Mixture examples - using a continuous method

ISFG Advanced topics in DNA interpretation

Specialist Science Solutions

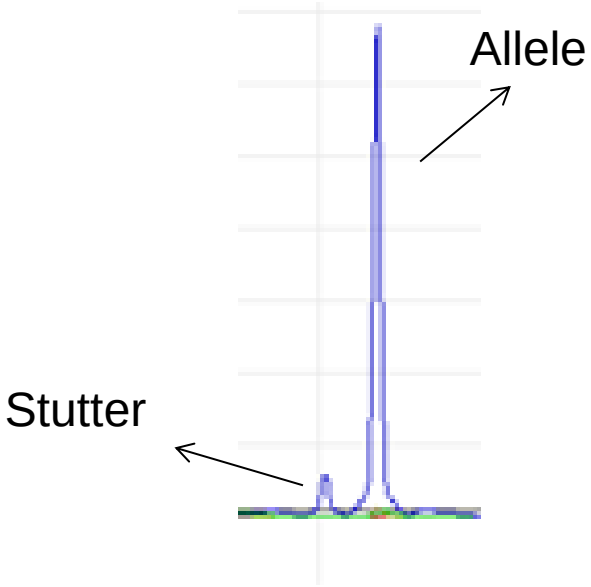
Manaaki Tangata Taiao Hoki
protecting people and their environment through science

Introduction

- Previously introduced a biological model
- Duncan touched on MCMC methods
- This talks combines both, demonstrating a how a continuous method of DNA interpretation works
- Includes a worked example

Total allelic product

- **STRmix models the ‘true’ (but unknown) amount of template DNA**
- **Total allelic product: allele plus stutter peak heights**
- **Modelled by mass parameters**
- **Exponential equation**



Modelling total allelic product

- **Mass of an allele at a locus is modelled by the mass parameters:**
  - **Slope d_n (degradation) and intercept t_n (template)**
- **Mass decreases with increasing molecular weight of an allele at a locus (m^l_a)**
- **Locus offset at each locus A^l (locus specific amplification efficiency)**

$$T^{\ell}_{an} = A^{\ell} t_n X^{\ell}_{an} \times e^{-d_n \times m^{\ell}_a}$$

Where X^l_{an} = dose, the count of allele a at locus l for contributor n:

Heterozygote = 1

Homozygote = 2

Estimating mass parameters

- **Determined by MCMC**
- **Starting state: randomly choose values for parameters**
  - **Genotype Set (Sj)**
  - **DNA amount (t_n)**
  - **Degradation (d_n)**
  - **Locus specific amplification efficiencies (A^l)**

Russian mathematician
Andrey Markov (1856-1922)

- **Calculate the probability of obtaining the observed profile given the genotype set and mass parameters; $\Pr(O|S_j,M)$ (Pr0)**

Estimating mass parameters

- A second set of parameters is proposed (step 1)
- Calculate $\Pr(O|S_j,M)$, the probability of obtaining the observed profile given the mass parameters (Pr1)
- If Pr1 >= Pr0 the proposed set of values are *accepted*
- If Pr1 < Pr0 then the proposed set of value is accepted only Pr1/Pr0 of the time
- If rejected, the proposed set of parameters are rejected and a new set of values are proposed

Estimating mass parameters

- For each step of the MCMC chain the mass parameters and a genotype set that differs at one locus are chosen
- Eventually the MCMC will reach ‘equilibrium’ where:
- DNA amount, degradation, and locus specific amplification efficiency are stable
- Limited number genotypes are chosen in proportion to their probability
- The amount of time the MCMC spends on each genotype is tallied and normalised to obtain *weightings* for use in the LR calculation

Peak height estimation

- **Use mass parameters to calculate total allelic product**

$$T_{an}^{\ell} = A^{\ell} t_n X_{an}^{\ell} \times e^{-d_n \times m_a^{\ell}}$$

- **The total allelic product from an allele is divided into stutter and allelic peak heights**
- **The height of the stutter and allelic peaks formed from allele *a* contributor *n* are calculated by:**

Allele

$$E_{an}^{\ell} = \frac{T_{an}^{\ell}}{1 + SR_a^{\ell}}$$

Stutter

$$E_{(a-1)n}^{\ell} = \frac{SR_a^{\ell}\left(T_{an}^{\ell}\right)}{1 + SR_a^{\ell}}$$

