Approaches to handling complex mixtures

ISFG basic mixture interpretation workshop

Jo-Anne Bright

Specialist Science Solutions

Manaaki Tangata Taiao Hoki
protecting people and their environment through science

Introduction

- The binary method of DNA interpretation has served us well for many years
- Interpretation methods have not kept pace with advances in technology
- More trace DNA, more mixtures
- Under certain circumstances the binary method can be extended to interpret mixtures where dropout is possible
- Application and limitations are discussed in this talk

Contents lists available at SciVerse ScienceDirect

Science and Justice

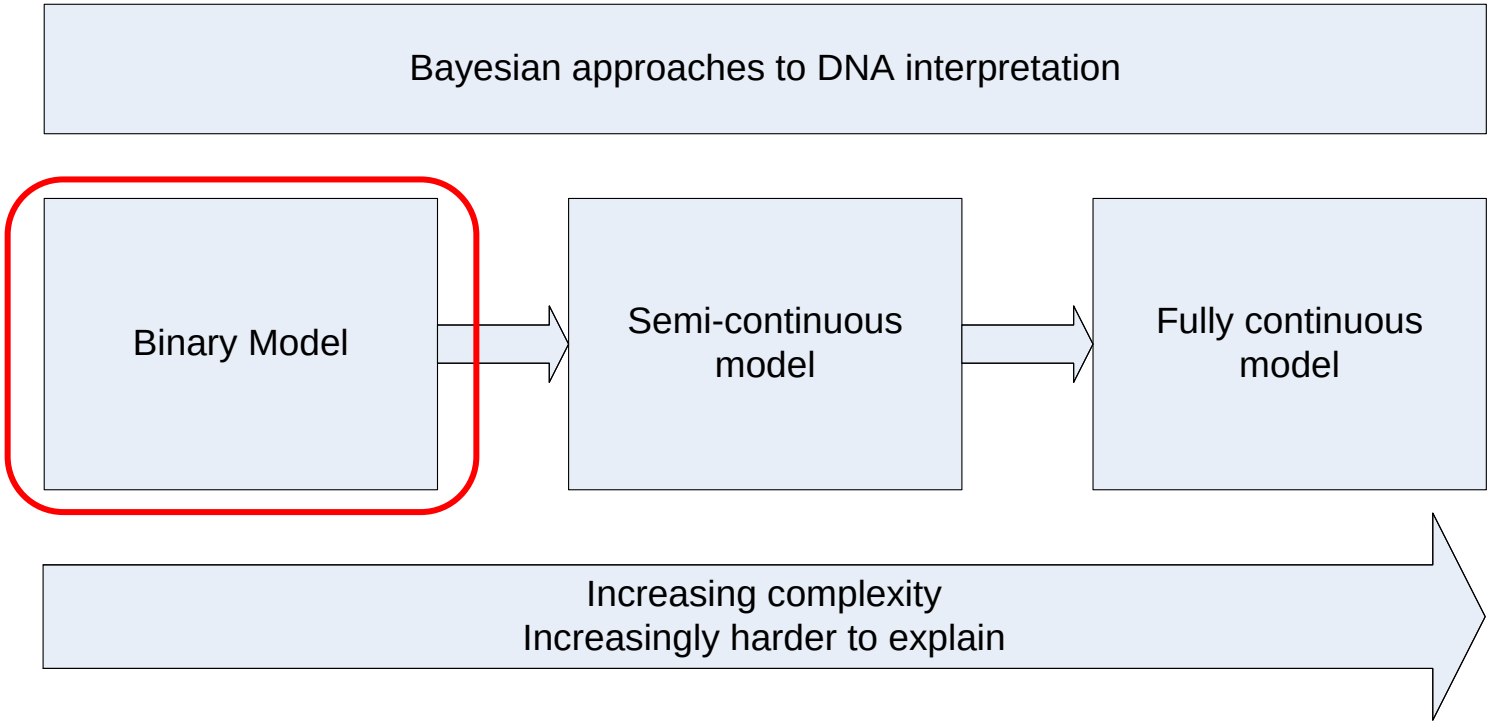
journal homepage: www.elsevier.com/locate/scijus

Emerging researcher article

A comparison of statistical models for the analysis of complex forensic DNA profiles

Hannah Kelly [a,b,*], Jo-Anne Bright [a], John Buckleton [a], James Curran [b]

[a] ESR, PB 92021, Auckland 1142, New Zealand
[b] Department of Statistics, University of Auckland, PB 92019, Auckland 1142, New Zealand


Bayesian approaches to DNA interpretation
Binary Model
Semi-continuous model
Fully continuous model
Increasing complexity
Increasingly harder to explain

The binary model

- Possible genotype combinations are considered either ‘in’ or ‘out’
- Manual method
- Can be extended to mixtures with 3 or more contributors
- Two subsets:
  - The constrained model
  - The unconstrained model

Qualitative binary method

- Most basic implementation of the binary model
- No peak height information taken into account
- Implemented in software:
  - POPSTATS
  - DNAMIX I
  - DNAMIX II (with 4.2 formulae)
  - DNAMIX III (with 4.2 and Beecham and Weir sampling uncertainty)

Unconstrained approach

Unconstrained method of mixture interpretation:

- **Write out all possible genotype combinations under H_2**
- **Do not rule any combinations out**

- **Less use of the information**
- **More efficient time wise**

Semi-quantitative binary method

- Making *partial* use of the profile data
- Empirical guidelines and expert judgement are used to exclude certain genotype combinations
- Heuristics such as:
  - Heterozygote balance
  - Mixture proportion
- The semi-quantitative model is mainly applied manually
  - An exception is GeneMapper® *ID-X*

