Meeting 17
Forensic applications, part II

DNA evidence

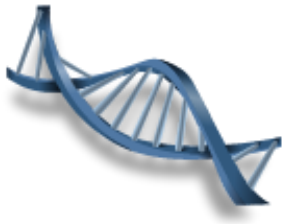
Was “discovered” as a useful tool in forensic investigation of bio-traces in the 1980:s (*Sir Alec Jeffreys*)

Crime forensic purposes: comparisons between DNA from a recovered trace with unknown origin and DNA from a suspect/from another recovered trace
Non-crime purposes: paternity disputes, disaster victim identification (DVI), kinship investigations

Has undergone an enormous development:

- Started with a few bio-markers that in clear cases could lead to likelihood ratios (crime cases) of magnitude 10 000 (very large at that time)
- Today, most “kits” in use can in clear cases lead to likelihood ratios of magnitude 10^{20}

The DNA Double Helix



Consists of so-called nucleobases: *adenine* (A), *thymine* (T), *cytosine* (C) and *guanine* (G) always in the pairs A-T, C-G.

Humans: genetic information comes in 23 *chromosome* pairs, where each chromosome is a double helix - referred to as the *genome* of a human being.

- Along one chromosome sequences of nucleobase pairs define so-called *markers* or *loci* (one locus).
- Can consist of one up to hundreds of nucleobases.
- A corresponding sequence of nucleobase pairs on the other chromosome but not necessarily of the same length.

--- CGATCGATCGATCGATCGATCGATCGATCGATCGATCGAT --- *10 repetitions of "CGAT"*
GCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTA
--- AATAATAATAATAATAAT --- *6 repetitions of "AAT"*
TTATTATTATTATTATTA

The two sequences are called *alleles* and together they form the *genotype* of that marker.

One of the alleles is inherited from the mother and the other from the father, but for most markers it is not possible to know which is what.

Each chromosome pair hosts a great variety of genotypes (or genes). One of the chromosome pairs defines the sex. The others are referred to as autosomes (autosomal DNA).

Most of the genome (more than 90%) is today not verified to have any other function than possibly "assisting" (*non-coding DNA*).

Forensic DNA analysis is mainly concerned with the non-coding part of the DNA – shows much more variation (polymorphism) than the coding part and therefore constitute "genetic fingerprints".

Several techniques are used to read off the information contained in the genome:

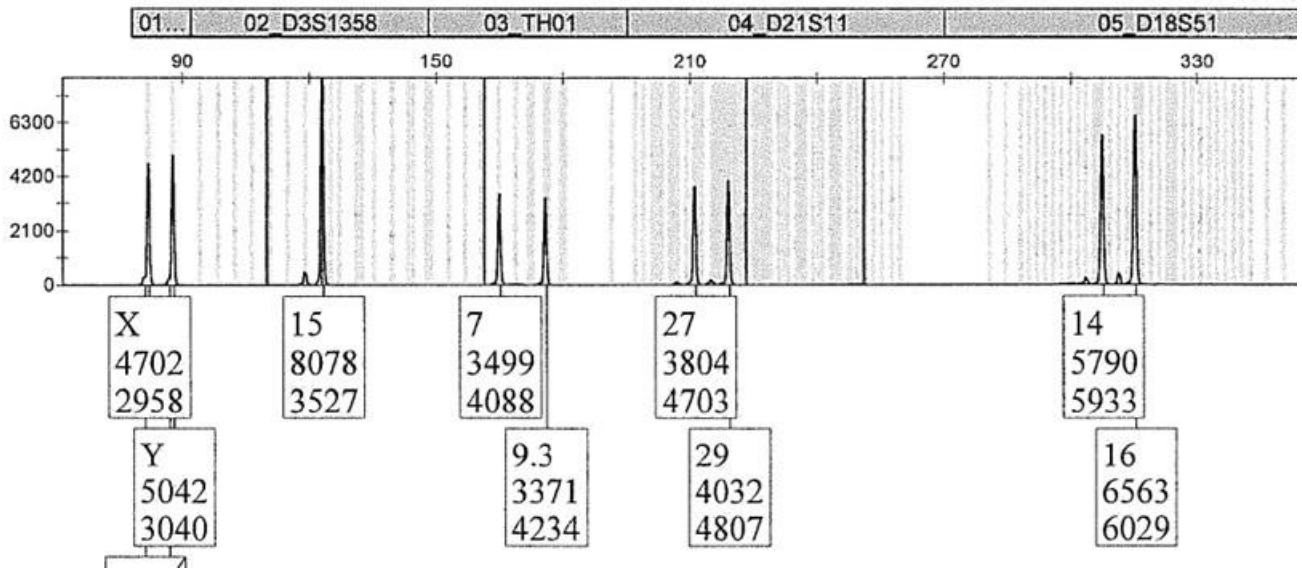
- PCR (investigates so-called short tandem repeats (STR) or single nucleotide polymorphisms (SNP) or other polymorphisms) by multiplying extracted molecules
- Sequencing (todays technique for microorganisms, possibly tomorrow's for humans) strives at projecting the whole genome

PCR for STR (still a consensus technique among European forensic science institutes (ENFSI) for crime investigation)

