Note 1 - Prioritizing in MPI cases: the role of preliminary investigation in Database Search

Franco Marsico

Buenos Aires, 4/1/2021

This note provides the analytical details related to the github repository https://github.com/MarsicoFL/CAPI. Childhood abduction preliminary investigation is a tool for computing prior probabilities in missing children identification cases through database search.

Setting up libraries

Dirichlet package provides the key functions of the code.

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Data set

There is a substantial lack of data in this type of cases. Mainly when the pregnant mother was abducted, the gender of the missing children is unknown. Even so, there is a possibility of deducing a birth date. These cases comprise different situations. When the newborn baby or a child was abducted, the birth date could be precisely known. When the pregnant mother was kidnapped, or there is contradictory or a lack of testimonies, it could be estimated within a range. On the other hand, the abductors give a false name and birth date to the children. We will refer to the children's real birth date as Actual Birth Date (ABD), and the documented or declared (eventually false birth date as Declared Birth Date (DBD). One of the main objectives is to propose who to formalize a relation between variables ABD and DBD. We consider H1: POI is the MP, and H2: POI is not related to MP. Here, we study the relationship of these two variables using as example a set of solved cases of the "Missing Grandchildren of Argentina" case. ABD and DBD dates of the 90 examples are modified in order to preserve anonymity but distance between ABD and DBD is maintained.

Descriptive statistics

Then, we analyzed some descriptive statistics regarding the distance between ABD and DBD. In the histogram we can see that in most of the cases ABD and DBD are close, so DBD minus ABD in near to zero.
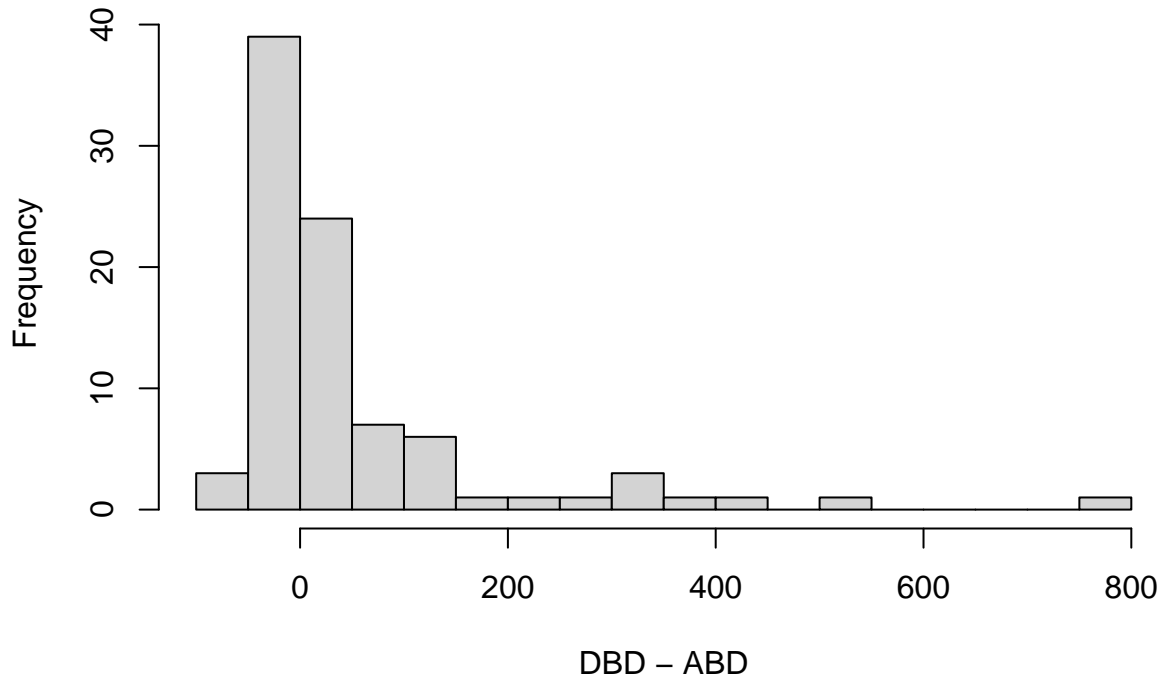
DBD − ABD

Sorting the cases by DBD minus ABD value we can see that few cases present an ABD > DBD (so DBD - ABD < 0). On the other hand some of them have a DBD much greater than ABD. It results on DBD - ABD > 730 days (more than two years!). One can intuitively think about the complications of having a two years child declared as a new born baby. This complication could ignite suspicion in the social environment. For example in scholar and health system. For these complication we can guess that this type of cases are infrequent.

