Evaluating Drop-out Probabilities

Hinda Haned
A three-person mixture

- sample: epithelial cells recovered from the victim of an assault
- two suspects are detained by the police
A three-person mixture

Hypotheses

- Hp: the victim, suspect 1 and one unknown contributed to the sample
- Hd: the victim and two unknowns contributed to the sample

Evaluation of the two hypotheses using likelihood ratios
Sensitivity analysis - LRMix

LRs range from $10^3$ to $10^7$:

- $10^3 \Rightarrow \text{PrD}=0.99$
- $10^7 \Rightarrow \text{PrD}=0.41$
Available methods

➢ Experimental mixtures (Perez et al, Coratian Med J, 2011)
  ▪ the levels of drop-out, based on large sets of DNA mixtures obtained in different conditions
  ▪ DNA quants
  ▪ Number of contributors
  ▪ Ratio of contribution

➢ Maximum likelihood principle LoComation software (Gill et al 2007)
  ▪ Drop-out probabilities that maximize the probability of observing the questioned epg
Available methods

▶ Experimental mixtures (Perez et al, Coratian Med J, 2011)
  - the levels of drop-out, based on large sets of DNA mixtures obtained in different conditions
  - DNA quants
  - Number of contributors
  - Ratio of contribution

▶ Maximum likelihood principle LoComation software (Gill et al 2007)
  - Drop-out probabilities that maximize the probability of observing the questioned epg

Methods derive estimates from empirical distributions
Qualitative approach to the estimation of PrD

Relies on:
- the number of alleles observed in the sample
- the genotypes of the hypothesized contributors under H

What are the probabilities of dropout that could have led to observing the same number of alleles recovered in the sample?
Qualitative approach to the estimation of PrD

Relies on:
- the number of alleles observed in the sample
- the genotypes of the hypothesized contributors under H

What are the probabilities of dropout that could have led to observing the same number of alleles recovered in the sample?

What is the distribution of the number of alleles for the questioned sample, conditioned on PrD?
Qualitative approach to the estimation of PrD

What are the probabilities of dropout that could have led to observing the same number of alleles recovered in the sample?

We don’t know the probabilities of drop-out, but we can evaluate the drop-out probabilities that could have led to a mixture similar to the one we are investigating.
Qualitative approach to the estimation of PrD

What are the probabilities of dropout that could have led to observing the same number of alleles recovered in the sample?

We don’t know the probabilities of drop-out, but we can evaluate the drop-out probabilities that could have led to a mixture similar to the one we are investigating.

Build the empirical distributions of the numbers of alleles, conditioned on the probabilities of dropout ranging in [0,1] using Monte-Carlo simulations.
Monte Carlo method

Any method which solves a problem by generating suitable random numbers and observing that fraction of the numbers obeying some property or properties.
Monte Carlo method

Any method which solves a problem by generating suitable random numbers and observing that fraction of the numbers obeying some property or properties.

Questioned sample properties:
- three-person mixture: SGM+
- 33 alleles observed in the epg
- profiles of victim and suspect available
Monte Carlo method

Any method which solves a problem by generating suitable random numbers and observing that fraction of the numbers obeying some property or properties.

Questioned sample properties:
- three-person mixture: SGM+
- 33 alleles observed in the epg
- profiles of victim and suspect available

Simulate a large number of mixtures that have these properties
Important note!

The hypothesized contributors change under Hp and under Hd:

- Derive distribution of the numbers of alleles under Hp and under Hd separately

- Yields two distributions, one under Hp and one under Hd
Monte-Carlo simulation procedure
Step 1: simulate 1000 mixtures

Victim

| Suspect | fixed | 36 alleles |

Unknown

| By randomly sampling from the Dutch DNA frequencies | 17 alleles |

Mixture #1

51 distinct alleles

Mixture #2

50 distinct alleles

Mixture #1000

53 distinct alleles

...
Step 2: Apply drop-out

<table>
<thead>
<tr>
<th>Mixture #1</th>
<th>51 alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Victim</td>
<td></td>
</tr>
<tr>
<td>Suspect</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pr(D)</th>
<th># surviving alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>51</td>
</tr>
<tr>
<td>0.02</td>
<td>50</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.50</td>
<td>25</td>
</tr>
<tr>
<td>...</td>
<td></td>
</tr>
<tr>
<td>0.99</td>
<td>1</td>
</tr>
</tbody>
</table>
Repeat the procedure 1000 times