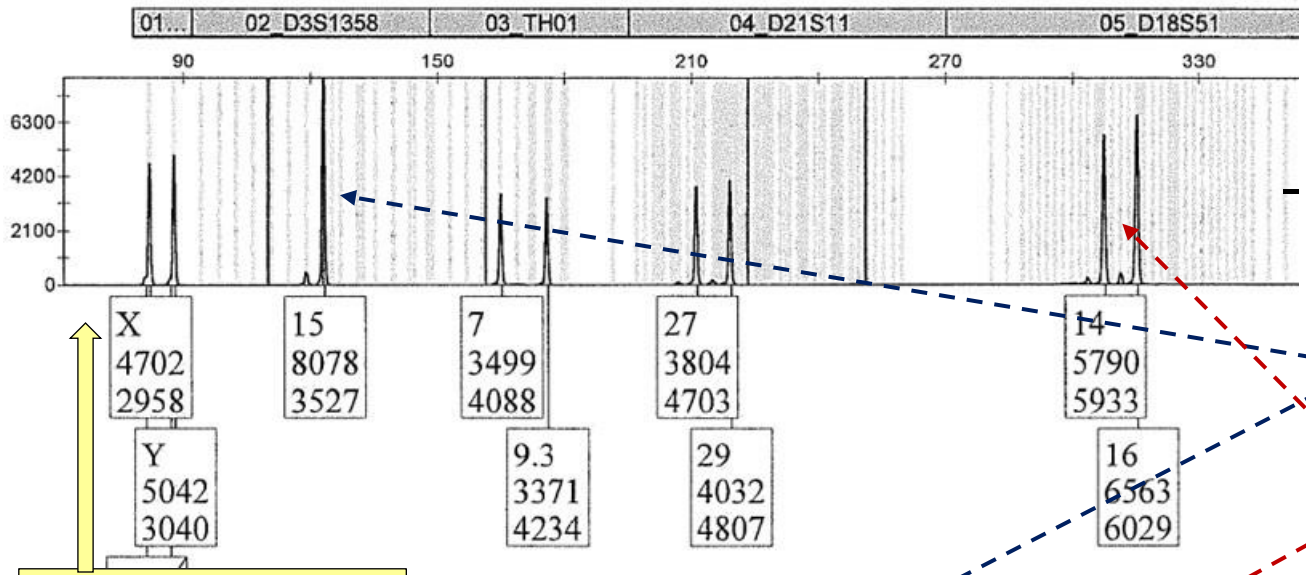
A number of so-called *kits* are available. At NFC, Linköping the kit ESX-16 is used:

16 markers are investigated, 15 autosomal and one sex chromosome marker (used to identify the sex of the human being the source of the DNA).

By *Capillary electrophoresis* the alleles of a marker can be detected as peaks in a so-called *electropherogram*



An example from typing (identifying the genotypes in each autosomal marker):



Sex chromosome marker

Homozygote genotype (two identical alleles- one peak)

Heterozygote genotype (two different alleles- two peaks)

Marker	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16
Allele 1	15	7	27	14	16	17.3	18	11	15	15	12	21	11	17.3	15
Allele 2	15	9.3	29	16	16	18.3	25	12	16	19	13	22.2	11	19	16

The allele codes are simply number of repeats of a certain sequence.

A complete set of 15 genotypes is referred to as a *DNA profile*.

DNA comparisons

In a criminal case there is a recovered trace from a crime scene:

- blood stain
- saliva stain
- semen stain
- hairs (with roots) or body tissues
- vaginal samples
- …

When there is a suspect, ordinary samples can be taken (today buccal swabs are standard, previously blood samples were taken) to recover DNA.

Typed DNA profiles are compared:

$\Rightarrow$ match or no match

Marker	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16
Allele 1	15	7	27	14	16	17.3	18	11	15	15	12	21	11	17.3	15
Allele 2	15	9.3	29	16	16	18.3	25	12	16	19	13	22.2	11	19	16

Match

Marker	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16
Allele 1	15	7	27	14	16	17.3	18	11	15	15	12	21	11	17.3	15
Allele 2	15	9.3	29	16	16	18.3	25	12	16	19	13	22.2	11	19	16

How to evaluate a match?

The rarity of the matching profile must be assessed.

Requires *population genetics models*.

Marker	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16
Allele 1	15	7	27	14	16	17.3	18	11	15	15	12	21	11	17.3	15
Allele 2	15	9.3	29	16	16	18.3	25	12	16	19	13	22.2	11	19	16

How rare is a particular genotype in a particular marker?

An allele (coded as the number of repetitions of a nucleobase sequence) has a marker-specific relative frequency in the population of interest.

For instance, in the profile above, the relative frequency of allele 15 in marker 02 is different from the relative frequency of allele 15 in marker 11.

For two alleles, *A* and *B* let f_A and f_B denote their relative frequencies in a particular marker.

Assuming random mating (so-called *Hardy-Weinberg equilibrium*) the *genotype* frequencies of genotypes (*A*, *A*) (homozygote) and (*A*, *B*) (heterozygote) can be calculated as

$$f_{A,A} = f_A^2 \qquad f_{A,B} = 2 \cdot f_A \cdot f_B$$

Many national populations almost satisfies Hardy-Weinberg (HW) equilibrium (at least such an hypothesis is hard to reject on basis of collected data)

Adjustment (Balding & Nichols, 1994) to take into account so-called *subpopulation effects* (meaning that mating is not random but alleles are structurally inherited along "lines" in the population):

$$f_{A,A} = \frac{(2 \cdot F_{ST} + (1 - F_{ST}) \cdot f_A) \cdot (3 \cdot F_{ST} + (1 - F_{ST}) \cdot f_A)}{(1 + F_{ST}) \cdot (1 + 2 \cdot F_{ST})}$$

$$f_{A,B} = \frac{2 \cdot (F_{ST} + (1 - F_{ST}) \cdot f_A) \cdot (F_{ST} + (1 - F_{ST}) \cdot f_B)}{(1 + F_{ST}) \cdot (1 + 2 \cdot F_{ST})}$$

where F_{ST} is the *co-ancestry coefficient* measuring the subpopulation effects (to what extent the mating is non-random).

In Sweden F_{ST} is close to 0.01.

Example

A study was made in a population where the coancestry coefficient is estimated to be around 3 % . The following results were obtained for marker TH01:

Allele	Relative frequency
6	0.295
7	0.147
8	0.184
9	0.232
9.3	0.026
10	0.116

Relative frequencies for the genotypes (7,8) and (8,8):

$$f_{7,8} = 2 \cdot 0.147 \cdot 0.184 \approx 0.054$$
$$f_{8,8} = 0.184^2 \approx 0.034$$

Assuming Hardy-Weinberg equilibrium

$$f_{7,8} = \frac{2 \cdot (0.03 + (1 - 0.03) \cdot 0.147) \cdot (0.03 + (1 - 0.03) \cdot 0.184)}{(1 + 0.03) \cdot (1 + 2 \cdot 0.03)} \approx 0.066$$
$$f_{8,8} = \frac{(2 \cdot 0.03 + (1 - 0.03) \cdot 0.184) \cdot (3 \cdot 0.03 + (1 - 0.03) \cdot 0.184)}{(1 + 0.03) \cdot (1 + 2 \cdot 0.03)} \approx 0.059$$

Assuming substructures

Marker	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16
Allele 1	15	7	27	14	16	17.3	18	11	15	15	12	21	11	17.3	15
Allele 2	15	9.3	29	16	16	18.3	25	12	16	19	13	22.2	11	19	16

How rare is the entire profile?

Linkage equilibrium:

Genotypes at different markers become less statistical dependent with the distance them between in the double helix – due to so-called *recombinations* at the *meiosis* phase upon conception.

