Module 1: Overview

HISTORY OF IDENTIFICATION

Legal v. Scientific Thinking

"The very goals of science and law differ.
Science searches for the truth and seeks to increase knowledge by formulating and testing theories. Law seeks justice by resolving individual conflicts, although this search often coincides with one for truth."

"Rules of decision that are not tailored to individual cases, such as those that turn on statistical reasoning, are often viewed as suspect."

Feinberg SE (Editor). 1989. *The Evolving Role of Statistical Assessments as Evidence in the Courts.* Springer.

Forensic Science Approach

"The central problem of the criminal investigator is the establishment of personal identity – usually of the criminal, sometimes of the victim."

Need to distinguish between identity and individualization. **Identity** refers to unique existence – no two different things can be identical. The DNA profiles from a suspect and a crime scene are different things.

Individualization points to a specific person. A fingerprint from a crime scene is not identical to a suspect's recorded fingerprint, but can be used to identify him and prove his individuality.

Kirk PL. 1974. *Crime Investigation, (Second Edition).* Krieger,

Uniqueness

"no two objects can ever be identical. They can and often do have properties that are not distinguishable. If enough of these properties exist ... *identity of source* is established."

"The criminalist of the future may well be able to individualize the criminal directly through the hair he has dropped, the blood he has shed, or the semen he has deposited. All these things are unique to the individual, just as his fingerprints are unique to him."

Kirk PL. 1974. *Crime Investigation, (Second Edition).* Krieger,

Forensic science question

Not: "Is this profile unique?" (it is).

Not: "Are these two profiles identical?" (they can't be).

But: " Is there sufficient evidence to demonstrate that these two profiles originate from the same source?"

Bertillonage

Alphonse Bertillon (1853-1914), French anthropometrist. Son and brother of statisticians. Used 11 measurements:

1. Standing height
2. Arm reach
3. Sitting height
4.* Head length
5.* Head breadth
6. Length of right ear
7. Cheek width
8.* Length of left foot
9.* Length of left middle finger
10. Length of left little finger
11. Length of the left forearm and hand to the tip of extended middle finger

Bertillonage

Searching was done on four categories 4, 5, 8, 9. Each measurement divided into three subdivisions (large, medium, small) i.e. $3^4 = 81$ categories per person. Filing cabinets with 81 drawers used.

Using all 11 characters, plus 7 eye colors, the number of possible profiles is $3^{11} \times 7 = 1,240,029$.

Wikipedia entry for Alphonse Bertillon

"Being an orderly man, he was dissatisfied with the ad hoc methods used to identify the increasing number of captured criminals who had been arrested before. This, together with the steadily rising recidivism rate in France since 1870, motivated his invention of anthropometrics. His road to fame was a protracted and hard one, as he was forced to do his measurements in his spare time. He used the famous La Sant Prison in Paris for his activities, facing jeers from the prison inmates as well as police officers.

He is also the inventor of the mug shot. Photographing of criminals began in the 1840s only a few years after the invention of photography, but it was not until 1888 that Bertillon standardized the process."

https://en.wikipedia.org/wiki/Alphonse_Bertillon

Coincidental match

Two different men at Leavenworth in 1903 had very similar Bertillon dimensions (lengths in mm):

	Will West	William West
1	19.7	19.8
2	15.8	15.9
3	12.3	12.2
4	28.2	27.5
5	50.2	50.3
6	178.5	177.5
7	9.7	9.6
8	91.3	91.3
9	187.0	188.0
10	6.6	6.6
11	14.8	14.8

http://www.globalsecurity.org/security/systems/biometrics-history.htm

Fingerprints

"The arrangement of skin ridges is never duplicated in two persons."
J.C.A. Mayer, 1783.

J.E. Purkinje established categories of fingerprints in early 19th century.

W. Herschel, a British administrator, used fingerprints in India in 1850's.

H. Faulds, a British physician, used fingerprints in Japan.

Francis Galton wrote the book "Fingerprints" in 1892, and gave some probabilities for coincidental matches.