number of cases
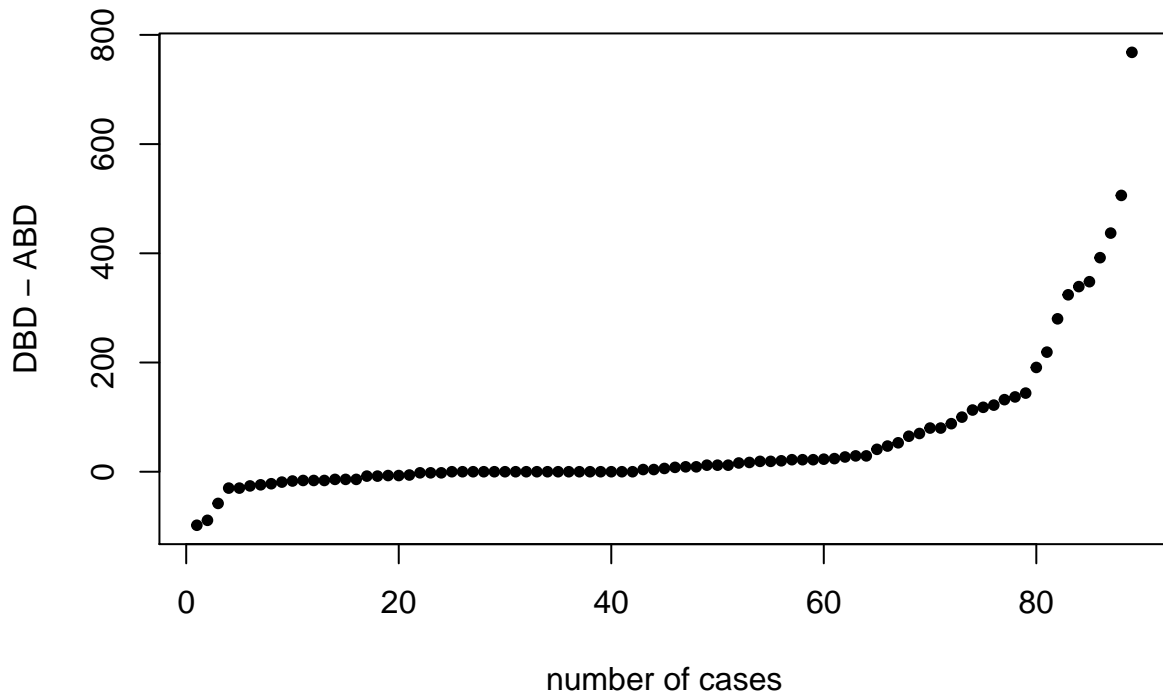
Prosecutor and legal system researchers tries to explain this process as inscription system based. For example, people related with persons of the legal inscription of with trafficking collaborators could "relax" the annotation birth date. In contrast, abductors with few or none contacts will have to be more cautious

and it will derive in closer ABD and DBD dates. Based on this assumptions, researchers divide the DBD - ABD distance (in days) in groups: Group A or 1: -inf;-120 Group B or 2: -120;-30 Group C or 3: -30;30 Group D or 4: 30;120 Group E or 5: 120;240 Group F or 6: 240;360 Group G or 7: 360;inf Below we analyze the frequency table of that groups in the solved-cases.csv dataset:

```
Freqs <- data.frame(table(Solved$Group))
Freqs
```

```
##   Var1 Freq
## 1    A    1
## 2    B    4
## 3    C   60
## 4    D   11
## 5    E    6
## 6    F    4
## 7    G    4
```

Following this classification, we propose to transform these frequencies in prior probabilities. The main assumption is that social and legal mechanism shaping the relationship between DBD - ABD remains unchanged. So, if we think in a database search, each pedigree will have an ABD (exactly known or calculated between a range). In contrast, each Person of Interest will have a DBD. Below we set the parameter for the diricihlet process.

Seven dimensions with alfa parameters obtained from the frequency table allows us to perform simulations and fitting obtained results for a dirichlet distribution. Maximum likelihood approach is used.

Below we show the prior probability of the defined groups. Each pedigree-POI pairs will correspond for a specific group based on DBD - ABD and, therefore, have a pior probability. This prior could be updated with genetic testing or more formalized preliminary investigation data. Particularly this approach could be useful when a subset of POI must be selected in order to obtain genetic data. New solved cases could be incorporated to the solved-cases.csv dataset and actualize the frequency table, modifying dirichlet alpha parameters and prior odds respectively.