Independence is empirically proven for markers situated on different chromosomes.

Markers chosen in forensic kits for typing short tandem repeats (STR) markers satisfy the assumption of (approximate) independence and are said to be in *linkage equilibrium* (LE).

A small set of single nucleotide polymorphisms (SNPs) may be in LE, but the benefit of using SNPs is that thousands can be analysed in one run. These do not satisfy LE.

With linkage equilibrium the relative frequency of a DNA profile can be calculated from the genotype relative frequencies:

$$f_{\text{profile}} = f_{A_1,B_1} \cdot f_{A_2,B_2} \cdot \ldots \cdot f_{A_L,B_L}$$

L = number of markers in the kit

(A_i , B_i) is the genotype of locus i ($A_i \neq B_i$ or $A_i = B_i$)

Linkage equilibrium implies that a profile relative frequency at a very fast rate goes towards zero when the number of markers used increases.

With a full 15-marker profile typical relative frequencies are of magnitude less than 10^{-14}.

Are these actually to be considered as relative frequencies?

Example

Consider the previously shown profile:

Locus	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Allele 1	15	7	27	14	16	17.3	18	11	15	15	12	21	11	17.3	15
Allele 2	15	9.3	29	16	16	18.3	25	12	16	19	13	22.2	11	19	16
$f_{A,B}$	0.085	0.140	0.016	0.051	0.020	0.028	0.026	0.192	0.254	0.017	0.099	0.011	0.152	0.008	0.026

The genotype relative frequencies have been calculated using allele relative frequencies obtained from a database from an average modern Swedish population and assuming subpopulation effects with $F_{ST} = 0.01$

The relative frequency of this profile is calculated to $4 \cdot 10^{-21}$

With a population of almost 10 million inhabitants this cannot be a profile belonging to that population if the value is to be taken for an observed relative frequency.

Actually, one estimates that just above $100 \cdot 10^9$ human beings have ever existed on earth. Even in this population the value cannot be an observed relative frequency.

The evaluation model used in a criminal case

Assume there is a stain left at a crime scene and there is a male suspect assumed to have been involved with the criminal activity. DNA is recovered from the stain and from a buccal swab of the suspect.

A full profile is obtained from the stain and (as expected) a full profile is obtained from the suspect.

The two profiles match in every marker.

Assume it is the profile previously discussed

Locus	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Allele 1	15	7	27	14	16	17.3	18	11	15	15	12	21	11	17.3	15
Allele 2	15	9.3	29	16	16	18.3	25	12	16	19	13	22.2	11	19	16

Hypotheses:

$\boldsymbol{H_m}$: "The suspect is the donor of the stain"

$\boldsymbol{H_a}$: "Someone else is the donor of the stain"

Evidence:

$\boldsymbol{E}$: "A match in DNA profile (matches in all 15 autosomal markers of an ESX16-profile and match in the sex-defining marker) "

Value of evidence (likelihood ratio):

$$V = \frac{P(\boldsymbol{E}|\boldsymbol{H_m})}{P(\boldsymbol{E}|\boldsymbol{H_a})}$$

How to find (estimates of) the numerator and the denominator?

$P(E|H_m)$

H_m : "The suspect is the donor of the stain"
H_a : "Someone else is the donor of the stain"

$$V = \frac{P(E|H_m)}{P(E|H_a)}$$

If the suspect actually left the stain we expect to obtain matches in all markers.

There is no genetic reason for any variation (besides mutations, but such interventions can usually be controlled).

There could be variation due to deficiencies with the equipment or with the operators (reading off the wrong values).

However it is generally non-debatable to set this probability to 1.

H_m : "The suspect is the donor of the stain"
H_a : "Someone else is the donor of the stain"

$$V = \frac{P(E|H_m)}{P(E|H_a)}$$

$P(E|H_a)$

If someone else left the stain, what is the probability of obtaining the match?

Sometimes things become clearer if we formulate the evidence in terms of the variables

E_c : DNA profile of crime stain

E_s : DNA profile of suspect

The evidence can then be written

$$E = (E_c = \Gamma, E_s = \Gamma)$$

where Γ is the profile obtained both with the stain and the suspect.

H_m : "The suspect is the donor of the stain"
H_a : "Someone else is the donor of the stain"

E_c : DNA profile of crime stain
E_s : DNA profile of suspect

$$\Rightarrow$$

$$V = \frac{P(E|H_m)}{P(E|H_a)} = \frac{P(E_c = \Gamma, E_s = \Gamma|H_m)}{P(E_c = \Gamma, E_s = \Gamma|H_a)}$$

$$= \frac{P(E_c = \Gamma|E_s = \Gamma, H_m) \cdot P(E_s = \Gamma|H_m)}{P(E_c = \Gamma|E_s = \Gamma, H_a) \cdot P(E_s = \Gamma|H_a)} =$$

$$= \left\langle \begin{matrix} \text{Suspect's profile (isolated) does} \\ \text{not depend on the hypotheses} \end{matrix} \right\rangle = \frac{P(E_c = \Gamma|E_s = \Gamma, H_m) \cdot P(E_s = \Gamma)}{P(E_c = \Gamma|E_s = \Gamma, H_a) \cdot P(E_s = \Gamma)} =$$

$$\left\langle \begin{matrix} \text{If someone else left the stain, the suspect's profile} \\ \text{cannot have any impact on the profile of the stain} \end{matrix} \right\rangle = \frac{P(E_c = \Gamma|E_s = \Gamma, H_m)}{P(E_c = \Gamma|H_a)}$$

Now, the denominator is the probability of obtaining the profile Γ of the stain if the stain was left by someone else than the suspect.

This probability should account for the rarity of this profile in the population of potential donors of the stain.

$\boldsymbol{H_m}$: "The suspect is the donor of the stain"	$\boldsymbol{E_c}$: DNA profile of crime stain
$\boldsymbol{H_a}$: "Someone else is the donor of the stain"	$\boldsymbol{E_s}$: DNA profile of suspect

$$V = \frac{P(\boldsymbol{E_c} = \Gamma | \boldsymbol{E_s} = \Gamma, \boldsymbol{H_m})}{P(\boldsymbol{E_c} = \Gamma | \boldsymbol{H_a})}$$

$$P(\boldsymbol{E_c} = \Gamma | \boldsymbol{H_a})$$

Is this probability higher for certain groups of the population of potential donors (i.e. is the population stratified with respect to the occurrence of this profile)?