Fingerprints

Galton considered that the chance that a random fingerprint would match a specified print was 2^{-36}. For a population of size 1.6×10^9, the odds were 1 to 39 that the print of any single finger would be exactly like the same finger of any other person.

[This is based on the probability of not finding the print in a sample of size 1.6 billion.]

Heritability of fingerprints

Galton looked at 105 sib-pairs:

Second sib	First sib		
	Arches	Loops	Whorls
Arches	**5**	12	2
Loops	4	**42**	15
Whorls	1	14	**10**

Galton noticed that the diagonal counts of 5, 42, 10 are larger than those (2, 40, 6) expected if the sibs had independent fingerprints, but not as great as they could be (10, 68, 27). He did not have the chi-square test available in 1892, but did conclude that there was an association.

He did not find racial differences.

Uniqueness of fingerprints

Probability arguments not used now. By 1924, textbooks would say "No two fingerprints are identical in pattern." In 1939 J.Edgar Hoover wrote that fingerprints were "a certain and quick means of identification."

Acceptance of uniqueness probably followed from "(i) striking visual appearance of fingerprints in court, (ii) a few dramatically successful cases, and (iii) a long period in which they were used without a single case being noted where two different individuals exhibited the same pattern."

Stigler SM. 1995. Galton and identification by fingerprints. Genetics 140:857-860.

Stigler anticipated the same growing acceptance of DNA profiles being unique.

Misuse of Fingerprints

Oregon attorney Brandon Mayfield was wrongly identified by the FBI as the source of a fingerprint on an item of evidence in the 2004 Madrid train bombings.

https://en.wikipedia.org/wiki/Brandon_Mayfield

A subsequent report by the FBI admitted the error

https://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/jan2005/special_report/2005_special_report.htm

Accuracy of Fingerprints

A subsequent study Ulery et al "Accuracy and reliability of forensic latent fingerprint decisions" was published by authors including FBI scientists:

"169 latent print examiners each compared approximately 100 pairs of latent and exemplar fingerprints from a pool of 744 pairs. ...Five examiners made false positive errors for an overall false positive rate of 0.1%. Eighty-five percent of examiners made at least one false negative error for an overall false negative rate of 7.5%."

Ulery BT, Hicklin RA, Buscaglia J, Roberts MA. 2011. Proc Natl Acad Sci USA 108: 7733-7738.

Statistical approach

Partial transfer evidence: physical material or impressions transferred from crime scene to perpetrator (or perpetrator's possessions), or vice versa.

PTE is characterized and assigned to an identity-set. Does a particular person (or their type) belong to the set? Does anyone else belong to the set?

"If it is *highly improbable* that another member could be found, we would be *reasonably sure* that the correct origin has been located. But if it is *quite probable* that other members exist, we would *not be so sure* that we have the correct origin."

Kingston CR. 1965. J Am Stat Assoc 60:70-80, 1028-1034.

Blood Typing

Human ABO blood groups discovered in 1900. ABO gene on human chromosome 9 has 3 alleles: A, B, O. Six genotypes but only four phenotypes (blood groups):

Genotypes	Phenotype
AA, AO	A
BB, BO	B
AB	AB
OO	O

ABO System

The possible offspring blood groups for each pair of parents:

	Mother			
Father	A	B	AB	O
A	A,O	A,B,AB,O	A,B,AB	A,O
B	A,B,AB,O	B,O	A,B,AB	B,O
AB	A,B,AB	A,B,AB	A,B,AB	A,B
O	A,O	B,O	A,B	O

ABO System

Blood group	Antigens in red blood cells	Antibodies in serum
O	None	Anti-A and Anti-B
A	A	Anti-B
B	B	Anti-A
AB	A and B	None

http://www.redcrossblood.org/learn-about-blood/blood-types

ABO System

For blood transfusions, recipient should not produce antibodies to the donor's antigens:

	Donor			
Recipient	O	A	B	AB
O	OK			
A	OK	OK		
B	OK		OK	
AB	OK	OK	OK	OK

Charlie Chaplin and ABO Testing

Relationship	Person	Blood Group	Genotype
Mother	Joan Berry	A	AA or AO
Child	Carol Ann Berry	B	BB or BO
Alleged Father	Charles Chaplin	O	OO

The obligate paternal allele was B, so the true father must have been of blood group B or AB.

