Pros and Cons Date

Pros

- Pros of Bayesian Statistics
  - Prior knowledge is useful
- Probability Statements
  - Bayesian statistics provide probability distributions for parameters, giving us a more intuitive understanding of uncertainty
- Smaller Sample sizes
- Iterative learning
  - With Bayesian statistics, we can updat our beliefs as we gather more data

Cons

- Subjectivity
- Complexity
- Limited Popularity

STATISTICS

* Check presentation *

AI INVOLVEMENT

Date [03/02/26]

- AI is used to improve halogroup classifications

* Halogroups are groups of individuals who share a common ancestor thraugh their maternal or paternal lineage

↳ Halogroup classification is used to help determine the ancestral origins of unidentified remains or individuals in criminal investigations

- AI helps significantly enhance the efficiency and accuracy of the classification pocess

↳ The classification process is where people had to manually analyze genetic data, which took too much time and effort

- AI can be very useful in the creation of automated DNA profile interpretation systems
- ML algorithms are used to analyze DNA profiles generated by STR typing

↳ This analyzes a vast amount of data and identifies patterns that may not be easily recognizable by human examiners

- Bayesian AI - the critic - logical, careful and grounded.

↳ It doesn't just ask what is possible, but how sure we are that its true

TECHNIQUES & TECHNOLOGY Date [03/01/26]

DNA Fingerprinting Techniques

- STR's
  - DNA sequences with repeating units of 2-6 base pairs
  - They vary between individuals making them excellent for differentiation
  - Uses fluorescent dyes for automated detection and analysis
  - Able to handle small and degraded samples
  - Used in databases like the Combined DNA Index System (CODIS), aiding law enforcements in identifying suspects by comparing crime scene DNA with database entries
- Restriction Fragment Length Polymorphism
  - Restriction Fragment Length Polymorphism was one of the first DNA fingerprinting techniques
  - Uses enzymes to cut DNA into fragments of varying lengths, seperated by gel electrophoresis
  - The resulting bands are compared between samples to determine similarities and differences in the genetics
  - Require large amounts of high-quality DNA
- Polymerase Chain Reaction
  - Amplification of specific DNA sequences, generating millions of copies from a small initial sample
  - Crucial when dealing with minute or degraded samples.
  - Repeated cycles of heating and cooling, seperating DNA analysis techniques

ERRORS AND CHALLENGES Date [02/26/26]

- These issues include human error and human bias linking innocent people to crimes, privacy rights and a surge in racial disparities
- Error of bringing people and families that have no link to the crime
- Some downfalls for Bayesian statistics include
  - Misunderstanding the prior probability
  - Misinterpreting the liklihood ratio
  - Failing to account for uncertainty
- Computational Complexity
  - Many Bayesian models need complex numerical methods like MCMC, which can be computationally expensive, especially for large data sets
- Choice of prior
  - The prior you choose can greatly affect the results. Picking a good prior is hard when you don't have much prior knowledge about the problem
- Model selection
  - It can be tough to figure out the best model or how complex it should be. Bayesian methods require assumptions about the model and the prior.
- Sample contamination, faulty preparation procedures, and mistakes in interpretation of results
- Living or dead specimen

Ethics and Privacy / Reliable Info

Date [02/21/26]

- ACCURACY AND RELIABILITY
  - Apart from internal inspections and audits to asses compliance with the QMS and the ISO standards, the QMS and testing methods must be exposed to regular external peer review through accreditation
  - DNA analysis using Bayesian inference:
    DNA Sample → DNA Profiling → Bayesian Inference → Likelihood-Ratio → Interpretation
  - Fingerprint Comparison & Identification
    - Used to evaluate the likelihood that a fingerprint found at a crime scene matches a particular individual
  - Understanding Likelihood Ratios and Posterior Probabilities
    - Calculated as the ratio of the likelihood of the evidence given one hypothesis to the likelihood of the evidence given an alternative hypothesis.
    - The posterior probability is the probability of a hypothesis given the evidence
      - → >1 = Evidence supports hypothesis
      - =1 = Evidence is neutral
      - <1 = Evidence contradicts the hypothesis
- ETHICS AND PRIVACY
  - Privacy rights are violated by familial seaches
  - Privacy rights critique has several dimensions:
    - Concerns for those tested, for the many family members who are caught in the larger crime response dragnet, and a distinct concern for the innocents whose lives might be destroyed by being under a cloud of suspicion

