Mathematical Application of DNA Profiling

Esthelle Chung

Chadwick International

Abstract

Adolescent sex crimes have impacted the world for millennia, particularly in South Korea, with the recent Jo Du Sun Case and the notorious Hwaseong Serial Murders, both of which have left numerous Koreans in terror. Of greater significance are cold cases and unsolved criminal investigations. Cold cases often make it difficult to punish their perpetrators, even when caught, due to a lack of evidence or accountability. These "cold cases" represent traumatic challenges for crime investigators, where traditional investigative methods have failed to identify the culprit and yielded limited progress. In the modern day, scientists search for innovative solutions that diverge from court testimonials and solid scientific data. This research paper explores mathematical modeling and DNA profiling in biological sciences as a solution to producing concrete evidence on cold adolescent sex crimes. Mathematical modeling is a technique that applies mathematical concepts and formulas by setting variables applicable to specific scenarios. Such models allow the quantification of the likelihood of specific genetic markers being present and verifying a committed crime. This study is two-parted, as it delves into a) a logical derivation of a conditional probability model deciphering the relationship of DNA evidence and an actual verdict and b) a graphical representation of the probabilities in a sex crime using graph theory.

Introduction

Currently, adolescent sexual crimes are on a surge, particularly a rise in unsolved cold cases due to a lack of confirmed evidence to determine a verdict. Such crimes not only violate the fundamental rights and dignity of victims but also produce profound psychological, physical, and social impacts. Studies from various countries indicate that a substantial proportion of adolescents experience sexual violence before age 17-18, with prevalence rates ranging from 20% to over 50% in different regions. Sexual crimes against adolescents can lead to long-term

mental health issues, such as post-traumatic stress disorder (PTSD), depression, anxiety, and substance abuse, including drugs or alcohol. Victims are also at an increased risk of engaging in risky sexual behaviors, experiencing revictimization, and developing adverse physical health concerns.[1]

In particular, cold cases pose significant challenges due to the passage of time, loss of evidence, and the vulnerability of victims during the initial investigation. As DNA evidence is substantial in most cases, serving as a strong indicator of a verdict, the failure of such evidence to accurately align with an actual verdict is detrimental to the investigation process. If it alone is not enough solid evidence, especially in cold cases, there is a need to find other methods to validate the existing evidence and hold perpetrators accountable.

While official statistics on cold cases involving adolescent sex crimes are limited, South Korea's overall crime rate has shown a declining trend in recent years.

According to MacroTrends, the country's crime rate per 100,000 population decreased from 0.57 in 2019 to 0.52 in 2021. However, the psychological and societal impact of unresolved cases cannot be understated, as they leave victims' families without closure and the perpetrators at large, posing an ongoing threat to public safety.

The Hwaseong serial murders occurred between 1986 and 1991 in the Gyeonggi Province, where ten women and girls, ranging in age from 13 to 71, were brutally raped and murdered. Despite a massive investigation involving over 2 million police officers, the case remained unsolved for years due to the lack of forensic technology and DNA analysis capabilities at the time. (Korean Herald) It was not until 2019 that advancements in DNA restoration techniques allowed investigators to identify Lee Chun-Jae as the perpetrator, who confessed to 14 murders and rapes in total.

On January 17th, 2024, a suspect who committed a sex crime two years ago was about to be released after serving their complete sentence, yet was re-arrested, identified as the perpetrator of an unsolved child sex crime from 18 years ago. (Chosun News) The Women and Children Crimes Investigation Department of the Seoul Southern District Prosecutors' Office identified A (42)'s DNA matching to a suspect who broke into a home in Seoul in 2006 and threatened and molested two children, 9 and 11 years old, at the time, with a weapon. A was already serving a sentence for another sex crime committed in 2022.

---

[1]Doerr CM, Hoeffler A, Goessmann K, Olorunlambe W, Hecker T. Sexual violence affects adolescents' health and prosocial behaviour beyond other violence exposure. Eur J Psychotraumatol. 2023;14(2):2263319. doi: 10.1080/20008066.2023.2263319. Epub 2023 Oct 16. PMID: 37843878; PMCID: PMC10580796.

