Repetition: Statistical interpretation of DNA mixtures

Thore Egeland

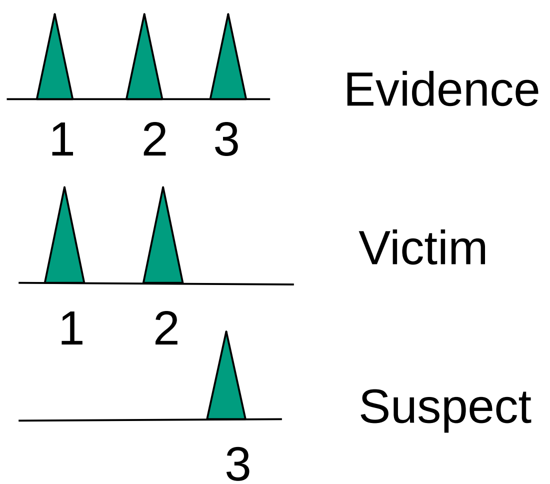
Copenhagen May 20-23 2014

Overview

- LR
  - Binary model
  - drop-in and drop-out
    - Precise mathematics:
      - Supplementary of Haned, Slooten, Gill (2012)
  - LRmixTK

Mixture, one locus

Forget about peak heights now



Likelihood Ratio (LR)

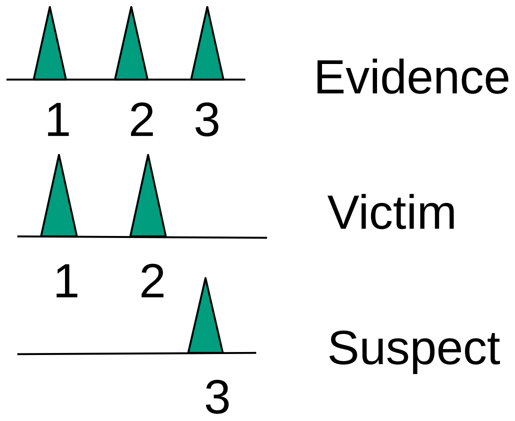
- Ratio of two probabilities
  - The probability of the **evidence (E)** given the prosecution hypothesis (H_p)
  - The probability of the **evidence (E)** given the defense hypothesis (H_d)

$$LR = \frac{P(E \mid H_p)}{P(E \mid H_d)}$$

- If LR > 1 → The evidence supports H_p
- If LR < 1 → The evidence supports H_d

Example 1: Likelihood Ratio

- Specify two hypotheses
  - H_p: Victim + Suspect
  - H_d: Victim + Unknown



- The likelihood ratio, with p_i=0.2

$$LR = \frac{P(E \mid H_p)}{P(E \mid H_d)} = \frac{1}{p_3^2 + 2p_1p_3 + 2p_2p_3} = \frac{1}{0.2} = 5$$

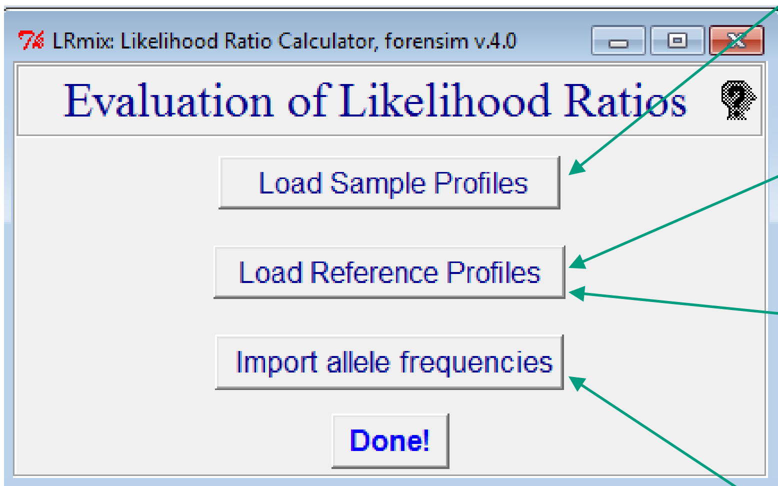
- The numerator is 1 because a mixture of the victim and suspect gives those evidence alleles with probability 1
- The denominator is the summed probability that the unknown will have one of the genotypes 3/3, 1/3 or 2/3

Likelihood Ratio

- What does an LR of 5 mean?
- *The evidence is 5 times more likely **IF** the victim and suspect are the contributors than **IF** the victim and an unrelated unknown are the contributors*

- It does <u>not</u> mean that it is 5 times more likely that the victim and the suspect are the contributors than the victim and an unrelated unknown
  - Then we would say something about $P(H_p|E)$ and $P(H_d|E)$, which we do not know!