Constrained approach

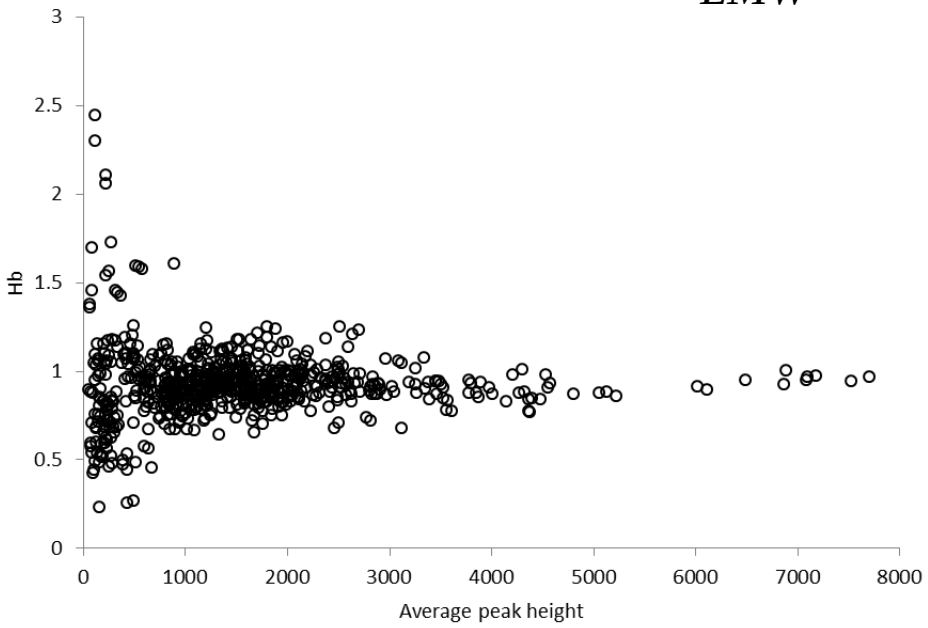
Constrained method of mixture interpretation:

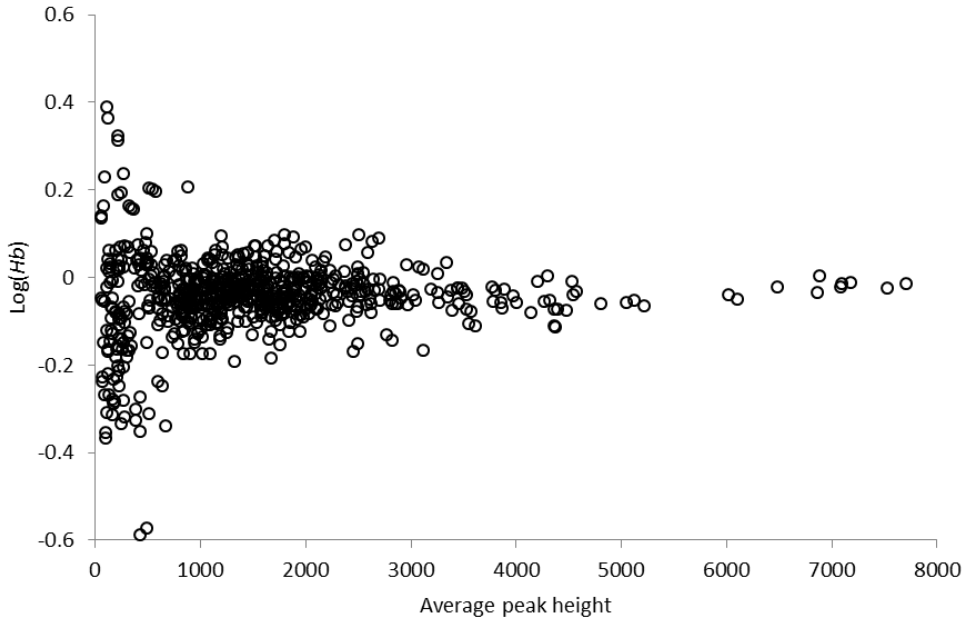
- **Write out all possible genotype combinations under H_2**
- **Exclude combos based on some set of heuristics:**
  - **Peak imbalance**
  - **Mixture proportion**
- **Simplify the H_2 (apply the sampling formula, 4.2)**