DNA Profiling Mechanism

1.1 Structure of DNA

DNA, or deoxyribonucleic acid, is a double-stranded helix stored in an individual's genome. DNA is essentially the "blueprint" of the human body, storing all genetic information for reproduction, body growth, and characteristic features. The DNA structure is a polymer of nucleotides, where each nucleotide is composed of a nitrogenous base, deoxyribose (five-carbon sugar), and phosphate. There are two nitrogenous bases: purines (adenine, guanine) and pyrimidines (cytosine, thymine). The bases form complementary pairs in each nucleotide link: adenine with thymine and guanine with cytosine. Groups of three bases form a codon, and one codon codes for 20 amino acids, the building blocks of protein. A genome is the entire set of DNA instructions in a human cell. There are two complete genome copies in each cell, and only a small portion of each genome carries genetically relevant information. Approximately 20,000-25,000 genes (1.5%) comprise the coding and regulatory regions that encode and regulate protein synthesis. 23.5% are enhancers, promoters, repressors, or polyadenylation signals responsible for gene regulation. The remaining 75% are extragenic DNA, including repetitive copies or interspersed repeats.

1.2 Profiling Conditions

DNA profiling is a technology that relies on direct biological evidence. As a result, samples must meet appropriate conditions. First, all samples must be highly polymorphic, where the sample size is large and carries at least two different DNA sequences. Second, the samples should be cheap and easy to characterize. Most preferably, DNA units are interspersed repetitive elements or satellites, as they have clear, recurring patterns that each carry a distinct characteristic. Third, samples should have low mutation rates, as all living samples are prone to modifications during an experimental period.

1.3 DNA Profiling Procedure

The current DNA profiling procedure in professional settings follows five main steps: initial collection, extraction, quantification, amplification, and analysis.

1.3.1 Collection, Characterization, and Storage

In forensic crime scenes, investigators collect direct biological evidence samples left at sites to identify their DNA. Sample types must be nucleated epithelial cells, except red blood cells. Sources can vary in their form from liquid or dry deposits (blood, saliva, semen) to hard tissues (bone, teeth) or hair (follicles).

All samples are collected with a sterile brush or bud, then wrapped in a plastic or paper envelope, and kept in a dry environment at room temperature. Maintaining the crime scene's integrity is vital, so scientists must collect relevant, unmodified samples and wear full protective gear to prevent cross-contamination. After collection, samples should be preserved in an anticoagulant (ethylenediamine tetra-acetic acid), initially at 4 degrees C, for 5-7 days. Afterward, they are stored at -20 degrees C for a few weeks before experimentation.

1.3.2 Extraction

The extraction process allows DNA to be separated from other elements in a biological sample, including cell membranes, proteins, or additional liquids that hinder the observation of raw genetic material. The most conventional extraction method is silica extraction.

The silica extraction method, a widely used DNA extraction technique, is divided into two approaches: silica matrices extraction and silica-column extraction. In the silica matrices extraction, the DNA sample is selectively bonded with a silica surface with positive ions. This simple process, which involves the electrical attraction causing the negatively charged DNA to bind to the silica matrix while other cellular contaminants are distilled through water or Tris-EDTA, a dilution buffer, is fast and cost-efficient. However, it's important to note that each silica matrix is not reusable, which means that creating new matrices for each extraction requires time and effort.

In the silica-column method, the sample is mixed with 1% sodium dodecyl sulfate (SDS) and 100 mg proteinase K, incubated at 60 degrees C for an hour, and inserted into a silica gel tube. Then, phenol and chloroform are added at a 1:1 ratio and centrifuged for 5 minutes. The mixture forms a biphasic emulsion, a separation of fluids, into the organic protein phase beneath the silica tube and the aqueous DNA phase above the gel polymerase. An advantage of this method is increased purity. The silica gel layer prevents cross-contamination and direct contact with toxic reagents, reducing modifications of the original sample. Also, it produces a 40% higher yield of extracted DNA than organic extraction methods.

1.3.3 Quantification

When analyzing DNA, accurately measuring the sample size and quality is crucial to gain ideal results, as different ratios can lead to difficult or impossible profiles to interpret. As a result, the quantification process is an essential part of enhancing experimental accuracy.

