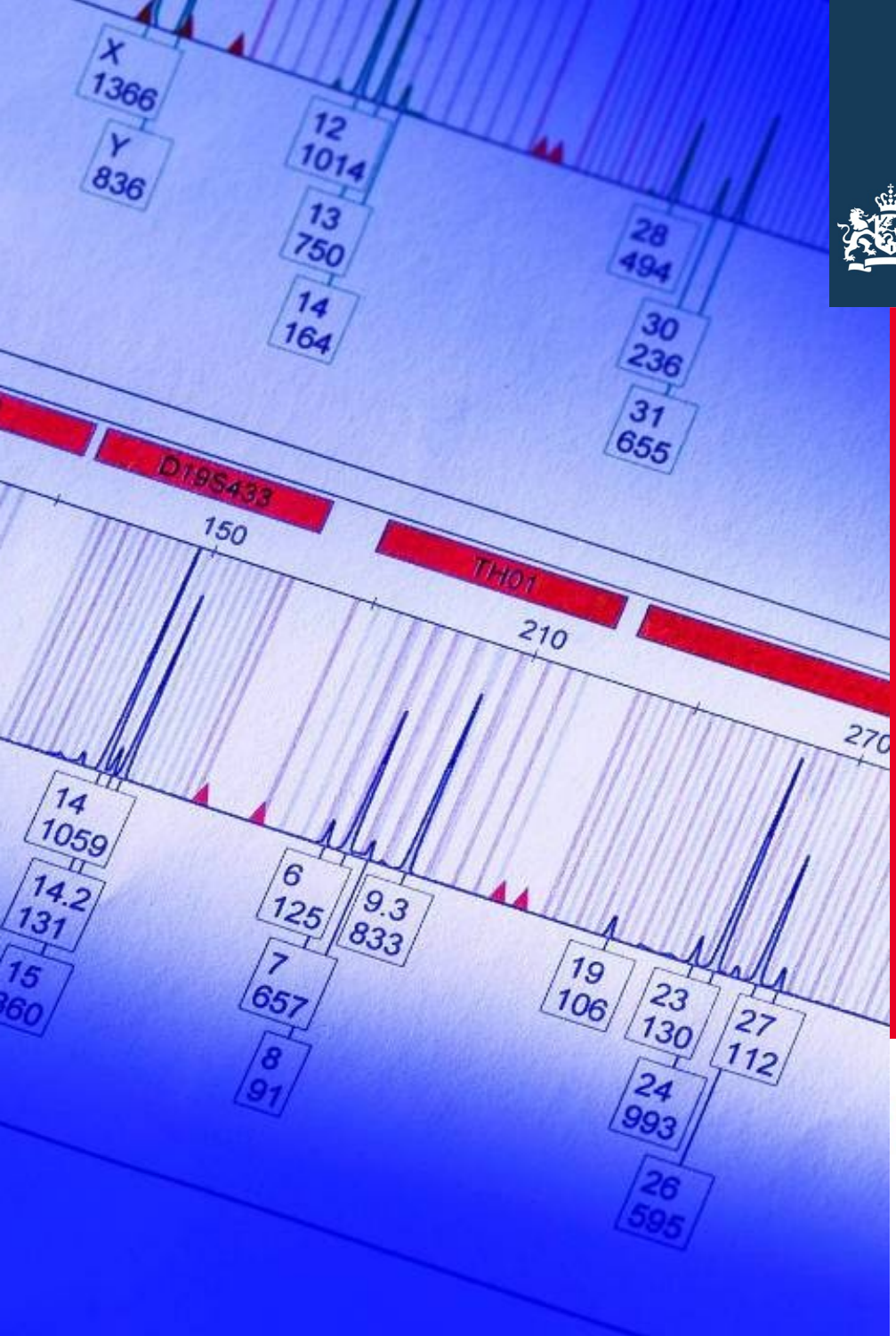


Netherlands Forensic Institute
Ministry of Security and Justice

The Likelihood Ratio model

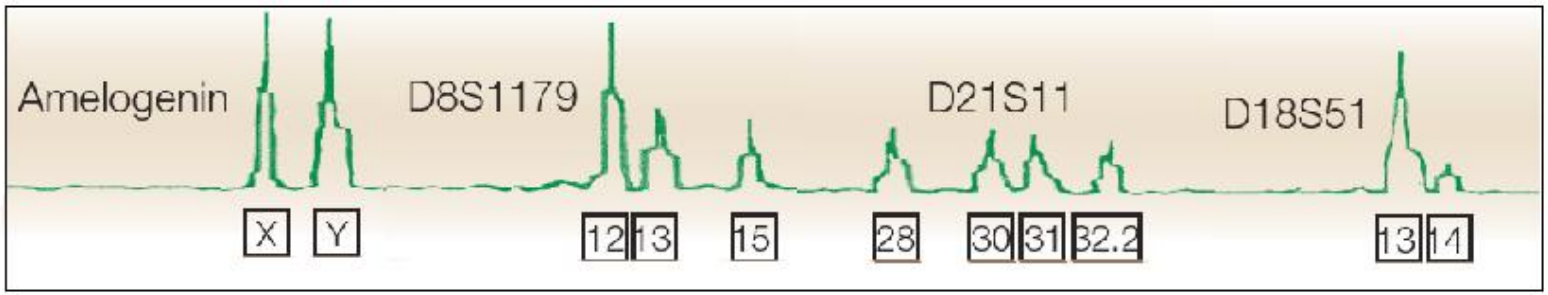
Hinda Haned
h.haned@nfi.minvenj.nl

Forensic Genomics
Consortium Netherlands

Outline

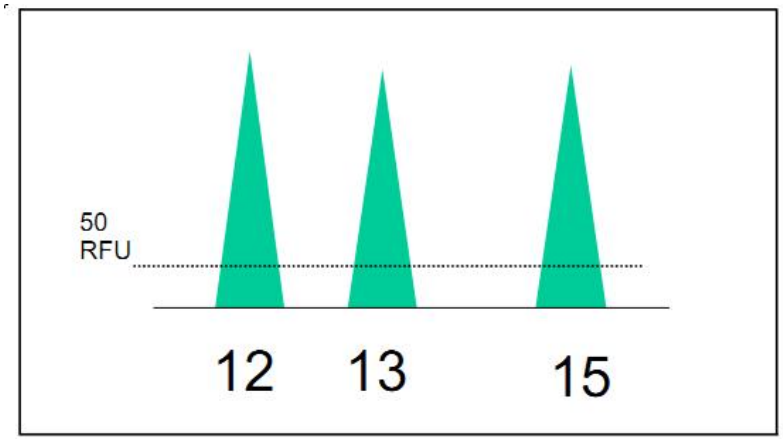
- Illustration of the LR principle applied to DNA mixtures
- Two-person mixtures to explain the principle (but no general formula is given!)
- Example with and without allelic dropout

DNA mixtures



- Two or more individuals contributing to the sample
- More than two peaks per locus

Why are mixtures challenging?