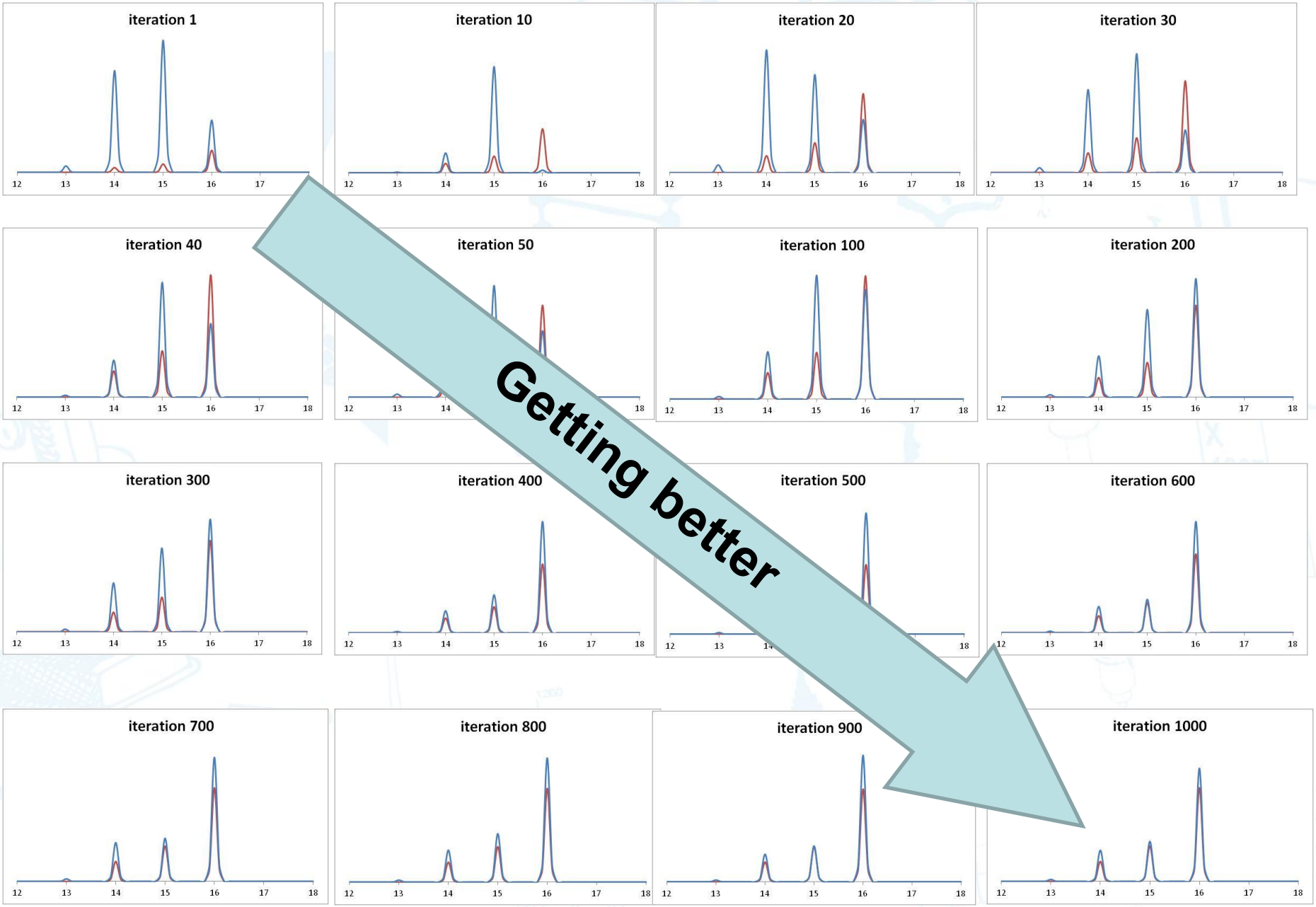
Model distribution

Assuming:

- **an approximate normal distribution,**
- **mean of zero,**
- **a variance = $\frac{c^2}{E_{an}^l}$ for the allele model,**
- **and a variance = $\frac{k^2}{E_{an}^l}$ for the stutter model, then:**

$$\log\left(\frac{O_{(a-1)}}{E_{(a-1)n}^l}\right) \sim N\left(0, \frac{k^2}{E_{an}^l}\right) \text{ for stutter}$$

$$\log\left(\frac{O_a}{E_{an}^l}\right) \sim N\left(0, \frac{c^2}{E_{an}^l}\right) \text{ for alleles}$$


iteration 1
iteration 10
iteration 20
iteration 30
iteration 40
iteration 50
iteration 100
iteration 200
iteration 300
iteration 400
iteration 500
iteration 600
iteration 700
iteration 800
iteration 900
iteration 1000
Getting better