Note! Since the stain is from a male (due to the match) the population only consists of males.

What about

- an identical twin of the suspect?
- a full brother of the suspect?
- the suspect's father?
- a son of the suspect?
- a half-brother of the suspect?
- the grand-fathers of the suspect?
- an uncle or a male cousin of the suspect?

If stratification should be taken into account we need to use a so-called full Bayesian approach and compute the value of evidence as the Bayes factor

$$B = \frac{P\left(\boldsymbol{E_c} = \Gamma \middle| \boldsymbol{E_s} = \Gamma, \boldsymbol{H_m}\right)}{\sum P(\boldsymbol{E_c} = \Gamma | \text{Individual } i \text{ is the donor}, \boldsymbol{H_a}) \cdot P\left(\text{Individual } i \text{ is the donor} \middle| \boldsymbol{H_a}\right)}$$

…where the sum is over all individuals in the population of possible donors except for the suspect.

However, this will need knowledge about the prior probabilities

$$P\left(\text{Individual } i \text{ is the donor} \middle| \boldsymbol{H_a}\right), \qquad i = 1,2, \ldots$$

of which the forensic scientist has no opinion (and should not have).

Hence, the evidentiary strength cannot be assessed without prior opinions about which persons could have been involved.

To be able to report measures of evidentiary strength we need to formulate different alternative hypotheses.

First choice: $\boldsymbol{H_a}$: "Someone else, not closely related to the suspect, left the stain"

$$V = \frac{P\left(\boldsymbol{E_c} = \Gamma \middle| \boldsymbol{E_s} = \Gamma, \boldsymbol{H_m}\right)}{P\left(\boldsymbol{E_c} = \Gamma \middle| \boldsymbol{H_a}\right)}$$

The denominator of V can now be estimated from a random sample of individuals from the population to which the donor is assumed to belong.

Such a random sample is (today) a kind of panel, i.e. a number of persons from a general population (covering the population of potential donors with negligible effects of over coverage)

$\Rightarrow$ DNA population database

Hence, $P(\boldsymbol{E_c} = \Gamma | \boldsymbol{H_a})$ is estimated by calculating the relative frequency of this profile using the database.

Less problematic that this relative frequency is not possible to physically obtain in the population, it is used to estimate a probability through a *model* of the population.

For the current profile we previously obtained a calculated relative frequency of $4 \cdot 10^{-21}$.

$$V = \frac{P(\boldsymbol{E_c} = \Gamma | \boldsymbol{E_s} = \Gamma, \boldsymbol{H_m})}{P(\boldsymbol{E_c} = \Gamma | \boldsymbol{H_a})} = \frac{1}{4 \cdot 10^{-21}} = 2.5 \cdot 10^{20}$$

The match is thus $2.5 \cdot 10^{20}$ times more probable to obtain if the suspect is the donor than if someone else, not closely related to the suspect, is the donor

Was it him?

Another alternative hypothesis may be

$\boldsymbol{H_{a,2}}$: "The stain was left by a full brother of the suspect"

We then need more population genetics to calculate the probability

$$P\left(\boldsymbol{E_c} = \Gamma \middle| \boldsymbol{H_{a,2}}\right)$$

For the current profile an estimate of this probability becomes $1.82 \cdot 10^{-7}$

Hence, the value of evidence is

$$V^{(2)} = \frac{P\left(\boldsymbol{E_c} = \Gamma \middle| \boldsymbol{E_s} = \Gamma, \boldsymbol{H_m}\right)}{P\left(\boldsymbol{E_c} = \Gamma \middle| \boldsymbol{H_{a,2}}\right)} = \frac{1}{1.82 \cdot 10^{-07}} = 5.5 \cdot 10^6$$

The match is thus 5.5 million times more probable to obtain if the suspect is the donor than if a full brother of the suspect is the donor.

Besides identical twins, full siblings of the same sex are the closest related individuals.

Changing the alternative hypothesis to something like

“The donor of the stain is a father or a son of the suspect”

will also render a higher relative frequency (however lower than with a full brother) – and as a consequence a lower value of evidence (against the suspect) than with no close relatives in the alternative hypothesis.

It has become more and more common for a suspect to “blame the brother”. The most obvious way to handle this situation is to swab the brother.

- A mismatch directly excludes the brother.
- However, with a (utterly unexpected) match the two brothers cannot be separated by the current DNA evidence

Challenges with DNA evidence

With today's technique very small amounts of DNA can be recovered and typed (with PCR: LCN-analysis (*L*ow *C*opy *N*umber))

Small amounts of DNA is typical for so-called touch-DNA (contact with skin)

Since several persons may have been in contact with a surface of interest (someone's garments, doorhandle, table, …) it is common to observe DNA from more than one person in a sample – so-called *DNA mixtures*.

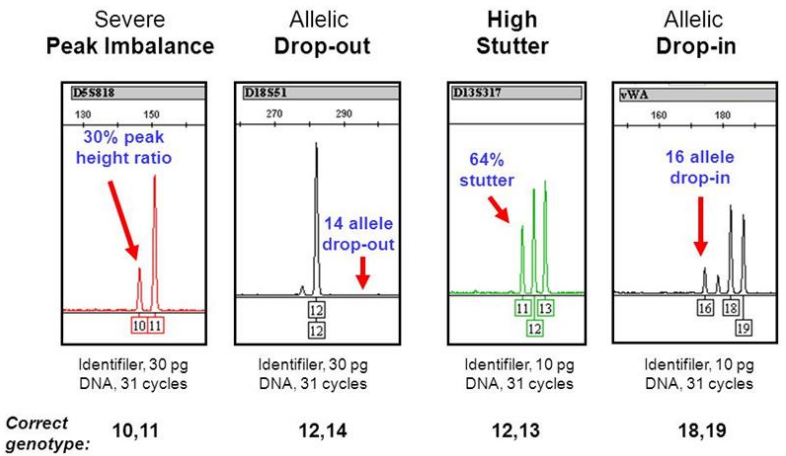
This is also often the case in sex crimes were body fluid samples contain DNA both from both the perpetrator and the victim (but sometimes also from a third or fourth person).

The hypothesis would comprise more than one person, e.g.