What genotypes created the mixture?

- 12,12/13,15
- 12,15/13,15
- 12,13/13,15
- ...

ISFG DNA commission recommendations

The likelihood ratio is the preferred approach to mixture interpretation. DNA commission 2005

Probabilistic approaches and likelihood ratio principles are superior to classical methods.

DNA commission 2012

The Bayesian framework: likelihood ratios

$$LR = \frac{Pr(data|H_{prosecution})}{Pr(data|H_{defence})}$$

- data: alleles and their peaks
- ratio of two probabilities or, ratio of two likelihoods

Interpretation

- Need for an interpretation framework that applies to all types of samples:
  - High template
  - Low template: PCR-related stochastic effects are exacerbated, creating uncertainty about the composition of the crime-sample

Reporting officers make pre-case assessments and formulate the propositions to be evaluated within the likelihood ratio framework.

Dropout/Drop-in definitions

Allele or locus dropout is defined as a signal that is below the limit of detection threshold, it occurs when one or both alleles of a heterozygote fail to PCR-amplify.

Allele drop-in is an allele that is not associated with the crime-sample and remains unexplained by the contributors under either Hp or Hd.

Low/High template DNA

High template DNA

- The epg reflects the composition of the sample:
  - no dropout
  - no drop-in

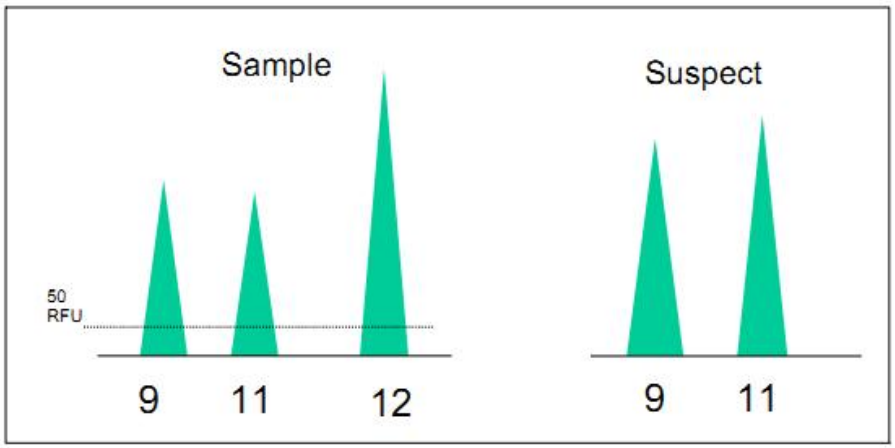
Low level DNA

- The epg does not reflect the composition of the sample:
  - allele dropout
  - allele drop-in
  - stutters
  - ...

Part 1: High template DNA, the epg reflects the composition of the sample.

Two-person mixture example

- Two-person mixture



Two-person mixture example

	Locus1
Evidence	9,11,12
Suspect	9,11
Victim	11,12

- H_p: Suspect + Victim contributed to the sample
- H_d: Victim + Unknown person (unrelated to the suspect) contributed to the sample

Two-person mixture: Under Hp

	Locus1
Evidence	9,11,12
Suspect	9,11
Victim	11,12

H_p: Suspect + Victim contributed to the sample

$$Pr(Evidence|H_p) = 1$$

Two-person mixture: Under Hd

	Locus1
Evidence	9,11,12
Victim	11,12
Unknown	?

H_p: Unknown + Victim contributed to the sample

Two-person mixture: Under Hd

- The victim's profile explains 11 and 12
- The unknown has to have allele 9: allele 9 is constrained

	Locus1
Evidence	9,11,12
Victim	11,12
Unknown	9,11
	9,12
	9,9

$$Pr(evidence|H_d) = 2p_9p_{11} + 2p_9p_{12} + p_9^2$$

Two-person mixure: LR

- H_p: Suspect + Victim contributed to the sample
- H_d: Victim + Unknown person (unrelated to the suspect) contributed to the sample

$$Pr(Evidence|H_p) = 1$$

$$Pr(evidence|H_d) = 2p_9p_{11} + 2p_9p_{12} + p_9^2$$

$$LR = \frac{1}{2p_9p_{11} + 2p_9p_{12} + p_9^2}$$

Two-person mixure: LR

- H_p: Suspect + Victim contributed to the sample
- H_d: Victim + Unknown person (unrelated to the suspect) contributed to the sample

$$Pr(Evidence|H_p) = 1$$

$$Pr(evidence|H_d) = 2p_9p_{11} + 2p_9p_{12} + p_9^2$$

$$LR = \frac{1}{2p_9p_{11} + 2p_9p_{12} + p_9^2}$$

What is the underlying model?

- LR is a function of the genotypic frequencies
- Assumes independent association of the alleles within loci: Hardy Weinberg equilibrium
- Multiply between loci: Linkage equilibrium

The product rule

Summary

- Derive the possible genotypes for the unknowns
- Determine the genotypic probabilities
- Sum up the probabilities for all plausible genotypes
- Calculate the ratio of the probabilities under Hp and under Hd

You should not do this by hand!

- usually, analysis of 15 or more loci simultaneously
- calculations get complicated with two or more unknowns

What happens if there are two unknowns under Hd?

- H_p: Suspect + Victim contributed to the sample
- H_d: Two Unknown individuals (unrelated to the suspect) contributed to the sample

	Locus1
Evidence	9,11,12
Unknown 1	?
Unknown 2	?

- Have to consider all the plausible genotypic combinations for the unknown that explain alleles 9,11,12 observed in the crime-sample.

Under Hd: two unknowns

Unknown 1	Unknown 2
9,9	11,12
11,11	9,12
12,12	9,11
9,11	9,12
9,11	11,12
9,12	11,12

$$Pr(Evidence|H_d) = 2(p_9^2 2p_{11}p_{12} + p_{11}^2 2p_9 p_{12} + p_{12}^2 2p_9 p_{11} + 2p_9 p_{11} 2p_9 p_{12} + 2p_9 p_{11} 2p_{11} p_{12} + 2p_9 p_{12} 2p_{11} p_{12})$$

LR: two unknowns

$$LR = \frac{1}{2(p_9^2 2p_{11}p_{12} + p_{11}^2 2p_9 p_{12} + p_{12}^2 2p_9 p_{11} + 2p_9 p_{11} 2p_9 p_{12} + 2p_9 p_{11} 2p_{11} p_{12} + 2p_9 p_{12} 2p_{11} p_{12})}$$

- Increasing the number of unknowns increases the number of terms under H_d

Part 2: Low template DNA, the epg does not reflect the composition of the sample.