LRmixTK()



evidence_example1.csv : (click Display…)

SampleName	Marker	Allele1	Allele2	Allele3
p16	M1	1	2	3

suspect_example1.csv :

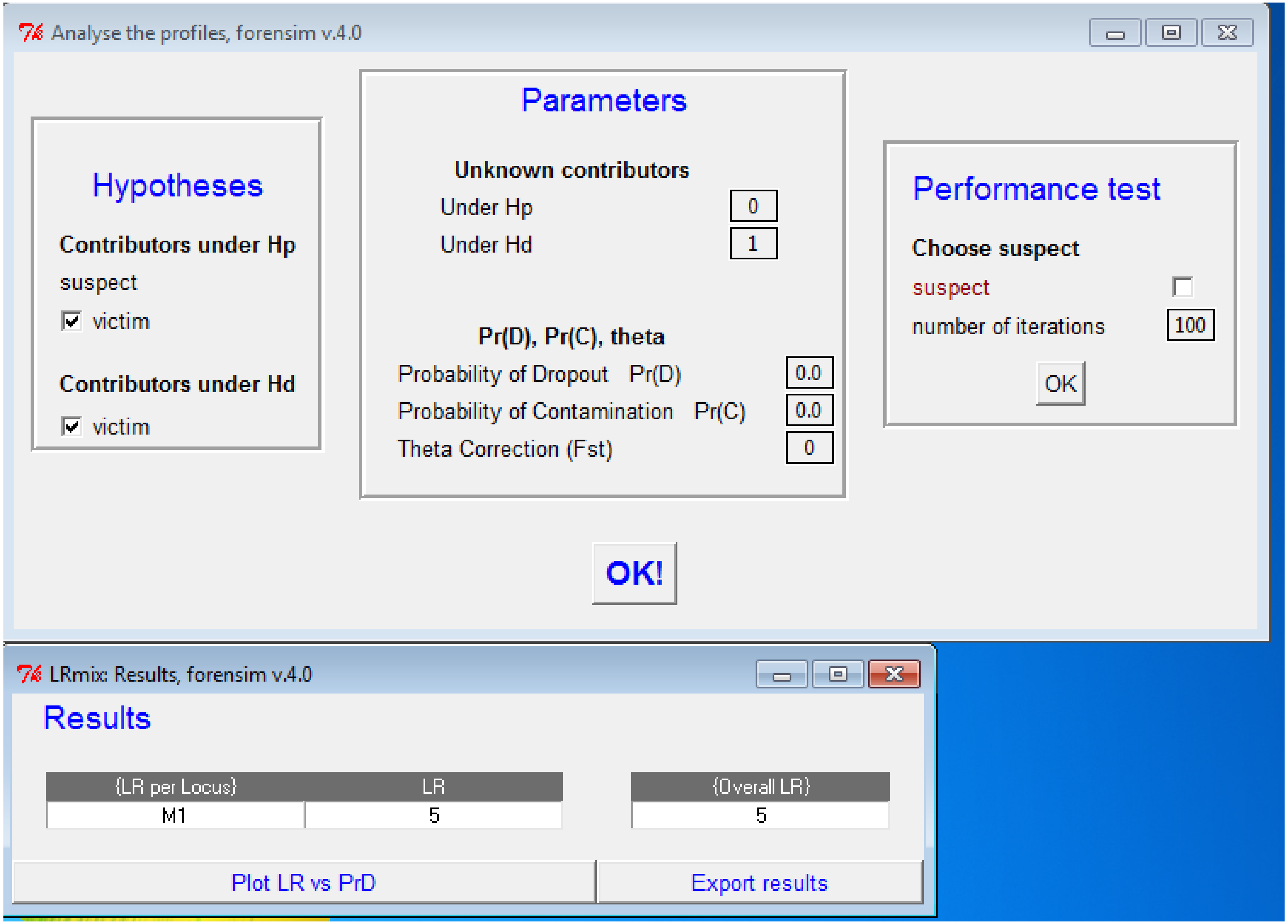
SampleName	Marker	Allele1	Allele2
suspect	M1	3	3

victim_example1.csv :

