Lessons from Lucia

gill@math.leidenuniv.nl
http://www.math.leidenuniv.nl/~gill

Overview of Lecture

- Theory (paradigms)
- Examples (Lucia, DNA mixture profile, Hariri)
- Conclusions

Theory: paradigms

- Bayes (one person statistics)
- Frequentist (two person, collaborative statistics)
- Likelihood (avoiding the issue)

Bayes' rule

- Posterior odds

  = prior odds * likelihood ratio

- Likelihood ratio

  = Prob(*data* | H_P) : Prob(*data* | H_D)

Bayesian/frequentist peaceful coexistence theorem:
{Decision theoretic admissible} = {Bayesian (for some prior)}

Examples:

1. Lucia

2. forensic DNA

3. Hariri assasination

Lucia: the data

JKZ MCU-1

Oct '00 – Sept '01

Lucia		incident: with	incident: without
	with	9	b133
	without	0	b887

RKZ-42

Aug – Nov '97

Lucia			
	with	b6	b52
	without	b9	272

RKZ-41

Aug – Nov '97

Lucia			
	with	1	bb0
	without	4	361

Lucia de B.

Reconstructie van een gerechtelijke dwaling

Ton Derksen

Lucia: time-line

- Sept. 4, 2001, "unexpected" death of Amber
- 2003: life sentence for 4 murders and 2 attempts; proof: statistical
- 2004: life sentence of 7 murders and 3 attempts; proof: medical
- 2006: confirmed by supreme court
- 2006: publication of book by Ton Derksen (philosopher of science)
- 2006: case submitted to special committee for review of exceptional possibly unsafe convictions

- 2008: CEAS reports death of Amber natural, recommends reopening
- 2008: "advocate-general" to supreme court admits there is no "novum", commissions further investigations
- 2009: AG recommends case is reopened (with "novum" if required: former key pathologist agrees with new findings – he had less information at his disposal
- 2009: supreme court accepts, case is reopened
- 2010: nonguilty verdict (all deaths natural; nurses behaviour exemplary; medical errors)

Lucia: likelihood ratio

- Hypothesis of the prosecution: (most of the) Lucia incidents are murders or attempted murders
- Hypothesis of the defence: the events are natural and would have happened anyway
- Prob(*data*|H_P):Prob(*data*|H_D)=1:1

Lucia: the original statistical analysis

- Frequentist approach; hypothesis test; null hypothesis = "balls in vases"
- For each of three data sets, court's statistician computed the "p-value" P(as extreme as Lucia or more | balls in vases model)
- For JKZ MCU-I, he multiplied by 26 (= # nurses worked on the ward that year)
- Product of three p-values = 1 in 342 million

Lucia: the defense

- Judge: "what is the probability the coincidence is due to chance?"
  - Defence 1. There are so many different probability models, you cannot compute a probability
    - Defence 2. Multiplying p-values is wrong (*reductio ad absurdam*)
- Judges: "we are not here to do thought experiments, but to determine facts"
- Judges: "The verdict of the court does not depend on a statistical computation of probabilities"

Lucia: the defense

- Judge: "what is the probability the coincidence is due to chance?"
  - Defence 1. There are so many different probability models, you cannot compute a probability
    - Defence 2. Multiplying p-values is wrong (*reductio ad absurdam*)
- Judges: "we are not here to do thought experiments, but to determine facts"
- Judges: "The verdict of the court does not depend on a statistical computation of probabilities"

No one checked the data!

- Three children responsible for multiple identical events, some in Lucia's shifts, some not
- No consistent definition of "incident"
- No consistent definition of "time of incident"
- The data suggested the hypothesis
- No-one studied the "normal" situation (clusters of events, clusters of shifts are normal)

Shifts

Court data

Corrected data

JKZ MCU-1

Oct '00 – Sept '01

Lucia

Lucia \ incident	with	without
with	9	b133
without	0	b887

Lucia \ incident	with	without
with	b7	b135
without	b4	b883

RKZ-42

Aug – Nov '97

Lucia

Lucia \ incident	with	without
with	b6	b52
without	b9	272

Lucia \ incident	with	without
with	b5	b53
without	10	271

RKZ-41

Aug – Nov '97

Lucia

Lucia \ incident	with	without
with	1	bb0
without	4	361

Lucia \ incident	with	without
with	1	bb2
without	4	359

Some p-values

- Cochran-Mantel-Haenszel test & Elffers' post-hoc correction
  1 in 916
- Ultimate stratification
  11 days at JKZ with both incident & Lucia on duty
  1 in 25
- Gamma(1) heterogeneity over Poisson intensity JKZ, RKZ pooled
  1 in 25

Aftermath

- Since 2010, no more media interest
- The legal system got the blaim, the taxpayer paid the bill
- There have been reforms, improvements, communication between legal and scientific communities
- Medical community is silent

Interview with president

Council for Justice

- “The system worked fine”
- “Murderers who escape conviction usually confess on their deathbed”

What really happened?

- In Dutch hospitals: 2000 deaths per year due to avoidable medical errors; culture of denial; frequent communication failures
- During 9 months up to 4 Sept. 2001, there was gossip about Lucia among nurses and specialists
- Medical errors by specialists were being associated with Lucia
- Director and top medical staff (but not all), under oath: there was no suspicion till 4 Sept. 2001

What really happened?

