Duncan Taylor

duncan.taylor@sa.gov.au

Continuous systems and their implementation

forensic science sa

In the beginning…

Before contact DNA and LtDNA – Forensics dealt with biological fluids, which most commonly gave simple single source profiles
•no dropout
•no complicating factors

If any loci were incomplete they were omitted from the calculation

Used basic profile frequencies based on Hardy-Weinberg Equilibrium and then later using subpopulation model

forensic science sa

In the beginning…

Then contact DNA came along and everything became complicated
(in SA this was largely due to the Snowtown murders in 90s)

•An abundance of new sample types were now available
•Higher proportion of these were mixtures
•There were many instances of incomplete profiles

Added to this:
•DNA profiling kits were getting more powerful and sensitive
•Lab hardware was getting faster and more sensitive
•Labs began experimenting with ways of enhancing profiles

forensic science sa

Two methods evolved to apply statistical weightings to this evidence:

•Random Man Not Excluded (RMNE) – a frequentist method that determined what proportion of the population would not be excluded from an observed mixed profile

•Likelihood Ratio (LR) – a Bayesian method that determined the probability of obtaining the observed mixed profile given two competing hypotheses

I will be talking about LR during this presentation

forensic science sa

Most LR methods can be thought of using the same formulae

They just use different methods of 'weighting' genotype combinations

Our ability to generate weightings has been refined as our ability to calculate them and our understanding of DNA profiles has improved

forensic science sa

Interpretation from binary to continuous

forensic science sa

Moving to a continuous method

The first part of this talk is about moving from “binary” to “continuous” systems of interpretation

Binary system:
A system of rules and thresholds where certain aspects of the profile where ruled as either ‘acceptable’ or ‘inacceptable’ – two opposite interpretations, hence binary

Subject profiles to this binary system of rules, then interpret

Analogous to subjecting an image to a series of rules that change it into black and white (binary) and then trying to make interpretations from it

forensic science sa

Moving to a continuous method

Sometimes, doing this will fit the original very well and interpretation will be easy

Reference

Binary evidence

Based on this evidence it would be pretty simple to conclude that the reference person is in the evidence image

Moving to a continuous method

It gets a lot more complicated to make interpretations when the evidence image becomes more complex

Reference

Binary based evidence

Now much less clearas to whether the reference person is in there or whether they have left the building

Moving to a continuous method

Adding to the complexity is the way in which the binary rules are applied

forensic science sa

Imagine writing a set of rules would allow people to draw an image in black and white

And that everyone had to follow those rules and apply them to every image they encounter regardless of:

- Colour depth
- Brightness
- Complexity
- etc

Moving to a continuous method

You are guaranteed to end up with a number of variations

And they could all be interpreted slightly differently

Moving to a continuous method

Ideally we want to carry out interpretations on the initial image

Using all available information

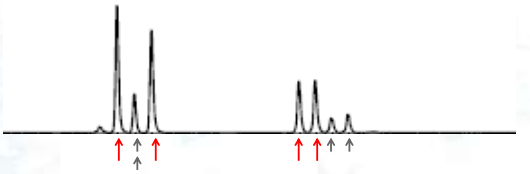
Without any rules

This is the idea behind moving to a continuous system

There are no rules that you need to fit your evidence into

Binary DNA profile interpretation

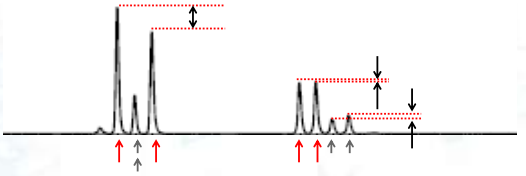
Now applying this idea to DNA profiles



We think these might be the profiles of the two contributors, but how can we make this judgement in an objective and consistent manner ?

Binary DNA profile interpretation

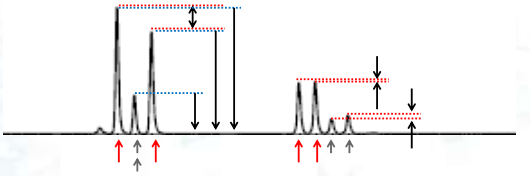
Notice that there are two contributors, a major and a minor and we want to interpret profiles for both



-Heterozygous balance

Binary DNA profile interpretation

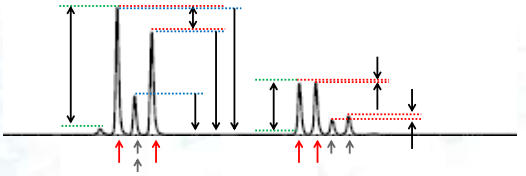
Notice that there are two contributors, a major and a minor and we want to interpret profiles for both



-Heterozygous balance
-Dropout / Homozygous

Binary DNA profile interpretation

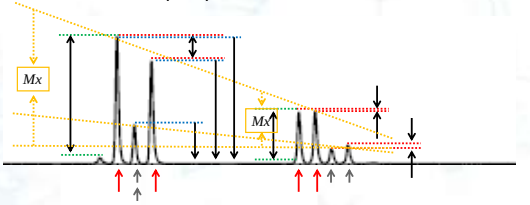
Notice that there are two contributors, a major and a minor and we want to interpret profiles for both



-Heterozygous balance
-Dropout / Homozygous
-Stutter ratios

Binary DNA profile interpretation

Notice that there are two contributors, a major and a minor and we want to interpret profiles for both