Berry v. Chaplin, 74 Cal. App. 2d 652

California Court of Appeals, 1946

"Concerning the immutability of the scientific law of blood-grouping, which we have no reason to question ..."

"Whatever claims the medical profession may make for blood tests to determine paternity, no evidence is by law made conclusive or unanswerable unless so declared by the Code of Civil Procedure of the State of California "

74 Cal.App.2d 652 (1946)

Outcome of Chaplin Trial

"The brouhaha surrounding Chaplin's case and similar paternity suits (like 1937's Arais v. Kalensnikoff and 1951's Hill v. Johnson) led to the reformation of paternity laws in the state of California, with other states eventually following suit. In 1953, along with Oregon and New Hampshire, California drafted the Uniform Act on Blood Tests to Determine Paternity, which in legalese states that: 'If the court finds that the conclusions of all the experts as disclosed by the evidence based upon the tests are that the alleged father is not the father of the child, the question of paternity shall be resolved accordingly.' "

http://mentalfloss.com/article/63158/how-charlie-chaplin-changed-paternity-laws-america

Spencer v Commonwealth of Virginia

From the Supreme Court of Virginia, September 22, 1989.

"Timothy Wilson Spencer was indicted for the capital murder of Susan Tucker, i.e., the willful, deliberate, and premeditated murder during the commission of, or subsequent to, rape. Spencer also was indicted for the rape of Tucker. ... a jury convicted Spencer of capital murder and fixed his punishment at death. The jury also convicted Spencer of rape and fixed his punishment at life imprisonment. Following a sentencing hearing, the trial court imposed the sentences fixed by the jury and entered judgments on the jury verdicts.

... We have considered all of Spencer's assignments of error and find no reversible error. We also have made the review of the death sentence mandated by Code 17-110.1 and conclude that the sentence should be affirmed. Accordingly, the judgments of the trial court will be affirmed."

Spencer v Commonwealth of Virginia (contd.)

"The parties stipulated that Spencer does not have an identical twin and that none of his blood relatives had committed the murder. **Therefore, the chances that anyone other than Spencer produced the semen stains was one in 135 million.** There are approximately 10 million adult black males in the United States."

Spencer was the first person executed after a conviction based on DNA evidence.

SPENCER v. COM 384 S.E.2d 775 (Va. 1989)
aw.justia.com/cases/virginia/supreme-court/1989/890579-1.html

Extreme Numbers: Robinson v. Mandell, 1868.

Two signatures matched at 30 downstrokes. The probability of a coincidental match was estimated to be 1 in 5. The probability of 30 coincidences in one pair of signatures was "once in 2,666 millions of millions of millions." (Mathematics professor Benjamin Pierce.)

"This number far transcends human experience. So vast an improbability is practically an impossibility. Such evanescent shadows of probability cannot belong to actual life. They are unimaginably less than those least things which the law cares not for."

Refers to chance of a coincidental match between two handwriting samples.

https://en.wikipedia.org/wiki/Howland_will_forgery_trial

No Extreme Numbers in Minnesota

"Schwartz contends that any probative value of statistical frequency evidence is outweighed by its prejudicial effect, as illustrated by the media exposure forensic DNA typing has received implying its infallibility. In dealing with complex technology, like DNA testing, we remain convinced that juries in criminal cases may give undue weight and deference to presented statistical evidence and are reluctant to take that risk."

447 N.W.2d 422 (1989)

Refers to matching DNA profile with a frequency reported as 1 in 33 billion.

Extreme numbers: Fingerprints

Chance of a match for a single finger print estimated to be less than 1 in 64 thousand million.

