Naïve Bayesians

Back to Basics Series

23 Jan 2021

Goal

Developing the Bayesian muscle to solve a wide range of problems

Naïve Bayesian Philosophy

Intuitive (Visual) Understanding of the Bayesian Reasoning

Ability to model real world problems in a Bayesian Setting

Starting from Simple Probabilistic modelling

Adapting it in a a Bayesian setting
And moving towards ML models

Fluency in the Calculus of Bayesian Stats & ML model

Season 2: Back to Basics

Ep 1	Bayes Theorem
Ep 2, Ep 3	Problems with Binomial Likelihoods
Ep 4	Disease Detection
Ep 5	Naive Bayes Classification
Ep 6	Gaussian Naive Bayes Classification
Ep 7	German Tank Problem
Ep 8	Waiting Times (Continuous Distributions)

Back to Basics

		Canonical Problem	Applications
Ep 1	Bayes Theorem	There are 2 boxes from which cookies can be taken from. Box A and Box B. Box A contains 10 chocolate cookies, Box B contains 5 ginger cookies. Given that you get a chocolate cookie which box was it taken from?	The Shy Librarian Problem Naive Bayes algorithm
Ep 2 Ep 3	Problems with Binomial Likelihoods	You have 2 coins C1 and C2. p(heads for C1) = .7 & P(heads for C2) = 0.6 You flip the coin 10 times. What is the probability that the given coin you picked is C1 given you have 7 heads and 3 tails?	A/B Testing
Ep 4	Disease Detection	A particular disease affects 1% of the population. There is an imperfect test for this disease: The test gives a positive result for 90% of people who have the disease, and 5% of the people who are disease-free. Given a positive test result – what is the probability of having the disease?	COVID Tests (PCR & Antibody)! Fraud Detection
Ep 5	Naive Bayes Classification	Given these words occur in this text what's the probability it's spam?	Any Classification Problem
Ep 6	Gaussian Naive Bayes Classification	Given the weights and heights of basketball players, what's the probability that person a is a basketball player given weight = w and height = h?	

Back to Basics

		Canonical Problem	Applications
Ep 7	German Tank Problem	Suppose tanks were given a serial number based on the order in which they were manufactured. Given that you've observed a tank with serial number "10", how many tanks were actually manufactured in total?	?
Ep 8	Waiting Times (Continuous Distributions)	Suppose you need to gather 10 patients for a trial. Each signup happens at time t_i (i=1, 10). How long do you have to wait after it took you 3 weeks to accrue 2 signups?	Planning Trials Estimating Queues

Canonical Problem

A particular disease affects 1% of the population.
There is an imperfect test for this disease:
The test gives a positive result for 90% of people who have the disease and 5% of the people who are disease-free.
Given a positive test result – what is the probability of having the disease?

$$P(C) =$$

$$P(+ \mid C) = 0.9$$

$$P(+ \mid \neg C) = 0.05$$

Bayes Rule

Posterior Likelihood Prior

$$P(\theta_i \mid D) = \frac{P(D \mid \theta_i)\, P(\theta_i)}{\sum_{all\ j} P(D \mid \theta_j)\, P(\theta_j)}$$

Normalising Constant

Canonical Problem

A particular disease affects 1% of the population.
There is an imperfect test for this disease:
The test gives a positive result for 90% of people who have the disease and 5% of the people who are disease-free.
Given a positive test result – what is the probability of having the disease?

Canonical Problem

A particular disease affects 1% of the population.
There is an imperfect test for this disease:
The test gives a positive result for 90% of people who have the disease and 5% of the people who are disease-free.