-Heterozygous balance
-Dropout / Homozygous
-Stutter ratios
-Mixture ratios

Then we make a judgement on whether we can interpret single component(s) from this profile

Systems of thresholds are limited

-OK for simple scenarios, but break down when profiles become complex

-Suffer from 'falling off the cliff' effect

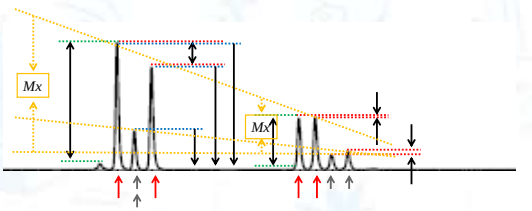
-Waste a lot of data

-Has difficulties handling non-concordances

-Cannot be applied consistently between analysts

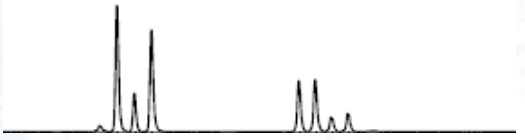
-Leads to inconsistencies in interpretations, potentially unequal justice outcomes

Do away with all thresholds



Do away with all thresholds

Easier said than done



How do you interpret a DNA profile when anything is possible ?

This is the challenge – we must make use of the fact that while everything is possible, they are not equally probable

Systems of interpretation

forensic science sa

simple mixture analysis

•Loci must be complete (i.e. given some threshold, dropout cannot be possible in the observed profile)

•Incomplete loci are omitted from the calculation

•Peak heights are not taken into account.

•All genotypes are weighted equally

forensic science sa

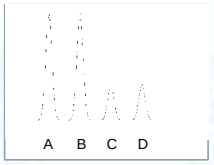
simple mixture analysis

•Only required a dropout threshold to designate loci as in or out

•A highly restrictive method that wastes an enormous amount of profile information

•This is the method used in DNAmix

forensic science sa

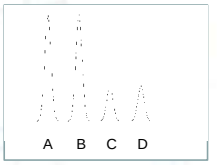


Using only qualitative information about which peaks are present the possible genotypes of unknowns are:
[A,B] & [C,D]
[A,C] & [B,D]
[A,D] & [B,C]
[C,D] & [A,B]
[B,D] & [A,C]
[B,C] & [A,D]

With each combination being weighted equally with 1

forensic science sa

If our suspects are [A,B] and [C,D] and we are considering the hypotheses:
Hp = 2 Suspects
Hd = 2 unknowns



$$LR = \frac{1}{\begin{aligned}&2\Pr(AB \mid ABCD)\times 2\Pr(CD \mid ABABCD)+\\&2\Pr(CD \mid ABCD)\times 2\Pr(AB \mid ABCDCD)+\\&2\Pr(AC \mid ABCD)\times 2\Pr(BD \mid ABACCD)+\\&2\Pr(BD \mid ABCD)\times 2\Pr(AC \mid ABBDCD)+\\&2\Pr(AD \mid ABCD)\times 2\Pr(BC \mid ABADCD)+\\&2\Pr(BC \mid ABCD)\times 2\Pr(AD \mid ABBCCD)\end{aligned}}$$

Normally the equation would be simplified by collecting and cancelling common elements, however I will leave them expanded as it will make later demonstrations easier

Clayton rules

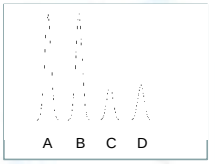
Use a set of predefined threshold to determine whether genotypes are included in the LR calculation
– effectively weighting them with 0 or 1.

-Loci still needed to be complete (i.e. no possibility of dropout)
- all accepted genotype combinations are weighted equally

The thresholds used can vary, however will typically include:
•Heterozygous balance and
•Mixture proportions

This method can be used in conjunction with other methods to remove some genotype combinations and weight the remainder in some way

Software – PENDULUM employs an advanced for of this method



Here the genotypes are either included or excluded from the list of unknowns depending on a set of rules. For Pr(E|Hd) the 2 unknowns could be:

[A,B] & [C,D] - accepted
[A,C] & [B,D] – rejected (Het imbalance)
[A,D] & [B,C] – rejected (Het imbalance)
[C,D] & [A,B] – rejected (ratio flip)
[B,D] & [A,C] – rejected (Het imbalance)
[B,C] & [A,D] – rejected (Het imbalance)

Now the first combination is weighted with 1
The other combinations are weighted with 0 – due to failing rules

$$LR = \frac{1}{2\Pr(AB \mid ABCD)\times 2\Pr(CD \mid ABABCD)}$$

2p rule

Gave the ability to handle dropouts
This meant loci no longer needed to be complete

POI - [A,A]

CS - [A]

Hypothesis 1 = POI is the source of DNA and dropout has not occurred

Hypothesis 2 = Someone else is the source of DNA and are either:
[A,A] – and dropout has not occurred
[A,Q] – and the Q has dropped out

$$LR = \frac{1}{2\Pr(A \mid AA)}$$

forensic science sa

2p rule

A big step forward – not as much information wasted

Problem handling non-concordances, e.g.
CS = [A]
POI = [A,B]
If POI is the source of the DNA then dropout must have occurred

-This is ok, when [A] is weak, but becomes a problem as [A] approached the homozygous threshold

-Studies have found that 2p rule is not always conservative

fore

Is the 2*p* rule always conservative?
John Buckleton[a,*], Christopher Triggs[b]

Stage 4 – Pr(D) and Pr(C)

Better way to deal with dropouts was developed
•Able to handle non-concordances
•Able to handle drop-in

In combination these two probabilities produced a type of weighting for each genotype combination.