One quantification technique is intact or degraded DNA agarose-gel electrophoresis. This method utilizes the electric current and size of DNA molecules to determine the number of fragments in the sample. The advantages are that it is relatively easy, quick to identify, and indicates the size of each fragment. However, the counting process of this method is subjective, and thus, DNA concentration can be overestimated.

1.3.4 Amplification

DNA amplification increases the number of sample copies for final analysis experiments. The most dominant method is polymerase chain reaction (PCR), where specific regions of the DNA are amplified by up to 1 billion nucleotides within 30 repetitive cycles. Each cycle consists of three stages: denaturation, annealing, and extraction. Often, expected cycle frequencies range from 28 to 32, but with low DNA samples, cycle numbers can increase to 34. However, currently, PCR is prone to contamination, so profilers are developing alternative methods, such as nucleic acid sequence-based amplification, strand displacement, or hybridization chain.

1.3.5 Detection & Analysis

The final and most crucial step of DNA profiling is analysis. Most investigators commonly profile autosomal short tandem repeats (STR). Alleles in STR loci are differentiated by the number of copies of repeat sequences within each STR locus, and scientists utilize this trait to compare genotypes and distinguish individuals. The more STR loci collected, the greater the discrimination value. This technique is often used for maternity/paternity testing or kinship testing, identifying an anonymous rape perpetrator, or discerning disaster victims by

comparing heterozygous and homozygous profiles. These profiles of DNA fragments are visualized as a pattern of bands like a barcode.

II. Comparison of DNA Samples

In particular cases, specific DNA profiling is initiated under special conditions. First, scientists analyze the Y chromosome in forensic medicine. The Y chromosome is only present in males. Because azoospermic or vasectomized rapists leave no measurable semen trace due to a lack of sperm or permanent sterilization, analyzing the chromosome can help identify the perpetrator in sexual assault cases instead of ineffective microscopic examinations. Second, scientists investigate the mitochondrial DNA (mt-DNA) to analyze severely degraded or old biological samples with increased proficiency. Mt-DNA is inherited from the mother, and because there are 200-1700 copies per cell, there is an abundant supply and increased probability of sample survival than nuclear DNA. Consequently, meaningful results can be derived even in samples with low amounts of DNA, like hair shafts.

Mathematical Applications

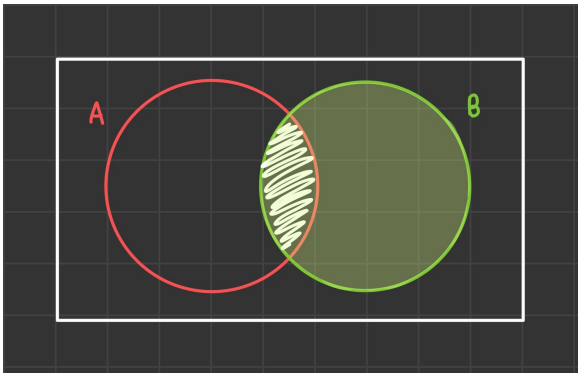
I. Math Model Development

This mathematical model consists of basic theory on conditional probabilities and the Bayes Theorem regarding the Bayesian posterior probability. The Bayesian posterior probability is the probability that the defendant is the trustworthy source of the DNA found at the crime scene, given that the defendant is a DNA match. (Ayres and Nalebuff, 2015)

Bayes Theorem is commonly expressed as $P(A|B)=P(AB)P(B)$, where $P(A|B)$ represents the probability of A given B has conditionally occurred, $P(AB)$ represents the probability of both A and B happening, and $P(B)$ is the probability of B.

Intuitively, we can use a Venn Diagram to understand this theorem:

Figure 1. Venn Diagram Representing Bayes Theorem through Events A and B



When B is valid or has occurred, the only possible way for A to be true is when both A and B are genuine, represented as the intersection of A and B in a Venn Diagram. Thus, the conditional probability of $P(A|B)$ is equal to the ratio between the likelihood when both A and B are true, $P(AB)$, under the probability of B being actual, $P(B)$.

Consider this theorem with more than two events.

Let A1, A2, A3, ..., An be a finite set of mutually exclusive events.