"When two fingers of each of two persons are compared, and found to have the same minutiae, the improbability [of 1 in 2^{36}] becomes squared, and reaches a figure altogether beyond the range of the imagination."

Galton F. 1892. Fingerpints.

DNA Profiling

Human Genome has about 3,000,000,000 elements (base pairs).

Any two people differ at about 3,000,000 of these.

Forensic profiles use 20 STR markers. Each of these markers as at least 10 variant forms, or at least 55 different combinations. Therefore there are about $55^{20} = 6.4 \times 10^{34}$ different profiles possible.

Only 1 in 10^{24} of the possible profiles can exist in the whole world.

Beyond Reasonable Doubt?

After forensic evidence is presented, a jury or judge may have to make a decision, based on the concept of "beyond reasonable doubt." What does that mean? A survey found:

Probability	Judges	Jurors	Students
0–50%	0	5	3
50%	1	6	2
55%	2	2	1
60%	8	4	1
65%	2	1	0
70%	14	2	1
75%	23	2	1
80%	58	8	9
85%	21	2	3
90%	68	9	20
95%	44	3	17
100%	106	25	30
Total	347	69	88

Source unknown.

People v. Collins

Another attempt to introduce probabilities into court:

Characteristic	Frequency
Girl with blond hair	1 in 3
Girl with ponytail	1 in 10
Man with mustache	1 in 4
Black man with beard	1 in 10
Yellow car	1 in 10
Interracial couple in car	1 in 1000
All six characteristics	1 in 12 million

https://en.wikipedia.org/wiki/People_v._Collins

Alec Jeffreys

For forensic applications, the work of Alec Jeffreys with on Restriction Fragment Length Polymorphisms (RFLPs) or Variable Number of Tandem Repeats (VNTRs) used electrophoresis. Different alleles now represented different numbers of repeat units and therefore different length molecules. Smaller molecules move faster through a gel and so move further in a given amount of time.

Initial work was on mini-satellites, where repeat unit lengths were in the tens of bases and fragment lengths were in thousands of bases. Jeffrey's multi-locus probes detected regions from several pats of the genome and resulted in many detectable fragments per individual. This gave high discrimination but difficulty in assigning numerical strength to matching profiles.

Jeffreys et al. 1985. Nature 316:76-79 and 317: 818-819.

Single-locus Probes

Next development for gel-electrophoresis used probes for single mini-satellites. Only two fragments were detected per individual, but there was difficulty in determining when two profiles matched.

The technology also required "large" amounts of DNA and was not suitable for degraded samples.

PCR-based STR Markers

The ability to increase the amount of DNA in a sample by the Polymerase Chain Reaction (PCR) was of substantial benefit to forensic science. The typing technology changed to the use of capillary tube electrophoresis, where the time taken by a DNA molecule to pass a fixed point was measured and used to infer the number of repeat units in an allele.

"Following multiplex PCR amplification, DNA samples containing the length-variant STR alleles are typically separated by capillary electrophoresis and genotyped by comparison to an allelic ladder supplied with a commercial kit. "
Butler JM. Short tandem repeat typing technologies used in human identity testing. BioTechniques 43:Sii-Sv (October 2007) doi 10.2144/000112582

Sequencing of STR Alleles

"STR typing in forensic genetics has been performed traditionally using capillary electrophoresis (CE). Massively parallel sequencing (MPS) has been considered a viable technology in recent years allowing high-throughput coverage at a relatively affordable price. Some of the CE-based limitations may be overcome with the application of MPS ... generate reliable STR profiles at a sensitivity level that competes with current widely used CE-based method."

Zeng XP, King JL, Stoljarova M, Warshauer DH, LaRue BL, Sajantila A, Patel J, Storts DR, Budowle B. 2015. High sensitivity multiplex short tandem repeat loci analyses with massively parallel sequencing. Forensic Science International: Genetics 16:38-47.

MPS also called NGS (Next Generation Sequencing.)