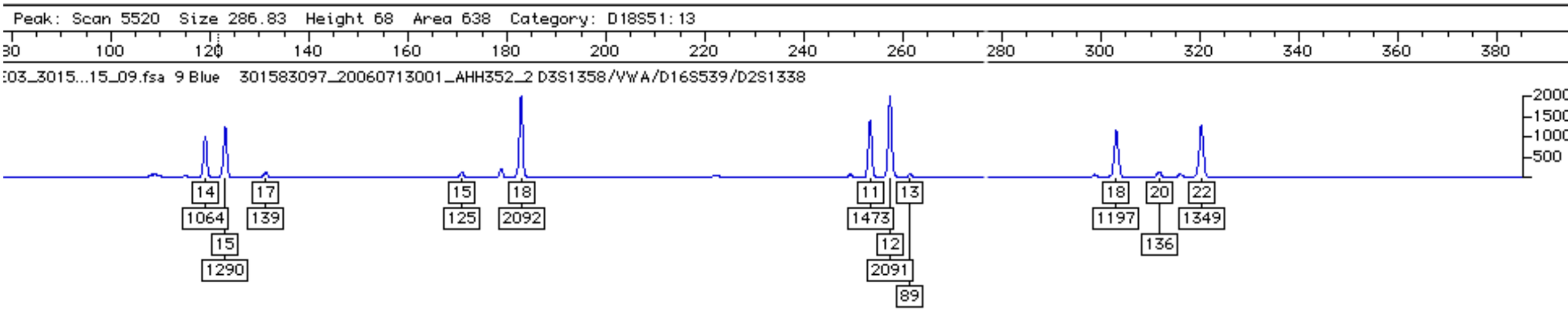
- No suspicion at all till 4 September, 2001? Director Paul Smits reported 10 unnatural deaths and suspicious reanimations, over last year, within 15 minutes of being informed of death of Amber, and on the very same day
- Strange fact: these 10 "incidents" were also reported to Health Inspectorate. Conclusion: nothing wrong.
- 4 medical specialists, it appears, have lied to police and to courts (and to one another) concerning the treatment of their own patients

Example 2: forensic DNA

Time for a success story...

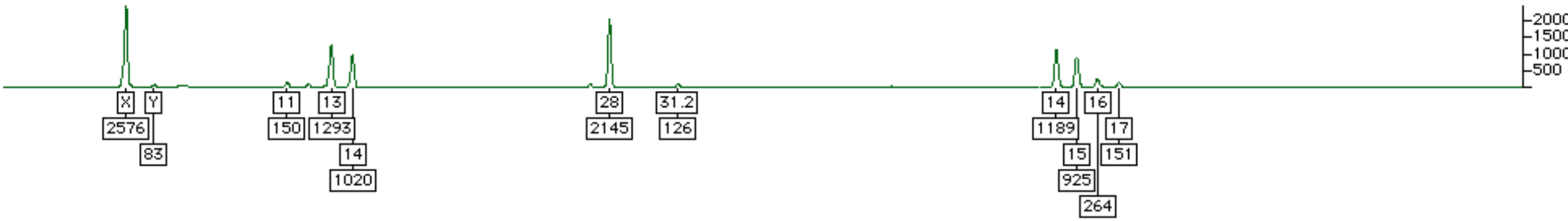
- Example: a problematic DNA profile
- First ever court application of recently developed "graphical models" (Bayes nets) for forensic DNA mixture profiles
- The methodology: Julia Mortera (Rome), Steffen Lauritzen (Oxford), Robert Cowell (London)
- The work: my master students Jasper van Wamelen, Giulia Cereda

Peak: Scan 5520 Size 286.83 Height 68 Area 638 Category: D18S51:13



:03_3015...15_09.fsa 9 Blue 301583097_20060713001_AHH352_2 D3S1358/VWA/D16S539/D2S1338

:03_3015...15_09.fsa 9 Green 301583097_20060713001_AHH352_2 AMEL/D8S1179/D21S11/D18S51

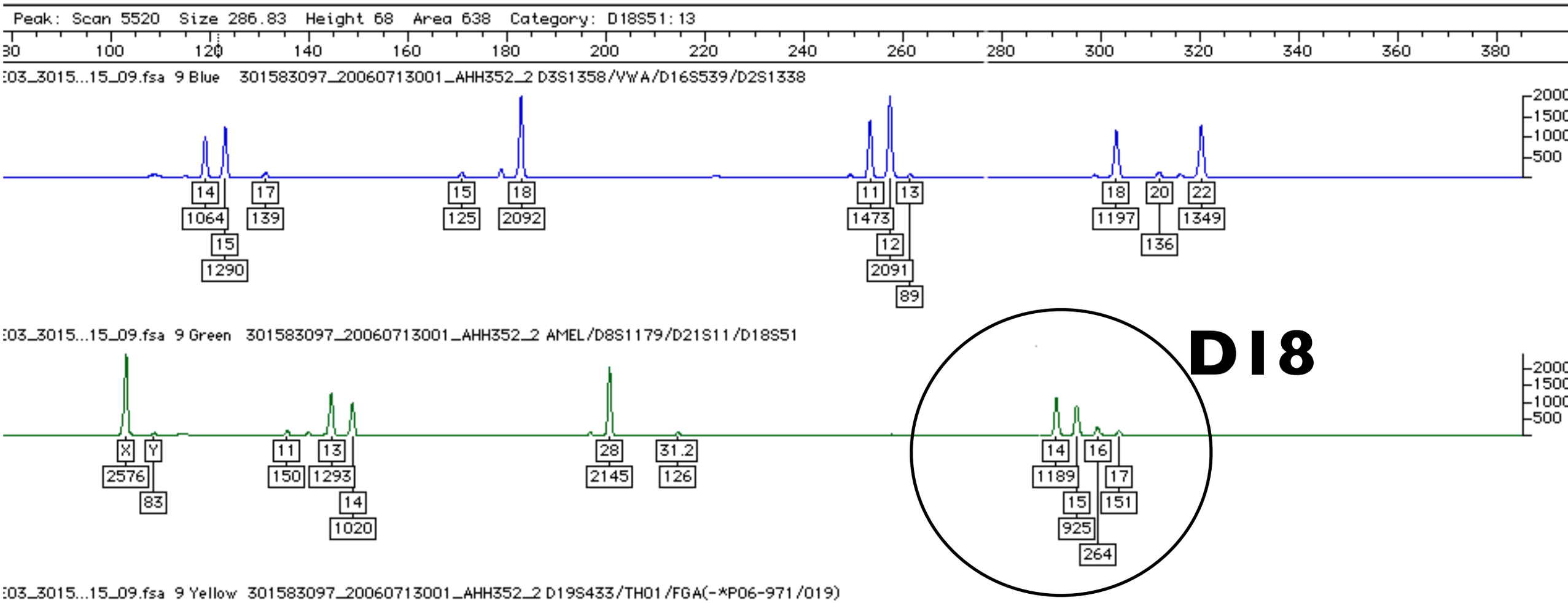


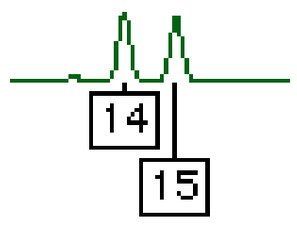
:03_3015...15_09.fsa 9 Yellow 301583097_20060713001_AHH352_2 D19S433/TH01/FGA(-*P06-971/019)

