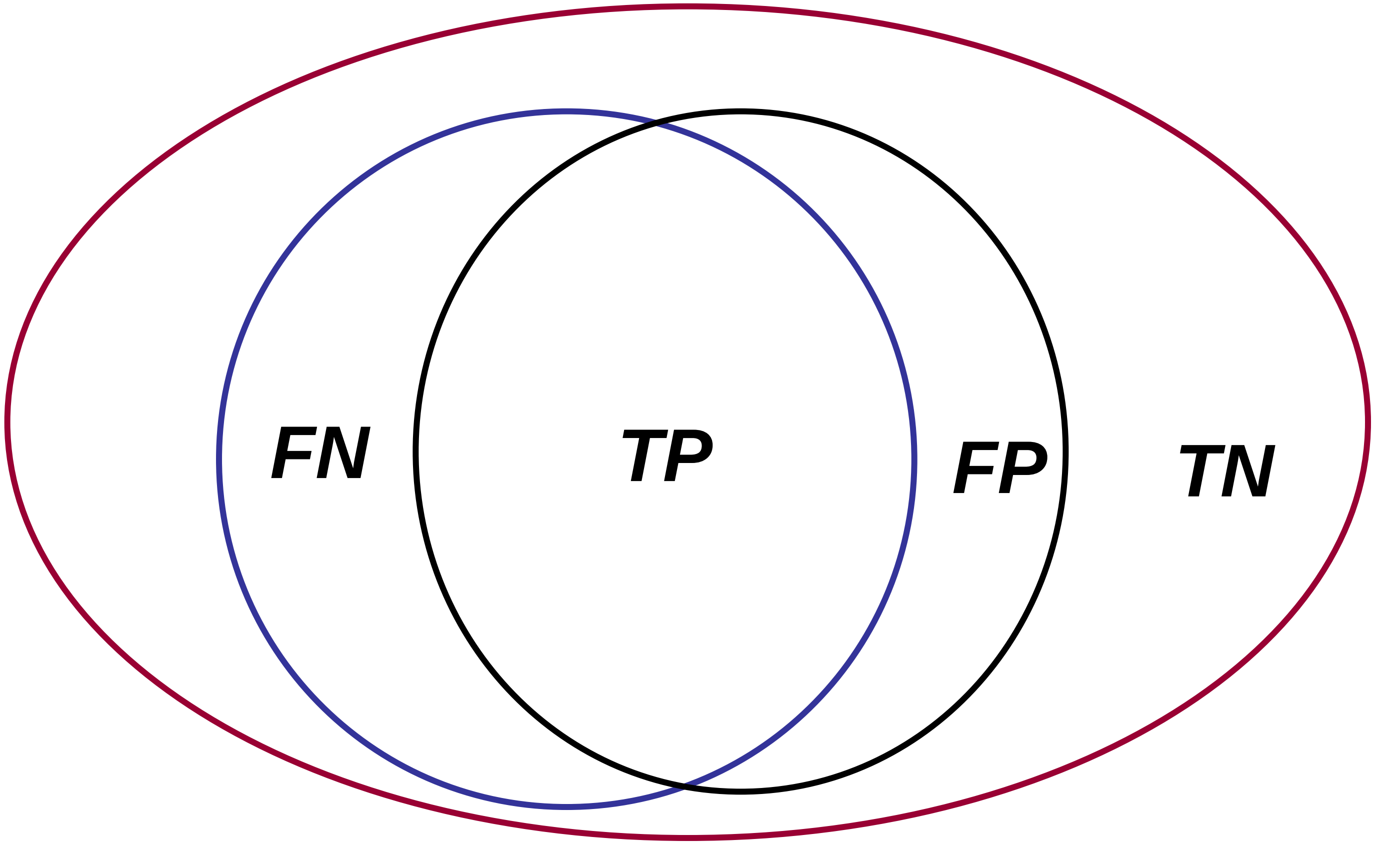
This is Bayes simplest equation for two linked events

Bayes ‘ equation can be extended to three or more linked events:

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

In this form, P(B) is now dependent on our knowledge of the probabilities of A and C occurring

Back to The Venn Diagram



The property we are looking for and the experimental test we are using both have linked probability functions, but we only have the test results (positive or negative). To distinguish the true results (positive or negative) from the false results we need to assign (guess?) some probabilities and establish a methodology:

		Property ‘A’		
		Yes	*No*	
Test	*Positive*	TP		FP
Results	*Negative*	FN		TN

P(YES) is the incidence of ‘A’, that is P(A)
P(TP) is the reliability of the test, P(B)

The Bayesian solution for the probability of a correct result is obtained by weighting the experimental results using this additional information:

Bibliography

1. http://plato.stanford.edu/entries/bayes-theorem/
2. http://en.wikipedia.org/wiki/Bayesian_probability
3. http://ba.stat.cmu.edu/
4. *Bayes and Empirical Bayes Methods for Data Analysis*, B.P. Carlin, T.A. Louis, (Chapman & Hall 1996) ISBN10: 0412056119
5. *Introduction to Bayesian Statistics*, William M. Bolstad, (John Wiley & Sons; 2nd Ed. 2007)
6. *Bayesian Statistics: An Introduction,* Peter M. Lee ,*(*Hodder Arnold*;* 3d Rev Ed 2004) ISBN-10: 0340814055 ISBN-13: 978-0340814055

Is it relevant?

1. Trace element analysis
2. Transient phase identification
3. Detecting observer bias
4. Estimating reliability of experimental method
5. Combining microstructural and property data
6. Determining the significance of results
7. Linking data sets: propertie, processsing and structure