Single Nucleotide Polymorphisms (SNPs)

"Single nucleotide polymorphisms (SNPs) are the most frequently occurring genetic variation in the human genome, with the total number of SNPs reported in public SNP databases currently exceeding 9 million. SNPs are important markers in many studies that link sequence variations to phenotypic changes; such studies are expected to advance the understanding of human physiology and elucidate the molecular bases of diseases. For this reason, over the past several years a great deal of effort has been devoted to developing accurate, rapid, and cost-effective technologies for SNP analysis, yielding a large number of distinct approaches. "

Kim S. Misra A. 2007. SNP genotyping: technologies and biomedical applications. Annu Rev Biomed Eng. 2007;9:289-320.

Phase 3 1000Genomes Data

- 84.4 million variants
- 2504 individuals
- 26 populations

www.1000Genomes.org

Whole-genome Sequence Studies

One current study is the NHLBI Trans-Omics for Precision Medicine (TOPMed) project. www.nhlbiwgs.org
For data freeze 5 of this study:

Sequence analysis identified 410,323,831 genetic variants (381,343,078 SNVs and 28,980,753 indels), corresponding to an average of one variant per 7 bp throughout the reference genome. Among all variant alleles, 46.0% were observed once across all samples (i.e. singletons).

There is an average of 3.78 million variants in each studied genome. Among these, an average of 30,207 were novel (0.8%) and 3,510 were singletons (0.1%). Thus while there are vast numbers of rare variants in humans, only a few of these are present in each genome.

Currently over 1 billion variants found from 140,000 whole-genome sequences.

Probability Theory

We wish to attach probabilities to different kinds of events (or hypotheses or propositions):

- Event A: the next card is an Ace.
- Event R: it will rain tomorrow.
- Event C: the suspect left the crime stain.

Probabilities

Assign probabilities to events: Pr(A) or p_A or even p means "the probability that event A is true." All probabilities are conditional on some information I, so should write Pr($A|I$) for "the probability that A is true given that I is known."

No matter how probabilities are defined, they need to follow some mathematical laws in order to lead to consistent theories.

First Law of Probability

$$0 \leq \Pr(A|I) \leq 1$$

$$\Pr(A|A, I) = 1$$

If A is the event that a die shows an even face (2, 4, or 6), what is I? What is $\Pr(A|I)$?

Second Law of Probability

If A, B are mutually exclusive given I

$$\Pr(A \text{ or } B|I) \;=\; \Pr(A|I) + \Pr(B|I)$$

$$\text{so } \Pr(\bar{A}|I) \;=\; 1 - \Pr(A|I)$$

($\bar{A}$ means not-A).

If A is the event that a die shows an even face, and B is the event that the die shows a 1, verify the Second Law.

Third Law of Probability

$$\Pr(A \text{ and } B|I) \;=\; \Pr(A|B,I) \times \Pr(B|I)$$

If A is event that die shows an even face, and B is the event that the die shows a 1, verify the Third Law.

Will generally omit the I from now on.

Independent Events

Events A and B are independent if knowledge of one does not affect probability of the other:

$$\begin{aligned} \Pr(A|B) &= \Pr(A) \\ \Pr(B|A) &= \Pr(B) \end{aligned}$$

Therefore, for independent events

$$\Pr(A \text{ and } B) = \Pr(A)\Pr(B)$$

This may be written as

$$\Pr(AB) = \Pr(A)\Pr(B)$$

Law of Total Probability

Because B and $\bar{B}$ are mutually exclusive and exhaustive:

$$\Pr(A) \;=\; \Pr(A|B)\Pr(B) + \Pr(A|\bar{B})\Pr(\bar{B})$$

If A is the event that die shows a 3, B is the event that the die shows an even face, and $\bar{B}$ the event that the die shows an odd face, verify the Law of Total Probability.

Odds

The odds $O(A)$ of an event A are the probability of the event being true divided by the probability of the event not being true:

$$O(A) = \frac{\Pr(A)}{\Pr(\bar{A})}$$

This can be rearranged to give

$$\Pr(A) = \frac{O(A)}{1 + O(A)}$$

Odds of 10 to 1 are equivalent to a probability of 10/11.