Likelihood ratios vs. Low template DNA

- Classical approach of the LR: the product rule
- Main source of uncertainty in previous examples: Genotypes of unknown contributors

We will now see how we can modify the classical LR approach to account for uncertainty in the data, due to low template DNA conditions

Uncertainty in the data: single-source example

	Locus1
Evidence	11
Suspect	9,11

- Hp: Suspect contributed to the sample
- Hd: Unknown person (unrelated to the suspect) contributed to the sample

- Classical LR: $Pr(Evidence|H_p) = 0$
- LR with dropout and drop-in: $Pr(Evidence|H_p) \neq 0$

Uncertainty in the data: single-source example

	Locus1
Evidence	11
Suspect	9,11

- Hp: Suspect contributed to the sample
- Hd: Unknown person (unrelated to the suspect) contributed to the sample

- Classical LR: $Pr(Evidence|H_p) = 0$
- LR with dropout and drop-in: $Pr(Evidence|H_p) \neq 0$

LR with dropout and drop-in

- Main theory described by:
  - Haned et al, FSIG, 2012
  - DNA commission ISFG, FSIG 2012
  - Gill et al, FSI 2007
  - Curran et al, FSI, 2005

- Two key parameters in the model
  - dropout: Heterozygote, Homozygote
  - drop-in: not treated here

Basic model: qualitative data only, also called the drop-model.

LR with dropout and drop-in

- An allele drops out with a probability of d
- An allele does not drop out with a probability of $1 - d$
- Allele dropout from a heterozygote: d
- Allele dropout from a homozygote: d'

Single-source example: Under Hp

- Hp: Suspect contributed to the sample

	dropout
Allele 9	yes
Allele 11	no

$$Pr(evidence|H_p) = Pr(\text{dropout of } 9) \times Pr(\text{non-dropout of } 11)$$
$$= d \times (1 - d)$$

Single-source example: Under Hd

- Unknown contributed to the sample

	Locus1
Evidence	11
Unknown	?

The Q alleles

- What are the possible genotypes for the unknown?
  - The dropped out alleles are gathered under a virtual alleles Q
  - Q is a 'place-holder' to all possible genotypes!
  - The Unknown's genotype has to explain allele 11 (no drop-in)

Under Hd

	Locus1
Evidence	11
Unknown	11,11
	11,Q

- Q can be anything except 11
- Unknown genotype must explain 11
- This leaves us with two possibilities:
  - Homozygote: 11, 11
  - Heterozygote 11, Q

Q allele

- Locus L has five alleles: $\{9, 10, 11, 12\}$
- $p_9 + p_{10} + p_{11} + p_{12} = 1$
- $p_Q = 1 - p_{11}$
- $p_Q = p_9 + p_{10} + p_{12}$

- 11,Q can be:
  - 9,11
  - 10,11
  - 11,11
  - 11,12

No need to worry about deriving all the genotypes!