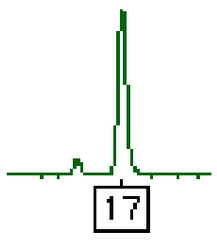
- Each group of peaks corresponds to one locus on a different chromosome
- Genotype of one person at one locus is pair (*m*,*n*), *m* ≤ *n* (numbers of repeats in two STR alleles), e.g. (7,9) or (18,18)
- *Relative* size of peak is roughly proportional to sum over contributors of:
  - 0, 1, or 2 depending on # alleles contributed
  - x proportion of contributor to mixture
- *Absolute size* of peak is (almost) irrelevant
- The peak sizes are definitely *random*, small peaks much more than large peaks

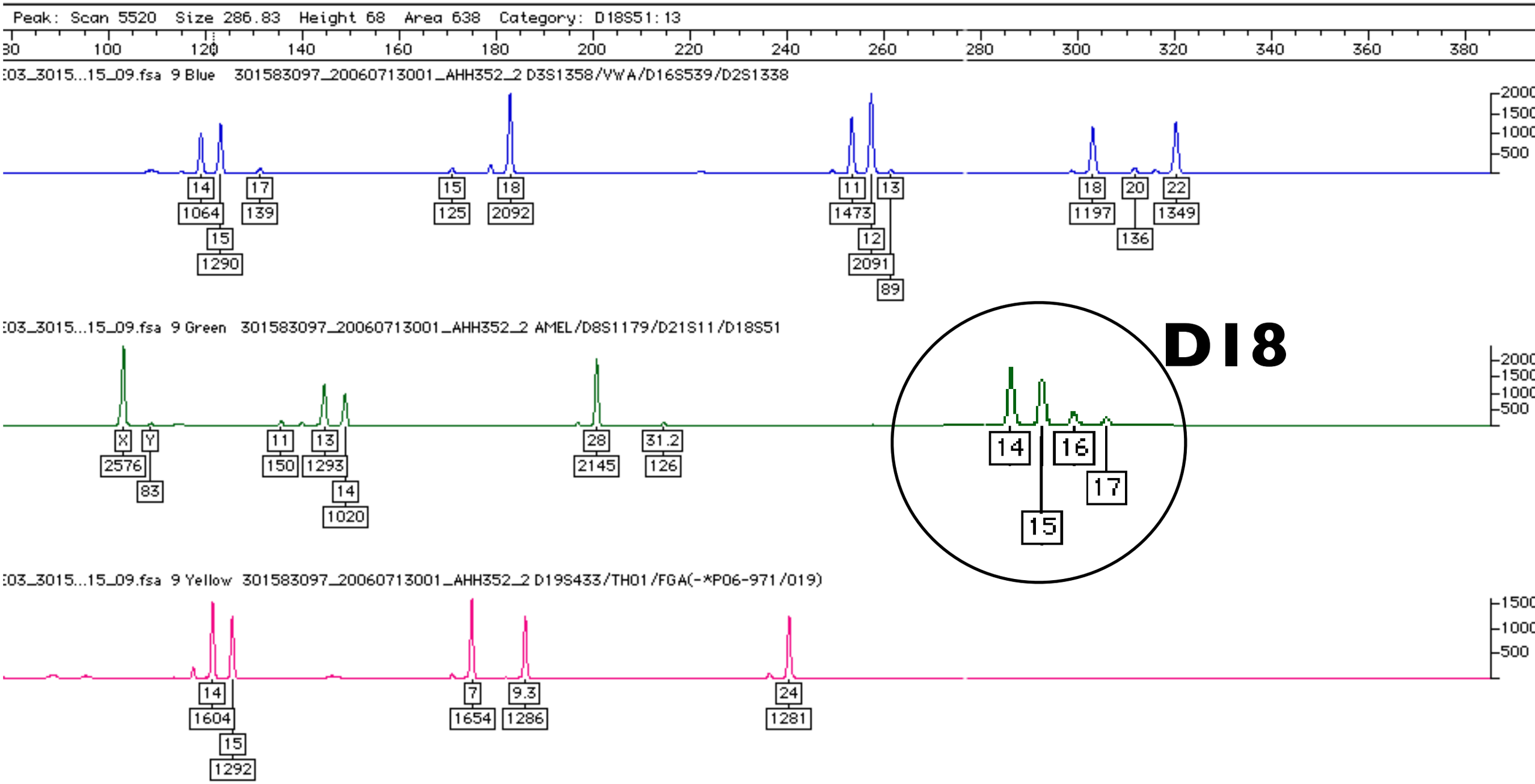
To bear in mind ...

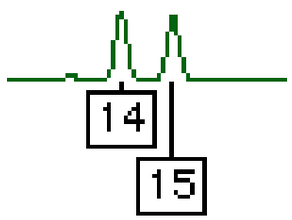
- Mendel's laws and relative frequencies of alleles in general population give us a fairly well understood model for the genotype of a random unknown person
- PCR procedure generates randomly sized peaks and suffers from "artefacts": stutter, dropout, silent alleles, mozaicism ...