Evolution

Date 02/01/2017

- Bayes, Gauss, Laplace (1986) - beta - binomial and normal - normal conjugate models with uniform priors
- Pearson's family (1916) - an important strain of statistics was the catogorization of parametric models of distributi
- Efron and Morris (1972) - The approach of estimating a prior from data could be seen as an extension, going outside the canonical Bayesian framework in which the prior is specified unconditionally on the data.
- Lindley and Smith (1972) - It can also be framed as Bayesia by considering the parameters of the prior as being "hyperparameters" that are estimated jointly with the parameters of the data model
- Tukey (1977) - Any discussion of the unexpected leads to the thoughts about "the expected", and this relates to statistical graphics and exploratory data analysis in two ways.
- Box (1980) AND Rubin (1984) - Statistical graphs can be folded into the Bayesian formalism

ABOUT

Date 01/08/25

Brief Description

- Our research project can allow us to use the Bayesian statistics to interpret DNA fingerprint matches more accurately. This method uses analysis, previous information and probability to determine how we can strengthen DNA fingerprinting evidence. By applying Baye's Theorem the project can demonstrate DNA fingerprinting to be a statistical tool which makes it more reliant on context, and not the match itself.

Purpose of Research

- The purpose of our research project is to determine how prior probabilities can influence the interpretation of DNA fingerprinting. This shows that only a DNA match does not provide full evidence and requires clear certainty which can be identified using the Bayesian analysis. This analysis can provide a more scientific and realistic method for forensic evidence during court cases. The main goal of this project is for others to understand the importance of using probability based reasoning to reduce the misconceptions of DNA evidence in real world investigations

Description of Research Study

- The goal of our project is to help improve the interpretation and reliability of DNA fingerprinting by using the Bayesian statistics. We will be exploring how DNA fingerprinting is evaluated, and how probability and prior probability influences the strength of DNA match. This project can examine the evolution of DNA matching and analyze if DNA evidence shold be considered with probability and statistical reasoning

Possible Topics For Reasearch

Date 01/03/25

TABLE OF CONTENTS

- Introduction
- What are Bayesian statistics
  - ↳ Bayes's theorum
- Application of Bayesian statistics in DNA fingerprinting
  - ↳ Bayesian Inference in DNA match probability
- Bayesian approaches to population genetics in DNA fingerprinting
  - ↳ estimate allele frequencies in different populations
- Intergenerational analysis of fingerprint patterns using Bayesian methods
  - ↳ Assess the influence of genetics versus environmental factors on fingerprint phenotypes
- Bayesian Methods for evaluating DNA evidence in Cold Cases
  - ↳ Bayesian statistics in re-evaluation of DNA evidence from cold cases
- Ethical Implications of Bayesian statistics in forensic DNA Profiling
  - ↳ Privacy, data security, and the potential for wrongful convictions
- Techniques / Technology
- Pros/Cons
- Reliable Information
- Advancements (AI)
- Challenges / Errors and future prospects

Brainstorming

Date [10/09/25]

- Drugs & genetics: Why do some people respond to drugs differently than others
- DNA Fingerprinting (~~maybe with~~ AI help)
- X Password Security: How easily your password can be hacked
- How waste can be turned into various types of energies that can be used?
- Stimulated restriction enzyme digestion demonstrates DNA modification effects
- X Multitasking: Brain drain or boost in Efficiency
- Exploring DNA damage: what effect do Ultraviolet Rays have on yeast colony growth?
- Using Bayesian statistics to interpret DNA fingerprinting evidence

- Formula:

$$P(\text{Guilty} \mid \text{Match}) = \frac{P(\text{Match} \mid \text{Guilty}) \times P(\text{Guilty})}{P(\text{Match})}$$

- P(Guilty | Match) - likelihood of how guilty someone is AFTER DNA match
- P(Match | Guilty) - How likely person is the suspect from match
- P(Guilty) - How likely person was guilty BEFORE DNA testing
- P(Match) -

a) True match

$$1 \times \frac{1}{50,000} = \frac{1}{50,000}$$

b) 49,999 innocent people from population

- False positive rate 1 in 1,000

$$49999 \times \frac{1}{1,000} \approx 50$$

- 1 true match 50 false matches = total matches = 51

$$P(\text{Match}) = \frac{51}{50,000}$$

$$P(\text{Guilty} \mid \text{Match}) = \frac{1 \times \frac{1}{50,000}}{\frac{51}{50,000}} = \frac{1}{51} \approx 0.0196$$

- Even though the DNA data matched, probability that the person is actually guilty is ~1.96%

- Comparison:
  - Without using Bayesian statistics, people would assume DNA match proves the person 100% guilty.
  - Bayesian analysis shows the truth behind how guilty a person is.
  - When false positives and database size were taken into consideration, the probability that the matched person was the true source of DNA was much lower.
  - Demonstrates how Bayesian statistics can actually be used when interpreting DNA fingerprinting evidence.