SampleName	Marker	Allele1	Allele2
victim	M1	1	2

allele_frequencies_example1.csv :

Allele	M1
1	0.2
2	0.2
3	0.2
4	0.4


Analyse the profiles, forensim v.4.0
Hypotheses
Contributors under Hp
suspect
victim
Contributors under Hd
victim
Parameters
Unknown contributors
Under Hp 0
Under Hd 1
Pr(D), Pr(C), theta
Probability of Dropout Pr(D) 0.0
Probability of Contamination Pr(C) 0.0
Theta Correction (Fst) 0
Performance test
Choose suspect
suspect
number of iterations 100
OK
OK!
LRmix: Results, forensim v.4.0
Results
{LR per Locus} LR
M1 5
{Overall LR}
5
Plot LR vs PrD
Export results

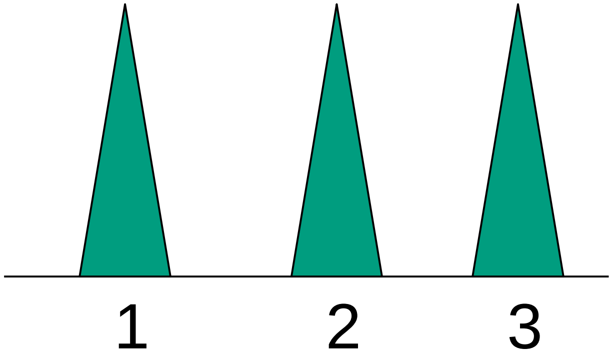
Drop-in and drop-out

- Drop-in and drop-out are stochastic phenomena that appears in low template DNA
- Drop-out: an allele fails to amplify (below detection level)

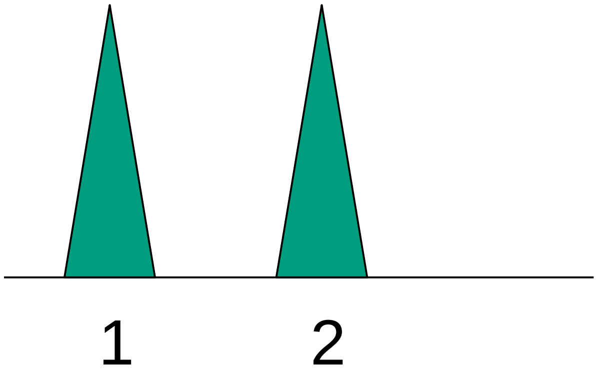
- Drop-in: False allele
  - Not reproducible
- Drop-in is not the same as contamination
  - Drop-in events are considered independent while contamination is dependent

Drop-in

- Evidence



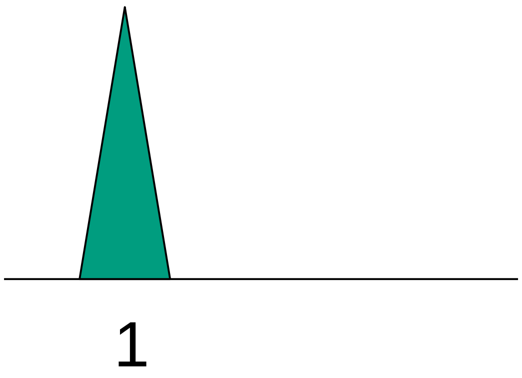
- Suspect



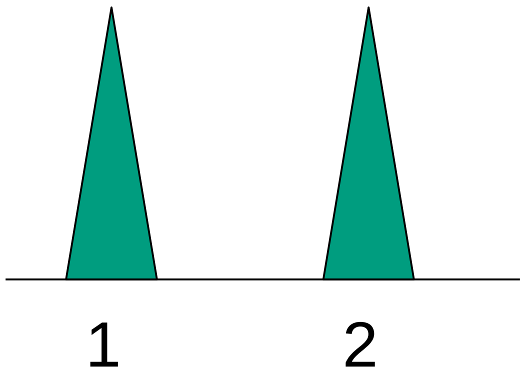
Match?
Then allele 3 must have dropped into the evidence