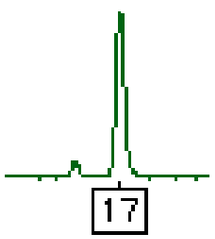


D18: victim 

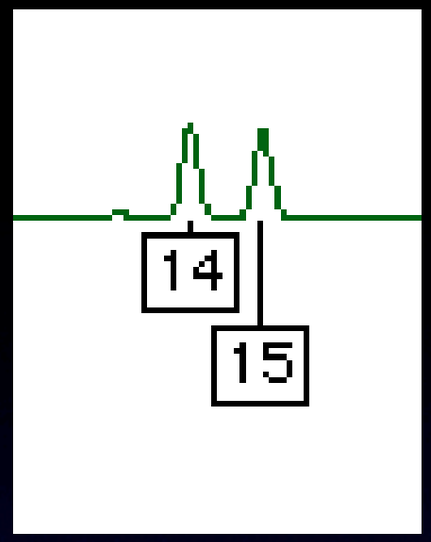
, suspect 



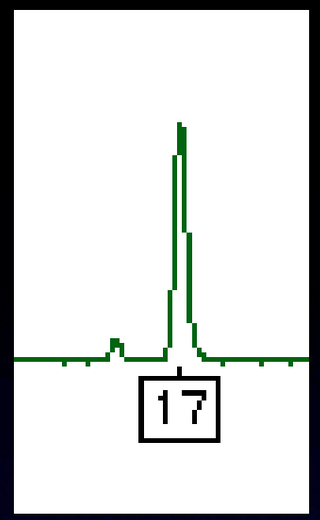
D18: victim 

, suspect 

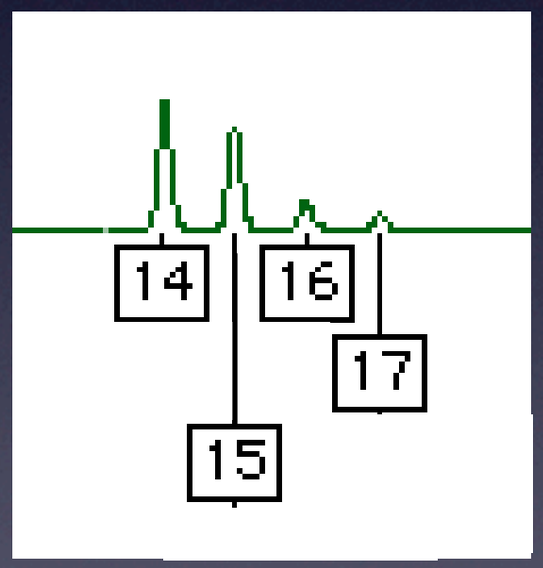
D18: victim



suspect



D18: "mixture profile" (fingernail scaping of victim)



Netherlands Forensic Institute: "If this trace is from two persons, we can exclude the suspect.
If it's from three, we can't"

NFI expert , following official (deterministic) interpretation rules:
“If this trace is from two persons, we can exclude the suspect. If it’s from three, we can’t”

The rest of the profile, visually, matches a 90-10 mixture victim-suspect rather well !

D18 is the most well-known locus for *mozaicism* which occurs there in *at least* 1 in 5000 persons !?

Review

Role of Short Tandem Repeat DNA in Forensic Casework in the UK—Past, Present, and Future Perspectives

BioTechniques 32:366-385 (February 2002)

Peter Gill
Forensic Science Service,
Birmingham, UK

ABSTRACT

The analysis of short tandem repeat (STR) DNA sequences is of fundamental importance to forensic science because they have become the recognized standard in constructing national public databases.

DEVELOPMENT OF MULTIPLEXED SYSTEMS

Early multiplexes consisted of few loci that were based on simple short tandem repeats (STRs). The four-locus "quadruplex" was probably the first to be widely used (44); because it consisted of few STRs, the match probability was consequently high—1 in 10 000. In 1996, a six-locus STR system (57,58) combined with the amelogenin sex test (61) was introduced—known as the

ous years, all six loci of the older SGM system were retained in the new Amp*Fl* STR SGM Plus system.

Development and Harmonization of National DNA Databases

The harmonization of STR loci has been achieved by collaboration at the international level. Notably, the European DNA profiling group (EDNAP) carried out a series of successful studies to identify and recommend STR loci for the

Somatic Mutation

If a somatic mutation occurs during embryological development, then two types of cells with different genotypes may coexist, and this leads to a three-banded profile (Figure 4). The peak areas will depend on the relative proportion of the mutant cell and will not be equivalent. This is arguably the most difficult condition to elucidate because it is possible that not all tissues will demonstrate somatic mutation. The incidence of somatic mutation is variable—out of 120000 samples, not one has been observed at the HUMTH01 locus, whereas the incidence is approximately 1 in 5000 at the D18S51 and HUMFIBRA loci. It is possible that some somatic mutations will be indistinguishable from stutters; therefore, these figures are probably underestimates because they are only recorded if unambiguous.

The genetic phenomena described (trisomy, translocation, and primer binding site mutations) can be verified by the analysis of the reference sample, which should also demonstrate the same anomaly unless a tissue-specific somatic mutation has occurred. In the latter case, confirmation may depend on a reference sample that has the same origin as the case stain, although we cannot completely rule out the possibility that the appearance of somatic mutations could vary over time within tissues such as the buccal lining, which consists of rapidly dividing cells.

To summarize, an understanding of the behavior of the DNA profiling system is important to assess potential mixtures. Loci will behave somewhat differently from each other, but it is possible to generalize. Here are some of the key features: (*i*) the smallest peak area of a heterozygote will usually be greater than 60% of the size of its partner (peak area or peak height); (*ii*) within the previous guideline, the high molecular weight peak is often smaller than the low mol-

Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International: Genetics Supplement Series 1 (2008) 640–642
www.elsevier.com/locate/FSIGSS

Research articles

Probabilistic modelling for DNA mixture analysis

R.G. Cowell [a,*], S.L. Lauritzen [b], J. Mortera [c]

[a] *Cass Business School, 106 Bunhill Row, London EC1Y 8TZ, UK*
[b] *Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK*
[c] *Dipartimento di Economia, Università Roma Tre, Via Silvio D'Amico, 77, 00145 Roma, Italy*

Received 13 August 2007; accepted 7 October 2007

Abstract

Taking peak area information into account when analysing STR DNA mixtures is acknowledged to be a difficult task. There have been a number of non-probabilistic approaches proposed in the literature, and some have been incorporated into computer systems, but comparatively little has been published from a probabilistic perspective. Here we briefly review our previous work on using Bayesian networks to analyse two-person mixtures within a probabilistic framework, and present preliminary results obtained for analysing two-person and three-person mixtures that combine peak area information from multiple independent samples.