- All thee genotypes are regrouped under 11Q with frequency: $2p_{11}p_Q$

Summary

- Two possible genotypes: 11,11 and 11Q

	Dropout	Genotype probability
11,11	$(1-d')$	p_{11}^2
11Q	$(1-d)d$	$2p_{11}p_Q$

$$LR = \frac{d(1-d)}{(1-d')p_{11}^2 + (1-d)d2p_{11}p_Q}$$

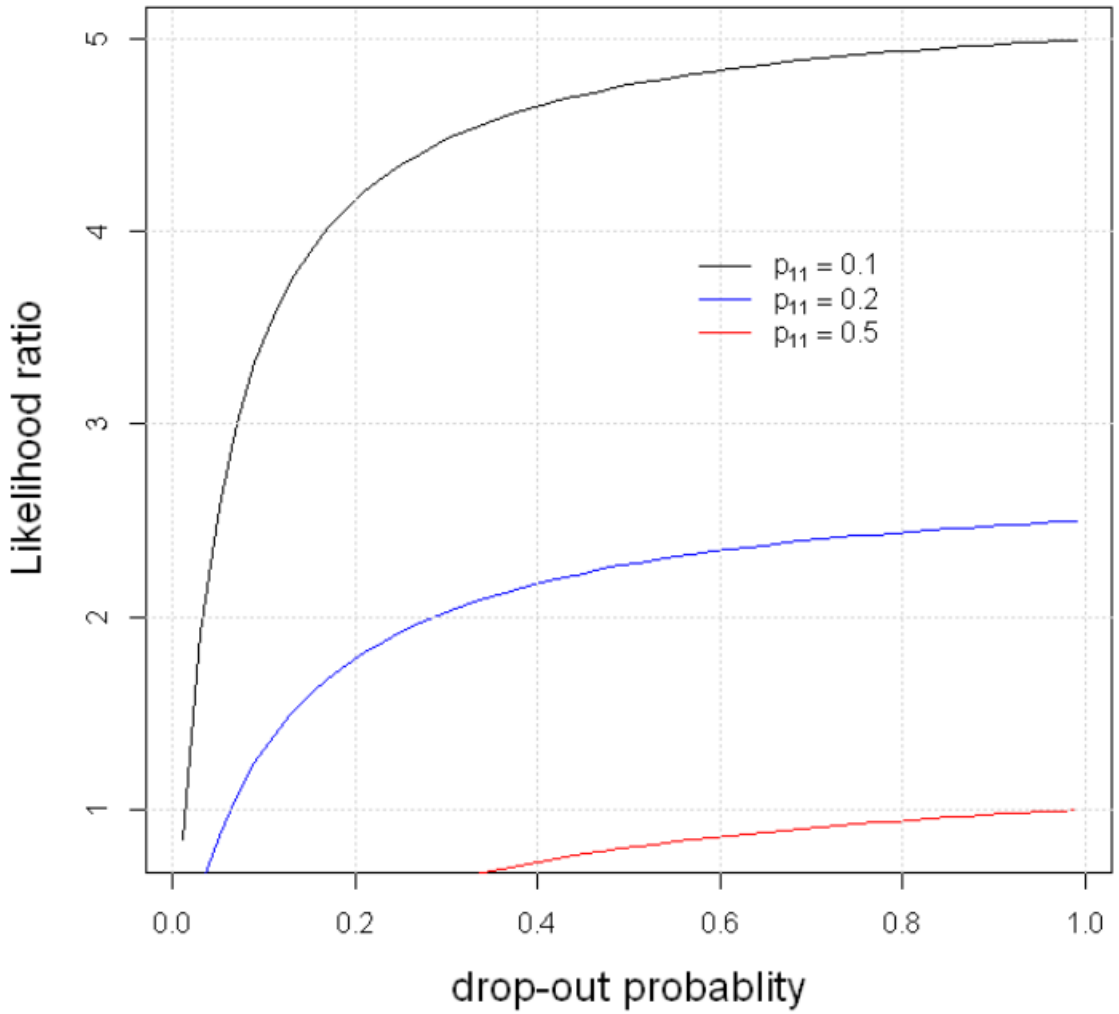
Summary

- Two possible genotypes: 11,11 and 11Q

	Dropout	Genotype probability
11,11	$(1-d')$	p_{11}^2
11Q	$(1-d)d$	$2p_{11}p_Q$

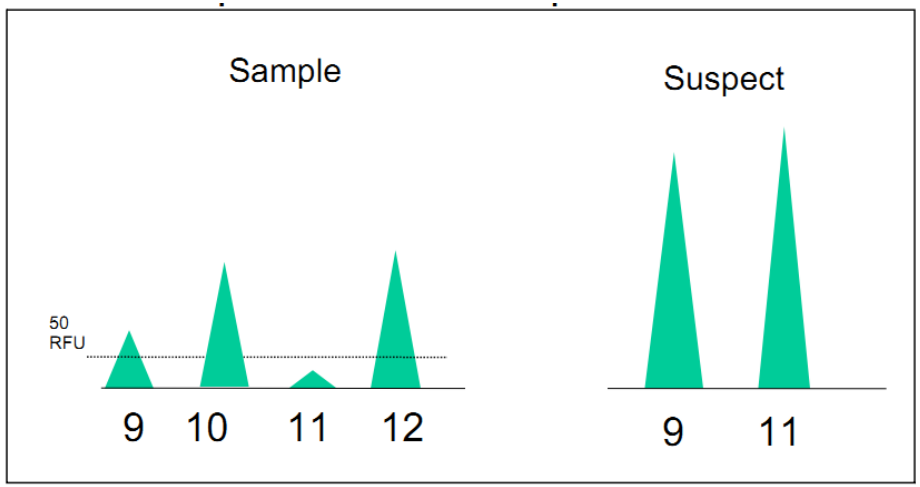
$$LR = \frac{d(1-d)}{(1-d')p_{11}^2 + (1-d)d2p_{11}p_Q}$$

LR vs. probability of dropout



Low-template mixture

- Low-template DNA mixture



Two-person mixture example: one dropout, no drop-in

	Locus1
Evidence	9,10,12
Suspect	9,11
Victim	10,12

- Hp: Suspect + Victim
- Hd: Two unknowns (unrelated to suspect/victim)

Under Hp: Dropout from the suspect

Suspect	9,11	d(1-d)
Victim	10,12	$(1\text{-d})^2$

$$Pr(Evidence|H_p) = d(1-d)^3$$

Under Hd: dropout is possible

- Hd: two unknowns
- Dropout is possible: Q allele, can be anything except 9, 10, 12

9,9	10,12	No-dropout
10,10	9,12	
12,12	9,10	
9,12	9,10	
9,12	10,12	
10,12	9,10	
9Q	10,12	One dropout
10Q	9,12	
12Q	9,10	

Under Hd: dropout is possible

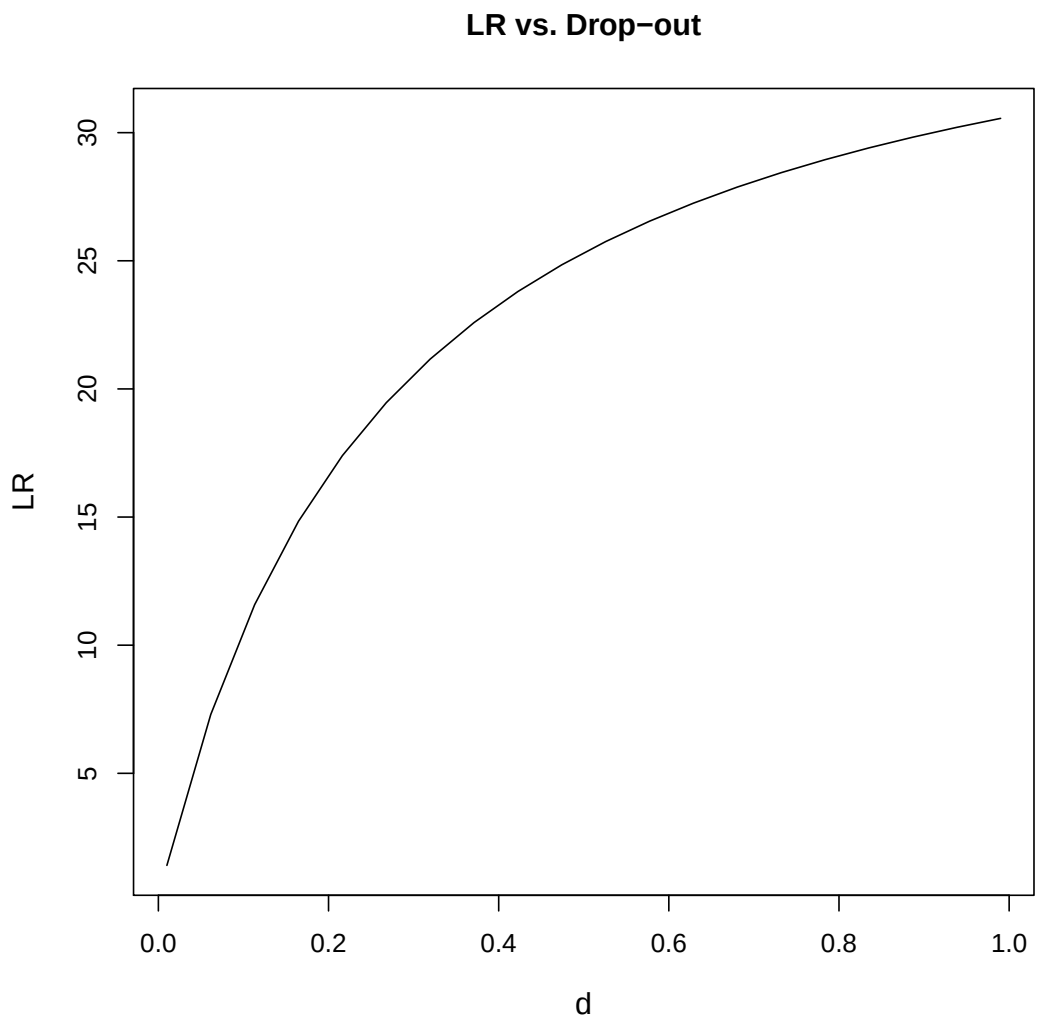
- Hd: two unknowns
- Dropout is possible: Q allele, can be anything execept 9, 10, 12