$\boldsymbol{H_m}$: The DNA originates from the victim and the suspect
$\boldsymbol{H_a}$: The DNA originates from the victim and an unknown person

When (very) small amounts of DNA are analysed there is appreciable risks that…

- alleles in one or several markers are not detected at all in the electropherogram (so-called *drop-out* alleles)
- peaks in a marker (if more than one) has substantially different heights – is it a heterozygote marker or alleles from more than one person?
- artefacts in forms of extra peaks (so-called *stutters*) aside the true peaks (a multiplying effect)
- residues from previous analyses – despite cleaning – may cause extra peaks in a marker (so-called *drop-in alleles*).



Several (commercial and non-commercial) software have been developed to handle these problems, especially for samples with DNA from more than one person (e.g. STRmix™, TrueAllele®, *EuroForMix*, *DNAxs*…)

Another challenge is when there is no longer a dispute on ***who's*** DNA it is.

Infancy of DNA evidence evaluation: Often sufficient to confront the suspect like "We've got your DNA!"

In course of time, culprits have learnt that there are loopholes in the interpretation of DNA evidence.

- *Blaming on a close relative* – will be less efficient as the amount of DNA analysed is increasing (more STR-markers, sequencing).

- *Questioning **how** the DNA was deposited* – Claiming a secondarily or even tertiarily DNA transfer from an innocent contact.

  Particularly common in sex crimes where the suspect denies having sexually assaulted the victim, but claims they had only social contact (e.g. drinking and/or dancing together).

In such cases the hypothesis are no longer about the source of the recovered DNA (since there is no dispute on that).

They must address *activities* (be formulated at *activity level*), e.g.

H_m : The suspect had sexual intercourse with the victim
H_a : The suspect and the victim had social contact during the party

When the evidence material is recovered body fluid (for instance DNA from the victim is found in the suspect's underwear), the amount of DNA recovered is important.

The probability of recovering substantial amounts of DNA cannot be well explained by the hypothesis H_a (e.g. pointing towards secondary transfer of saliva), but very well by hypothesis H_m (vaginal secretion).

When the amounts recovered are small it is not possible to discriminate between H_m and H_a with sufficient confidence. Sometimes it is possible to find out which type of cells (saliva secretion or vaginal secretion) the sample consisted of, but a potential secondary transfer cannot be easily rejected.

Note that the immensely high likelihood ratios obtained with the source attribution of the DNA are completely worthless in this dispute.

The probably hardest challenge is when there is no longer a dispute on the source or how the DNA was deposited, just ***when*** it was.

There is (yet) no method to determine the age of DNA.

A real case example

On 18 February 2015 a robbery against a money transport was committed in one of the southern districts, BC, of Stockholm.

In a couple of cars parked nearby where the robbery took place the Police found garments suspected to have been used by the perpetrators.

One suspect, S, was caught and arrested, and from a comparison of DNA from S with DNA recovered from a stain on one of the garments, a construction worker jacket, a match was obtained between the DNA from S and the major part of a DNA mixture obtained from the stain.

The suspect S has an explanation to “his DNA being on the jacket”: He had occasionally worn jackets of this type during “his entire life” until a month before the robbery (i.e. in January 2015)

The forensic problem: The suspect (S) did wear the jacket J either when taking part in a robbery against a money transport at BC on 18 February 2015, or for innocent purposes at least once in a period up to two years back in time.

Hypothesies (at activity level):

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.

$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.
$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

Findings

E_1 : The jacket J was found in an abandoned car (Volvo) near the crime scene
E_2 : On the jacket J one stain of secretion was recovered [backside of right collar] but no other biotraces
E_3 : The stain showed a mixture of DNA from two persons DNA (major and minor parts)
E_4 : The major part of the mixture showed a profile identical with the profile of S (Γ_S)

$$V = \frac{P\left(E_1, E_2, E_3, E_4 \middle| \boldsymbol{H_m}\right)}{P\left(E_1, E_2, E_3, E_4 \middle| \boldsymbol{H_a}\right)}$$

$$= \frac{P\left(E_4 \middle| E_1, E_2, E_3, \boldsymbol{H_m}\right) \times P\left(E_3 \middle| E_1, E_2, \boldsymbol{H_m}\right) \times P\left(E_2 \middle| E_1, \boldsymbol{H_m}\right) \times P\left(E_1 \middle| \boldsymbol{H_m}\right)}{P\left(E_4 \middle| E_1, E_2, E_3, \boldsymbol{H_a}\right) \times P\left(E_3 \middle| E_1, E_2, \boldsymbol{H_a}\right) \times P\left(E_2 \middle| E_1, \boldsymbol{H_a}\right) \times P\left(E_1 \middle| \boldsymbol{H_a}\right)}$$

$$= \frac{p_4 \times p_3 \times p_2 \times p_1}{q_4 \times q_3 \times q_2 \times q_1}$$

Numerator of V

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.
$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

$$V = \frac{P(E_4|E_1,E_2,E_3,\boldsymbol{H_m}) \times P(E_3|E_1,E_2,\boldsymbol{H_m}) \times P(E_2|E_1,\boldsymbol{H_m}) \times P(E_1|\boldsymbol{H_m})}{P(E_4|E_1,E_2,E_3,\boldsymbol{H_a}) \times P(E_3|E_1,E_2,\boldsymbol{H_a}) \times P(E_2|E_1,\boldsymbol{H_a}) \times P(E_1|\boldsymbol{H_a})} = \frac{p_4 \times p_3 \times p_2 \times p_1}{q_4 \times q_3 \times q_2 \times q_1}$$

$$p_4 = P(E_4|E_1,E_2,E_3,\boldsymbol{H_m})$$

E_4 : The major part of the mixture showed a profile identical with the profile of S (Γ_S)

If $\boldsymbol{H_m}$ is true then E_4 may have been observed if

- S deposited secretion on J when taking part in the robbery and no DNA from S was on J before ($\boldsymbol{C_1}$)
- S did not deposit secretion on J when taking part in the robbery, but DNA from S was on J before ($\boldsymbol{C_2}$)
- S did not deposit secretion on J when taking part in the robbery, DNA from S was ***not*** on J before, but DNA from someone else matching the DNA from S was ($\boldsymbol{C_3}$)

Other combinations of deposition of DNA from S when taking part in the robbery and possible presence of DNA from S or someone else before are discarded since their contribution will be negligibly small.