Has been around for over a decade
– LikeLTD, Rudin and Lollemueller

forensic science sa

Stage 4 – Pr(D) and Pr(C)

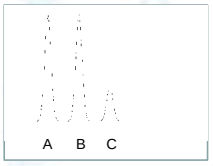
Pr(D) and Pr(C) are constants in the equations and do not vary with peak heights

This means that a peak expected at 1000rfu has the same dropout probability as a peak at 50rfu, however as an approximation the model works very well.

It can be extended to have differing probabilities based on heterozygous or homozygous peaks, or even to assign differing probabilities to differing individuals (in the case of a major/minor mix)

Gill & Haned have developed software that uses this method - LRmix

LRmix-style calculation



Each genotype is included in the model, but weighted by probabilities of dropout or drop-in:

Take genotype [A,B] & [C,Q] where [Q] is any allele other than [A], [B] or [C] (and so has dropped out)

This requires:
3 non-dropouts (each with probability D)
1 dropout (with probability D)
3 non-drop-ins (with probability C)

And so genotype [A,B] & [C,Q] has weighting:

$$\overline{D}^3 D \overline{C}^3$$

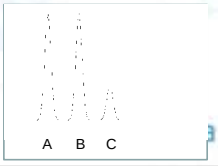
forensic science sa

LRmix-style calculation

Again with suspects [A,B] and [C,D] and hypotheses:
Hp = 2 Suspects
Hd = 2 unknowns

Good as it doesn't eliminate any possible genotypes

$$LR = \frac{\overline{D}^3 D \overline{C}^3}{\begin{array}{l}\overline{D}C^2\overline{C} \times \Pr(AA \mid ABCD) \times \Pr(AA \mid AAABCD) + \\ \overline{D}^2 C \overline{C}^2 \times \Pr(AA \mid ABCD) \times 2\Pr(AB \mid AAABCD) + \\ \overline{D}^2 C \overline{C}^2 \times \Pr(AA \mid ABCD) \times 2\Pr(AC \mid AAABCD) + \\ \overline{D}DC^2\overline{C} \times \Pr(AA \mid ABCD) \times 2\Pr(AF \mid AAABCD) + \\ \ldots. \\ D^4 C^3 \times \Pr(FF \mid ABCD) \times \Pr(FF \mid ABCD)\end{array}}$$



Usually the probability of drop-in is quite small so genotype combinations that require drop-in get a low weighting and do not have much effect on the overall LR (unless the scenario under Hp requires a drop-in to have occurred)

'drop' method

The first method that utilises peak height information and so a big step forward again in information usage

Probabilities are assigned which vary with each instance of dropout or non-dropout depending on observed and expected peak heights

Deals with non-concordances in a very robust manner

forensic science sa

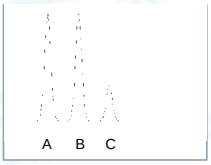
'drop' method

Simple to implement for single source cases but becomes complex to implement with mixtures without taking into account other factors (like heterozygous balance).

Software that uses this idea is LoComatioN – which is not generally available

forensic science sa

The Drop method



Using the same scenario as previous

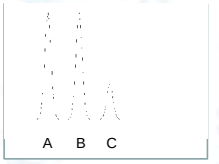
All the same genotypes are considered

Each genotype will have associated probabilities of dropin and dropout

This time the probabilities are functions rather than constants

forensic science sa

The Drop method



e.g. moving from LRmix to Drop model, the weighting for genotype [A,B] & [C,D] goes from this:

$$\overline{D}^3 D \overline{C}^3$$

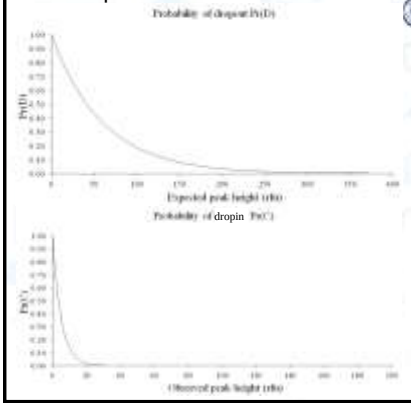
to this $\Pr(\overline{D}_{HA})\Pr(\overline{D}_{HB})\Pr(\overline{D}_{HC})\Pr(D_{HD})\Pr(\overline{C}_{HA})\Pr(\overline{C}_{HB})\Pr(\overline{C}_{HC})$

Where $Pr(D_{HD})$ is the probability of dropout for peak [D] expected to have been present at intensity HD

$Pr(D_X)$ and $Pr(C_X)$ are functions of expected peak height

forensic science sa

The Drop method

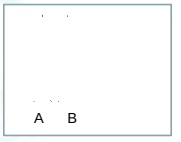


Example graphs of Pr(D) and Pr(C) vs peak height

Pr(C) graph is for 28 cycle standard work. It is set so that dropin is heavily penalised

The Drop method

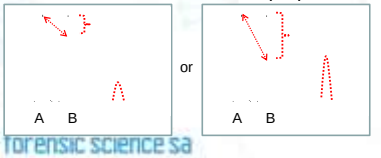
It is difficult to use this system beyond single source:



Suspect 1 [A,B]
Suspect 2 [B,Q]

For drop model we need to know the expected height of the dropped out allele

To do this – need Hb and mix proportions



One may be favoured over the other depending of mix proportions

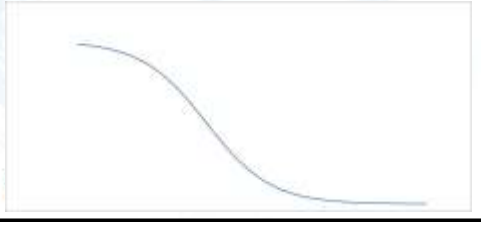
forensic science sa

The Drop method

You then must fit a curve to the observed peak heights to obtain expected peak heights

Then use expected peak heights for any dropouts

Logistic curves have been used for this:



Probabilistic methods

Weightings for genotype combinations are generated probabilistically.