Then, the following are exhaustive events:

1) P(A1A2A3 ... An) = P(A1) + P(A2) + P(A3) + ... + P(An)
2) A1A2A3 ... An = U where U represents the total set of events.

Axiom 1 of fundamental probabilities states P(U) = 1, and B is an arbitrary event.

Then, for i=1,2,3,...,n,

$$P(A_i|B) = \frac{P(A_i) \times P(B|A_i)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) + \ldots + P(B|A_n)P(A_n)}$$

This theorem can be applied in a crime setting to generate a relevant model for my global issue.

First, let's define all variables:

E: event that DNA evidence at the crime site matches the DNA profile in the database

G: an event that the identified suspect is guilty

Then, the conditional probability of the suspect being guilty, given that their DNA evidence is a match, can be represented by P(G|E).

Model 1.

$$P(G|E) = \frac{P(E|G) \times P(G)}{P(E|G) \times P(G) + P(E|\sim G) \times P(\sim G)}$$

In this equation, event ~G is the complementary event to event G.

We can assume P(E|G)=1, as it is the probability that the DNA evidence is a match given the suspect is guilty. P(E|~G) is the probability that the DNA evidence matches, given that the suspect is not guilty, meaning the game is random and unexpected. P(E|G)P(E|~G) can be used as a likelihood ratio (LR) and is equal to the reciprocal of the probability of the match being coincidentally equal.

Figure 2. Data From Soongsil University Confirming Likelihood Ratio

STR (Short Tandem Repeats) Locus	Evidence at Crime Site		Suspect's DNA		Occurrence Frequency
	Allele 1	Allele 2	Allele 1	Allele 2	
CSF1PO	13	12	13	12	0.056
D5S818	13	13	13	13	0.051
D7S820	12	8	12	8	0.052
D12S317	12	8	12	8	0.089
TH01	7	7	7	7	0.061
TPOX	12	8	12	8	0.025
vWA	18	18	18	18	0.033
D3S1358	18	17	18	17	0.033
D8S1179	15	11	15	11	0.024
D16S3253	9	9	9	9	0.004
D18S51	16	13	16	13	0.032
D21S11	32.2	30	32.2	30	0.057
FGA	23	21	23	21	0.063
PentaD	12	11	12	11	0.050
PentaE	12	10	12	10	0.006
D12S391	20	17	20	17	0.043
D14S608	12	11	12	11	0.066

Cho, Yoonjung, et al. "A Review of Extended STR Loci and DNA Database." Journal of Experimental & Biomedical Sciences/Biomedical Science Letters, vol. 28, no. 3, 30 Sept. 2022, pp. 157–169, www.bslonline.org/journal/view.html?doi=10.15616/BSL.2022.28.3.157, https://doi.org/10.15616/bsl.2022.28.3.157. Accessed 30 May 2024.

The Bayes Rule states "the posterior odds of an event occurring will equal the prior odds (not conditioned on evidence of any matches) of someone in the database being the source multiplied by the relative likelihood ratio of observing M matches.

Posterior Odds = Prior Odds x Relative

Likelihood of Observing M Matches.[2]

Applying this rule to Model 1,

$$\frac{P(G|E)}{P(\sim G|E)} = \frac{p}{1-p} \times \frac{P(E|G)}{P(E|\sim G)}$$

In this probability model, all events have two possible outcomes: guilty or not-guilty (G) and match or non-match (E). Thus, we can express all events as a binomial distribution.

A binomial distribution is a specific discrete probability distribution where there only exist two precise outcomes to an event. This predictability is key: let G=0 for success and G=1 for failure. Then, P(G=0) = p and P(G=1) = 1-p by the binomial distribution definition.

We can apply this logic into our Bayes Rule to simplify our conditional probabilities.

$$\frac{P(G|E)}{P(\sim G|E)} = \frac{p}{1-p} \times \frac{P(E|G)}{P(E|\sim G)}$$

II. Graph Theory Application

Graph theory can be employed to create a visual representation of the probabilities between a DNA match and a guilty suspect as an alternative approach towards mathematical DNA profiling.