		Dropout	Genotype Prob.
9,9	10,12		$p_9^2 \times 2p_{10}p_{12}$
10,10	9,12	$(1-d')(1-d)^2$	$p_{10}^2 \times 2p_9p_{12}$
12,12	9,10		$p_{12}^2 \times 2p_9p_{10}$
9,12	9,10		$2p_9p_{12} \times 2p_9p_{10}$
9,12	10,12	$(1-d)^4$	$2p_9p_{12} \times 2p_{10}p_{12}$
10,12	9,10		$2p_{10}p_{12} \times 2p_9p_{10}$
9Q	10,12		$2p_9p_Q \times 2p_{10}p_{12}$
10Q	9,12	d(1-d)^3	$2p_{10}p_Q \times 2p_9p_{12}$
12Q	9,10		$2p_{12}p_Q \times 2p_9p_{10}$

Likelihood ratio

$$LR = \frac{d(1-d)\times(1-d)^2}{2\left\{\begin{array}{l}(1-d')(1-d)^2\left[p_9^2\times 2p_{10}p_{12}\times + p_{10}^2\times 2p_9p_{12} + p_{12}^2\times 2p_9p_{10}\right] + \\ (1-d)^4\left[2p_9p_{12}\times 2p_9p_{10} + 2p_9p_{12}\times 2p_{10}p_{12} + 2p_{10}p_{12}\times 2p_9p_{10}\right] + \\ d(1-d)^3\left[2p_9p_Q\times 2p_{10}p_{12} + 2p_{10}p_Q\times 2p_9p_{12} + 2p_{12}p_Q\times 2p_9p_{10}\right]\end{array}\right\}}$$

LR vs. dropout probability



How about drop-in probability?

Under Hp: Dropout from the suspect

Suspect	9,11	d(1-d)
Victim	10,12	$(1\text{-d})^2$

- If drop-in=0 $Pr(Evidence|H_p) = d(1-d)^3$
- If drop-in $\neq 0$: $Pr(Evidence|H_p) = d(1-d)^3 \times (1-c)$
- c is the probability of drop-in

Under Hd: two unknowns

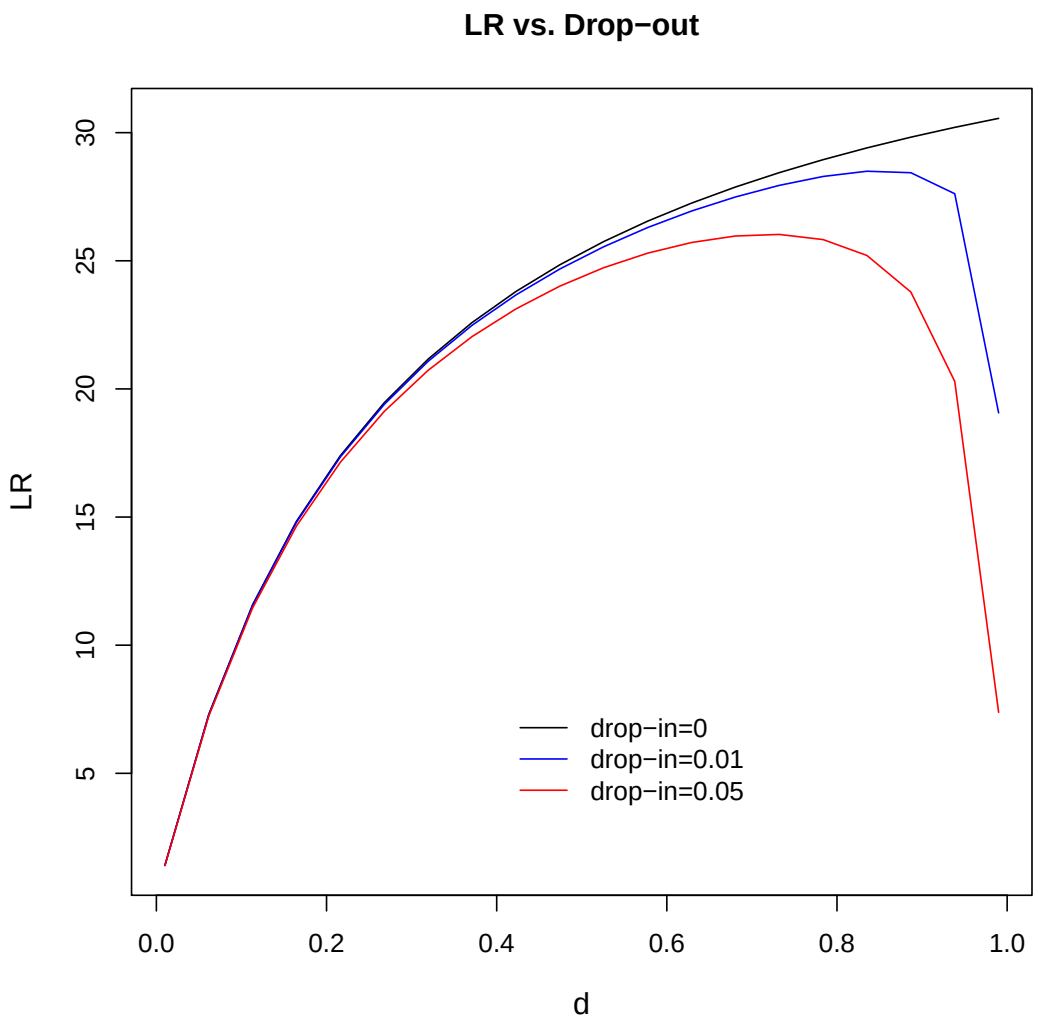
- Dropout is possible, no drop-in: Q allele, can be anything except 9, 10, 12

- If drop-in is possible: Q allele can be anything!
- So the genotypes of the unknown have no longer to explain alleles 9, 10, 12.
- This increases the number of terms under Hd

Think of drop-in as a scaling factor

- If an allele is a drop-in: multiply by $c\times$ frequency of allele i.
- If an allele is not a drop-in, multiply by $(1 - c)$