Similar to the drop model, except that probabilities can be based on:
- Heterozygous peak balances
- Mixture proportions
- Dropout
- Stutter
- Degradation
- Other factors if you are willing to make the model more complex.

forensic science sa

Probabilistic methods

- Very robust

- Can deal with many situations

- Must make simplifying assumptions before the calculation can be carried out for complex profiles with many variables

DNA INSIGHT – works using this method

forensic science sa

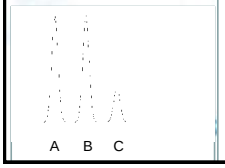
Probabilistic methods

In fully probabilistic system weightings are a combination of DNA profile measurements

Can be as complex or as simple as desired

Again no genotypes are discounted.

No genotypes are certainties i.e. 0 < weighting($w_{x.}$) < 1



$$LR = \frac{w_{x.}}{\begin{array}{l} w_{1.} \times \Pr(AA \mid ABCD) \times \Pr(AA \mid AAABCD) + \\ w_{2.} \times \Pr(AA \mid ABCD) \times 2\Pr(AB \mid AAABCD) + \\ w_{3.} \times \Pr(AA \mid ABCD) \times 2\Pr(AC \mid AAABCD) + \\ w_{4.} \times \Pr(AA \mid ABCD) \times 2\Pr(AF \mid AAABCD) + \\ .... \\ w_{n.} \times \Pr(FF \mid ABCD) \times \Pr(FF \mid ABCD) \end{array}}$$

Continuous method

Uses MCMC to generate a sampling from the posterior distribution of the DNA profile properties.

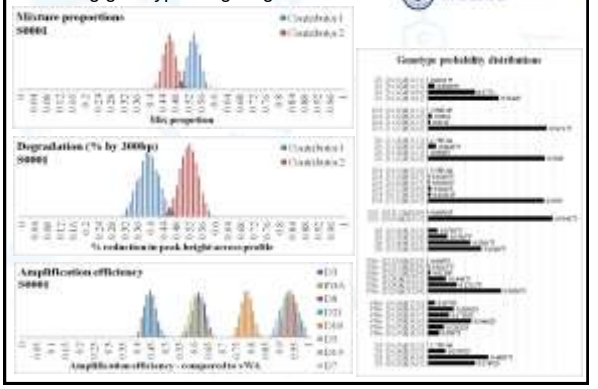
Very robust and in theory can handle any profile of any quality or complexity

Also can handle very complex modelling as it is easily extendable

Software that uses this method are - True Allele & STRmix

forensic science sa

Provides distributions for parameters including genotype weightings



LR is still constructed in the same way with each genotype combination being given a weighting

$$LR = \frac{w_{x.}}{\begin{array}{l} w_{1.} \times \Pr(AA \mid ABCD) \times \Pr(AA \mid AAABCD) + \\ w_{2.} \times \Pr(AA \mid ABCD) \times 2\Pr(AB \mid AAABCD) + \\ w_{3.} \times \Pr(AA \mid ABCD) \times 2\Pr(AC \mid AAABCD) + \\ w_{4.} \times \Pr(AA \mid ABCD) \times 2\Pr(AF \mid AAABCD) + \\ .... \\ w_{n.} \times \Pr(FF \mid ABCD) \times \Pr(FF \mid ABCD) \end{array}}$$

Weightings are complex multiple integrals

Formula below shows the weighting formula used by STRmix as an example

$$w_{q.} = \int_{\mathrm{M}} \prod_{l} \prod_{a} \prod_{r} \Pr(E_{ar}^{l} \mid S_q, \mathrm{M}) \Pr(\mathrm{M}) d\mathrm{M}$$

More easily written as $\Pr(G_c|S_q)$ - Probability of crime stain genotypic data given a genotype set

Law of total probability

$\Pr(A)$ What if you need to know B to work out Pr(A)

$\Pr(A|B)$ x $\Pr(B)$

What if there was a range of values that B could take ?

Then you would need to step to integration

Can write this as $\Pr(A) = \int_i \Pr(A \mid B_i) \Pr(B_i) di$

forensic science sa

Law of total probability

Imagine we wanted to know what the probability was of wearing size 13 shoes

$\Pr(Sz=13)$

But we know that height and shoe size are related. So we split up the probability into people above and below 150cm

$\Pr(Sz=13|H>150cm)\Pr(H>150cm) + \Pr(Sz=13|H<150cm)\Pr(H<150cm)$

We could just as validly split this into 10cm height groups:

$\Pr(Sz=13|H =50\text{-}60cm)\Pr(H =50\text{-}60cm) +$
$\Pr(Sz=13|H =60\text{-}70cm)\Pr(H =60\text{-}70cm)+...+$
$\Pr(Sz=13|H =220\text{-}230cm)\Pr(H =220\text{-}230cm)$

forensic science sa

Law of total probability

Ultimately we could step from ever diminishing discrete values to the continuous equivalent (integral with respect to height)

$$\int_{\text{height}} \Pr(Sz = 13|H = i)\Pr(H = i)dH$$

We don't really care what people's height is, we are just interested in the probability of them having size 13 shoes.