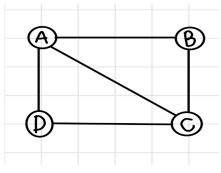
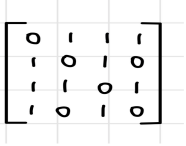
Discrete mathematics is one significant mathematical field that studies relationships between countable, distinct, and separate objects. Of its key components, graph theory plays a critical role. Graph theory is employed to abstract and analyze relationships and connectivity between different objects, making it a suitable tool for processing discrete data and events. Graphs consist of vertices and edges that represent variables in a situation and possess discrete properties.

Let's define terms in discrete mathematics:

Nodes (Vertices): connection points in a network; represents different entities in a crime scene, including suspects, pieces of evidence (DNA sample), victims, crime locations, witnesses, etc

Edges: lines that connect points in a network; represents relationships or connections between entities

This method will utilize an adjacency matrix based on a hypothetical case to display how graph theory may be employed in criminal settings. An adjacency matrix is a square matrix representing adjacent vertices in a finite graph.

Figure 3. Random Example of Graph (ABCD)	*Figure 4. Adjacency Matrix of Figure 3*
	

[2] Ayres, Ian, and Barry Nalebuff. "The Rule of Probabilities: A Practical Approach for Applying Bayes' Rule to the Analysis of DNA Evidence." Stanford Law Review, vol 67: 1447, June 2015, https://ianayres.yale.edu/sites/default/files/files/The%20Rule%20of%20Probabilities.pdf.

Figure 4 is an example of an adjacency matrix based on Figure 3, a randomly generated graph. There are four nodes in Figure 3; therefore, the adjacency matrix is a 4x4 matrix. Each row and column represents A to D, from top to bottom or left to right, respectively. For instance, (1,1) in the matrix would be (A, A), or (1,2) would be (B, A). We can observe that the input values of (1,1) are 0 and (1,2) are 1; there exist 0 edges between A and A (self), and there exists one edge (path) between A and B.

If n nodes exist in a graph, the adjacency matrix will be n x n. Each cell (i,j) in the matrix contains either one if an edge exists between node i and node j or 0 if no edge exists.

Here is a hypothetical scenario for analysis:

A teenage girl aged 13 has been raped in Location A. 2 DNA samples have been put to analysis, both collected at the site, and the police have identified two different suspects, both free of alibi. Analysis shows the two suspects match one of the DNA samples each.

Entities:

- Suspects: S1, S2
- DNA Samples: D1, D2
- Victim: V1
- Crime Site: C1

Relationships:

- Suspect S1's DNA matches DNA sample D1.
- Suspect S2's DNA matches DNA sample D2.
- DNA sample D1 was collected at crime site C1.
- DNA sample D2 was collected at crime site C1.
- Victim V1 was found at crime site C1.

This graph would consist of 6 nodes: S1, S2, D1, D2, V1, and C1. Edges exist in the following cells: (S1,D1), (S2,D2), (D1,C1), (D2,C1), and (V1,C1).

Figure 5. Graphical Representation of Hypothetical Case Stud

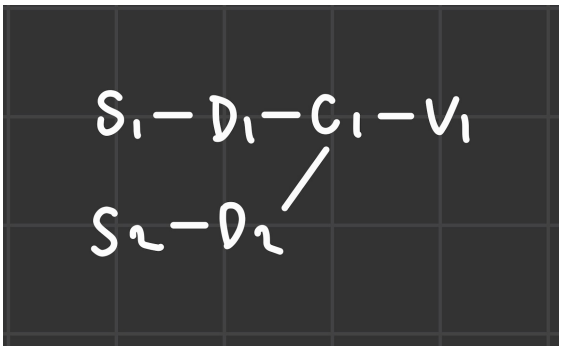


Figure 6. Adjacency Matrix of Figure 5

	S1	S2	D1	D2	V1	C1
S1	0	0	1	0	0	0
S2	0	0	0	1	0	0
D1	1	0	0	0	0	1
D2	0	1	0	0	0	1
V1	0	0	0	0	0	1
C1	0	0	1	1	1	0

Figure 5 and Figure 6 are graphical representations and adjacency matrices respectively of this hypothetical case. In analyzing an adjacency matrix, we may assess the following:

1) Direct paths with consecutive edges- walk between a suspect and a victim
2) Clusters or subgroups - closely connected entities
3) Nodes with high degrees of connectivity

This case study shows a clear path between S1 to V1, including C1. Therefore, we may infer that Suspect 1 is the perpetrator of the crime. An application to graph theory is the use of probabilities. By setting prior probability values at each edge and node and calculating the aggregate probabilities in each path, it is possible to determine the most likely outcome of a graph.