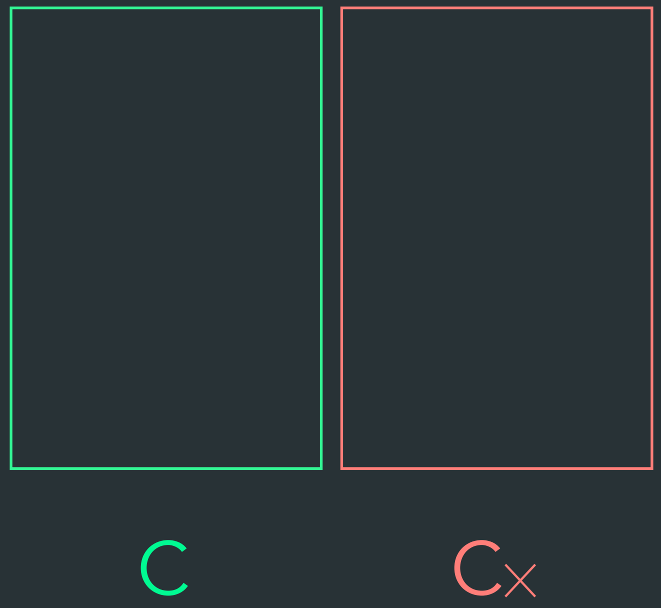
Given a positive test result – what is the probability of having the disease?

Canonical Problem

A particular disease affects 1% of the population.
There is an imperfect test for this disease:
The test gives a positive result for 90% of people who have the disease and 5% of the people who are disease-free.
Given a positive test result – what is the probability of having the disease?

C	$C_{\times}$
Disease	No Disease

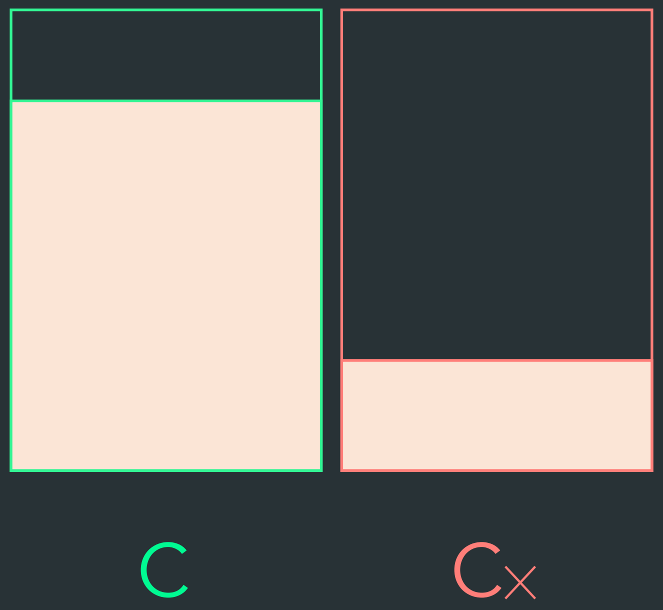
Visual representation of Bayes Rule for 2 Hypotheses



A particular disease affects 1% of the population.
There is an imperfect test for this disease:
The test gives a positive result for 90% of people who have the disease
and 5% of the people who are disease-free.

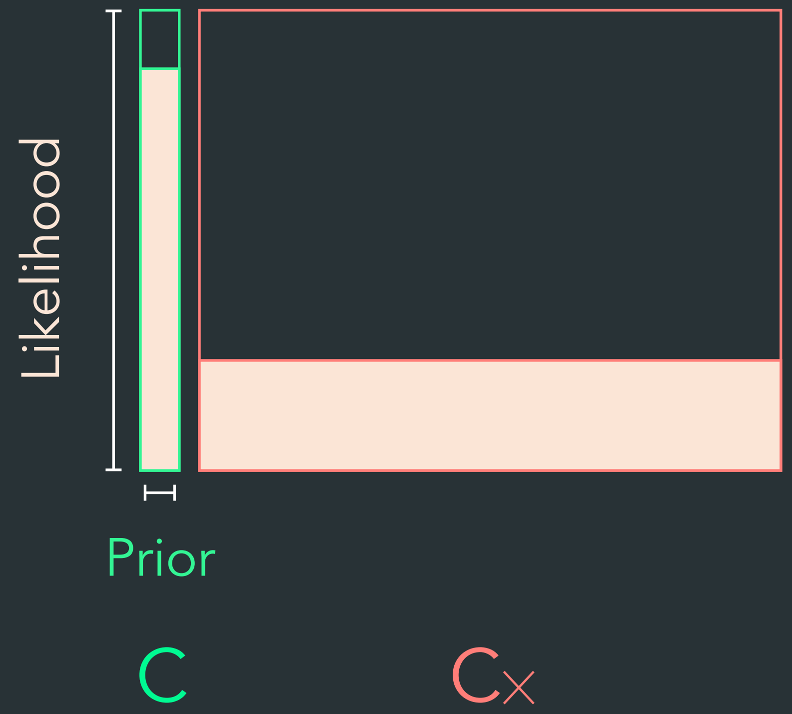
Given a positive test result – what is the probability of having the disease?