Bayes' Theorem

The third law of probability can be used twice to reverse the order of conditioning:

$$\begin{aligned} \Pr(B|A) &= \frac{\Pr(B \text{ and } A)}{\Pr(A)} \\ &= \frac{\Pr(A|B)\Pr(B)}{\Pr(A)} \end{aligned}$$

Odds Form of Bayes' Theorem

From the third law of probability

$$\begin{aligned} \Pr(B|A) &= \Pr(A|B)\Pr(B)/\Pr(A) \\ \Pr(\bar{B}|A) &= \Pr(A|\bar{B})\Pr(\bar{B})/\Pr(A) \end{aligned}$$

Taking the ratio of these two equations:

$$\frac{\Pr(B|A)}{\Pr(\bar{B}|A)} = \frac{\Pr(A|B)}{\Pr(A|\bar{B})} \times \frac{\Pr(B)}{\Pr(\bar{B})}$$

Posterior odds = likelihood ratio $\times$ prior odds.

Birthday Problem

Forensic scientists in Arizona looked at the 65,493 profiles in the Arizona database and reported that two profiles matched at 9 loci out of 13. They reported a "match probability" for those 9 loci of 1 in 754 million. Are the numbers 65,493 and 754 million inconsistent?
(Troyer et al., 2001. Proc Promega 12th Int Symp Human Identification.)

To begin to answer this question suppose that every possible profile has the same profile probability P and that there are N profiles in a database (or in a population). The probability of at least one pair of matching profiles in the database is one minus the probability of no matches.

Birthday Problem

Choose profile 1. The probability that profile 2 does not match profile 1 is $(1-P)$. The probability that profile 3 does not match profiles 1 or 2 is $(1-2P)$, etc. So, the probability P_M of at least one matching pair is

$$P_M = 1-\{1(1-P)(1-2P)\cdots[1-(N-1)P]\}$$

$$\approx 1-\prod_{i=0}^{N-1} e^{-iP} \approx 1-e^{-N^2P/2}$$

If $P=1/365$ and $N=23$, then $P_M=0.51$. So, approximately, in a room of 23 people there is greater than a 50% probability that two people have the same birthday.

Birthday Problem

If $P = 1/(754 \text{ million})$ and $N = 65,493$, then $P_M = 0.98$ so it is highly probable there would be a match. There are other issues, having to do with the four non-matching loci, and the possible presence of relatives in the database.

If $P = 10^{-16}$ and $N = 300$ million, then $P_M =$ is essentially 1. It is almost certain that two people in the US have the same rare DNA profile.

Statistics

- Probability: For a given model, what do we expect to see?

- Statistics: For some given data, what can we say about the model?

- Example: A marker has an allele A with frequency p_A.
  - Probability question: If $p_A = 0.5$, and if alleles are independent, what is the probability of AA?
  - Statistics question: If a sample of 100 individuals has 23 AA's, 48 Aa's and 29 aa's, what is an estimate of p_A?

Transfer Evidence

Relevant Evidence

Rule 401 of the US Federal Rules of Evidence:

"Relevant evidence" means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.

Single Crime Scene Stain

Suppose a blood stain is found at a crime scene, and it must have come from the offender. A suspect is identified and provides a blood sample. The crime scene sample and the suspect have the same (DNA) "type."

The prosecution subsequently puts to the court the proposition (or hypothesis or explanation):

H_p: The suspect left the crime stain.

The symbol H_p is just to assist in the formal analysis. It need not be given in court.

Transfer Evidence Notation

G_S, G_C are the DNA types for suspect and crime sample. $G_S = G_C$. I is non-DNA evidence.

Before the DNA typing, probability of H_p is conditioned on I.

After the typing, probability of H_p is conditioned on G_S, G_C, I.