Drop-out

- Evidence

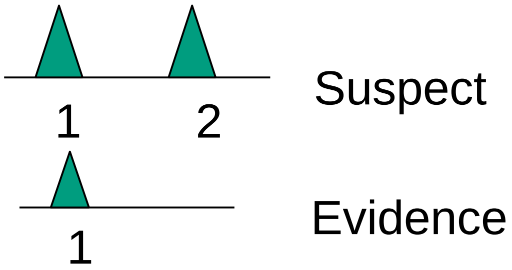


- Suspect



Match? Then allele 2 must have dropped out of the evidence

Drop-out model

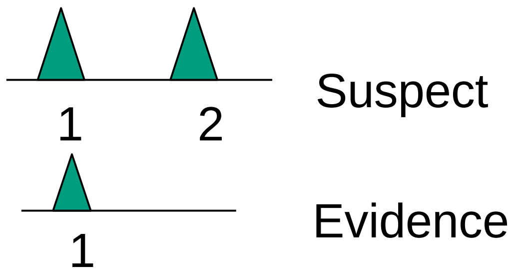


- H_p: Suspect
- H_d: Unknown

- We define
  - probability of dropout for heterozygotes = *d*
    - *both alleles drop out with probability d^2*
  - probability of dropout for homozygotes = *D* where $D \leq d^2$ *(equality in forensim and hereafter)*

Example 2: LR with dropout (Case B Haned et al., 2012)

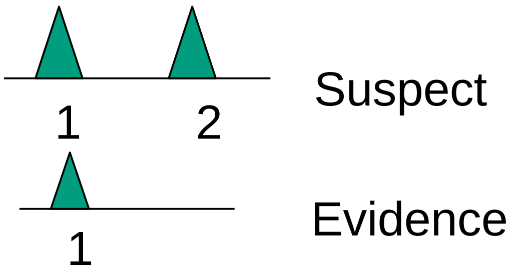
- H_p: Suspect
- H_d: Unknown



- For H_p to be true
  - allele 2 in the suspect's profile has dropped out
  - allele 1 in the suspect's profile has not dropped out

- We calculate by hand so we can check forensim

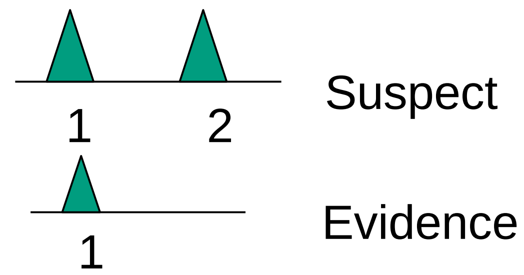
- H_p: Suspect
- H_d: Unknown



- For H_d to be true the unknown must have a genotype that contains the allele 1
- Let Q denote any other allele than the allele in the evidence:

$$p_Q = 1 - p_1$$

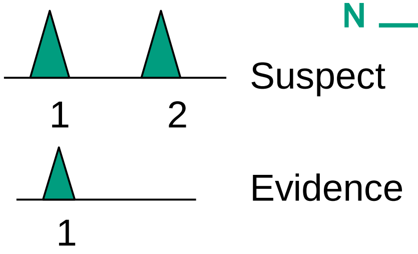
- H_p: Suspect
- H_d: Unknown



$$LR = \frac{P(E \mid H_p)}{P(E \mid H_d)} = \frac{d(1-d)}{p_1^2(1-d^2) + 2p_1 p_Q d(1-d)}$$