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.
$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

$$V = \frac{P(E_4|E_1,E_2,E_3,\boldsymbol{H_m}) \times P(E_3|E_1,E_2,\boldsymbol{H_m}) \times P(E_2|E_1,\boldsymbol{H_m}) \times P(E_1|\boldsymbol{H_m})}{P(E_4|E_1,E_2,E_3,\boldsymbol{H_a}) \times P(E_3|E_1,E_2,\boldsymbol{H_a}) \times P(E_2|E_1,\boldsymbol{H_a}) \times P(E_1|\boldsymbol{H_a})} = \frac{p_4 \times p_3 \times p_2 \times p_1}{q_4 \times q_3 \times q_2 \times q_1}$$

E_4 : The major part of the mixture showed a profile identical with the profile of S (Γ_S)

C_1: S deposited at robbery, no DNA from S before
C_2: S did not deposit at robbery, DNA from S before
C_3: S did never deposit, matching DNA from someone else before

$$\Rightarrow p_4 = P(E_4|C_1,E_1,E_2,E_3,\boldsymbol{H_m}) \times P(C_1|E_1,E_2,E_3,\boldsymbol{H_m}) +$$
$$P(E_4|C_2,E_1,E_2,E_3,\boldsymbol{H_m}) \times P(C_2|E_1,E_2,E_3,\boldsymbol{H_m}) +$$
$$P(E_4|C_3,E_1,E_2,E_3,\boldsymbol{H_m}) \times P(C_3|E_1,E_2,E_3,\boldsymbol{H_m})$$

$$= 1 \times P(C_1|E_1,E_2,E_3,\boldsymbol{H_m}) + 1 \times P(C_2|E_1,E_2,E_3,\boldsymbol{H_m}) + \gamma_{\Gamma_S} \times P(C_3|E_1,E_2,E_3,\boldsymbol{H_m})$$

where γ_{Γ_S} is the random match probability of the DNA profile Γ_S. However, compared to $P(C_1|E_1,E_2,E_3,\boldsymbol{H_m})$ and $P(C_2|E_1,E_2,E_3,\boldsymbol{H_m})$, γ_{Γ_S} is negligible.

$$\Rightarrow p_4 \approx P(C_1|E_1,E_2,E_3,\boldsymbol{H_m}) + P(C_2|E_1,E_2,E_3,\boldsymbol{H_m}) = p_{41} + p_{42}$$

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.
$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

$$V = \frac{P(E_4|E_1,E_2,E_3,\boldsymbol{H_m}) \times P(E_3|E_1,E_2,\boldsymbol{H_m}) \times P(E_2|E_1,\boldsymbol{H_m}) \times P(E_1|\boldsymbol{H_m})}{P(E_4|E_1,E_2,E_3,\boldsymbol{H_a}) \times P(E_3|E_1,E_2,\boldsymbol{H_a}) \times P(E_2|E_1,\boldsymbol{H_a}) \times P(E_1|\boldsymbol{H_a})} = \frac{p_4 \times p_3 \times p_2 \times p_1}{q_4 \times q_3 \times q_2 \times q_1}$$

E_4 : The major part of the mixture showed a profile identical with the profile of S (Γ_S)

C_1: S deposited at robbery, no DNA from S before
C_2: S did not deposit at robbery, DNA from S before

$$p_4 = P(E_4|E_1,E_2,E_3,\boldsymbol{H_m}) \approx P(C_1|E_1,E_2,E_3,\boldsymbol{H_m}) + P(C_2|E_1,E_2,E_3,\boldsymbol{H_m}) = p_{41} + p_{42}$$

E_1, E_2 and E_3 are not irrelevant for C_1 and C_2 but compared to the relevance of $\boldsymbol{H_m}$ their relevancies are negligible.

$$\Rightarrow p_{41} \approx P(C_1|\boldsymbol{H_m}) \text{ and}$$

$$p_{42} \approx P(C_2|\boldsymbol{H_m})$$

E_1 : The jacket J was found in an abandoned car (Volvo) near the crime scene
E_2 : On the jacket J one stain of secretion was recovered but no other biotraces.
E_3 : The stain showed a mixture of DNA from two persons DNA

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.
$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

$$V = \frac{P(E_4|E_1, E_2, E_3, \boldsymbol{H_m}) \times P(E_3|E_1, E_2, \boldsymbol{H_m}) \times P(E_2|E_1, \boldsymbol{H_m}) \times P(E_1|\boldsymbol{H_m})}{P(E_4|E_1, E_2, E_3, \boldsymbol{H_a}) \times P(E_3|E_1, E_2, \boldsymbol{H_a}) \times P(E_2|E_1, \boldsymbol{H_a}) \times P(E_1|\boldsymbol{H_a})} = \frac{p_4 \times p_3 \times p_2 \times p_1}{q_4 \times q_3 \times q_2 \times q_1}$$

$\boldsymbol{C_1}$: S deposited at robbery, no DNA from S before
$\boldsymbol{C_2}$: S did not deposit at robbery, DNA from S before

Let

- $k \cdot d$ = probability to deposit secretion (here assumed to be saliva) on a jacket worn when taking part in an activity that affects k. When activity is a theft/robbery, then $k = 1 \Rightarrow 0 \leq d \leq 1$
- $m \cdot r$ = probability that DNA deposited on a jacket persists a period of time that affects $m \leq 1$ where $m \approx 1$ when time period is a few hours or less $\Rightarrow 0 \leq r \leq 1$

$$\Rightarrow p_{41} \approx P(C_1|\boldsymbol{H_m}) = d \cdot r \text{ and}$$

$$p_{42} \approx P(C_2|\boldsymbol{H_m}) = (1-d) \cdot k \cdot d \cdot m \cdot r$$

not at robbery
at latest in January 2015
persisted until 18 February 2015

$$\Rightarrow p_4 \approx d \cdot r + (1-d) \cdot k \cdot d \cdot m \cdot r$$
$$= d \cdot r \cdot (1 + (1-d) \cdot k \cdot m)$$

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.
$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

$$V = \frac{P(E_4|E_1, E_2, E_3, \boldsymbol{H_m}) \times P(E_3|E_1, E_2, \boldsymbol{H_m}) \times P(E_2|E_1, \boldsymbol{H_m}) \times P(E_1|\boldsymbol{H_m})}{P(E_4|E_1, E_2, E_3, \boldsymbol{H_a}) \times P(E_3|E_1, E_2, \boldsymbol{H_a}) \times P(E_2|E_1, \boldsymbol{H_a}) \times P(E_1|\boldsymbol{H_a})} = \frac{p_4 \times p_3 \times p_2 \times p_1}{q_4 \times q_3 \times q_2 \times q_1}$$

$$p_3 = P(E_3|E_1, E_2, \boldsymbol{H_m})$$

E_3 : The stain showed a mixture of DNA from two persons DNA (major and minor parts)

Neither E_1 nor $\boldsymbol{H_m}$ is assumed relevant for E_3

E_1 : The jacket J was found in an abandoned car (Volvo) near the crime scene
E_2 : On the jacket J one stain of secretion was recovered but no other biotraces.

$$\Rightarrow p_3 = P(E_3|E_2)$$

which corresponds to how expected a DNA mixture is upon analysing a (single) secretion stain recovered from the backside of the right collar of a jacket.