Updating Uncertainty

Method of updating uncertainty, or changing $\Pr(H_p|I)$ to $\Pr(H_p|G_S, G_C, I)$ uses Bayes' theorem:

$$\begin{aligned}\Pr(H_p|G_S, G_C, I) &= \frac{\Pr(H_p, G_S, G_C|I)}{\Pr(G_S, G_C|I)} \\ &= \frac{\Pr(G_S, G_C|H_p, I)\Pr(H_p|I)}{\Pr(G_S, G_C|I)}\end{aligned}$$

We can't evaluate $\Pr(G_S, G_C|I)$ without additional information, and we don't know $\Pr(H_p|I)$.

Can proceed by introducing alternative to H_p.

First Principle of Evidence Interpretation

To evaluate the uncertainty of a proposition, it is necessary to consider at least one alternative proposition.

The simplest alternative explanation for a single stain is:

H_d: Some other person left the crime stain.

Evett IW, Weir BS. 1998. "Interpreting DNA Evidence." Can be downloaded from
`www.biostat.washington.edu/ bsweir/InterpretingDNAEvidence`

Updating Odds

From the odds form of Bayes' theorem:

$$\frac{\Pr(H_p|G_S, G_C, I)}{\Pr(H_d|G_S, G_C, I)} = \frac{\Pr(G_S, G_C|H_p, I)}{\Pr(G_S, G_C|H_d, I)} \times \frac{\Pr(H_p|I)}{\Pr(H_d|I)}$$

i.e. Posterior odds = LR $\times$ Prior odds

where

$$\text{LR} = \frac{\Pr(G_S, G_C|H_p, I)}{\Pr(G_S, G_C|H_d, I)}$$

Questions for a Court to Consider

The trier of fact needs to address questions of the kind

- What is the probability that the prosecution proposition is true given the evidence,
  $\Pr(H_p|G_C, G_S, I)$?

- What is the probability that the defense proposition is true given the evidence,
  $\Pr(H_d|G_C, G_S, I)$?

Questions for Forensic Scientist to Consider

The forensic scientist must address different questions:

- What is the probability of the DNA evidence if the prosecution proposition is true,
  $\Pr(G_C, G_S | H_p, I)$?

- What is the probability of the DNA evidence if the defense proposition is true,
  $\Pr(G_C, G_S | H_d, I)$?

Important to articulate H_p, H_d. Also important not to confuse the difference between these two sets of questions.

Second Principle of Evidence Interpretation

Evidence interpretation is based on questions of the kind 'What is the probability of the evidence given the proposition.'

This question is answered for alternative explanations, and the ratio of the probabilities presented. It is not necessary to use the words "likelihood ratio". Use phrases such as:
'The probability that the crime scene DNA type is the same as the suspect's DNA type is one million times higher if the suspect left the crime sample than if someone else left the sample.'

Third Principle of Evidence Interpretation

Evidence interpretation is conditioned not only on the alternative propositions, but also on the framework of circumstances within which they are to be evaluated.

The circumstances may simply be the population to which the offender belongs so that probabilities can be calculated. Forensic scientists must be clear in court about the nature of the non-DNA evidence I, as it appeared to them when they made their assessment. If the court has a different view then the scientist must review the interpretation of the evidence.

Example

"In the analysis of the results I carried out I considered two alternatives: either that the blood samples originated from Pengelly or that the ... blood was from another individual. I find that the results I obtained were at least 12,450 times more likely to have occurred if the blood had originated from Pengelly than if it had originated from someone else."

Example

Question: "Can you express that in another way?"

Answer: "It could also be said that 1 in 12,450 people would have the same profile ... and that Pengelly was included in that number ... very strongly suggests the premise that the two blood stains examined came from Pengelly."

[Testimony of M. Lawton in R. v Pengelly 1 NZLR 545 (CA), quoted by
Robertson B, Vignaux GA, Berger CEH. 2016.*Interpreting Evidence (Second Edition)*. Wiley.

