

# Looking for Victims Theory

Fabio Corradi

Rome, 21 May 2008

# Speech Outline

- Basics on identification
- Alternative composite hypotheses
- Inference : Configuration
- Inference : Priors
- Inference : Likelihood
- Inference : Posteriors

# Basics on Identification

- Let's start from the “standard” paternity identification case.
- Evidence=Mother, Child and Alleged father DNA
- Usually results are given in terms of a likelihood ratio

$$LR = \frac{\Pr(E | H = 1)}{\Pr(E | H = 2)}$$

- Identification consists on the computation of the Evidence probability assuming:
- The AF paternity ( $H=1$ )

Versus

- Someone else is the father of the Child ( $H=2$ )

# Basics on identification

- H=1 has a precise meaning, not necessarily H=2
- Often, neither the mother nor the AF are interested in better specifying H=2
- Very often H=2 is assumed to be a “random man”
- Very very often belonging to the same ethnic group as the AF
- NOTE: H=1,2 are **not** random variables in a LR, they are just two mutually excluding conjectures
- RESULT: The probability to observe the evidence if AF is the father is  $x \in [0, \infty]$  times the evidence probability, if H=2 holds true

# Basics on identification

- PROBLEM: Can I use the LR approach if more alternatives to  $H=1$  exist, i.e. if an alternative composite hypothesis arises?
- NO: at most it is possible to evaluate a LR for each of the alternative hypothesis' components. Each ratio assumes the alternative component as the only existing one :
  - many ratios = high confusion
- If we want to compare the main hypothesis with all the possible alternatives they must be considered all together

# Basics on Identification

- Solution: introduce prior probabilities on the each of the alternative component hypothesis: the overall set of hypotheses becomes a random variable

# Basics on Identification

Two main results

- 1) We can compute via Bayes theorem the posterior probability for each conjecture taking into account the others:

$$\Pr(H = 1 | E) = \frac{\Pr(E | H = 1) \Pr(H = 1)}{\Pr(E | H = 1) \Pr(H = 1) + \Pr(E | H = 2a) \Pr(H = 2a) + \dots + \Pr(E | H = 2c) \Pr(H = 2c)}$$

- 2) We can compute the Weight of Evidence (WE), i.e. the Bayesian version of the LR

$$\frac{\Pr(H = 1 | E)}{\Pr(H = 2a \cup \dots \cup 2c | E)} = \frac{\Pr(E | H = 1)}{\Pr(E | H = 2a \cup \dots \cup 2c)} \frac{\Pr(H = 1)}{\Pr(H = 2a \cup \dots \cup 2c)}$$

# Alternative composite hypotheses in paternity identification

Consider:

- just one diallelic locus:  $A=(A,a)$
- Evidence: Mother="Aa"; Child="aa"; Alleged Father="aa"

Population 1:

- $\Pr(a)=0.1$ ;  $\Pr(A)=0.9$ ;  $\Pr(aa)=0.01$ ;  $\Pr(aA)=0.18$ ;  $\Pr(AA)=0.81$

Population 2

- $\Pr(a)=0.9$ ;  $\Pr(A)=0.1$ ;  $\Pr(aa)=0.81$ ;  $\Pr(aA)=0.18$ ;  $\Pr(AA)=0.01$

Hypotheses:

- $H=1$ : "Alleged father" is the true father
- $H=2-a$  A generic member of Population 1 is the true father
- $H=2-b$  A generic member of Population 2 is the true father
- $H=2-c$  The brother of alleged father is the true father
- .....

# Alternative composite hypotheses

|         | H=1:AF | H=2a:Pop1 |
|---------|--------|-----------|
| Pr(H)   | 0.5    | 0.5       |
| Pr(H E) | 0.91   | 0.09      |

$$LR = \frac{\Pr(E | H = 1)}{\Pr(E | H = 2a)} = 10,11$$

# Alternative composite hypotheses

|         | H=1:AF | H=2a: Pop1 | H=2b: Pop2 |
|---------|--------|------------|------------|
| Pr(H)   | 0.5    | 0.25       | 0.25       |
| Pr(H E) | 0.666  | 0.033      | 0.30       |

$$WE = \frac{\Pr(E | H = 1)}{\Pr(E | H = 2a \cup 2b)} = \frac{0.666}{0.333} = 2$$

$$LR_a = \frac{\Pr(E | H = 1)}{\Pr(E | H = 2a)} = \frac{0.9091}{0.09} = 10.11$$

$$LR_b = \frac{\Pr(E | H = 1)}{\Pr(E | H = 2b)} = \frac{0.5263}{0.4737} = 1.11$$

WE= Weight of Evidence (i.e. the Bayesian version of the LR, essentially an “average” using prior probabilities as weights)