Worked example

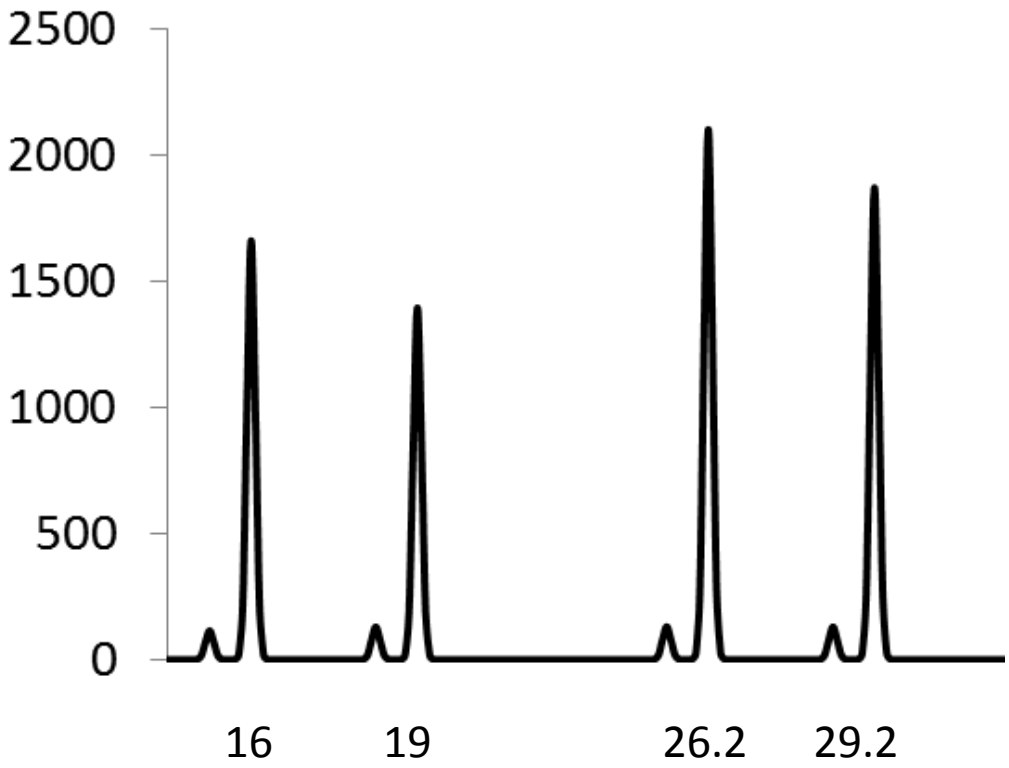
- A single locus profile example
- 8 peaks: 4 alleles, 4 stutter peaks
- I will provide mass parameters:
  - Slope d_n (degradation) and intercept t_n (template) for each contributor
  - Locus offset A^l (locus specific amplification efficiency)
- I will provide observed profile parameters:
  - Observed peak heights
  - Molecular weight all alleles (m^l_a)
  - Allele specific stutter ratios

Simplifications

- A single locus profile example
- 8 peaks: 4 alleles, 4 stutter peaks
- We'll use the same variance constant for alleles and stutter
- We won't change the mass parameters between steps – only the genotype combination

SE33

- **Two person mixture**



Allele	Height	m'_a
15	115	353
16	1659	357
18	129	365
19	1393	369
25.2	130	396
26.2	2100	400
28.2	129	408
29.2	1869	412

Possible genotype combinations

Number	Contributor 1		Contributor 2	
1	19	29.2	16	26.2
2	16	19	26.2	29.2
3	16	29.2	19	26.2
4	19	26.2	16	29.2
5	16	26.2	19	29.2
6	26.2	29.2	16	19

Other MCMC optimised parameters

	d_n	t_n
Contributor 1	0.0015	850
Contributor 2	0.001	800

A^{SE33}
0.8669

Variance
4

Complete the worksheet

In small groups, using your assigned genotype combination, complete the worksheet:

- **Calculate total allelic product for C1 and C2**
- **Calculate the expected stutter and allele heights**
- **Calculate $\Pr(O|S_j,M)$**
- **Report the resulting product**

MCMC process

- Start with genotype combination GC1
- Randomly propose new step (another GC)
  - By rolling six sided die
- Consider new step
- Is Pr(1) > Pr(0)? Accept new step. Add to tally, propose new step and repeat
- Is Pr(1) < Pr(0)? Accept new step only a fraction of the time when Pr(1)/Pr(0).
  - Roll the ‘probability die’
  - Add to relevant tally, propose new step and repeat
- Repeat thousands or millions of times!