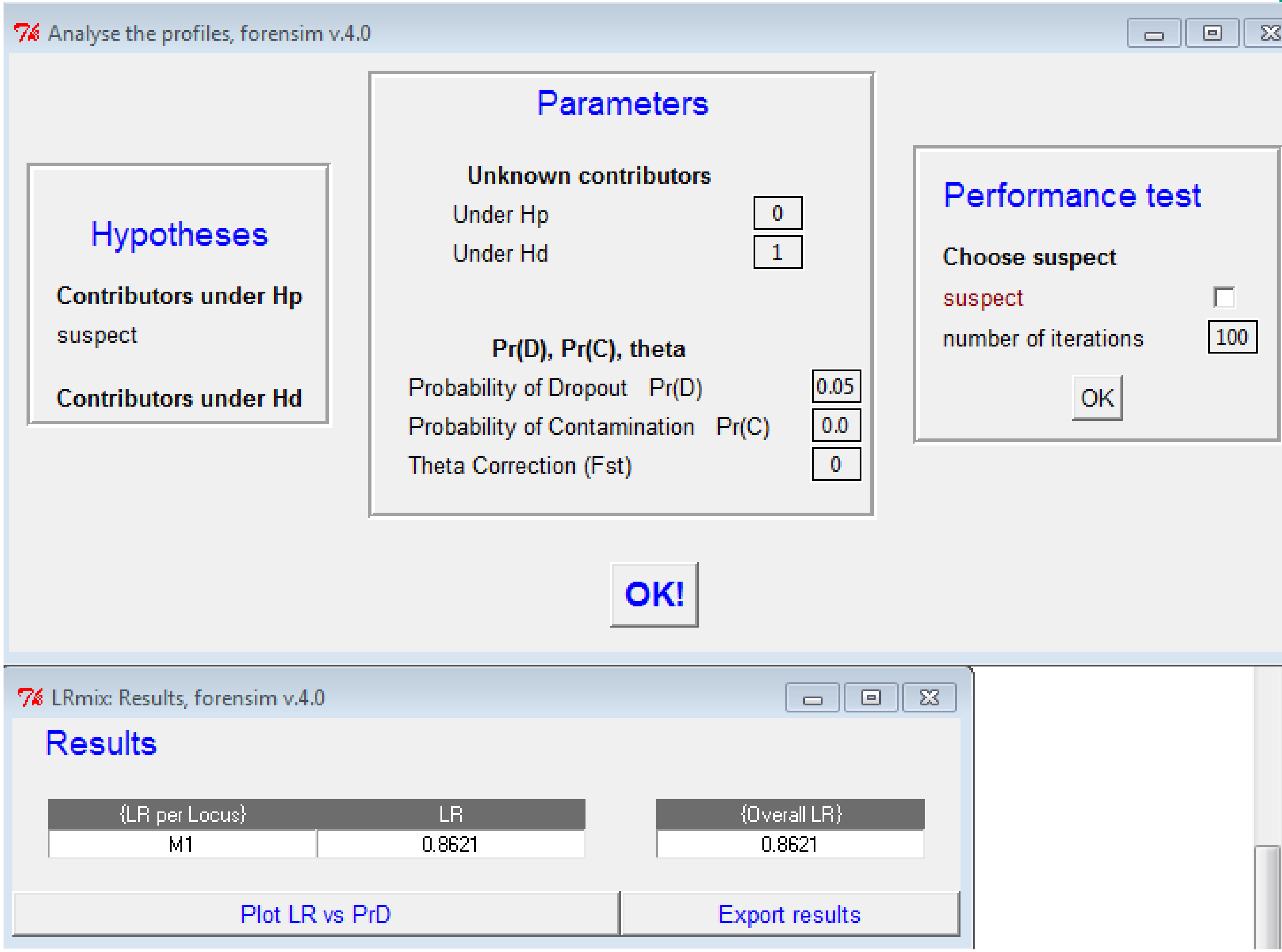
$$= \frac{d}{p_1^2(1+d) + 2p_1(1-p_1)d}$$

LR



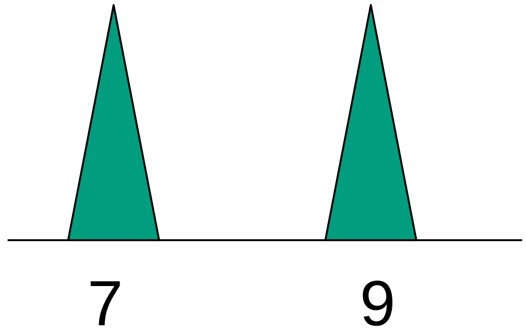
- What is the effect of dropout probability and allele frequency on LR?
- p_1=0.2, p_Q=0.8, d=0.05

$$LR = \frac{0.05(1-0.05)}{0.2^2(1-0.05^2)+2\times 0.2\times 0.8\times 0.05(1-0.05)} = 0.86$$