Visual representation of Bayes Rule for 2 Hypotheses



A particular disease affects 1% of the population.
There is an imperfect test for this disease:
The test gives a positive result for 90% of people who have the disease
and 5% of the people who are disease-free.
Given a positive test result – what is the probability of having the disease?

Whats the probability that the coin is unfair?



A particular disease affects 1% of the population.
There is an imperfect test for this disease:
The test gives a positive result for 90% of people who have the disease
and 5% of the people who are disease-free.
Given a positive test result – what is the probability of having the disease?

Whats the probability that the coin is unfair?

$$P(C \mid +) = \frac{P(+|C)\,P(C)}{P(+|C)\,P(C) + P(+|C_{\times})\,P(C_{\times})}$$

$\boxed{P(C) = 0.01}$

$P(C_{\times}) = 0.99$

$\boxed{P(+|C) = 0.9}$

$\boxed{P(+|C_{\times}) = 0.05}$

Whats the probability that the coin is unfair?

$$P(C \mid +) = \frac{P(+|C)\,(0.01)}{P(+|C)\,(0.01) + P(+|C_{\times})\,(0.99)}$$

Whats the probability that the coin is unfair?

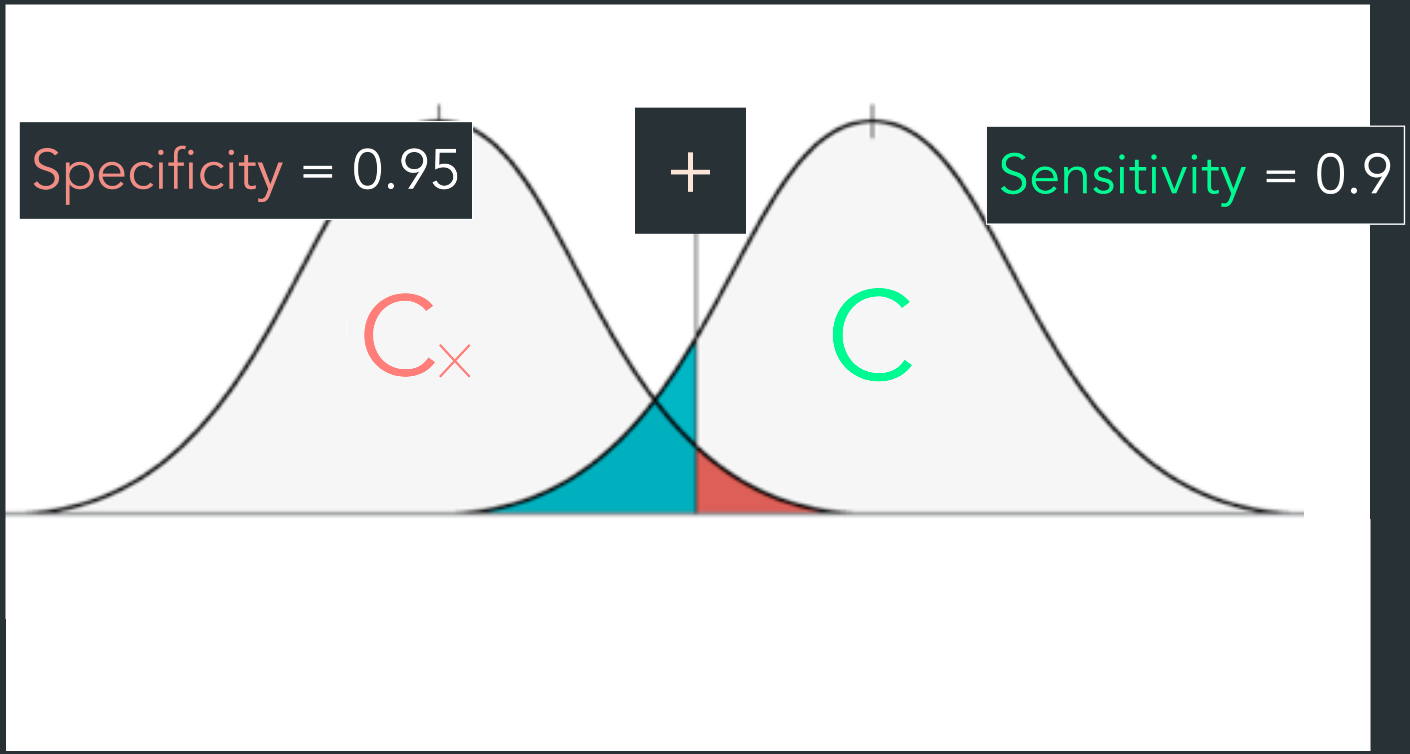
$$P(C \mid +)$$

$$= \frac{P(+|C)\,(0.01)}{P(+|C)\,(0.01) + P(+|C_\times)\,(0.99)}$$

$$= \frac{(0.9)\,(0.01)}{(0.9)\,(0.01) + (0.05)\,(0.99)}$$