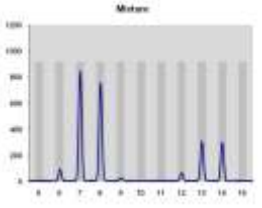
So we integrate the conditional probability of having size 13 shoes across the full possible range of heights

forensic science sa

What is $\Pr(G_C|S_q)$?

We do exactly this same process with the parameters in our weighting formula – We use M for Mass

$$\Pr(G_c \mid S_q) = \int_M \Pr(G_c \mid S_q, M)\Pr(M)dM$$

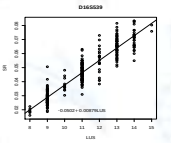


To calculate the LR we are not really interested in the probabilities of the various mass parameters, however they are needed to calculate the probabilities of the evidence given genotype sets

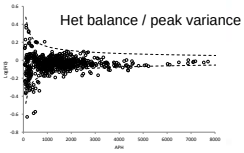
So we integrate the probability of the evidence given the genotype with respect to the mass parameters

Continuous systems

Stutter graphs



Het balance / peak variance



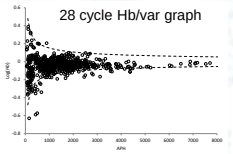
Earlier we saw drop-in and dropout graphs

Now we need biological models for other behaviours

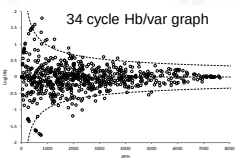
forensic science sa

Continuous systems

28 cycle Hb/var graph



34 cycle Hb/var graph



As methods for DNA profile generation change so to will constants within the models

But the models stay the same

e.g. increasing PCR cycles means peak height is more variable, but still contains information

forensic science sa

Using MCMC for continuous models

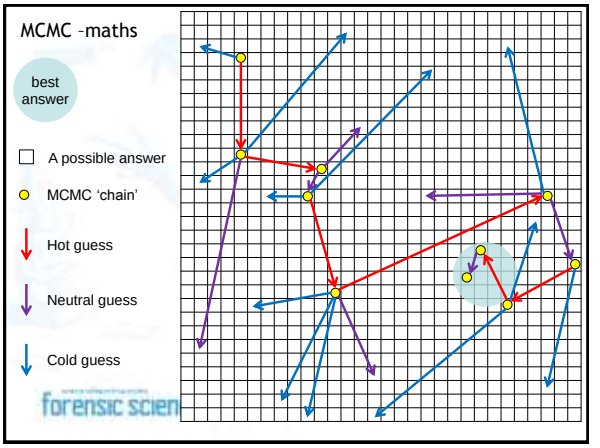
forensic science sa

Continuous models

- DNA profiles problems can be really really complex
- So complex that even with modern computers, it would be impossible to test every possible combination of answers…so we don't
- Instead the computer uses a process similar to the game of 'hot and cold' with the DNA profile
- This mathematical process is called Markov Chain Monte Carlo – or MCMC
- MCMC allows the computation of complex problems with standard computers

forensic science sa

MCMC –maths



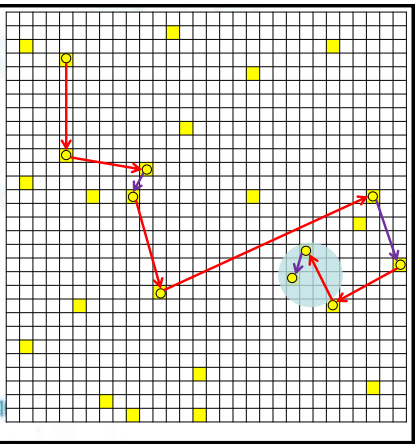
forensic scien

MCMC – maths

We ran MCMC chain for 10 moves

Notice that it only had to test a small fraction of all the possible answers to get there

Imagine if we ran it again for 10 moves it would still get into good space, but probably a different square



MCMC – the maths

- The same thing happens when we analyse DNA profiles using STRmix – except that we run this 'game' for hundreds of thousands to billions of moves
- Each time we run the same problem, STRmix gives a different answer
- Importantly these answers are all clustered around each other and the amount that they would vary is small in relation to the size of the number itself

forensic science sa

MCMC – the maths

- E.g. we could run the same problem through five times and get:
  - 4 million
  - 2 million
  - 10 million
  - 6 million
  - 1.5 million
- Importantly all these results convey the same strength of evidence, i.e. something is millions of times more likely than something else
- This is unique to MCMC, and does not occur for probabilistic modeling

forensic science sa

MCMC – the maths

Two drawbacks

1) The programs that are produced can be 'black-boxes'
2) More dimensions → more computing power → longer runtime

The first issue can be addressed in the design of the results and education

The second issue can be addressed by using an adaptive Bayesian computing algorithms or advanced computational methods.