Discussion

Both the probabilistic approach using conditional probabilities and the graphical approach using graph theory offer distinct perspectives to calculating the relationship between a DNA match and a verdict, serving as potentially solid, scientific evidence in court.

Method 1 provides a quantitative framework for assessing the likelihood of events based on available evidence. One strength is that conditional probabilities can adapt to incorporate various factors and evidence into a single numerical probability value by adding multiple conditions into one situation. This adaptability can be particularly reassuring in cases with different types of evidence, providing a versatile tool for legal and forensic professionals.

On the other hand, Method 2 offers a visual and intuitive approach to modeling causal relationships and dependencies between variables. The node-edge graphs and adjacency matrices can effectively represent the multi-factored situation in a crime scene that all influences the probability of a verdict. Such visual representations are often simple and easy to comprehend, as they simplify complex situations into “nodes and vertices.” Therefore, its strength lies in its ability to capture complex relationships and facilitate reasoning. However, a limitation is that while simplicity may be beneficial in comprehending the situation at first sight, the simple nature may downgrade some critical aspects of the crime scenario, possibly undermining key events or possibilities.

Both approaches have clear strengths and limitations. However, the choice between these approaches may depend on the specific case, the available evidence, and the expertise of the analysts involved. This investigation suggests the most optimal method is a hybrid of Method 1 and Method 2: using graphs to analyze situations and producing numerical results through conditional probabilities. Importantly, cross-checking with both approaches will be a merit for crime investigators, providing a

reliable way to assess the verdict of a particular crime.

This investigation offers valuable insights into the potential of integrating mathematical and biological theories in legal contexts where court proceedings traditionally lean towards qualitative evidence. At the same time, there are significant limitations that reduce its applicability in real-world scenarios. First, the biological model compared samples collected at optimal conditions - where the evidence is uncontaminated and easily observable. In reality, especially in sexual crime sites, it is challenging to preserve evidence of high quality. Factors like contamination, degradation, or environmental exposure can lead to distorted or defective DNA samples, thereby reducing the accuracy of the biological and mathematical models used in the investigation. This raises concerns about extrapolation, as drawing conclusive statements from these suboptimal samples, significantly when evidence quality is compromised, leads to significant inaccuracies in probability derivation and overall reliability.

Second, both mathematical models in the study were assessed using a relatively small sample size of the database. As databases expand, especially with diverse and heterogeneous datasets, the accuracy of these models tends to diminish due to increased uncertainty. Similarly, applying generalizations derived from simpler models may lead to the extrapolation of particular DNA samples at a larger scale. The comprehensive nature of the referenced database also introduces more significant uncertainties in probability calculations. It is vital to acknowledge that while scientific theories always hold, they may not be entirely accurate or representative of real-life examples, as they are romantic hypotheses. Especially in a crime scenario as complicated and ponderous as adolescent sex crimes, there may be countless exceptions that impede the accuracy of the developed models. Therefore, this paper may be a starting point for further research and adaptations in each case regarding mathematical applications.

Furthermore, cold cases generally make it difficult to collect evidence in the first place, which may restrict the potential of statistics. Moreover, sexual crime processing must be completed quickly, as collected evidence may become insignificant if time elapses. In addition, as the victims are young children, investigators and scientists prioritize minimizing the pain and fear victims may be facing throughout the investigation; thus, all court processes should occur speedily. Utilizing the mathematical model, I developed may hinder such an objective, as it takes time to process the data and convert it into numerical standards that can be analyzed. If such approaches are to be developed further, it may be necessary to involve technology or AI to compute probabilities more accurately and quickly.