# Alternative composite hypotheses

|         | H=1:AF | H=2a: Pop1 | H=2b:Pop2 | H=2c: Brother |
|---------|--------|------------|-----------|---------------|
| Pr(H)   | 0.5    | 0.1666     | 0.1666    | 0.1666        |
| Pr(H E) | 0.659  | 0.022      | 0.1978    | 0.1208        |

$$WE = \frac{\Pr(E | H = 1)}{\Pr(E | H = 2a \cup 2b \cup 2c)} = 1.93$$

$$LR_a = \frac{\Pr(E | H = 1)}{\Pr(E | H = 2a)} = 10.11$$

$$LR_b = \frac{\Pr(E | H = 1)}{\Pr(E | H = 2b)} = 1.11$$

$$LR_c = \frac{\Pr(E | H = 1)}{\Pr(E | H = 2c)} = 1.818$$

# Alternative composite hypotheses : conclusions

- The Bayesian approach (WE) is attractive since it provides support to the main hypothesis taking into account all the alternatives at the same time
- Since it gives more, it demands more: prior probabilities on the hypotheses are required
- Identification of the victims of a mass disaster: we need to take into account many alternatives for each main hypothesis. Fortunately we can use some convincing non-informative prior probabilities

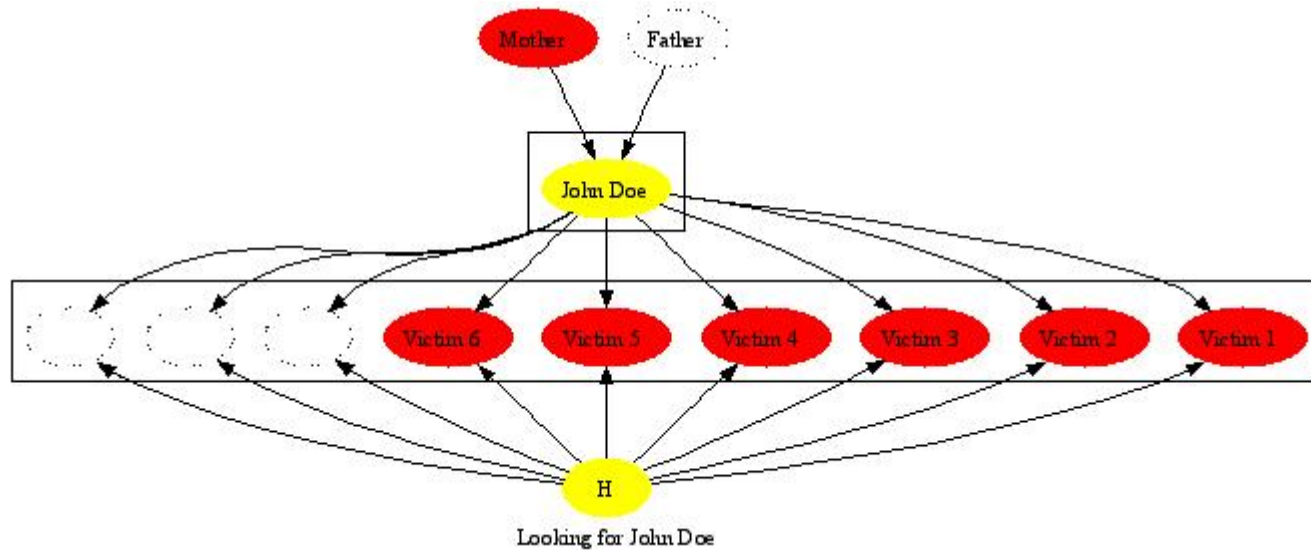
# Why are many alternatives relevant in the Mass Disaster Identification problem?