Need to be carful not to violate point 1

Fact remains that for really complex problems there is just no way the results can be instant

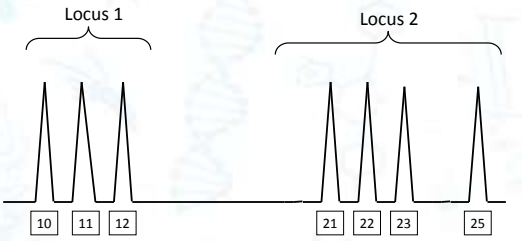
MCMC – the maths

- The following is a two-locus example of how an expected DNA profile can be built up with only four paramaters:

1) Genotypes
2) Amount of DNA
3) Degradation
4) Locus specific amplification

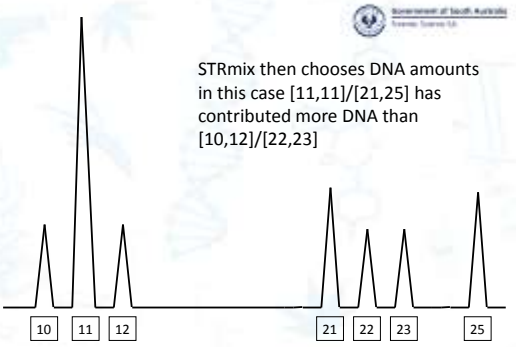
forensic science sa

Start by building up the profile given the genotypes:
[11,11]&[10,12] at locus 1 and [21,25]&[22,23] at locus 2



At this point only genotypes have been chosen. DNA amounts and therefore peak heights have not been incorporated yet

forensic science sa



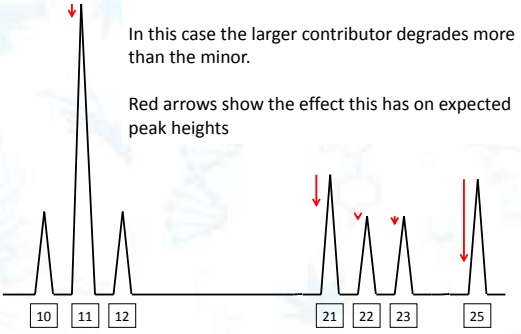
STRmix then chooses DNA amounts in this case [11,11]/[21,25] has contributed more DNA than [10,12]/[22,23]

forensic science sa

Degradation is then added for each contributor.

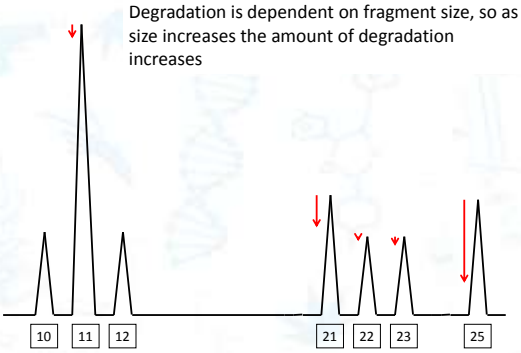
In this case the larger contributor degrades more than the minor.

Red arrows show the effect this has on expected peak heights



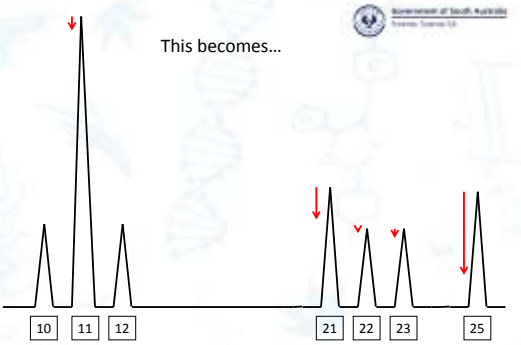
forensic science sa

Degradation is dependent on fragment size, so as size increases the amount of degradation increases



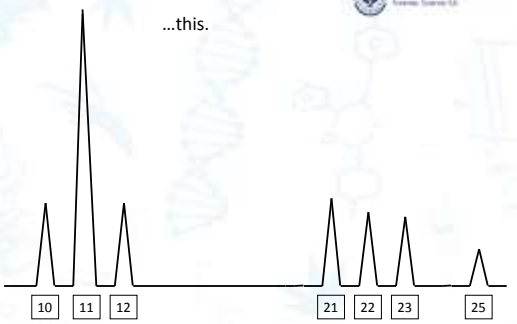
forensic science sa

This becomes…



forensic science sa

…this.

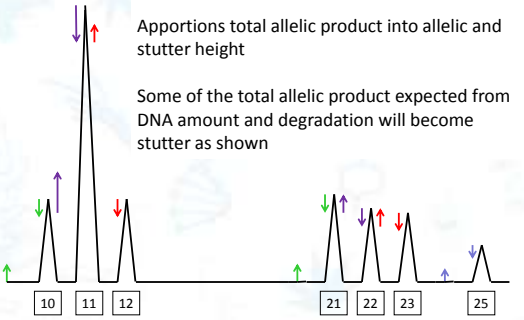


forensic science sa

Then stutter taken into account.