Conclusion

Applying Bayes's Theorem and statistical analysis to DNA evidence can be a powerful tool for reinvestigating and potentially solving cold cases of adolescent sexual crimes in South Korea. By correctly calculating the conditional probabilities of whether a DNA match always means the suspect is guilty while accounting for factors including database size, random match probabilities, and the possibility of contamination, investigators can gain valuable insights into determining the likelihood of guilt. It is essential to consider the drawbacks that this method may currently have. Estimating prior probabilities requires careful empirical analysis. There may also be concerns about properly conveying complex statistical concepts to judges to be admitted as worthy evidence.

Nonetheless, a Bayesian framework provides a quantitative method for weighing DNA evidence. When utilized responsibly, it has the potential to supersede testimonial evidence, a shift that can instill confidence in the reliability of court decisions. It also helps overcome human cognitive biases that subjectively influence court decisions, allowing a more objective evaluation of guilt or innocence. While not a panacea, statistical analysis through both conditional probabilities or graph theories are worthwhile investigations in the pursuit of justice for solving this global issue.

Works Cited

1. Ayres, Ian, and Barry Nalebuff. “The Rule of Probabilities: A Practical Approach for Applying Bayes’ Rule to the Analysis of DNA Evidence.” Stanford Law Review, vol 67: 1447, June 2015, https://ianayres.yale.edu/sites/default/files/files/The%20Rule%20of%20Probabilities.pdf. Accessed 31 May, 2024.
2. Bart Verheij, Floris Bex, Sjoerd T. Timmer, Charlotte S. Vlek, John-Jules Ch. Meyer, Silja Renooij, Henry Prakken, Arguments, scenarios, and probabilities: connections between three normative frameworks for Evidential reasoning, Law, Probability and Risk, Volume 15, Issue 1, March 2016, Pages 35–70, https://doi.org/10.1093/lpr/mgv013.
3. Doerr CM, Hoeffler A, Goessmann K, Olorunlambe W, Hecker T. Sexual violence affects adolescents' health and prosocial behavior beyond other violence exposure. Eur J Psychotraumatol. 2023;14(2):2263319. Doi: 10.1080/20008066.2023.2263319. Epub 2023 Oct 16. PMID: 37843878; PMCID: PMC10580796.
4. Doyle, Ken. “Bayes’ Theorem: Can Statistics Help Guide a Verdict in the Courtroom?” ISHI News, Apr 22, 2021, https://www.ishinews.com/bayes-theorem-can-statistics-help-guide-a-verdict-in-the-courtroom/. Accessed May 5, 2024.
5. Foreman, Lindsey A., et al. “Bayesian Analysis of DNA Profiling Data in Forensic Identification Applications.” Journal of the

Royal Statistical Society. Series A (Statistics in Society), vol. 160, no. 3, 1997, pp. 429-69. JSTOR, https://www.jstor.org/stable/2983264. Accessed May 5, 2024.

6. "MA1020 - Discrete Mathematics | University of Moratuwa." Uom. Lk, 2024, Com. lk/math/ma1020-discrete-mathematics. Accessed May 30, 2024.
7. Min-kyung, Jung. "The Korea Herald." The Korea Herald, May 26, 2017, www.koreaherald.com/view.php?ud=20170526000720. Accessed May 30, 2024. Panneerchelvam, S, and M N Norazmi. "Forensic DNA profiling and database." The Malaysian Journal of Medical Sciences: MJMS vol. 10,2 (2003): 20-6, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3561883/.
8. Rodriguez, Jae Joseph Russel B, et al. "An integrated system for forensic DNA testing of sexual assault cases in the Philippines." Forensic Science International. Synergy vol. 3 100133. Jan 11, 2021, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7848663/.
9. "South Korea Crime Rate & Statistics 1960-2024." Macrotrends.net, 2024, www.macrotrends.net/global-metrics/countries/KOR/south-korea/crime-rate-statistics. Accessed May 30 2024.
10. Yoon, L. "South Korea: Number of Arrests for Sexual Assault 2021 | Statista." Statista, 26 Jan. 2022, https://www.statista.com/statistics/629302/south-korea-number-of-arrests-for-sexual-assault/. 조선일보. "출소 하루 앞두고 다시 구속…18년 전 "미제" 아동 성추행 피의자였다." 조선일보, 2024, www.chosun.com/national/national_general/2024/01/17/VKPLY6X7RNDOBJTONWIT5ZJ6IE/. Accessed 30 May 2024.