LR vs. dropout and drop-in probability



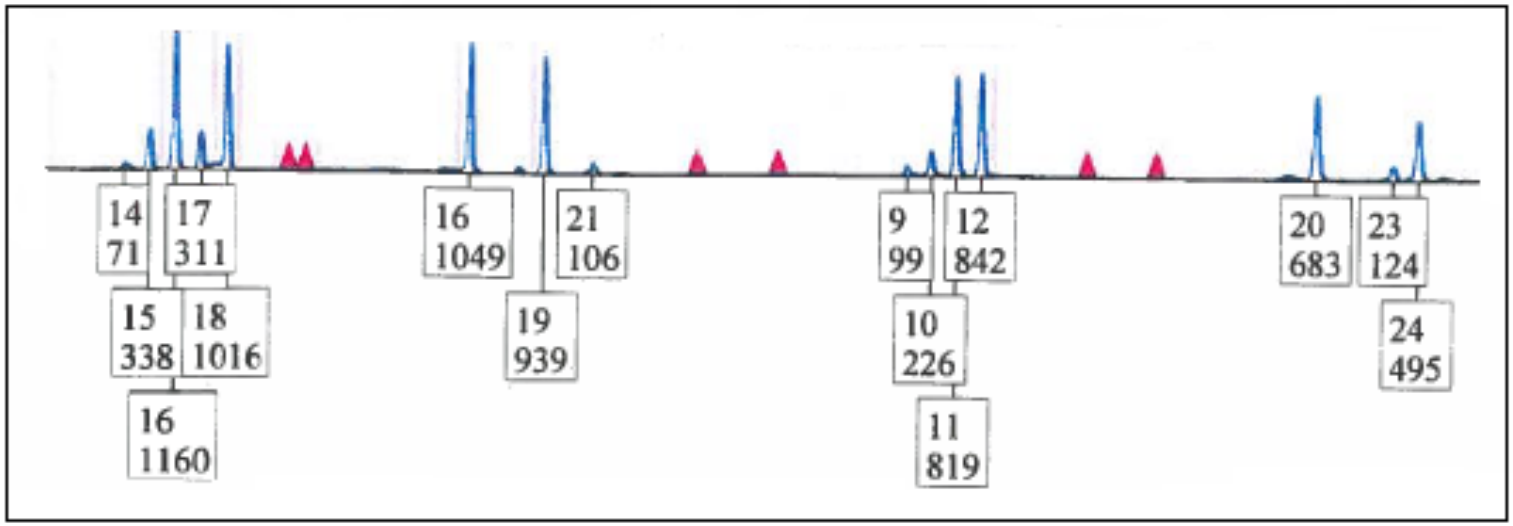
Summary

- Derive the possible genotypes for the unknowns
- Determine the genotypic probabilities
- Sum up the probabilities for all plausible genotypes
- Determine the corresponding dropout probabilities
- Calculate the ratio of the probabilities under Hp and under Hd

Software

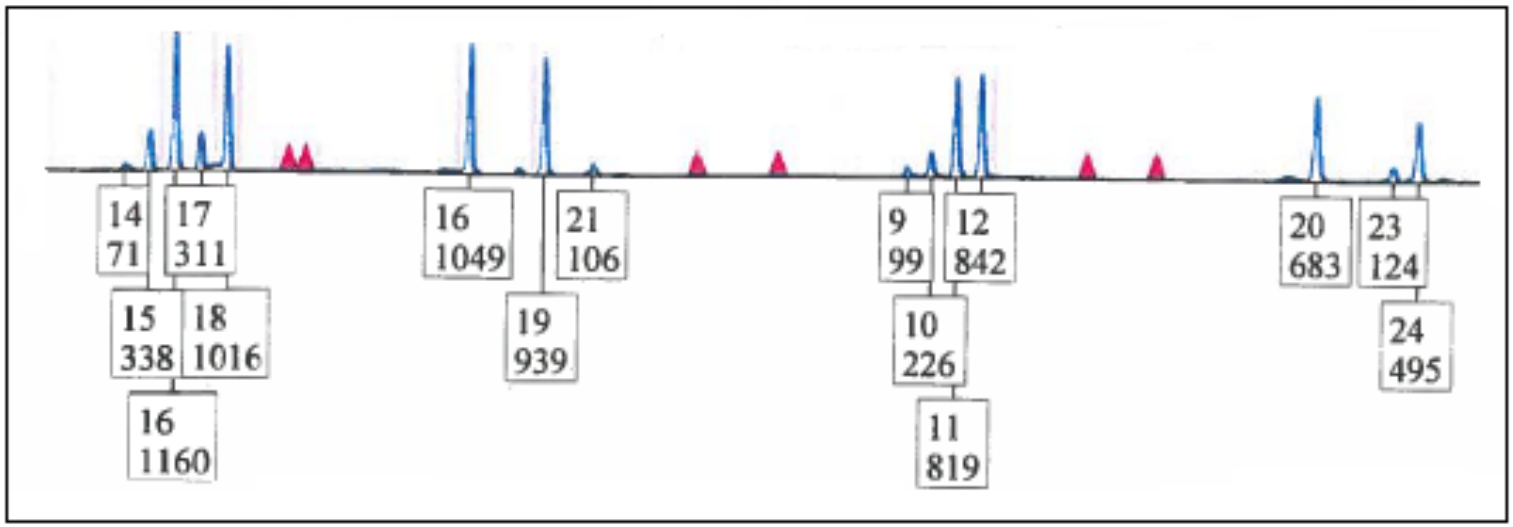
- Derive genotypes of the unknowns is the key issue
- Assign genotype probability to each genotype
- The number of possibilities increases with the number of contributors, deriving LRs for mixtures by hand is not realistic!