- Assume a disaster has occurred
- $N$  persons were involved
- $n(V)$  victims were recovered
- $n(M)$  missing persons were named
- A question of possible interest arises...

# A question of interest

- The family of John Doe is interested in finding JD among the victims, so they ask:
- “Is victim 12 John Doe?”
- What is behind the alternative “John Doe is not Victim 12”? There are some specific alternatives:
  - 1) JD is one of the other  $n(V)-1$  recovered bodies
  - 2) JD has still not been recovered

# A graph representing the problem



# A question of interest

- To derive the posterior identification probability for JD with respect to all the recovered victims (including Victim 12), we need a prior probability on each of them. Since we assume to have only DNA evidence, a non-informative prior probability is a sensible choice

$$\Pr(H_{JW} = 1) = \cdots \Pr(H_{JW} = n(V)) = \frac{1}{N}$$

$$\Pr(H_{JW} = \text{Rest}) = \frac{N - n(V)}{N}$$

# A question of interest

- After a non-informative prior probability is given to the hypothesis, the posterior probabilities, conditionally on the evidence, can be provided

| $H_{JW}$        | V1   | ....  | V12  | V13  | Rest |
|-----------------|------|-------|------|------|------|
| $\Pr(H_{JD})$   | 0.01 | 0.01  | 0.01 | 0.01 | 0.87 |
| $\Pr(H_{JD} E)$ | 0.1  | ..... | 0.88 | 0.02 | 0.01 |

Note: to evaluate if JD is Victim 12, we used information on all the other victims obtaining a probability for each of the components of the hypothesis

# A question of interest

- If required, it is possible to derive the WE from prior and posterior hypothesis probabilities

$$\begin{aligned} WE &= \frac{\Pr( E | H = 12 )}{\Pr( E | H = 1 \cup \dots \cup H = 13 \cup H = rest )} \\ &= \frac{\Pr( H = 12 | E )}{1 - \Pr( H = 12 | E )} \frac{1 - \Pr( H = 12 )}{\Pr( H = 12 )} \end{aligned}$$

# Another question of interest

- Also the family of Mario Rossi is interested in finding MR among the victims
- Is it correct to proceed performing the same analysis, i.e. starting from a non-informative prior probability and deriving the posterior identification probability for all the victims????
- A possible result of such an approach