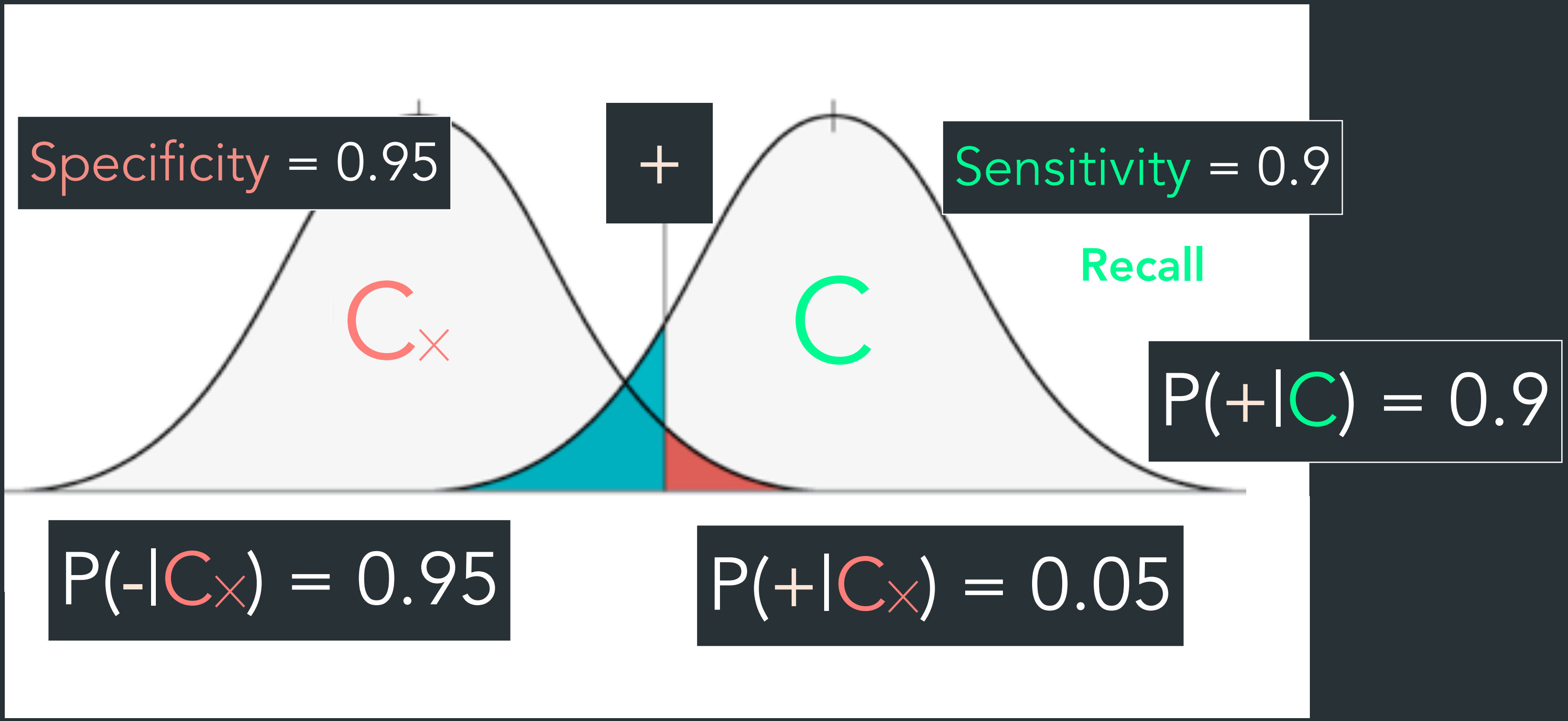
$$\approx \quad 15\%$$

Test Performance



Modified from: https://www.abtasty.com/blog/type-1-and-type-2-errors/

Test Performance



Modified from: https://www.abtasty.com/blog/type-1-and-type-2-errors/

Alternative: Bayes Factor

Odds $a : b$

Probabilities: $a/(a + b)$

$1 : 99$

$P(C) = 0.01$ $P(C_{\times}) = 0.99$

$$P(C) : P(C_{\times})$$

Alternative: Bayes Factor

$$P(C \mid +) = \frac{P(+|C)\,P(C)}{P(+|C)\,P(C) + P(+|C_{\times})\,P(C_{\times})}$$

Alternative: Bayes Factor

$$P(C \mid +) = \frac{P(+|C)\,P(C)}{P(+|C)\,P(C) + P(+|C_{\times})\,P(C_{\times})}$$

$$\text{Bayes Factor, } B = \frac{P(+|C)}{P(+|C_{\times})}$$

Alternative: Bayes Factor

$$P(C \mid +) = \frac{P(+|C)\,P(C)}{P(+|C)\,P(C) + P(+|C_{\times})\,P(C_{\times})} \quad \begin{matrix} \div \; P(+|C_{\times}) \\[2em] \div \; P(+|C_{\times}) \end{matrix}$$

Alternative: Bayes Factor

$$P(C \mid +) = \frac{P(+|C)\,P(C)}{P(+|C)\,P(C) + P(+|C_\times)\,P(C_\times)} \quad \begin{matrix} \div\; P(+|C_\times) \\ \div\; P(+|C_\times) \end{matrix}$$

$$\text{Bayes Factor, } B = \frac{P(+|C)}{P(+|C_\times)}$$

$$P(C \mid +) = \frac{B\,P(C)}{B\,P(C) + P(C_\times)}$$

$$B\,P(C) : P(C_\times)$$

Alternative: Bayes Factor