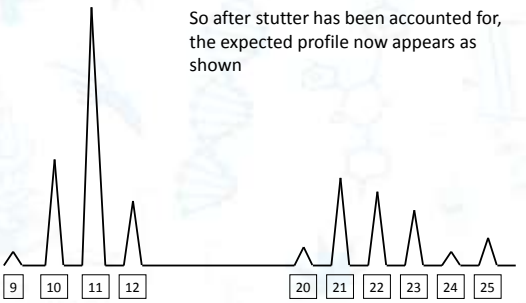
Apportions total allelic product into allelic and stutter height

Some of the total allelic product expected from DNA amount and degradation will become stutter as shown



forensic science sa

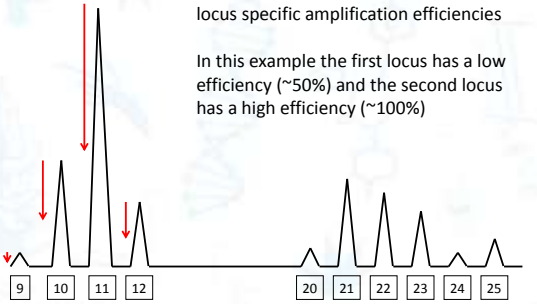
So after stutter has been accounted for, the expected profile now appears as shown



forensic science sa

locus specific amplification efficiencies

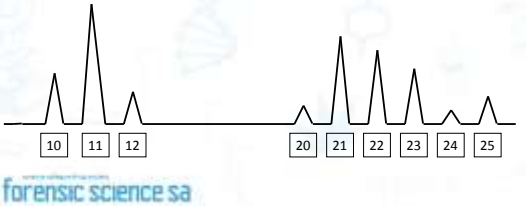
In this example the first locus has a low efficiency (~50%) and the second locus has a high efficiency (~100%)



forensic science sa

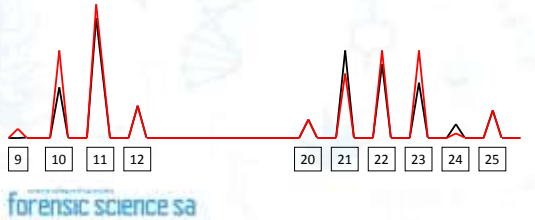
Now have a complete 2 locus expected DNA profile that can be compared to an observed profile

This will determine likelihood of observed profile given parameter values used to generate it – Most importantly this includes the genotypes



forensic science sa

Expected Observed

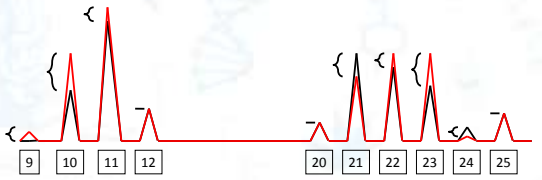


forensic science sa

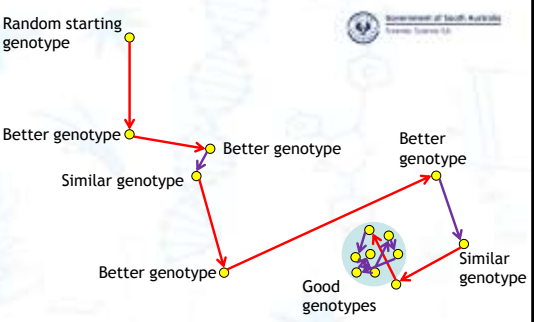
Expected Observed

Peak height differences between the expected and observed.

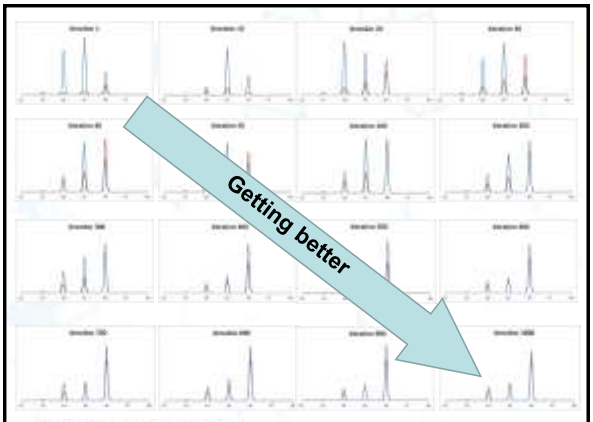
If the test values for parameters are true then these differences must be accounted for by stochastic peak variation - which can be assigned likelihoods



forensic science sa



forensic science sa



The amount of time that the MCMC spends on each genotype set then becomes the 'weighting' for that genotype set

Is then incorporated into the LR e.g. From before

$$LR = \frac{w_{x.}}{\begin{array}{l} w_{1.} \times \Pr(AA \mid ABCD) \times \Pr(AA \mid AAABCD) + \\ w_{2.} \times \Pr(AA \mid ABCD) \times 2\Pr(AB \mid AAABCD) + \\ w_{3.} \times \Pr(AA \mid ABCD) \times 2\Pr(AC \mid AAABCD) + \\ w_{4.} \times \Pr(AA \mid ABCD) \times 2\Pr(AF \mid AAABCD) + \\ .... \\ w_{n.} \times \Pr(FF \mid ABCD) \times \Pr(FF \mid ABCD) \end{array}}$$

Why use continuous methods ?

forensic science sa

As these systems of weighting become more refined they require:

- More complex mathematics
- More optimisation through modelling of stochastic effects
- More powerful computers to run

The question arises:

Why shift from the simple methods ?

forensic science sa

1. The simple methods require a system of thresholds to function

<u>Could use a very simple system that has minimal thresholds</u>

Pros:
- Easy to implement/understand/teach
- No speciality software or hardware required
- Will be reasonably consistent between practitioners

Cons:
- Rules are often set very conservatively so will waste most of the information in the profile
- Will not be able to be used on a large number of profiles
- Are usually not a good representation of reality

forensic science sa

1. The simple methods require a system of thresholds to function

<u>Or could use a system that has numerous thresholds</u>

Pro:
- Much better representation of reality
- Uses more of the information
- Depending on complexity may still not require speciality software or hardware