- **Uses more of the profiling data**
- **More time consuming**

Hb versus average peak height

$$Hb = \frac{O_{HMW}}{O_{LMW}}$$

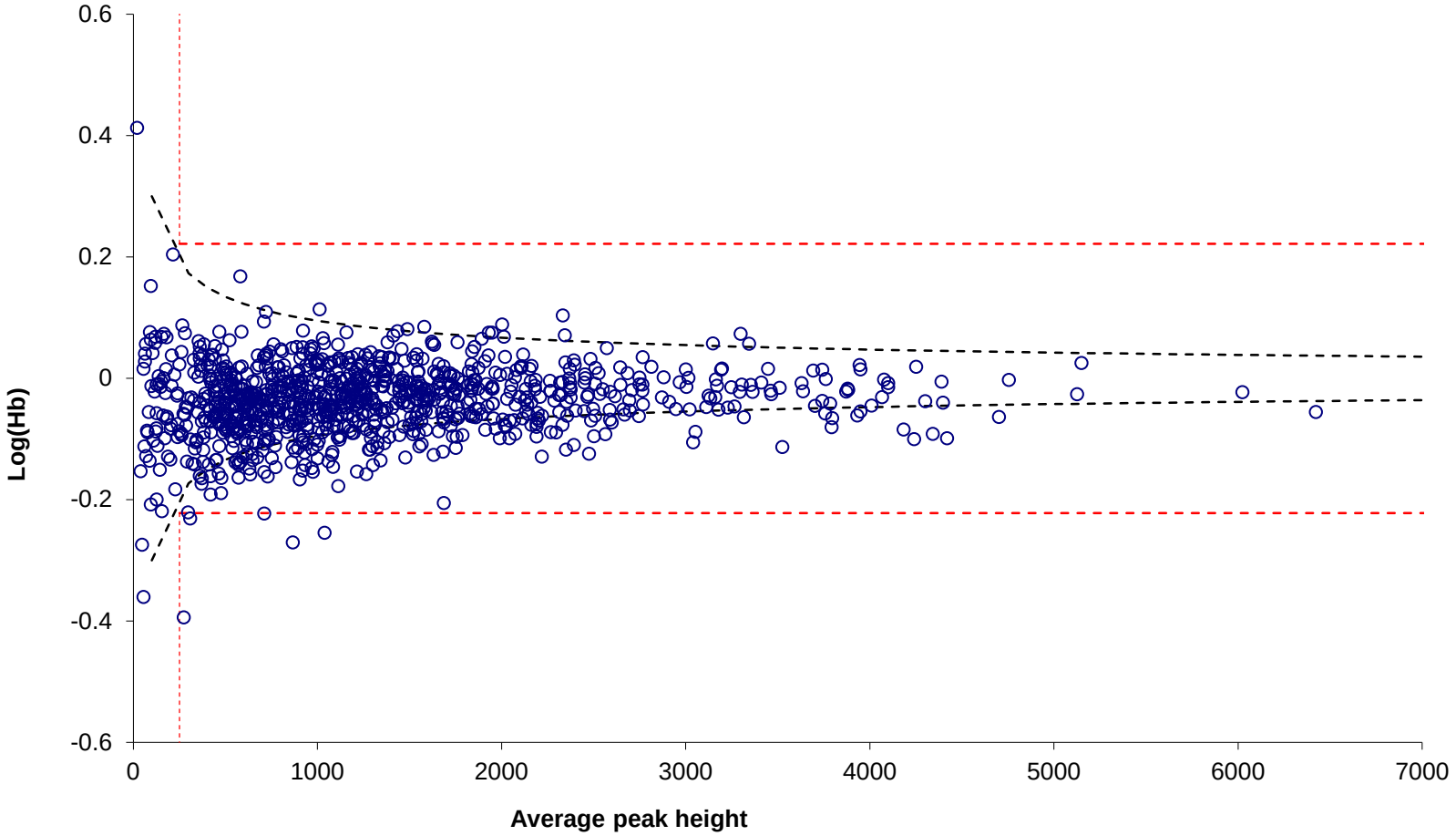




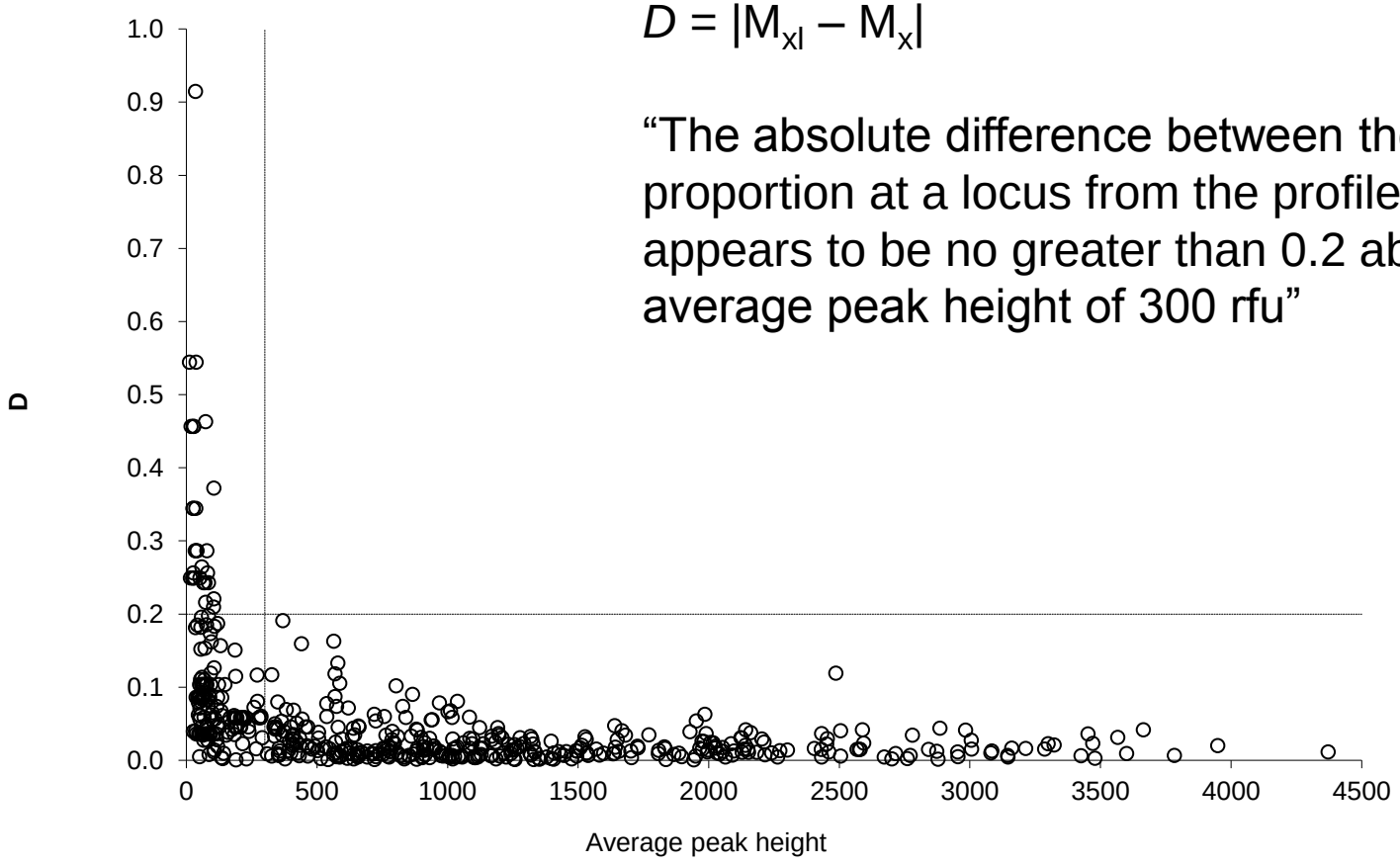
Variability of Hb

Conventional thresholds

95% intervals



Variability in mixture proportion



$D = |M_{xl} - M_x|$

"The absolute difference between the mixture proportion at a locus from the profile average appears to be no greater than 0.2 above an average peak height of 300 rfu"