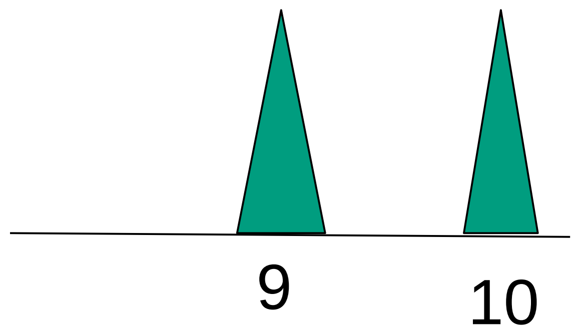

Analyse the profiles, forensim v.4.0
Hypotheses
Contributors under Hp
suspect
Contributors under Hd
Parameters
Unknown contributors
Under Hp 0
Under Hd 1
Pr(D), Pr(C), theta
Probability of Dropout Pr(D) 0.05
Probability of Contamination Pr(C) 0.0
Theta Correction (Fst) 0
Performance test
Choose suspect
suspect
number of iterations 100
OK
OK!
LRmix: Results, forensim v.4.0
Results
{LR per Locus} LR
M1 0.8621
{Overall LR}
0.8621
Plot LR vs PrD
Export results

Example 3 (Case C, Haned et al.): Drop-out and drop-in

- Suspect



- Evidence



H_p: Suspect
H_d: Unknown

Exact and approximate LR

- Formula included only to explain that LR is not exact with both drop-out and drop-in