Cons:
- Experience has shown that complex systems of rules cannot be applied consistently by practitioners.
- Takes a lot of time for an analyst to interpret and apply rules, and then just as much time again for a second analyst to review them
- Increases false exclusion rate

forensic science sa

1. The simple methods require a system of thresholds to function

- Experience
- Level of conservatism
- Training
- Error

will mean inconsistencies exist between practitioner (or the same practitioner at different times)

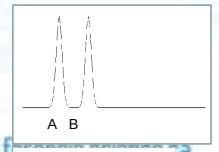
Experience has shown that multifaceted, interacting networks of rules are difficult and impractical to apply to complex profiles.

forensic science sa

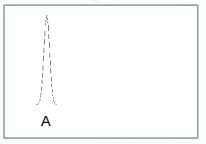
2. Thresholds suffer from 'falling off the cliff' effect

Refers to a phenomenon caused by the arbitrary application of a threshold whereby the difference of a single unit leads to diametrically opposed interpretations

For example – if homozygous threshold was 200rfu and we had the following example:



Suspect's - [A,B]



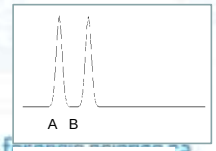
Crime Stain - [A]

2. Thresholds suffer from 'falling off the cliff' effect

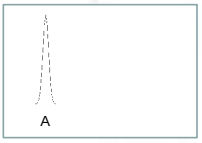
If [A] in the crime stain was 199rfu then the suspect is not-excluded

If [A] is 200rfu then the suspect is excluded

In reality the difference in support for the proposition that the suspect is the source of DNA between these two scenarios is miniscule



Suspect's - [A,B]



Crime Stain - [A]

3. Every system has its limit

Every system has a limit to its capabilities (even continuous systems)

However threshold based systems usually have a much larger pool of profiles they are unable to handle than continuous systems.

As level of expected stochastic variation hinders or inhibits interpretation, thresholds tend to break down

e.g. at low levels it is common for heterozygous balance thresholds to be dropped.

forensic science sa

4. Sometimes simple methods are anti-conservative

Whilst simple systems seek to set conservative thresholds, they sometimes do the exact opposite

Three common misconceptions are:

Omitting a locus from a calculation is always conservative

-This is not necessarily the case if the locus being omitted contains non-concordances with the reference

forensic science sa

4. Sometimes simple methods are anti-conservative

The 2p rule is always conservative when taking dropout into account

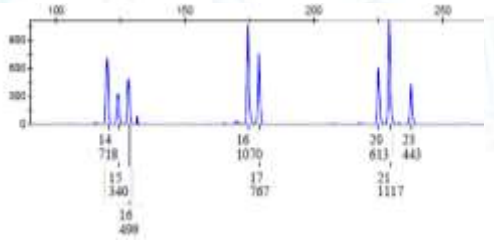
-Shown previously this is not the case when the required dropout approaches the homozygous threshold

Using Pr(D) and Pr(C) method is always conservative

- Not necessarily the case if the inclusion of the suspect requires a large heterozygote imbalance to have occurred

forensic science sa

Example of a continuous analysis



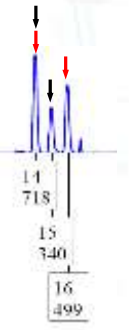
Two person mixed profile

Continuous systems estimates mixture proportions of approximately 47% and 53%

Weightings as a human diagnostic

D3S1358

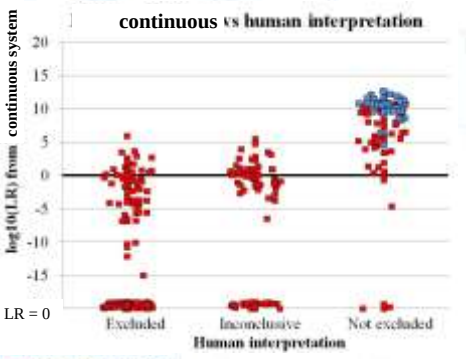
Genotypes	weights	
[15,16] [14,14]	0.172	
[14,16] [14,15]	0.252	←
[15,16] [14,15]	1.40E-5	
[16,16] [14,15]	0.007	Most supported genotypes
[14,16] [15,15]	3.91E-5	
[14,15] [14,16]	0.352	←
[15,15] [14,16]	2.74E-4	
[15,16] [14,16]	0.003	Known genotypes included
[14,14] [15,16]	0.209	←
[14,16] [15,16]	0.002	
[14,15] [16,16]	0.003	



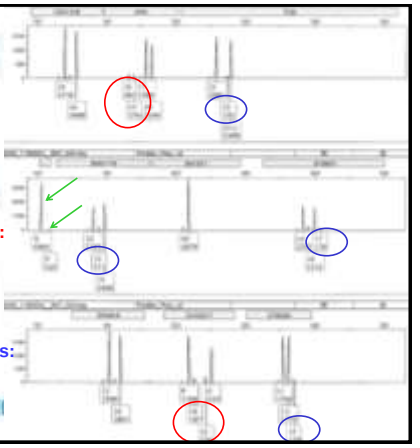
Continuous system in practice

forensic science sa

How it works in practise



Sexual Assault
Sample: tapelift from victims underwear



Using information at 2 DNA markers: LR = 900

Using information at all DNA markers: LR = 22 million

forensic scien