Forensic Science International: Genetics

Volume 4, Issue 2, February 2010, Pages 111–114

Examination of the variability in mixed DNA profile parameters for the Identifiler™ multiplex

Jo-Anne Bright, Jnana Turkington, John Buckleton

ESR, 120 Mt Albert Road, PB 92021, Auckland, New Zealand

Forensic Science International: Genetics

Volume 6, Issue 6, December 2012, Pages 729–734

Analysis and biostatistical interpretation of complex and low template DNA samples

Modelling heterozygote balance in forensic DNA profiles

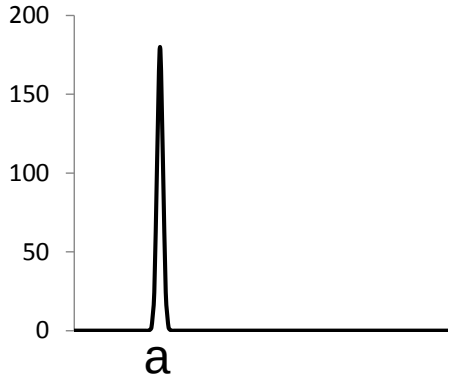
Hannah Kelly[a, b], Jo-Anne Bright[a], James M. Curran[b], John Buckleton[a]

[a] ESR, PB 92021, Auckland, New Zealand

[b] Department of Statistics, University of Auckland, PB 92019, Auckland, New Zealand

Dropout in a semi-quantitative method

- Traditionally, handled by dropping the locus or using the 2p rule
- The 2*p* rule assigns the probability $2p_a$ to the following profile



- Where p_a is the probability of allele *a*
- Assumed to be conservative in all circumstances…

ELSEVIER

Forensic Science International

Volume 159, Issues 2–3, 2 June 2006, Pages 206–209

Is the 2*p* rule always conservative?

John Buckleton[a], Christopher Triggs[b]

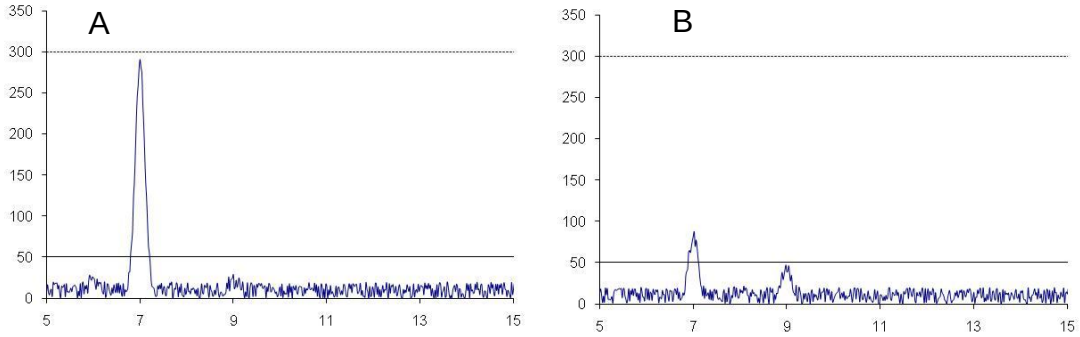
[a] ESR, Mount Albert Science Center, Private Bag 92021, Auckland, New Zealand

[b] Department of Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand

- … however this has proved a false assumption
- No longer recommended for use.

Non-concordance

Consider the following:



POI = 7,9

- **Two extremes**
- **A - large concordant 7 allele with no 9 peak observed (non-tolerable non-concordance)**
- **B - small concordant 7 allele with a non-concordant 9 peak visible sub-threshold (tolerable non-concordance)**

Non-concordances

- A locus where at least one allele of the POI is not observed in the profile
- Binary models cannot deal with a locus showing a non-concordance
- Motivator for change
- Also, how do we interpret 3 and 4 person mixtures?

About 2009…

- It was known that the binary method was not the most appropriate method
- Approaching end of “best before” date
- Very hands on – operator in control
- What were our options?
- Off the shelf solutions:
  - Expensive
  - Loss of control
  - Loss of expertise
- Could we extend the life of the binary?

Extensions of the binary model

- **Methods to extend the binary method to complex mixtures that have *no non-concordant alleles***
  - **There is no modification of the binary method that can deal with a non-concordant allele in a universally conservative manner**
- **Uses an unconstrained quantitative methods with F or Q alleles**
- **‘F ’ designation denotes an allele that may have dropped out or ‘failed’**
  - **Any allele at the locus in question, including alleles already observed**
- **Q designation represents any allele at the locus except for those alleles already present**

ELSEVIER

Forensic Science International: Genetics

Volume 6, Issue 2, March 2012, Pages 191–197

The interpretation of low level DNA mixtures

Hannah Kelly[a], Jo-Anne Bright[a], James Curran[b], John Buckleton[a]

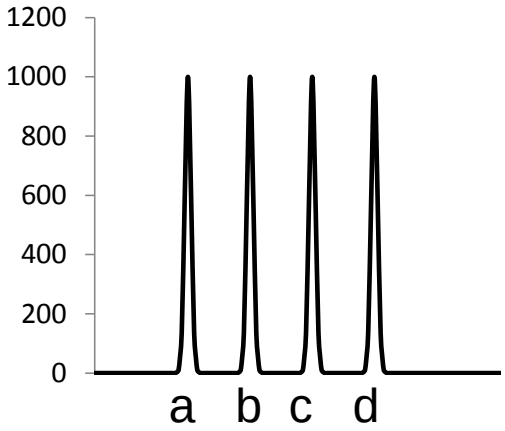
[a] ESR, PB 92021 Auckland, New Zealand

[b] Department of Statistics, University of Auckland, PB 92019 Auckland, New Zealand