Keywords: DNA mixtures; Bayesian networks; Multiple traces; Multiple contributors

1. Introduction

In a recent series of papers [1–3] we have presented a probabilistic methodology for analysing peak area information from DNA mixtures based on Bayesian networks. A representative fragment of these networks is shown in Fig. 1

apparatus *after* amplification of the mixture sample. We model the stochastic variations in these areas by Gamma distributions, where the Gamma distribution of the area for allele a depends on the mean μ_a and has expectation proportional to μ_a; similarly for alleles b and c. For further details of the Gamma model and Bayesian networks, and how the probability

Basis model: one locus, three adjacent STR numbers

R.G. Cowell et al. / Forensic Science International: Genetics Supplement Series 1 (2008) 640–642

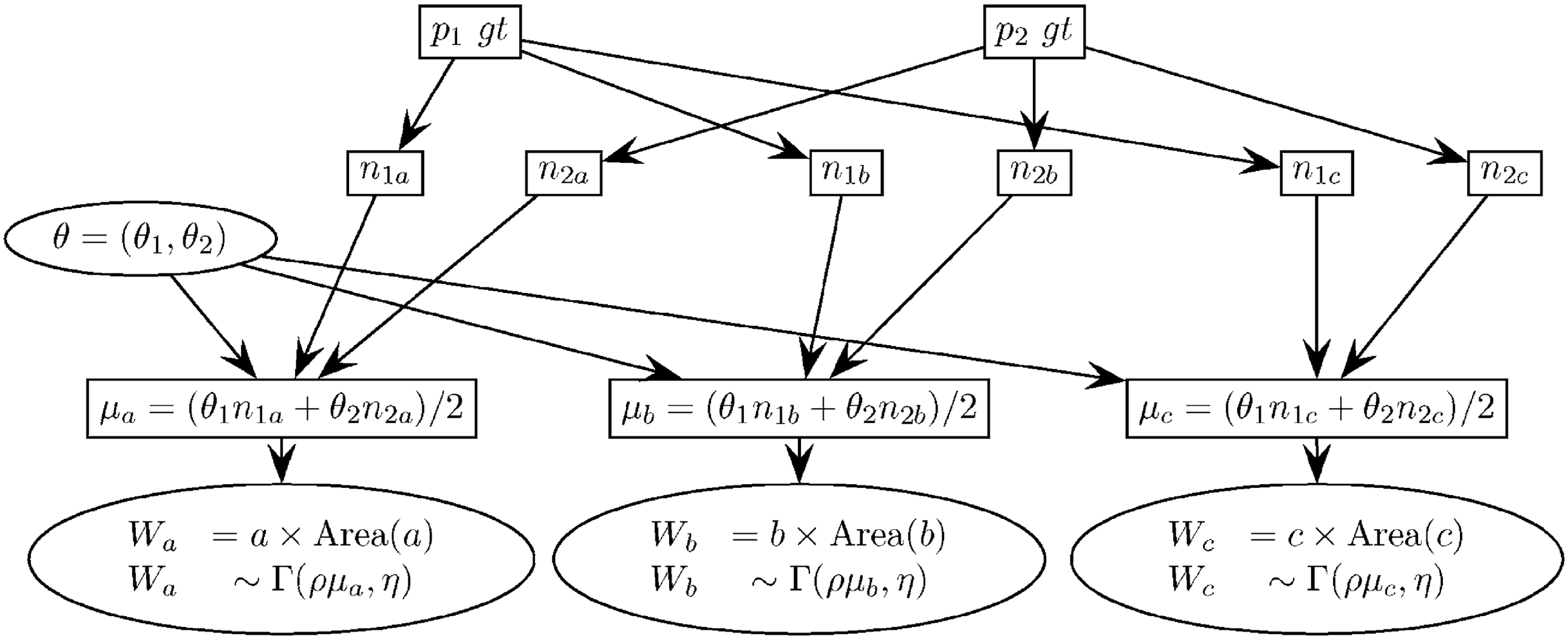
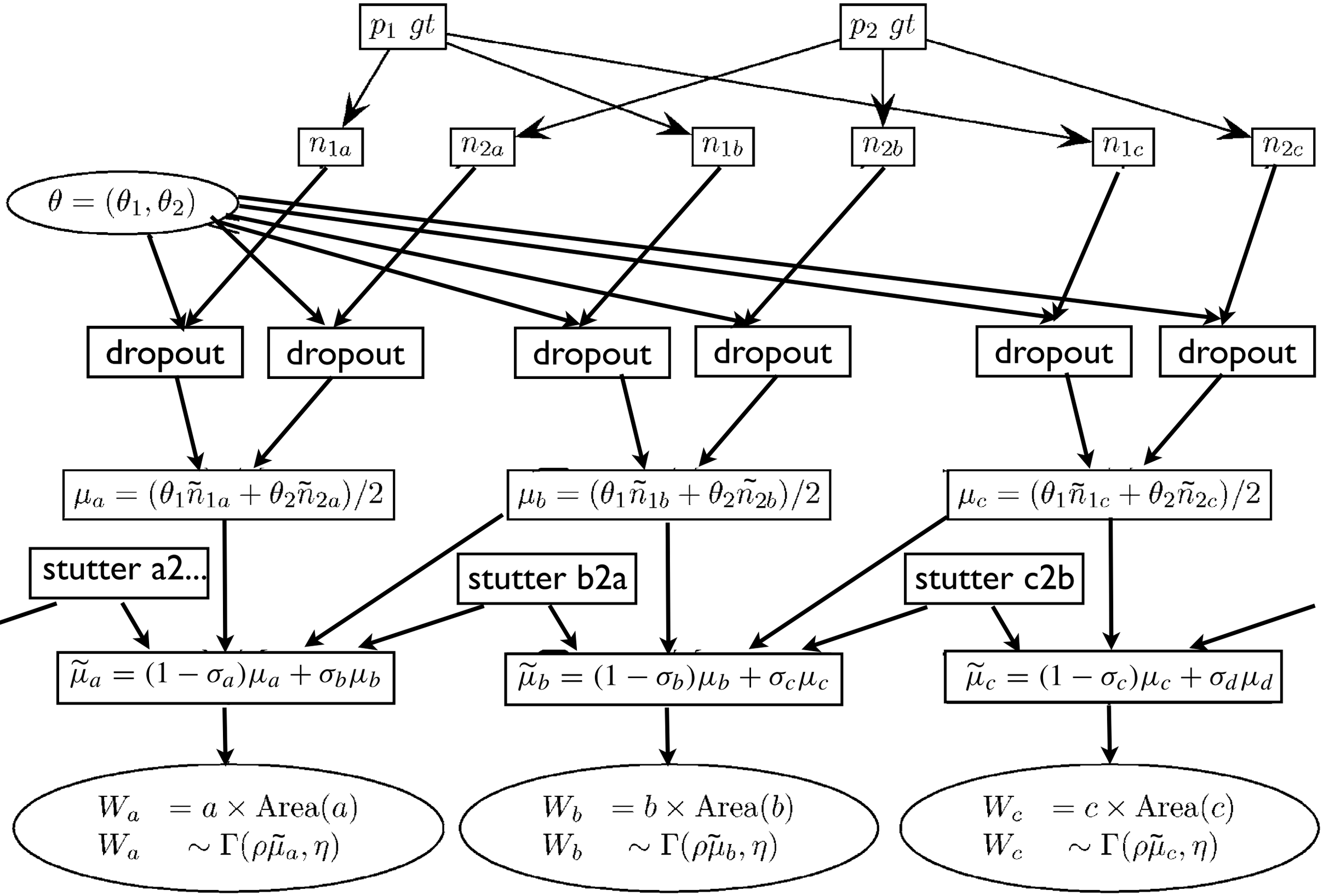
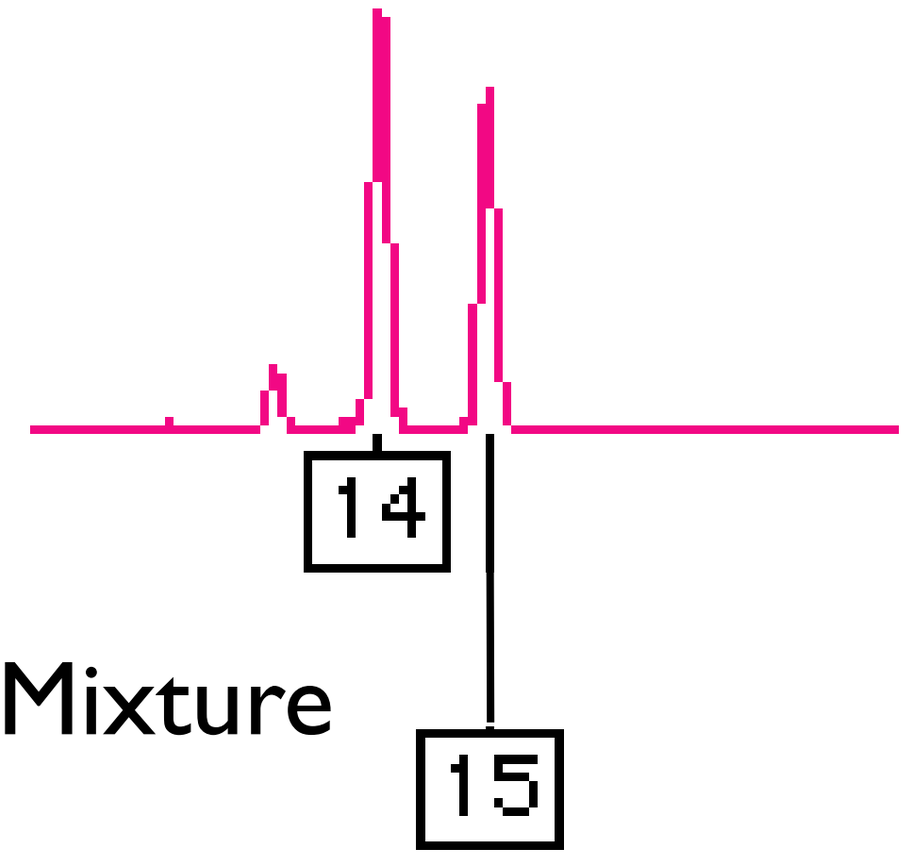