Casework example 1: A 3-person mixture



- Victim is major contributor
- At least two minor contributors

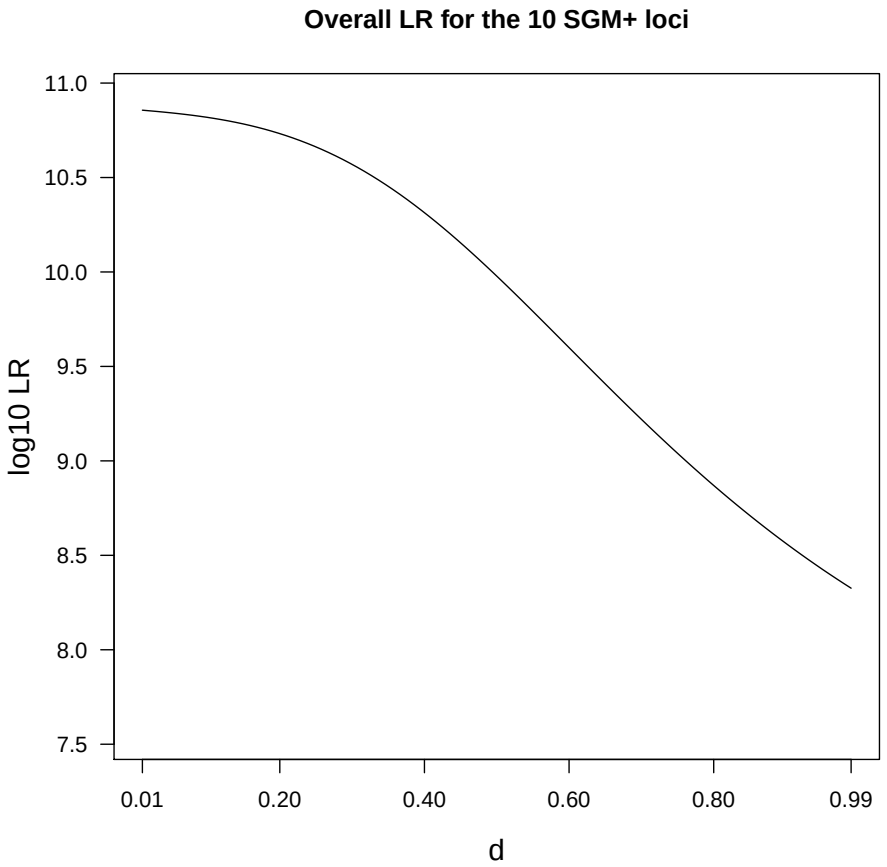
Casework example 1: A 3-person mixture



- Hp: Victim + Suspect + Unknown
- Hd: Victim + two unknowns

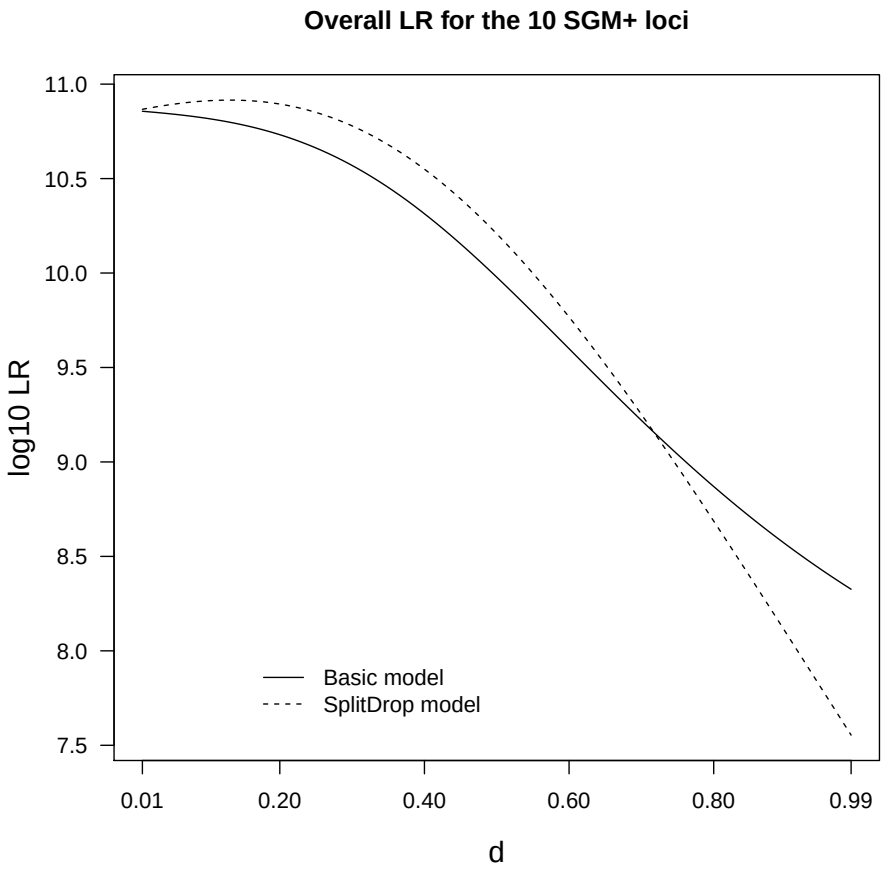
Sensitivity analysis: Overall LR

Same dropout probability for all contributors



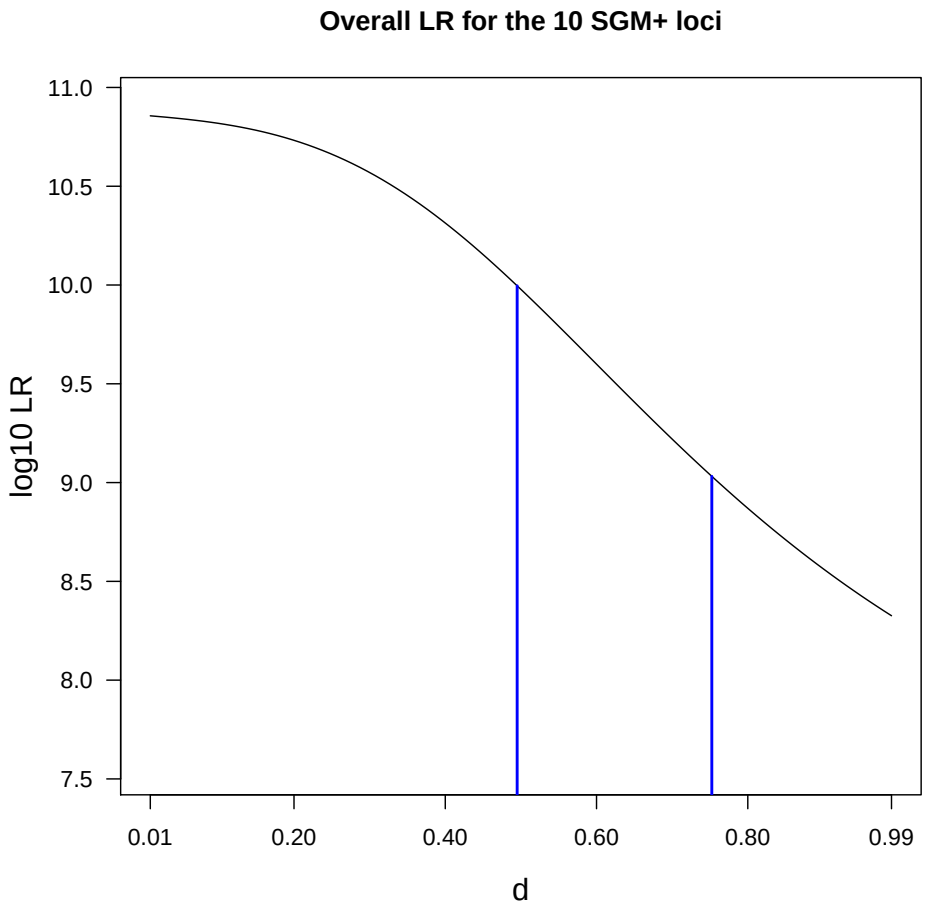
Sensitivity analysis: Overall LR

Average probability vs. Splitting dropout/contributor $\Longrightarrow$ No significant differences between the models!