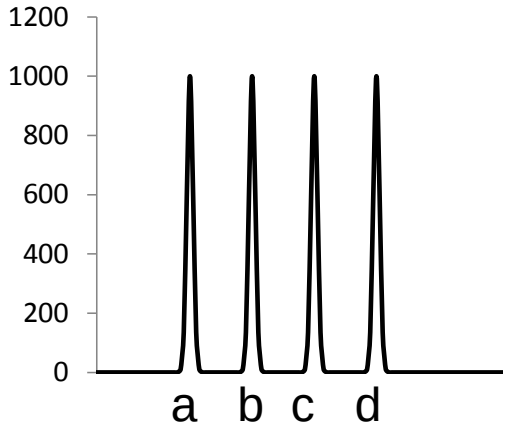
Introduction to concepts

- Consider 2 person mixture
- All peak heights above threshold
- Two reference samples from POIs
- H_1: POI 1 and POI 2
- H_2: Two unknowns

- Locus 1; 4 peaks; a b c d
- POI 1 = a,b POI 2 = c,d



Locus 1, 4 peaks



- Locus 1; 4 peaks; a b c d
- POI 1 = a,b POI 2 = c,d
- H_1: $Pr(E|H_1) = 1$
  - The hypothesis is fully explained by the evidence
  - The two POIs are contributors to the stain
- H_2: $Pr(E|H_2)$ = all possible combinations of alleles a, b, c, d

→ Write out all possible combinations

Locus 1 $Pr(E|H_2)$

C1	C2			Multipliers for reverse options	Product	Sum of products $Pr(E\|H_2)$
ab	cd	2 x $p_a p_b$	2 x $p_c p_d$	X 2	8 x $p_a p_b p_c p_d$	24Pr ($p_a p_b p_c p_d$)
ac	bd	2 x $p_a p_c$	2 x $p_b p_d$	X 2	8 x $p_a p_b p_c p_d$	
ad	bc	2 x $p_a p_d$	2 x $p_b p_c$	X 2	8 x $p_a p_b p_c p_d$	

Permutations and factorials

- **The number of possible permutations for a set of elements (alleles) can be determined using factorials**
- **Where:**

$$\frac{[\text{Total number of alleles}]!}{[\text{Individual allele count } a]!\,[\text{Individual allele count } b]!\,\text{etc}}$$

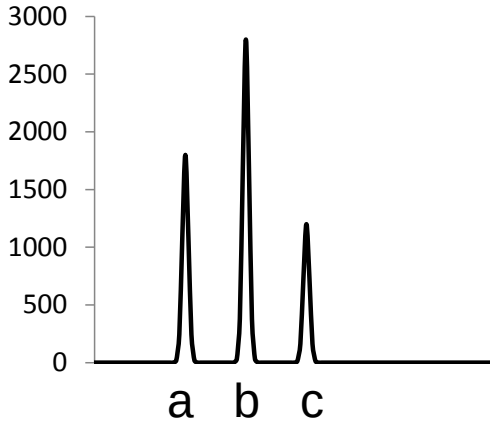
- **Locus 1 example:**

$$\frac{N!}{[n_a]!\,[n_b]!\,[n_c]!\,[n_d]!}$$

$$\frac{4!}{1!1!1!1!} = 24$$

→ $Pr(E|H_2) = 24Pr(p_a p_b p_c p_d)$

Locus 2, 3 peaks, 4 alleles



- **Crime profile: a, b, c**
- **POI 1: a,b POI 2: b,c**
- **$Pr(E|H_1) = 1$**
  - **The hypothesis is fully explained by the evidence**
  - **The two POIs are contributors to the stain**
- **$Pr(E|H_2)$ = all possible combinations of alleles a, b, c**

  → Write out all possible combos or use the permutation approach

Locus 2, combination approach

C1	C2			Multipliers for reverse options	Product	Sum of products $Pr(E\|H_2)$
aa	bc	p_a^2	$2 \times p_b p_c$	x2	$4p_a^2p_bp_c$	$12p_a^2p_bp_c$
ab	ac	$2 \times p_a p_b$	$2 \times p_a p_c$	x2	$8p_a^2p_bp_c$	
bb	ac	p_b^2	$2 \times p_a p_c$	x2	$4p_ap_b^2p_c$	$12p_ap_b^2p_c$
ab	bc	$2 \times p_a p_b$	$2 \times p_b p_c$	x2	$8p_ap_b^2p_c$	
cc	ab	p_c^2	$2 \times p_a p_b$	x2	$4p_ap_bp_c^2$	$12p_ap_bp_c^2$
ac	bc	$2 \times p_a p_c$	$2 \times p_b p_c$	x2	$8p_ap_bp_c^2$	

$$Pr(E|H_2) = 12p_a^2p_bp_c+12p_ap_b^2p_c+12p_ap_bp_c^2$$

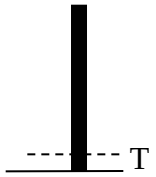
Locus 2, permutation approach

- 3 peaks, 4 alleles
- One of the a, b, or c alleles is shared
- Either aabc or abbc or abcc

$$\Pr(E \mid H_2) = \frac{4!}{2!1!1!} p_a^2 p_b p_c + \frac{4!}{1!2!1!} p_a p_b^2 p_c + \frac{4!}{1!1!2!} p_a p_b p_c^2$$

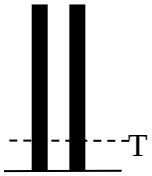
$$= 12 p_a^2 p_b p_c + 12 p_a p_b^2 p_c + 12 p_a p_b p_c^2$$

Peaks versus Alleles



One peak, 2 alleles (assuming one contributor)

One peak, 4 alleles (assuming two contributors, no D)



Two peaks, 2 alleles (assuming one contributor)

Two peaks, 4 alleles (assuming two contributors, no D)

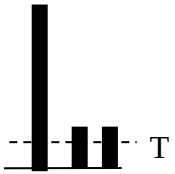
One peak, 1 allele

Three peaks, 3 alleles

Where T = stochastic threshold

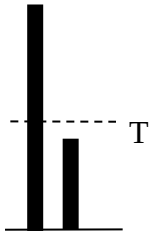
Peaks versus Alleles

- **When converting peaks to alleles can use a constrained approach**
- **Take into account imbalance**
- **Try for yourselves:**

1. 

