# Bayes’ Theorem with Application

Vani K Borooah
University of Ulster

# Thomas Bayes

- The Reverend Thomas Bayes – an 18th century Presbyterian minister – proved what is, arguably, the most important theorem in statistics (see "In Praise of Bayes", The *Economist*, 28 September 2000)

# Bayes’ Theorem

- We have a particular theory (T) which may be true or false
  - The car being offered to me is a good car
  - The candidate being interviewed is a good worker
  - A patient has cancer
  - The person at the security barrier is a terrorist
- In all these cases the theory cannot be directly verified

# Data

- Instead what we can observe is some data (D)
  - The car is being offered with a warranty
  - The candidate has a MBA from Harvard
  - The screening shows a tumour
  - The person at the barrier looks “Middle Eastern”

# Confronting Theory with Data

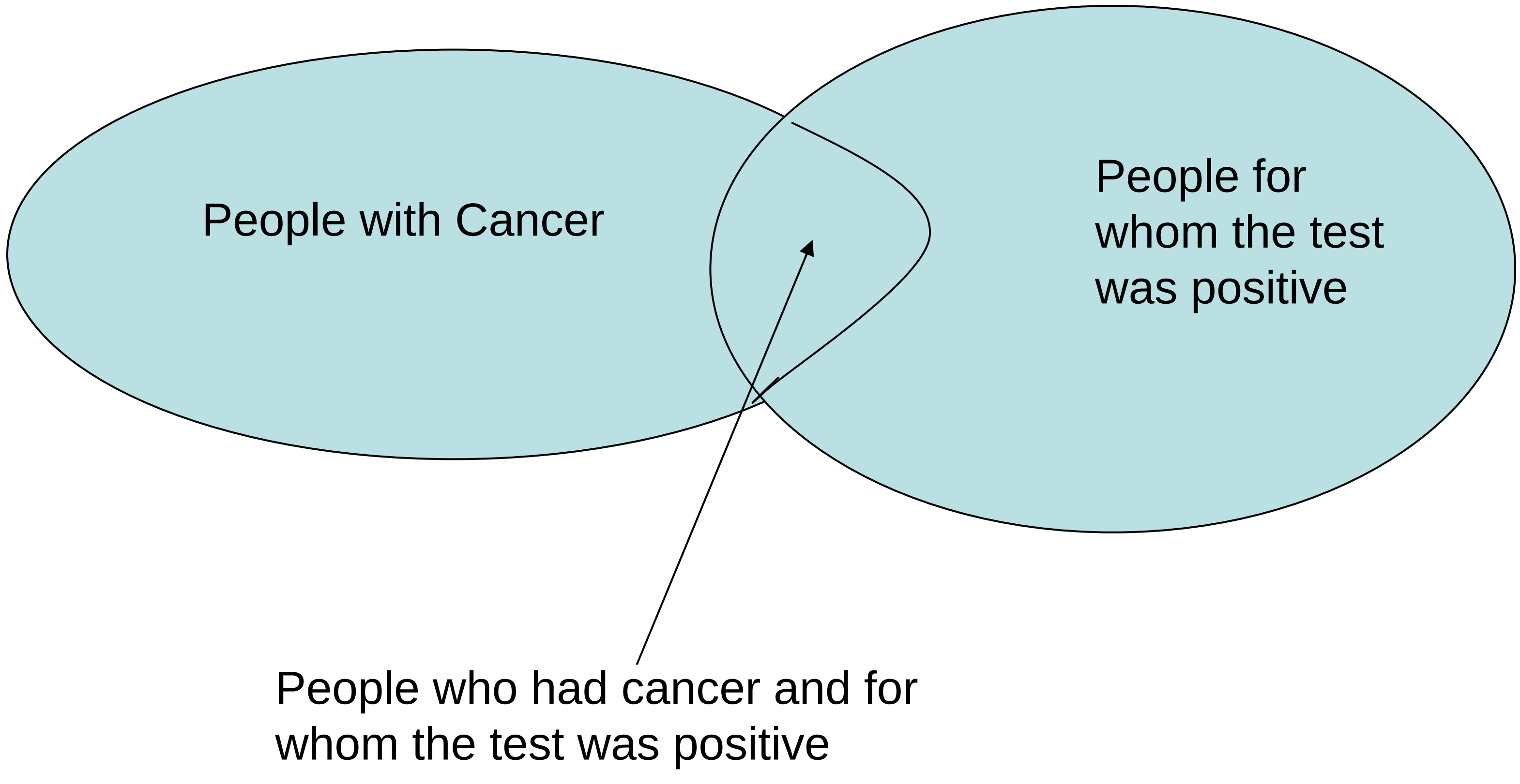
- We are interested in the probability that the theory is true, given that the data has been observed: P(T|D)
- What is the probability the car is good, given that it has a warranty?
- What is the probability the candidate is a good worker, given that he has a Harvard MBA?
- What is the probability the patient has cancer, given the screening reports a tumour?
- What is the probability the man is a terrorist, given he looks “Middle Eastern”?

# But…

- We do not know this!!
- We do know, however, $P(D|T)$
  - Given that the car is good, there is a 90% chance it will have a guarantee
  - Given that he has a Harvard MBA, there is a 80% chance he will be a good worker
  - Given the person has cancer, there is a 95% chance the screening will report a tumour
  - Given the person is a terrorist, there is a 98% chance he looks “Middle Eastern”

# Bayes' Theorem

- Can we deduce what P(T|D) is, given that we know what P(D|T) is?
- Bayes showed how this could be done through his Bayes' Theorem
- $P(T|D)=[P(D|T)*P(T)]/P(D)$


People with Cancer
People for whom the test was positive
People who had cancer and for whom the test was positive

# So…

- P(T|D)=P(T and D)/P(D)
- But P(D|T)=P(T and D)/P(T)
- So, P(T and D)=P(D|T)*P(T)
- P(T|D)=[P(D|T)*P(T)]/P(D)
- Interpretation:
- P(T) is our *prior* belief that the theory is true: the probability the patient has cancer
- P(T|D) is our *posterior* belief after the data has been observed: the probability the patient has cancer, *given that the test is positive*
- So, on the basis of the evidence, we update our prior belief to arrive at a posterior belief

# Example: Cars

- My prior belief is that only half the cars offered for sale are good: P(T)=0.5
- But, I observe that observe that if a car is good, it will carry a warranty: P(D|T)=1
- And, if a car is bad it will not carry a warranty: P(~D|~T)=1
- So, since half the cars are good and half bad, P(D)=0.5
- P(T|D)=P(T)*[P(D|T)/P(D)=1
- So, a warranty represents a separating equilibrium

# Example: Cancer Testing

- We know P(D|T)=0.95 and P(~D|~T)=0.95
- Assume P(T)=0.005
- So, P(T|D) = [P(D|T)*P(T)]/P(D)=[0.95*P(T)]/P(D)
- P(D)=P(D and T) + P(D and ~T)

= P(D|T)*P(T) + P(D|~T)*P(~T)

=0.95*P(T) + 0.05*P(~T) = 0.0545

So, P(T|D)=0.95*0.005/0.0545=0.087