Fig. 1. Bayesian network fragment for modelling peak areas in a mixture.

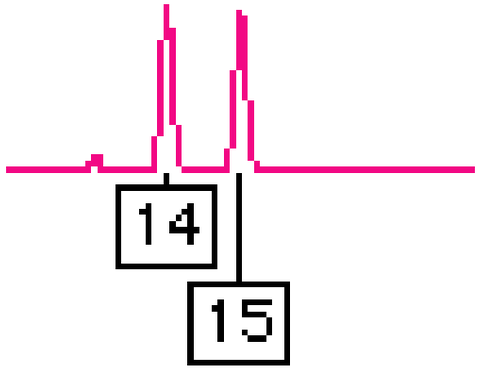
Extended model (with dropout and stutter)



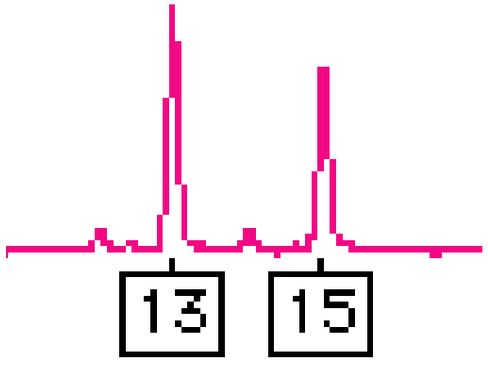
Locus D19 (not problematic...?)