$$P(C \mid +) = \frac{P(+|C)\,P(C)}{P(+|C)\,P(C) + P(+|C_{\times})\,P(C_{\times})} \begin{matrix} \div \; P(+|C_{\times}) \\ \div \; P(+|C_{\times}) \end{matrix}$$

$$\text{Bayes Factor, } B = \frac{P(+|C)}{P(+|C_{\times})}$$

$$P(C \mid +) = \frac{B\,P(C)}{B\,P(C) + P(C_{\times})}$$

$$B\,P(C) : P(C_{\times})$$

Alternative: Bayes Factor

Odds $a : b$

Probabilities: $a/(a + b)$

$P(C) : P(C_{\times})$

$$B = \frac{P(+|C)}{P(+|C_{\times})}$$

$B\ P(C) : P(C_{\times})$

Alternative: Bayes Factor

Odds: $a : b$

Probabilities: $a/(a + b)$

$$1 : 99$$

$$B = \frac{0.9}{0.05} = 18$$

$$18 \times 1 : 99$$

$$18 : 99$$

Alternative: Bayes Factor

Odds $a : b$

Probabilities: $a/(a + b)$

$1 : 99$

$$B = \frac{0.9}{0.05} = 18$$

$18 \times 1 : 99$

$18 : 99$

Alternative: Bayes Factor

Odds a : b

Probabilities: a/(a + b)

1 : 99

$$B = \frac{0.9}{0.05} = 18$$

18 x 1 : 99

18 : 99

Alternative: Bayes Factor

Odds $a : b$

Probabilities: $a/(a + b)$

$1 : 99$

$$B = \frac{0.9}{0.05} = 18$$

$18 \times 1 : 99$

$18 : 99$

$$P(C \mid +) = 18/(18+99)$$
$$\approx 15\%$$

Takeaways

Bayes Theorem for Disease Detection

Bayes Factor for an Alternative Formulation

References

3 Blue 1 Brown: Medical Test Paradox
https://www.youtube.com/watch?v=lG4VkPoG3ko