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.
$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

$$V = \frac{P(E_4|E_1, E_2, E_3, \boldsymbol{H_m}) \times P(E_3|E_1, E_2, \boldsymbol{H_m}) \times P(E_2|E_1, \boldsymbol{H_m}) \times P(E_1|\boldsymbol{H_m})}{P(E_4|E_1, E_2, E_3, \boldsymbol{H_a}) \times P(E_3|E_1, E_2, \boldsymbol{H_a}) \times P(E_2|E_1, \boldsymbol{H_a}) \times P(E_1|\boldsymbol{H_a})} = \frac{p_4 \times p_3 \times p_2 \times p_1}{q_4 \times q_3 \times q_2 \times q_1}$$

$$p_2 = P(E_2|E_1, \boldsymbol{H_m})$$

E_2 : On the jacket J one stain of secretion was recovered [backside of right collar] but no other biotraces

E_1 : The jacket J was found in an abandoned car (Volvo) near the crime scene

Presuming the jacket J was recently worn [by S], this presumption supported by finding E_1 [and by presuming the truth of $\boldsymbol{H_m}$], the number of biotraces recovered from the jacket can be assumed to follow a Poisson distribution. The probability p_2 may then be expressed

$$p_2 \approx \lambda_1 \cdot e^{-\lambda_1}$$

where λ_1 is the average number of biotraces found upon searching such traces on a jacket of this type that was recently worn by someone.

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.
$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

$$V = \frac{P(E_4|E_1, E_2, E_3, \boldsymbol{H_m}) \times P(E_3|E_1, E_2, \boldsymbol{H_m}) \times P(E_2|E_1, \boldsymbol{H_m}) \times P(E_1|\boldsymbol{H_m})}{P(E_4|E_1, E_2, E_3, \boldsymbol{H_a}) \times P(E_3|E_1, E_2, \boldsymbol{H_a}) \times P(E_2|E_1, \boldsymbol{H_a}) \times P(E_1|\boldsymbol{H_a})} = \frac{p_4 \times p_3 \times p_2 \times p_1}{q_4 \times q_3 \times q_2 \times q_1}$$

$$p_1 = P(E_1|\boldsymbol{H_m})$$

E_1 : The jacket J was found in an abandoned car (Volvo) near the crime scene

p_1 is approximately equal to the probability that a jacket worn during a criminal activity is left behind at the crime scene.

Denominator of V

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.
$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

$$V = \frac{P(E_4|E_1,E_2,E_3,\boldsymbol{H_m}) \times P(E_3|E_1,E_2,\boldsymbol{H_m}) \times P(E_2|E_1,\boldsymbol{H_m}) \times P(E_1|\boldsymbol{H_m})}{P(E_4|E_1,E_2,E_3,\boldsymbol{H_a}) \times P(E_3|E_1,E_2,\boldsymbol{H_a}) \times P(E_2|E_1,\boldsymbol{H_a}) \times P(E_1|\boldsymbol{H_a})} = \frac{p_4 \times p_3 \times p_2 \times p_1}{q_4 \times q_3 \times q_2 \times q_1}$$

$$q_4 = P(E_4|E_1,E_2,E_3,\boldsymbol{H_a})$$

E_4 : The major part of the mixture showed a profile identical with the profile of S (Γ_S)

If $\boldsymbol{H_a}$ is true then E_4 may have been observed if

- DNA from S was on J before ($\boldsymbol{C_4}$) – very similar to $\boldsymbol{C_2}$
- DNA from S was not on J before, but DNA from someone else matching DNA from S was ($\boldsymbol{C_5}$)

$$\Rightarrow q_4 = P(E_4|C_4,E_1,E_2,E_3,\boldsymbol{H_a}) \times P(C_4|E_1,E_2,E_3,\boldsymbol{H_a}) +$$
$$+ P(E_4|C_5,E_1,E_2,E_3,\boldsymbol{H_a}) \times P(C_5|E_1,E_2,E_3,\boldsymbol{H_a})$$
$$= 1 \times P(C_4|E_1,E_2,E_3,\boldsymbol{H_a}) + \gamma_{\Gamma_S} \times P(C_4|E_1,E_2,E_3,\boldsymbol{H_a}) \approx P(C_4|E_1,E_2,E_3,\boldsymbol{H_a})$$

RMP of Γ_S - negligible

$$\Rightarrow q_4 = k \cdot d \cdot m \cdot r$$

$k \cdot d$ = probability to deposit secretion (here assumed to be saliva) on a jacket worn when taking part in an activity that affects k. When the activity is a theft/robbery k is = 1
$m \cdot r$ = probability that DNA deposited on a jacket persists a period of time that affects $m \leq 1$ where $m \approx 1$ when the time period is a few hours or less

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.
$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

$$V = \frac{P(E_4|E_1, E_2, E_3, \boldsymbol{H_m}) \times P(E_3|E_1, E_2, \boldsymbol{H_m}) \times P(E_2|E_1, \boldsymbol{H_m}) \times P(E_1|\boldsymbol{H_m})}{P(E_4|E_1, E_2, E_3, \boldsymbol{H_a}) \times P(E_3|E_1, E_2, \boldsymbol{H_a}) \times P(E_2|E_1, \boldsymbol{H_a}) \times P(E_1|\boldsymbol{H_a})} = \frac{p_4 \times p_3 \times p_2 \times p_1}{q_4 \times q_3 \times q_2 \times q_1}$$

$$q_3 = P(E_3|E_1, E_2, \boldsymbol{H_a})$$

E_3 : The stain showed a mixture of DNA from two persons DNA (major and minor parts)

Neither E_1 nor $\boldsymbol{H_a}$ is assumed relevant for E_3

$$\Rightarrow q_3 = P(E_3|E_2)$$

E_1 : The jacket J was found in an abandoned car (Volvo) near the crime scene
E_2 : On the jacket J one stain of secretion was recovered but no other biotraces.

which corresponds to how expected a DNA mixture is upon analysing a secretion stain recovered from the backside of the right collar of a jacket, and hence the same as the corresponding probability in the numerator (p_3).