Mixture

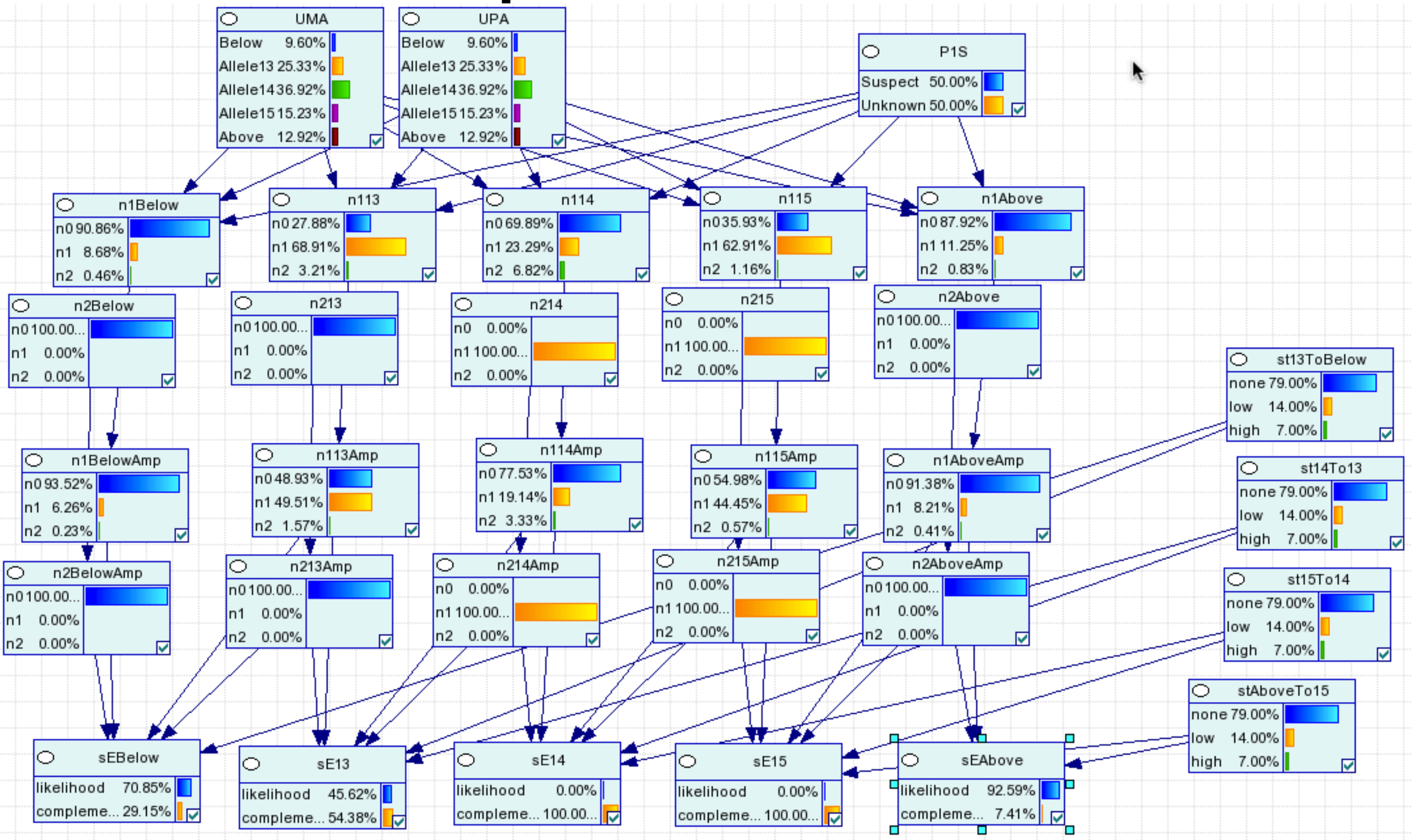


Victim



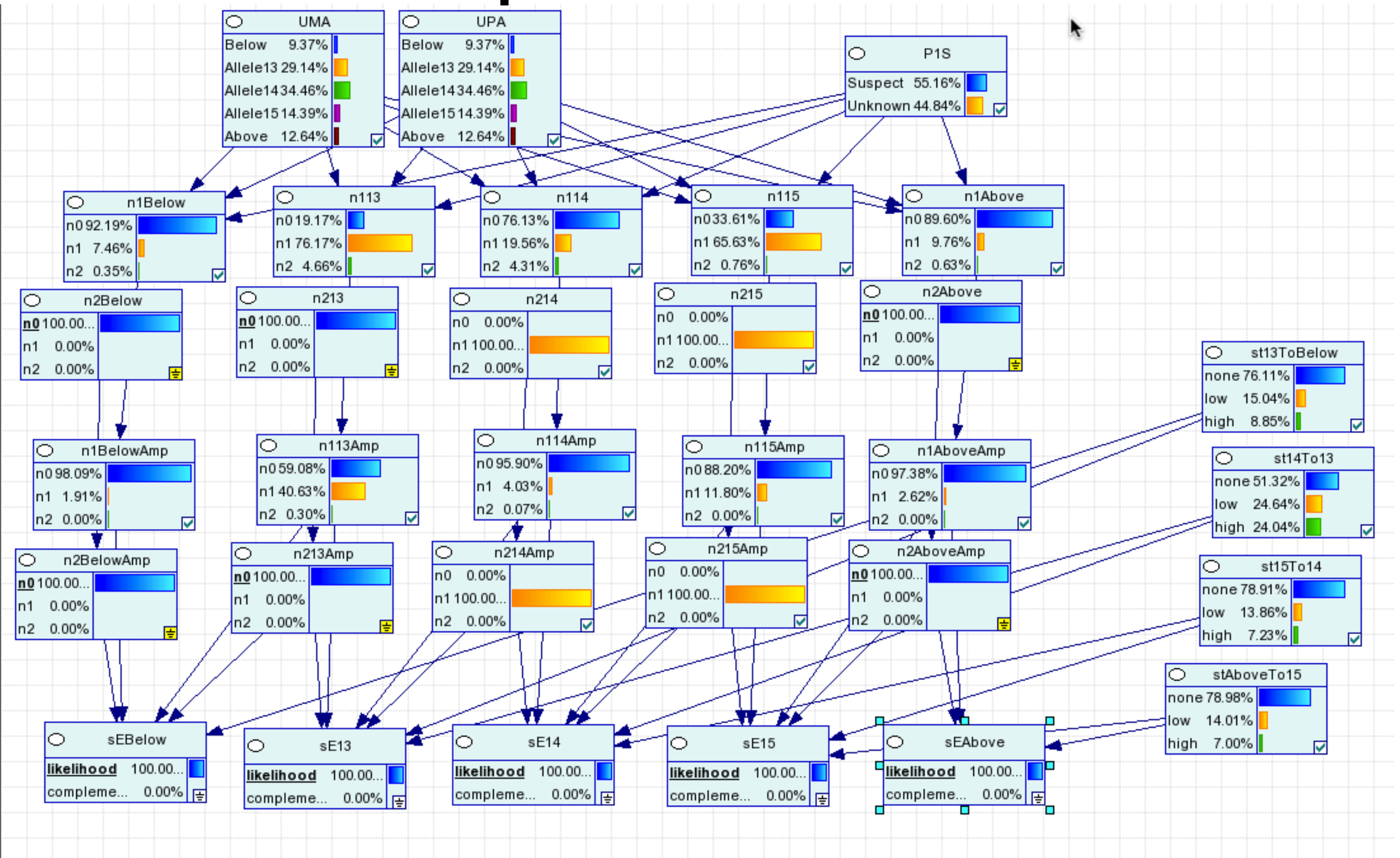
Suspect

Locus D19: prior



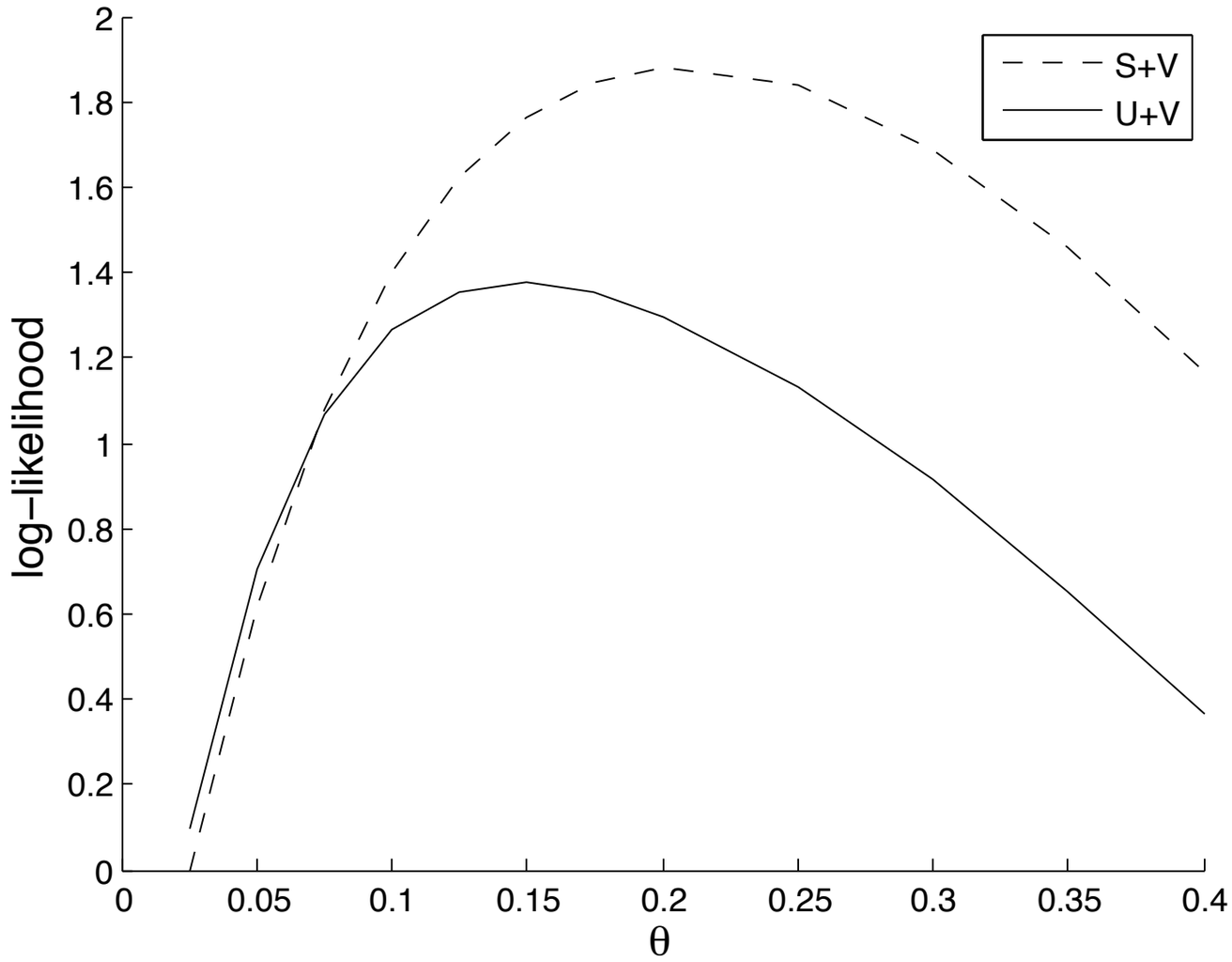
Assume 90:10 mixture

Locus D19: posterior



Assume 90:10 mixture

What does D19 say about the mixture proportion?



Defence would like $\theta < 0.08$, prosecution $\theta > 0.25$

Combining three replicate mixture profiles:

Evidential value for
victim+suspect+mosaicism versus victim+unknown
is around 10 000 : 1 ("very strong evidence")

Evidential value for
victim+suspect+unknown versus victim+unknown
+unknown
is also around 10 000 : 1 ("very strong evidence")

Free tools: GeNIe, matlab (octave?), R

GeNIe: http://genie.sis.pitt.edu/

Experience

- Important to over-estimate "noise parameters" in model to compensate for misspecification – otherwise you will overestimate weight of evidence
- Once you have done that, our results were pretty robust to unknown nuisance parameters
- The courts, and the DNA specialists, are not ready for this

New technologies

- New technologies are generating new forensic data of frightening dimension, little scientific understanding
- Plug-in methods (fit models using training data, then “estimate” likelihood ratio) tends to grossly overestimates weight of evidence
- Present research plans: tune fitting to task
- Courts won’t be ready for this for many years to come (and shouldn’t be)

Example 3: Hariri assasination

Hariri assasination

- UN Lebanon tribunal has published indictment of alleged members of terrorist gang
- Evidence: apparent cell phone colocation
- Trial is about to commence
- Question: what is chance of chance (apparent/false/...) colocation?

First idea (not mine)

- Make model of random people moving around Lebanon making random phone calls to one another
- Very simple models: chance of chance colocation falls off exponentially fast to zero, as time period grows longer
- But: how many pairs of cell phone users are there????
- And anyway: how relevant to actual case?

Present ideas

- The more false leads, and the faster they can be rejected, the stronger is the evidence
- The more interesting the movements and the more frequent the calls, the stronger is the evidence
- Smart data interogation allowed investigators to generate hypotheses, and narrow leads to one, expending only couple days of data. We have (e.g.) five more weeks to confirm ... by doing our best to reject!!!

Conclusions

- Forensic statistics is in its infancy
- It requires non-standard paradigms and will need new methodology
    - Multiparty statistics
    - Nuisance parameters
    - Model the forensic investigation process
- Communication of statistical ideas to non-statisticians is going to be the bottle-neck

gill@math.leidenuniv.nl
http://www.math.leidenuniv.nl/~gill