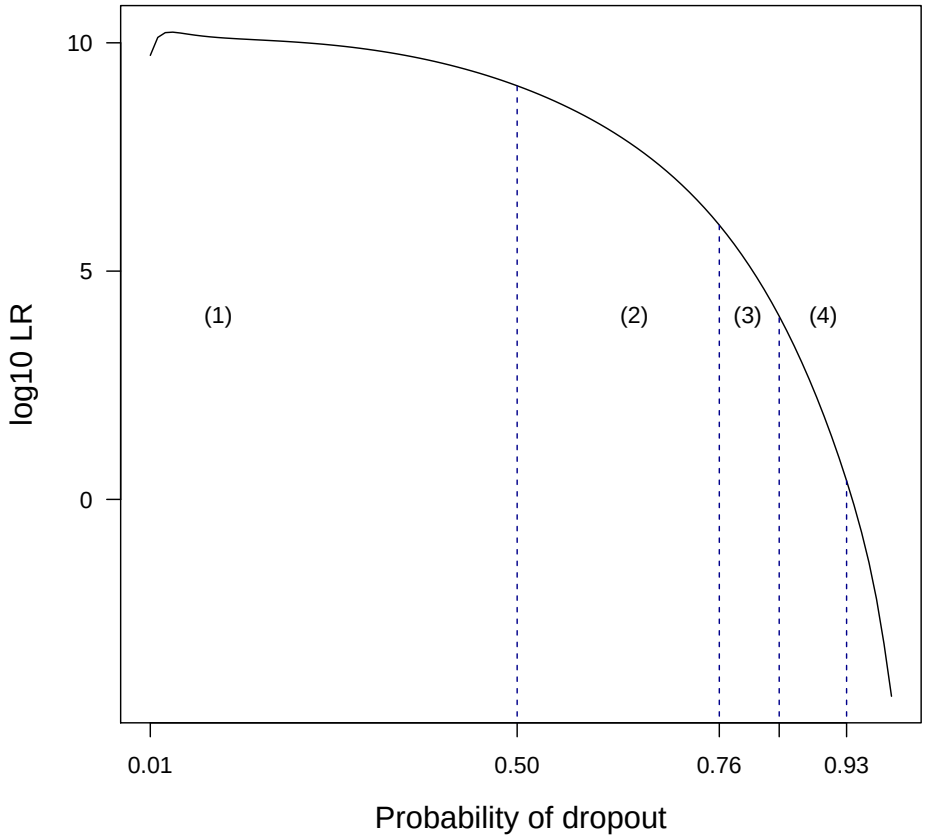
Plausible ranges for PrD?

LR	dropout
$\leq 10^{10}$	$0.01 \leq D \leq 0.50$
$[10^9, 10^8]$	$0.50 < D \leq 0.99$



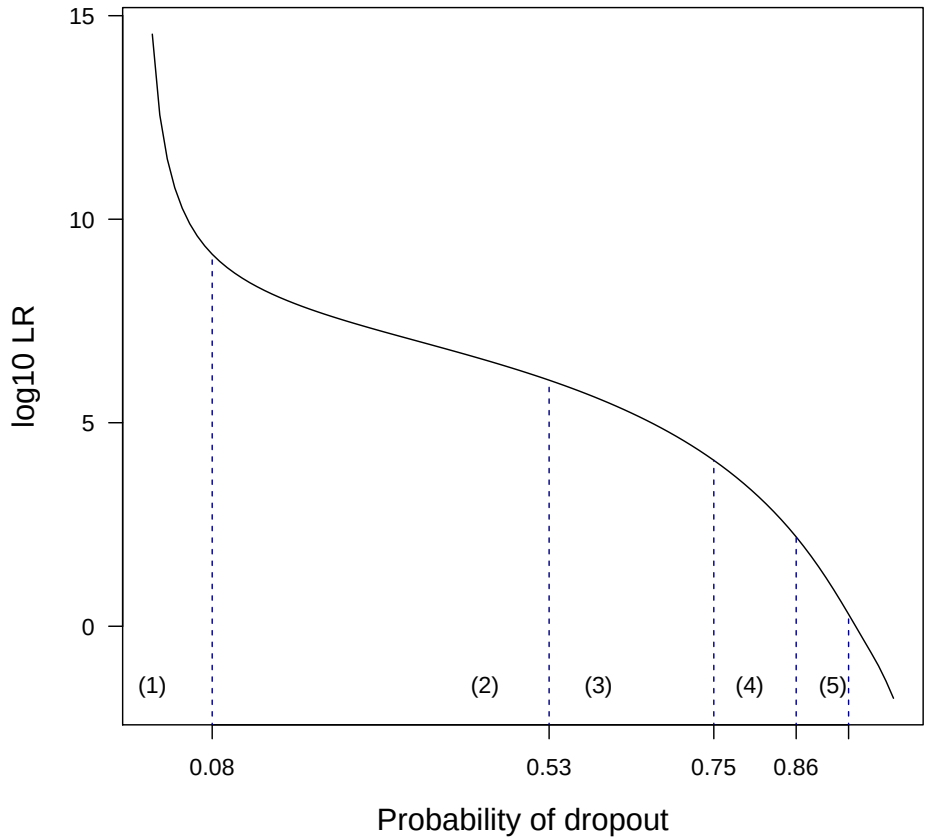
Casework example 2: two-person mixture

	LR	dropout
(1)	$[10^{10}, 10^{9}]$	$0 \leq D \leq 0.50$
(2)	$[10^{9}, 10^{6}]$	$0.50 < D \leq 0.76$
(3)	$[10^{6}, 10^{4}]$	$0.76 < D \leq 0.84$
(4)	$[10^{4}, 1]$	$D > 0.84$



Casework example 3: three-person mixture

	LR	dropout
(1)	$[10^{14}, 10^{9}]$	$0 \leq D \leq 0.08$
(2)	$[10^{9}, 10^{6}]$	$0.08 < D \leq 0.53$
(3)	$[10^{6}, 10^{4}]$	$0.53 < D \leq 0.75$
(4)	$[10^{4}, 100]$	$0.75 < D \leq 0.86$
(5)	$[100, 1]$	$0.86 < D \leq 0.93$



All models are wrong...

- Continuous models are expected to extract more information from the data, but their implementation is difficult and tedious in practice
- semi-continuous methods are easier to implement and can serve as a good approximation

How to inform dropout probabilities?

- *Estimate* dropout probabilities via logistic regression
  - difficult to extended to $>$ 2-person mixtures
- Define *plausible ranges* of dropout
  - based on expert belief
  - based on maximum likelihood principle
- Bayesian approach: combine prior belief and likelihood to yield a posterior *distribution*