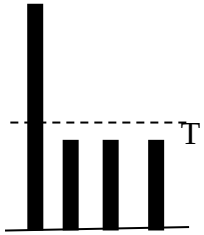
Assuming two contributors:

3 peaks, _____ alleles

2. 

Assuming two contributors:

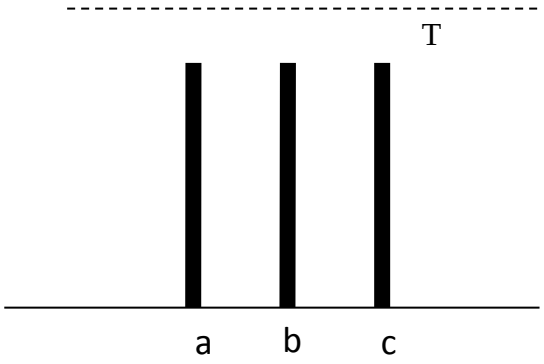
2 peaks, _____ alleles

3. 

Assuming three contributors:

4 peaks, _____ alleles

Harder examples



Example 1 - Considering dropout

- Assuming two person mixture
- Three peaks observed
- Given two contributors we're expecting to see four peaks…
- Introduce a Q allele:
  - aabc or abbc or abcc or abcQ
- Or introduce an F allele:
  - Could be an a, b, c, or any other
  - abcF

Example 1

$$\Pr(abcF) = \frac{4!}{2!1!1!} p_a^2 p_b p_c + \frac{4!}{1!2!1!} p_a p_b^2 p_c + \frac{4!}{1!1!2!} p_a p_b p_c^2 + \frac{4!}{1!1!1!1!} p_a p_b p_c p_Q$$

$$= 12 p_a p_b p_c [p_a + p_b + p_c + 2p_Q]$$

Where $p_Q = 1 - p_a - p_b - p_c$

Substitution $= 12 p_a p_b p_c \underbrace{[2 - p_a - p_b - p_c]}_{= <2}$

Then as a conservative approximation:

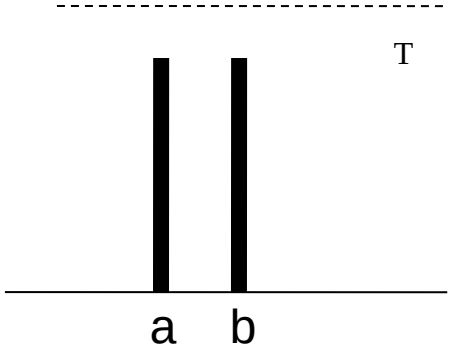
$$p_a p_b p_c F \approx 24 p_a p_b p_c$$

Other dropout examples – F approximation

Follow the steps:

1. **Ensure no non concordances**
2. **Convert peaks to alleles**
3. **Add in the required number of F alleles to make up the difference**
4. **Use permutation 'formula' (factorials) to determine the multipliers**
5. **Add in the ordinal and then cross out the Fs**

Example 2



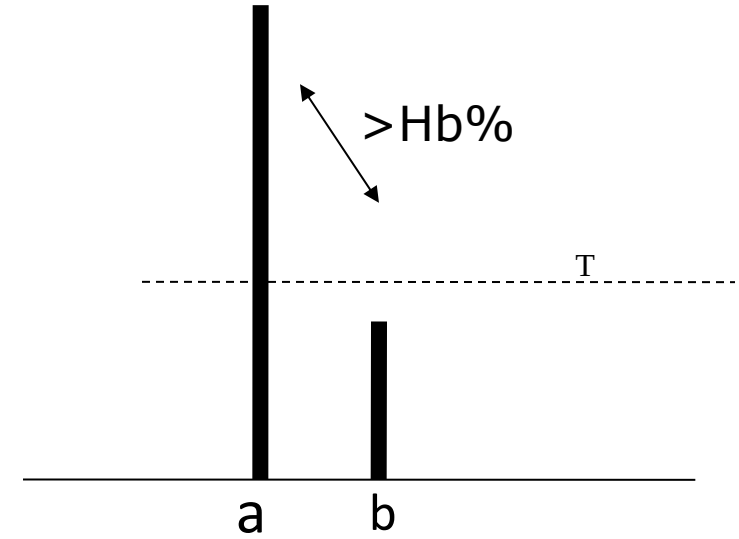
- Assuming two person mixture
- Convert peaks to alleles:
  - At least one a and one b allele
- Add in the required number of F alleles:
  - Two possible drops – FF
- Use permutation 'formula' (factorials) to determine the multipliers

$$abFF \approx \frac{4!}{1!1!2!}$$

- Add in the ordinal and then cross out the Fs

$$\approx 12\, p_a p_b$$

Example 3



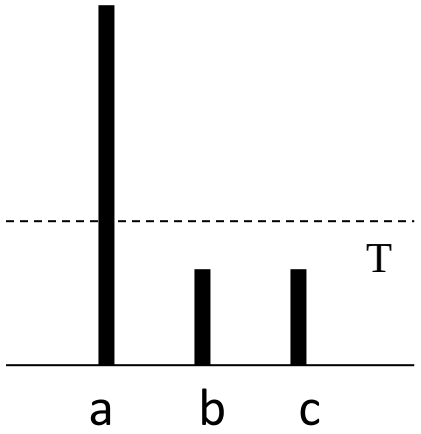
Follow the steps:

1. **Convert peaks to alleles**
2. **Add in the required number of F alleles to make up the difference**
3. **Use permutation 'formula' (factorials) to determine the multipliers**
4. **Add in the ordinal and then cross out the Fs**

$$aabF \approx \frac{\_!}{\_!\_!\_!} \_\_\_\_\_ \approx \_\_\_\_\_$$

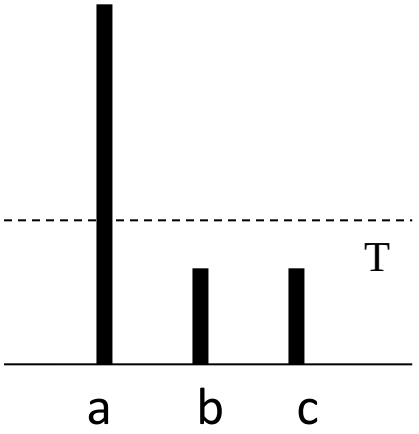
Example 4 with likelihood ratio

- Assume 2 contributors
- One suspect reference: ab
- Apply the ‘rules’.
- Peaks to alleles:
  - aabc
- Under H_1, unknown must be ac
- Under H_2, all combinations of aabc



Example 4, likelihood ratio

- Apply factorials
- Cancel where appropriate



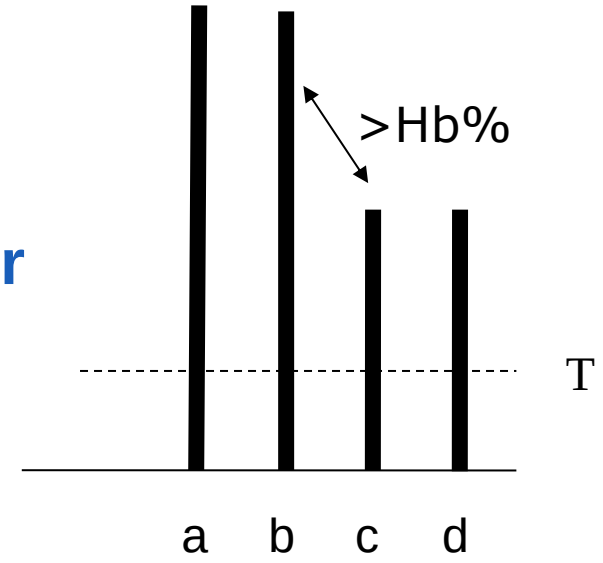
$$LR = \frac{\frac{2!}{1!1!} p_a p_c}{\frac{4!}{2!1!1!} p_a^2 p_b p_c}$$

$$= \frac{2 p_a p_c}{12 p_a^2 p_b p_c}$$

$$= \frac{1}{6 p_a p_b}$$

Example 5, 4 peaks, 4 alleles

- One POI: cd
- Assume 2 contributors, clear major
- Alleles: abcd
- Under H_1 unknown must be a,b
- Consider 2 unknowns under H_2



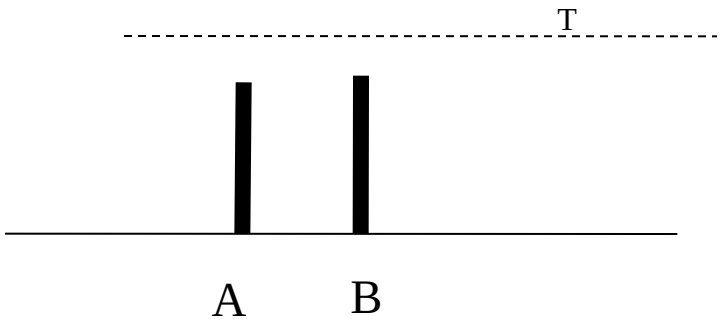
Example 5, likelihood ratio

- Apply factorials
- Cancel where appropriate

$$LR = \frac{\frac{2!}{1!1!} p_a p_b}{\frac{2!}{1!1!} p_a p_b \frac{2!}{1!1!} p_c p_d}$$

$$= \frac{1}{2 p_c p_d}$$

Example 6, considering dropout

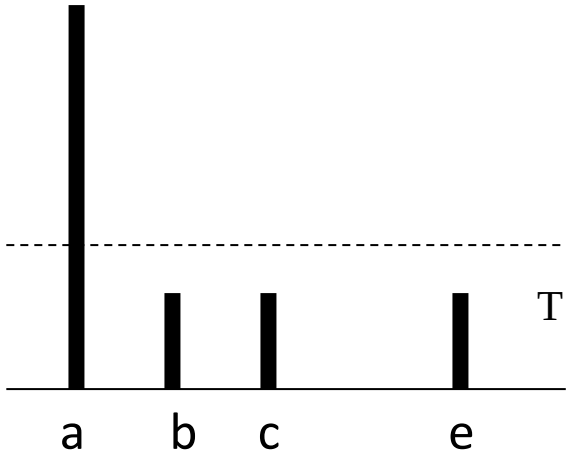


- One POI: ab
- Assume 2 contributors
- Alleles: abFF

Example 6, LR

$$LR = \frac{\frac{2!}{2!}FF}{\frac{4!}{1!1!2!}p_a p_b FF} = \frac{1}{12 p_a p_b}$$

Example 7, 3 contributors



- Assume 3 contributors
- Allele set: aabceF (total 6 alleles, possible dropout)
- If suspect = ab, factorials:

Example 7, likelihood ratio

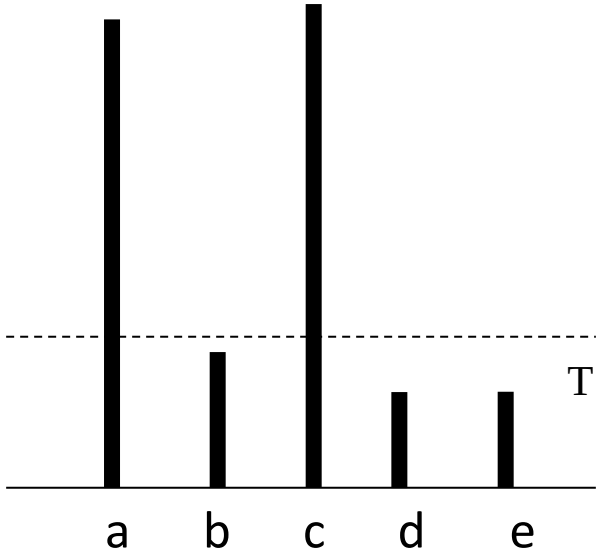
$$LR = \frac{\frac{4!}{1!1!1!1!} p_a p_c p_e F}{\frac{6!}{2!1!1!1!1!} p_a^2 p_b p_c p_e F}$$

$$= \frac{24 p_a p_c p_e}{360 p_a^2 p_b p_c p_e}$$

$$= \frac{p_a p_c p_e}{15 p_a^2 p_b p_c p_e}$$

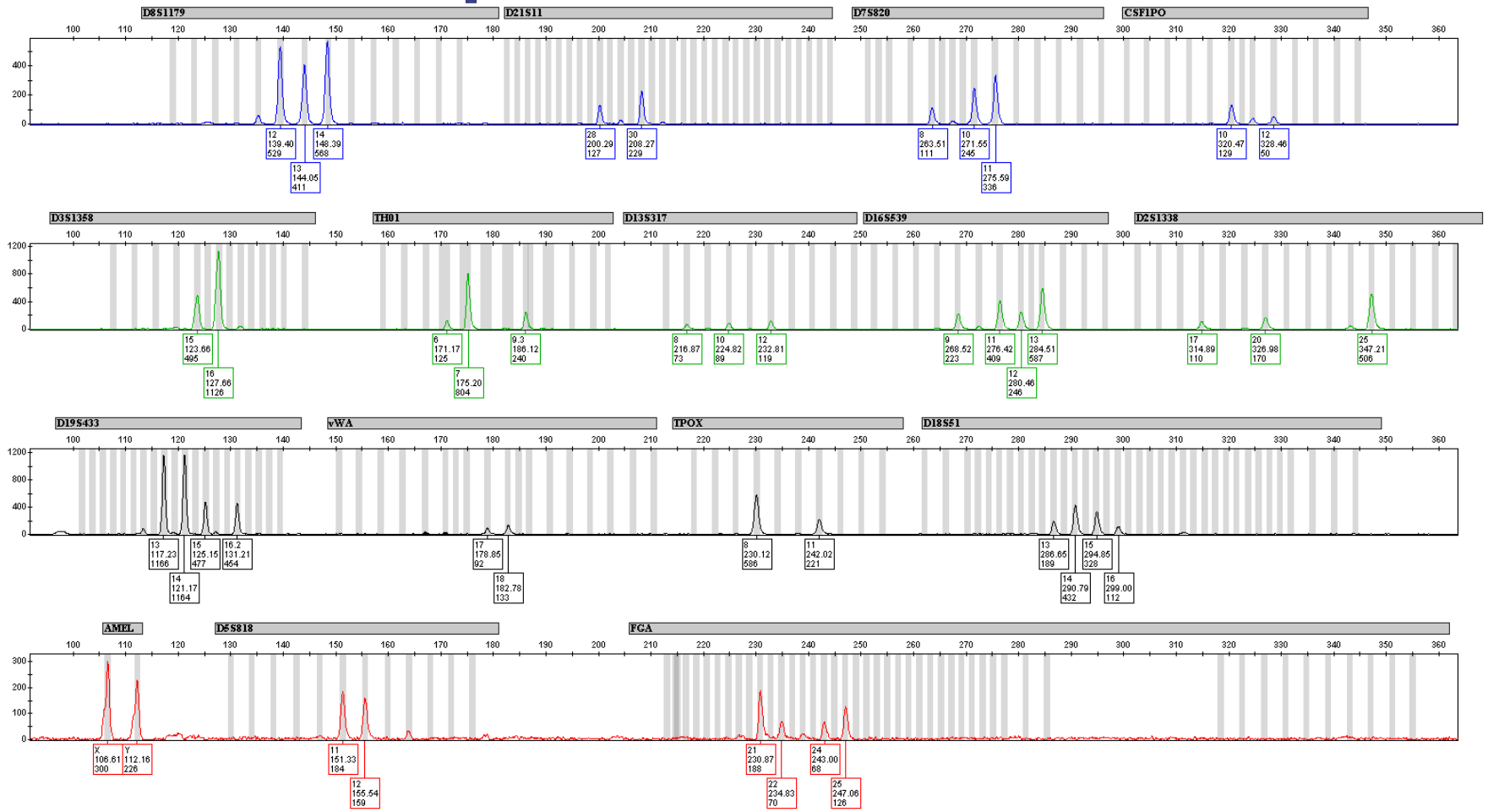
$$= \frac{1}{15 p_a p_b}$$

Example 8



1. Assume 3 contributors
2. Allele set: _____
3. If suspect = cd factorials:

Case example



Likelihood ratio

Marker	F model	Continuous model
D8S1179	3.04	13.61
D21S11	2.05	6.82
D7S820	1.62	1.20
CSF1PO	0.61	1.43
D3S1358	2.32	11.83
TH01	4.68	16.20
D13S317	10.66	7.49
D16S539	1.51	1.13
D2S1338	13.97	2.81
D19S433	0.98	5.80
vWA	1.49	5.90
TPOX	0.59	1.97
D18S51	3.47	0.76
D5S818	0.64	3.47
FGA	5.98	3.33
Total	1.71E+05	4.28E+08

Conclusion

- Can incorporate Fst correct for population substructure
- F (and Q) formula provided as appendices to Kelly et al. paper
- Easy to implement
- Wasteful of information
- Accounts for dropout but does not calculate the probability of dropout
- Recommend a model that makes more use of the profile data
  - Semi continuous or fully continuous model

Approaches to handling complex mixtures

ISFG basic mixture interpretation workshop

Jo-Anne Bright

Specialist Science Solutions

Manaaki Tangata Taiao Hoki
protecting people and their environment through science