Hilroy

Real world example of Bayesian statistics:

- Scenario:
  - A crime occurs, and a knife is found near by the scene with DNA evidence.
  - DNA is found as skin cells contain nuclear DNA (genetic blueprint found inside the nucleus of the Eukaryotic cells) which can be used for DNA fingerprinting.
  - DNA present on the knife could either be the victims, the suspects or someone else within the area, this piece of evidence does not prove the real suspect.
- DNA fingerprinting:
  - The lab analzyes STR marks (Short Tandem Repeats) to produce a DNA fingerprint from the evidence
  - DNA fingerprinting only observes a section of the entire genome so there are still chances of a rare match occurring.
- DNA database search:
  - DNA fingerprint is compared to a large database of people's DNA profile.
    - ↳ In our situation total population is 50,000 people
    - ↳ False positive rate is 1 in 1,000
- Even if only one person committed the crime, testing many people means there is a small chance that someone else would also have a match
- After doing the sample search, there would be one person's DNA that would match the crime scene sample. This does not prove that they were 100% right because DNA could have been transferred accidentally or the test could have had human error associated with it.

3. How accurate is this method and is it reliable?

- There can also be errors that arise from sample handling or test interpretations which can lead to false positives in DNA testing. The Bayesian model can demonstrate how this false information can affect the value of DNA evidence in a legal case that is strong enough to convict someone.
- Using this model, we can determine the likelihood of random DNA match occurring and the chances of an actual false positive occuring when looking into DNA evidence.
- Undermining the importance or understanding the risk of a false positive can lead to serious mistakes in the future, most especially if the suspect was discovered through a DNA database search. Not knowing the real error rate adds uncertainty to how much we can trust DNA evidence.
- Bayesian analysis, typically implemented through probabilistic genotyping (PG) software, can reliably distinguish between individuals in complex DNA forms where manual analysis (CPI - Combined probability of inclusion) fails.
  - Probabilistic genotyping → method used to analyze complex DNA. Uses mathematical models & computer algorithms to evaluate DNA mixtures and provide a probability of a match.

- Purines and Pyrimidines are nitrogenous bases that make up two different kinds of nucleotides bases in the DNA and RNA.
- Purines: are natural chemicals that are found in the body's cells and many foods which build DNA and RNA (adenine and guanine) and are vital for energy (ATP).
- Pyrimidines are organic compounds which contain nitrogen with a single-ring structure that form nucleobases of DNA and RNA. They mainly form cytosine and thymine (in DNA), and uracil (in RNA).

2. What is the Bayesian statistics method?
- Bayesian statistics is a method of analyzing data where you begin with an initial belief and then change your beliefs as you get new information. This combines what you know with new evidence to make more accurate decisions.
- Baye's Theorem is a mathematical formula used in probability that changes the probability of a hypothesis when given new evidence.

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

A, B = events

$P(A|B)$ = probability of A given B is true

$P(B|A)$ = probability of B given A is true

$P(A), P(B)$ = independant probability of A and B

In a forensics case:
- A = "suspect is the source of DNA."
- B = "The DNA profile found at the scene matches suspect profile."
- Bayesian statistics combines prior knowledge (suspect present at scene) with DNA evidence probabilities to calculate a clear likelihood.

Main Revolving Question:
- If everyone has unique DNA, how can DNA evidence still be wrong and how can statistics help explain this?

Why Bayestian statistics?
- Bayesian statistics updates the confidence of evidence in a suspect after seeing new evidence instead of assuming that the evidence in a suspect is accurate. It measures the uncertainty of an event occurring rather than ignoring the importance of it.

Why is it Beneficial?
- This is beneficial because DNA evidence is seen to be perfect and unquestionable. When there is a DNA match, scientists believe that their theory is correct, but misinterpreting this can cause innocent lives to be put in danger and being convicted in a court case. This form of analysis makes the justice system much more accurate and fair.

1. What is DNA fingerprinting?
- DNA fingerprinting is a technique used to determine the identity of a person based on the nucleotide sequence of specific parts of the human DNA which are unique to every individual.
- A DNA strand is built as two long strands, double helix, which are made up of repeating units of nucleotides. There are three main parts to the DNA strands: phosphate group, a deoxyribose sugar, and one of the following 4 nitrogen bases: Adenine (A), Thymine (T), Guanine (G), Cytosine (C).
- Polynucleotide chains are formed when nucleotides are linked together through bonds of sugar of one and the phosphate of the next which forms a sugar-phosphate backbone.

Hilroy