Likelihood Ratio

$$\text{LR} = \frac{\Pr(G_C, G_S | H_p, I)}{\Pr(G_C, G_S | H_d, I)}$$

Apply laws of probability to change this into

$$\text{LR} = \frac{\Pr(G_C | G_S, H_p, I) \Pr(G_S | H_p, I)}{\Pr(G_C | G_S, H_d, I) \Pr(G_S | H_d, I)}$$

Likelihood Ratio

Whether or not the suspect left the crime sample (i.e. whether or not H_p or H_d is true) provides no information about his genotype:

$$\Pr(G_S|H_p, I) = \Pr(G_S|H_d, I) \;=\; \Pr(G_S|I)$$

so that

$$\mathsf{LR} \;=\; \frac{\Pr(G_C|G_S, H_p, I)}{\Pr(G_C|G_S, H_d, I)}$$

Likelihood Ratio

$$\text{LR} = \frac{\Pr(G_C|G_S, H_p, I)}{\Pr(G_C|G_S, H_d, I)}$$

When $G_C = G_S$, and when they are for the same person (H_p is true):

$$\Pr(G_C|G_S, H_p, I) = 1$$

so the likelihood ratio becomes

$$\text{LR} = \frac{1}{\Pr(G_C|G_S, H_d, I)}$$

This is the reciprocal of the probability of the *match probability*, the probability of profile G_C, conditioned on having seen profile G_S in a different person (i.e. H_d) and on I.

Likelihood Ratio

$$\text{LR} \;=\; \frac{1}{\Pr(G_C|G_S, H_d, I)}$$

The next step depends on the circumstances I. If these say that knowledge of the suspect's type does not affect our uncertainty about the offender's type when they are different people (i.e. when H_d is true):

$$\Pr(G_C|G_S, H_d, I) \;=\; \Pr(G_C|H_d, I)$$

and then likelihood ratio becomes

$$\text{LR} \;=\; \frac{1}{\Pr(G_C|H_d, I)}$$

The LR is now the reciprocal of the *profile probability* of profile G_C.

Profile and Match Probabilities

Dropping mention of the other information I, the quantity $\Pr(G_C)$ is the probability that a person randomly chosen from a population will have profile type G_C. This profile probability usually very small and, although it is interesting, it is not the most relevant quantity.

Of relevance is the match probability, the probability of seeing the profile in a randomly chosen person after we have already seen that profile in a typed person (the suspect). The match probability is bigger than the profile probability. Having seen a profile once there is an increased chance we will see it again. This is the genetic essence of DNA evidence.

Likelihood Ratio

The estimated probability in the denominator of LR is determined on the basis of judgment, informed by I. Therefore the nature of I (as it appeared to the forensic scientist at the time of analysis) must be explained in court along with the value of LR. If the court has a different view of I, then the scientist will need to review the interpretation of the DNA evidence.

Random Samples

The circumstances I may define a population or racial group. The probability is estimated on the basis of a sample from that population.

When we talk about DNA types, by "selecting a person at random" we mean choosing him in such a way as to be as uncertain as possible about their DNA type.

Convenience Samples

The problem with a formal approach is that of defining the population: if we mean the population of a town, do we mean *every* person in the town at the time the crime was committed? Do we mean some particular area of the town? One sex? Some age range?

It seems satisfactory instead to use a convenience sample, i.e. a set of people from whom it is easy to collect biological material in order to determine their DNA profiles. These people are not a random sample of people, but they have not been selected on the basis of their DNA profiles.

Meaning of Likelihood Ratios

There is a personal element to interpreting DNA evidence, and there is no "right" value for the LR. (There is a right answer to the question of whether the suspect left the crime stain, but that is not for the forensic scientist to decide.)

The denominator for LR is conditioned on the stain coming from an unknown person, and "unknown" may be hard to define. A relative? Someone in that town? Someone in the same ethnic group? (What is an ethnic group?)

Meaning of Frequencies

What is meant by "the frequency of the matching profile is 1 in 57 billion"?

It is an estimated probability, obtained by multiplying together the allele frequencies, and refers to an infinite random mating population. It has nothing to do with the size of the world's population.

The question is really whether we would see the profile in two people, given that we have already seen it in one person. This conditional probability may be very low, but has nothing to do with the size of the population.