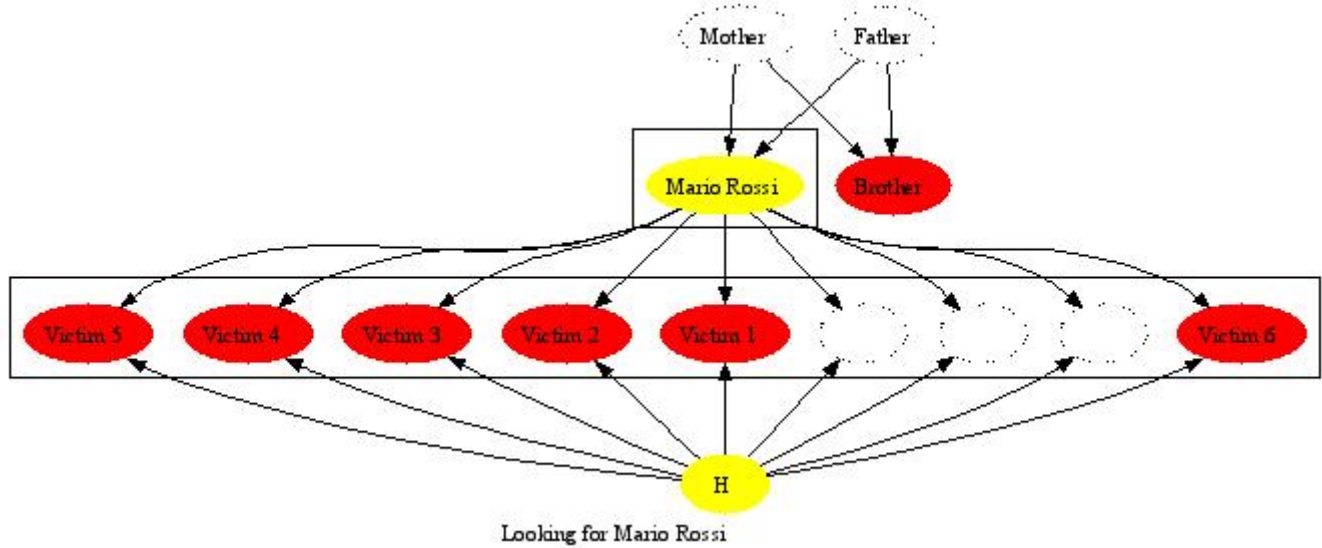
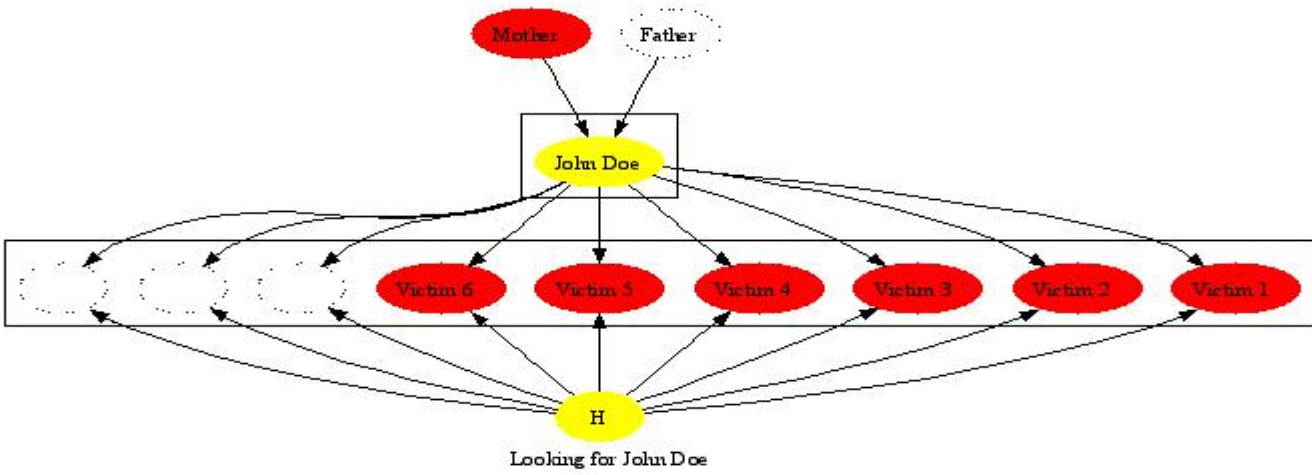
# Another question of interest

| H               | V1   | ....  | V12  | V13  | Rest |
|-----------------|------|-------|------|------|------|
| $\Pr(H_{JD})$   | 0.01 | ..... | 0.01 | 0.01 | 0.87 |
| $\Pr(H_{JD} E)$ | 0.11 | ..... | 0.82 | 0.02 | 0.01 |
| $\Pr(H_{MR})$   | 0.01 | ..... | 0.01 | 0.01 | 0.87 |
| $\Pr(H_{MR} E)$ | 0.94 | ..... | 0.01 | 0.02 | 0.01 |