- Simulate 1000 mixtures
- For each mixture, record the number of alleles obtained after the simulated drop-out procedure
Repeat the procedure 1000 times

<table>
<thead>
<tr>
<th>PrD</th>
<th>sim1</th>
<th>sim2</th>
<th>sim3</th>
<th>sim4</th>
<th>sim5</th>
<th>sim6</th>
<th>sim7</th>
<th>sim8</th>
<th>sim9</th>
<th>sim10</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>51</td>
<td>50</td>
<td>49</td>
<td>50</td>
<td>53</td>
<td>54</td>
<td>45</td>
<td>47</td>
<td>51</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>0.03</td>
<td>50</td>
<td>50</td>
<td>44</td>
<td>44</td>
<td>51</td>
<td>39</td>
<td>47</td>
<td>51</td>
<td>44</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>46</td>
<td>44</td>
<td>44</td>
<td>40</td>
<td>44</td>
<td>41</td>
<td>41</td>
<td>50</td>
<td>41</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>0.09</td>
<td>38</td>
<td>48</td>
<td>43</td>
<td>41</td>
<td>42</td>
<td>33</td>
<td>40</td>
<td>41</td>
<td>40</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>0.17</td>
<td>38</td>
<td>37</td>
<td>33</td>
<td>44</td>
<td>37</td>
<td>51</td>
<td>46</td>
<td>39</td>
<td>36</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>0.89</td>
<td>7</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>7</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>0.91</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>0.93</td>
<td>8</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>0.95</td>
<td>5</td>
<td>0</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>0.97</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>0.99</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
We look at the distributions of the numbers of alleles obtained in the simulation procedure
We are only interested in those where we obtained 33 alleles (the number observed in the epg).
Now we look at the distribution of the numbers of alleles and the corresponding drop-out probabilities
We are only interested in those where we obtained 33 alleles (the number observed in the epg).
The same procedure is carried out under Hd

- Victim: fixed alleles = 19
- Unknown 1: by randomly sampling from Dutch DNA frequencies, alleles = 35
- Unknown 2: by randomly sampling from Dutch DNA frequencies, alleles = 35

Mixture #1: 49 distinct alleles
Mixture #2: 50 distinct alleles
Mixture #1000: 48 distinct alleles
Apply drop-out

<table>
<thead>
<tr>
<th>Mixture #1</th>
<th>49 alleles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Victim</td>
<td></td>
</tr>
<tr>
<td>Suspect</td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
</tr>
</tbody>
</table>

Pr(D) | # surviving alleles  
0.01  | 48  
0.02  | 47  
...   | ... 
0.50  | 25  
...   |     
0.99  | 1   

The simulation procedure is repeated on a 1000 mixtures
Distribution of the numbers of alleles among the 1000 mixtures

We are only interested in those where we obtained 33 alleles (the number observed in the epg).
5% - 95% percentiles of the distributions

<table>
<thead>
<tr>
<th></th>
<th>Under Hp</th>
<th></th>
<th>Under Hd</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>0.19</td>
<td>5%</td>
<td>0.09</td>
</tr>
<tr>
<td>95%</td>
<td>0.45</td>
<td>95%</td>
<td>0.41</td>
</tr>
</tbody>
</table>

The drop-out estimates are given as a range: lowest-highest value

[0.09, 0.45]
Sensitivity analysis
Sensitivity analysis vs. plausible ranges for PrD

$$LR \cong 10^7$$
Summary

➢ the ranges of the drop-out probability can be evaluated separately under Hp and Hd

➢ avoid reporting values of drop-out that are supported by one hypothesis but not by its alternative

➢ qualitative data only

➢ peaks slightly under threshold not taken into account

Assess the uncertainty the data!
LRmix module
Available in Forensim 4.1