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.
$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

$$V = \frac{P(E_4|E_1,E_2,E_3,\boldsymbol{H_m}) \times P(E_3|E_1,E_2,\boldsymbol{H_m}) \times P(E_2|E_1,\boldsymbol{H_m}) \times P(E_1|\boldsymbol{H_m})}{P(E_4|E_1,E_2,E_3,\boldsymbol{H_a}) \times P(E_3|E_1,E_2,\boldsymbol{H_a}) \times P(E_2|E_1,\boldsymbol{H_a}) \times P(E_1|\boldsymbol{H_a})} = \frac{p_4 \times p_3 \times p_2 \times p_1}{q_4 \times q_3 \times q_2 \times q_1}$$

$$q_2 = P(E_2|E_1,\boldsymbol{H_a})$$

E_2 : On the jacket J one stain of secretion was recovered [backside of right collar] but no other biotraces

E_1 : The jacket J was found in an abandoned car (Volvo) near the crime scene

Presuming the jacket J was (probably) relatively recently worn (by someone), this presumption supported by finding E_1, the number of biotraces recovered from the jacket can (like previously) be assumed to follow a Poisson distribution. The probability q_2 may then be expressed

$$q_2 \approx \lambda_2 \cdot e^{-\lambda_2}$$

where λ_2 is the average number of biotraces found upon searching such traces on a jacket of this type that was (probably) relatively recently worn by someone.

Compared to λ_1 (cf. p_2) we can reasonably assume $0.5\ \lambda_1 \leq \lambda_2 \leq \lambda_1$

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.
$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

$$V = \frac{P(E_4|E_1, E_2, E_3, \boldsymbol{H_m}) \times P(E_3|E_1, E_2, \boldsymbol{H_m}) \times P(E_2|E_1, \boldsymbol{H_m}) \times P(E_1|\boldsymbol{H_m})}{P(E_4|E_1, E_2, E_3, \boldsymbol{H_a}) \times P(E_3|E_1, E_2, \boldsymbol{H_a}) \times P(E_2|E_1, \boldsymbol{H_a}) \times P(E_1|\boldsymbol{H_a})} = \frac{p_4 \times p_3 \times p_2 \times p_1}{q_4 \times q_3 \times q_2 \times q_1}$$

$$q_1 = P(E_1|\boldsymbol{H_a})$$

E_1 : The jacket J was found in an abandoned car (Volvo) near the crime scene

This is the probability that a jacket that was worn by S at latest in January 2015 is found in a car (an Volvo) near the crime scene just after the robbery. For the present we denote this probability t .

$$q_1 = t$$

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.
$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

$$V = \frac{P(E_4|E_1,E_2,E_3,\boldsymbol{H_m}) \times P(E_3|E_1,E_2,\boldsymbol{H_m}) \times P(E_2|E_1,\boldsymbol{H_m}) \times P(E_1|\boldsymbol{H_m})}{P(E_4|E_1,E_2,E_3,\boldsymbol{H_a}) \times P(E_3|E_1,E_2,\boldsymbol{H_a}) \times P(E_2|E_1,\boldsymbol{H_a}) \times P(E_1|\boldsymbol{H_a})} = \frac{p_4 \times p_3 \times p_2 \times p_1}{q_4 \times q_3 \times q_2 \times q_1}$$

$$V = \frac{p_4 \times p_3 \times p_2 \times p_1}{q_4 \times q_3 \times q_2 \times q_1} \approx$$

$$\approx \frac{[d \cdot r + (1-d) \cdot k \cdot d \cdot m \cdot r] \times [P(E_3|E_2)] \times [\lambda_1 \cdot e^{-\lambda_1}] \times [P(E_1|\boldsymbol{H_m})]}{[k \cdot d \cdot m \cdot r] \times [P(E_3|E_2)] \times [\lambda_2 \cdot e^{-\lambda_2}] \times [t]}$$

$$= \frac{(1 + (1-d) \cdot k \cdot m) \cdot \lambda_1 \cdot e^{-\lambda_1} \cdot P(E_1|\boldsymbol{H_m})}{k \cdot m \cdot \lambda_2 \cdot e^{-\lambda_2} \cdot t} \geq \quad / \lambda_1 \text{ is assumed} \geq \lambda_2 /$$

$$\geq \frac{(1 + (1-d) \cdot k \cdot m) \cdot \lambda_1 \cdot P(E_1|\boldsymbol{H_m})}{k \cdot m \cdot \lambda_1 \cdot t} \times e^{\lambda_2 - \lambda_1} \geq \quad / \lambda_1 \text{ is assumed} \leq 2\lambda_2 /$$

$$\geq \left(\frac{1}{k \cdot m} + 1 - d\right) \times \frac{P(E_1|\boldsymbol{H_m})}{t} \times e^{-\lambda_2} \geq \quad / d \leq 1, m \leq 1 / \quad \geq \frac{P(E_1|\boldsymbol{H_m})}{k \cdot t} \times e^{-\lambda_2}$$

$$V \geq \frac{P(E_1|\boldsymbol{H_m})}{k \cdot t} \times e^{-\lambda_2}$$

$\boldsymbol{H_m}$: The latest time S wore jacket J was when taking part in a robbery against a money transport at BC on 18 February, 2015.
$\boldsymbol{H_a}$: The latest time S potentially wore jacket J was in January 2015.

E_1 : The jacket J was found in an abandoned car (Volvo) near the crime scene

$t = P(E_1|\boldsymbol{H_a})$

$k \cdot d$ = probability to deposit secretion (here assumed to be saliva) on a jacket worn when taking part in an activity that affects k. When the activity is a theft/robbery $k = 1$.

λ_2 is the average number of biotraces found upon searching such traces on a jacket of this type that was (probably) relatively recently worn by someone.

$t = P(E_1|\boldsymbol{H_a})$ would have the greatest impact on V.

Why would a jacket that S wore at latest in January 2015 show up in this particular Volvo on 18 February 2015?

In February 2015 there were about 900 000 cars registered in the county of Stockholm. Each car may also be parked at different places. Some cars may not have been in the county that evening, while cars from other counties would…

What is the magnitude of the likelihood ratio V?