Calculate likelihood ratio

Number	Contributor 1		Contributor 2		Product $\Pr(O\|S_j,M)$	Expected weight
1	19	29.2	16	26.2	26	0.208
2	16	19	26.2	29.2	76	0.608
3	16	29.2	19	26.2	7	0.056
4	19	26.2	16	29.2	12	0.096
5	16	26.2	19	29.2	3	0.024
6	26.2	29.2	16	19	1	0.008

$$LR_C = \frac{\sum_j w_j \Pr(S_j \mid H_1)}{\sum_u w_u \Pr(S_u \mid H_2)}$$

Likelihood ratio

- **Assuming person of interest was 16,19**
- **Assume that according to other (unseen) loci POI must be contributor 1**

Number	Contributor 1		Contributor 2		Product $\Pr(O\|S_j,M)$	Expected weight
1	19	29.2	16	26.2	26	0.208
2	16	19	26.2	29.2	76	0.608
3	16	29.2	19	26.2	7	0.056
4	19	26.2	16	29.2	12	0.096
5	16	26.2	19	29.2	3	0.024
6	26.2	29.2	16	19	1	0.008

Likelihood ratio

$$\Pr(E \mid H_1) = 0.608 \times 2 \times f_{26.2} \times f_{29.2}$$

$$\begin{aligned}\Pr(E \mid H_2) = {} & 0.208 \times 2 \times f_{19} \times f_{29.2} \times 2 \times f_{16} \times f_{26.2} + \\ & 0.608 \times 2 \times f_{16} \times f_{19} \times f_{26.2} \times f_{29.2} + \\ & 0.056 \times 2 \times f_{16} \times f_{29.2} \times 2 \times f_{19} \times f_{26.2} + \\ & 0.096 \times 2 \times f_{19} \times f_{26.2} \times 2 \times f_{16} \times f_{29.2} + \\ & 0.024 \times 2 \times f_{16} \times f_{26.2} \times 2 \times f_{19} \times f_{29.2} + \\ & 0.008 \times 2 \times f_{26.2} \times f_{29.2} \times 2 \times f_{16} \times f_{19} + \\ = {} & 4 f_{16} f_{19} f_{26.2} f_{29.2}\end{aligned}$$

Likelihood ratio, product rule

Allele	Frequency
16	0.0456
19	0.0659

$$LR = \frac{0.608 \times 2 f_{26.2} f_{29.2}}{4 f_{16} f_{19} f_{26.2} f_{29.2}}$$

$$= \frac{0.608}{2 f_{16} f_{19}}$$

$$= 101.2$$

Likelihood ratio, sampling formula

The sampling formula
(Balding and Nichols, 1994)

$$\frac{\left[\left(x\theta+\left(1-\theta\ \right)p_a\right)\right]}{\left[1+\left(n-1\ \right)\theta\right]}$$

$$LR=\frac{0.608\times 2\,f_{26.2}\,f_{29.2}}{4\,f_{16}\,f_{19}\,f_{26.2}\,f_{29.2}}$$

$$=\frac{0.608}{\frac{2(\theta+(1-\theta)\,f_{16})(\theta+(1-\theta)\,f_{19})}{(1+\theta)(2+\theta)}}$$

$$=8.5$$

Forensic Science International: Genetics

Volume 7, Issue 5, September 2013, Pages 516–528

Forensic population genetics – original research

The interpretation of single source and mixed DNA profiles

Duncan Taylor[a], Jo-Anne Bright[b], John Buckleton[b,]

[a] Forensic Science South Australia, 21 Divett Place, Adelaide, SA 5000, Australia

[b] ESR Ltd, Private Bag 92021, Auckland 1142, New Zealand

Forensic Science International: Genetics

Volume 7, Issue 2, February 2013, Pages 296–304

Developing allelic and stutter peak height models for a continuous method of DNA interpretation

Jo-Anne Bright[a, b], Duncan Taylor[c], James M. Curran[b], John S. Buckleton[a]

[a] ESR Ltd, Private Bag 92021, Auckland, New Zealand

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

[c] Forensic Science South Australia, 21 Divett Place, SA 5000, Australia

Mixture examples - using a continuous method

ISFG Advanced topics in DNA interpretation

Specialist Science Solutions

Manaaki Tangata Taiao Hoki
protecting people and their environment through science