$$LR = \frac{d}{2p_9(1-d)}c$$

$$+\frac{d(-p_7 - p_9 + d(-4 + p_7 + 3p_9))}{4(d-1)^2 p_9}c^2$$

$$+O(c^3)$$

$$\approx \frac{d}{2p_9(1-d)}c \text{ for small } c \text{ and } d$$

Quality of approximation

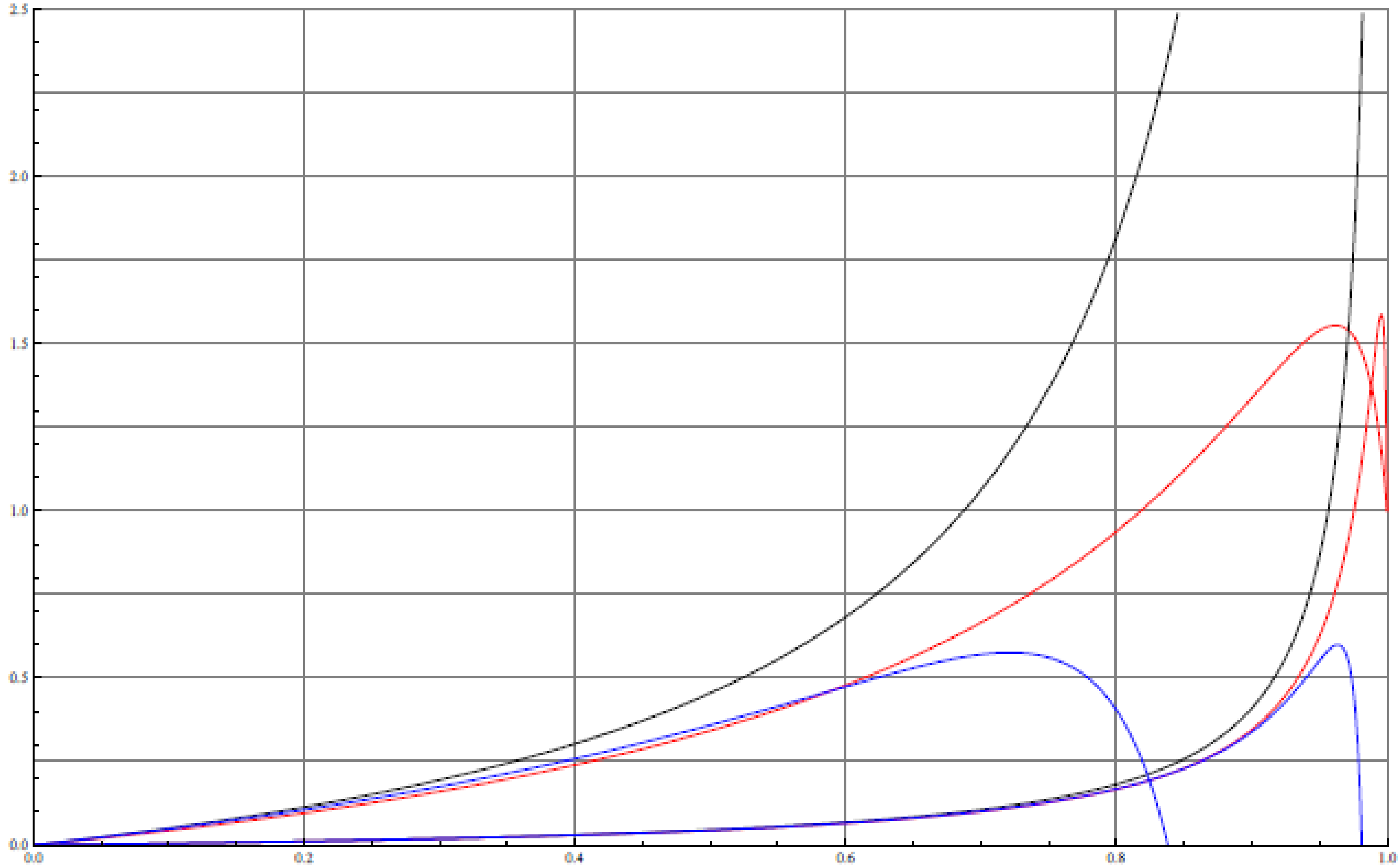


Figure 2: Linear (black) and quadratic (blue) approximations to the LR as well as exact LR (red) for $c = 0.01$ (lower three lines) and $c = 0.1$ (upper three lines).

2.2.4. Comparison to Curran model

Finally, we compare the likelihood ratio as obtained in Table 3, to the exact likelihood ratio calculated according to (4). The result is, for $c = 0.1$, presented in the Figure below, from which it is clear that for small values of d the likelihood ratios almost coincide, but for large d the likelihood ratio obtained from Table 3 (obtained from the Curran model) tends to slightly overestimate the evidence.

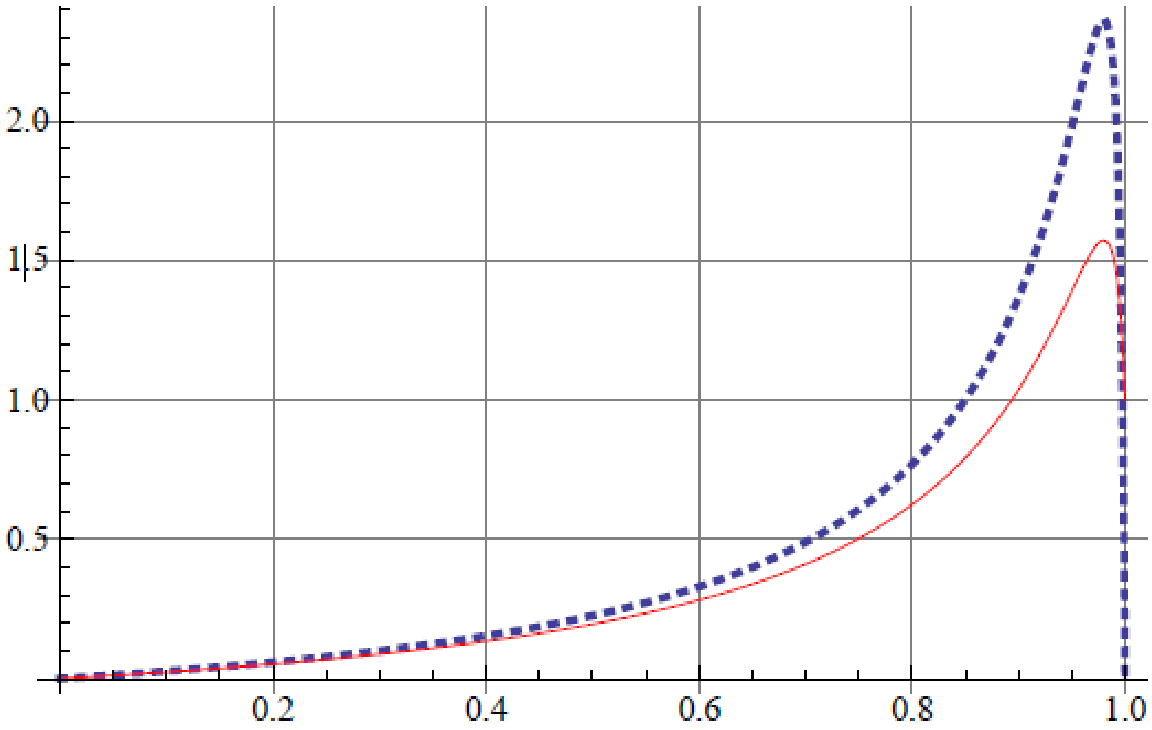


Figure 3: Exact (red) and Curran's LR (blue, dotted) for $c = 0.1$.