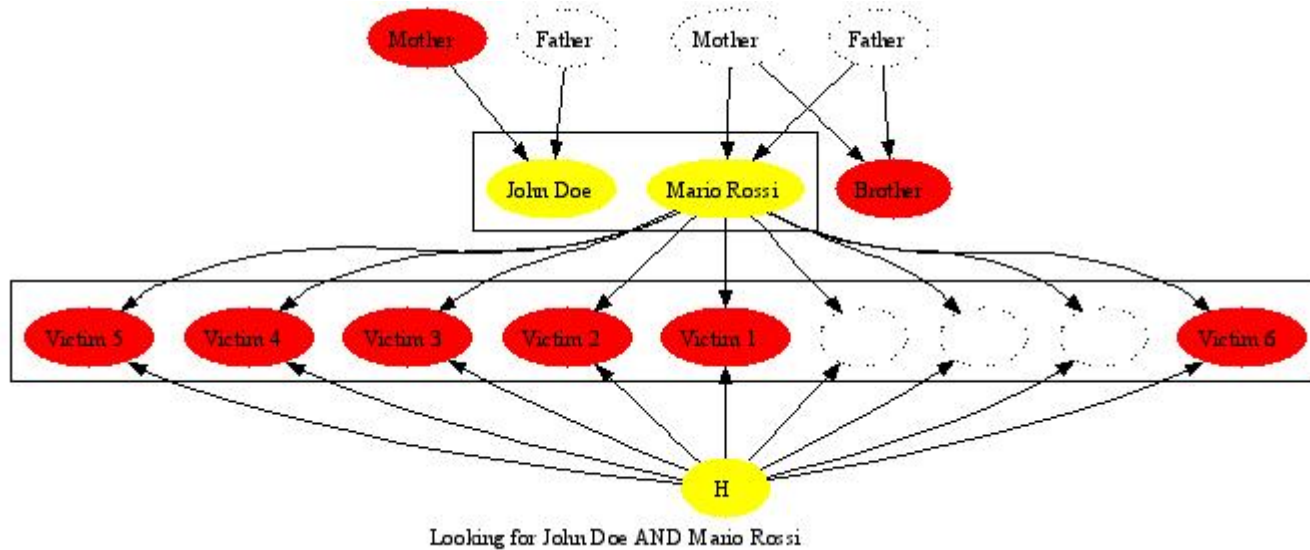
# Another question of interest

- Note: victim “1” has the probability of being one of the two missing persons with a probability  $>1!!!$
- This incoherency is because the hypothesis random variables are actually dependent, since they consider the same victims with respect to different missing individuals

# AQI: two separate analyses



# AQI: One integrated analysis



# Other questions of interest

- Obviously, other questions of interest concern all the missing persons included in the search
- Note: In this way the dual problem is also considered, i.e. the probability that a corpse belongs to one of the missing individuals or to someone else still nameless. The possible question could be:
- “What is the probability that victim 7 is one of the missing persons searched for by their families or is among those still unnamed??”

# Inference: the configuration hypothesis random variable

- What about the hypothesis random variable when several victims and missing persons are considered all together?

| The Missing H | 1   | 2   | 3   | 4   | n(M) | ... | ... | ... | N   |
|---------------|-----|-----|-----|-----|------|-----|-----|-----|-----|
| 1             | V1  | V7  | ... | ... | V5   | ... | V16 | ... | V76 |
| 2             | ... | ... | ... | ... | ...  | V5  | V8  | V7  | V32 |
| ....          | V2  | V4  | V8  | V44 | V77  | ... | ... | ... | ... |

# Inference: the main steps (1)

- Recognize that the number of possible identifiable victims are in this range:

$$[(\max(0, n(M) + n(V)), \min(n(M) + N(V)))]$$

- Produce all the possible states of the configuration hypothesis NOT assigning the same victim to more than one missing person
- Assume a non-informative prior probability for each of these states

# Inference: the main steps (2)

- Provide posteriors on the configuration hypothesis random variable via:
  - ❖ A propagation in a Probabilistic Expert System (PES)
  - ❖ Algebraic solution as provided in Cavallini-Corradi (2008)

# Inference: the main steps (3)

- Derive marginal probabilities for each victims and each missing persons summing up all over the posterior configuration hypothesis probabilities.

| Missings<br>H | 1   | 2   | 3   | 4   | n(M) | ... | ... | N   | Pr(H E) |
|---------------|-----|-----|-----|-----|------|-----|-----|-----|---------|
| 1             | V2  | ... | ... | V7  | V12  | V71 | V88 | ... | 0.15    |
| 2             | V1  | V7  | ... | ... | V5   | ... | V16 | ... | 0.09    |
| ....          | ... | ... | ... | ... | ...  | V5  | V8  | V7  | 0.01    |
| N(H)          | V2  | V4  | V8  | V44 | V77  | ... | ... | ... | 0.06    |

# Conclusions

L4V provides:

- for all the missing persons the probabilities to be one of the recovered victims or to be still not recovered
- for all missing persons the probabilities to be one of the recovered victims or to be still not recovered