NUŠL
národní
úložiště
šedé
literatury

**Stochastic Approaches to Identification Process in Forensic Medicine and Criminalistics**

Slovák, Dalibor
2011

Dostupný z http://www.nusl.cz/ntk/nusl-55975

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 30.09.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .

# Stochastic Approaches to Identification Process in Forensic Medicine and Criminalistics

*Post-Graduate Student:*
MGR. DALIBOR SLOVÁK
Department of Medical Informatics
Instutite of Computer Science of the ASCR, v. v. i.
Pod Vodárenskou věží 2
182 07 Prague 8, CZ
slovak@euromise.cz

*Supervisor:*
PROF. RNDR. JANA ZVÁROVÁ, DRSC.
Department of Medical Informatics
Instutite of Computer Science of the ASCR, v. v. i.
Pod Vodárenskou věží 2
182 07 Prague 8, CZ
zvarova@euromise.cz

Field of Study:
Biomedical informatics

This work was supported by The Centre of Biomedical Informatics and MŠMT ČR project 1M06014.

**Abstract**

In this paper we study an identification of culprit and assesment of evidence against him. We define a simple model called the island problem and we derive the weight-of-evidence formula in its basic form. We find how we can deal with uncertainty about basic parametres of model, like size of population. We investigate possibility of inclusion of relatedness and subpopulation structure into model through beta-binomial formula, we enlarge DNA mixtures of DNA and at the close we present brief overview about DNA databases.

## 1. Introduction

Technological progress that allows the use of DNA has caused a revolution in criminology. It helps convict the perpetrators of those crimes that once appeared irresolvable and also helps prove the innocence who have already been convicted. DNA analysis is now accepted by the broad public as a completely standard procedure, which reliably convicts the offender. Here, however, hides one of the main problems that results from using DNA, for even DNA evidence is not foolproof.

Several possibilities keep DNA from being completely reliable: for example there may be a false location of the trace (more specifically, the offender may have discarded a cigarette butt which had previously been smoked by someone else); the wrong take of biological samples or damage to the samples could have occurred; or there may have been secondary transfer of biological material. However, mathematicians do not deal with any of these things. Rather, they are faced with the following task: if all of the above options are excluded, what is the probability that a particular offender is a detained person, given that the perpetrator's DNA and the DNA of the suspect are available?

In forensic practice, genetic profiles consisting of the short tandem repeat (STR) polymorphisms are currently used. The number of polymorphisms varies from country to country, with the smallest being seven used in Germany and a maximum of sixteen used in the Czech Republic. The probability of correct identification depends on the number of comparisons of polymorphisms (or loci where studied polymorphisms lie) and their genetic variability. The more we investigate loci and the greater the variability between individual loci, the smaller the probability that the other person will have the same configuration (and therefore the same genetic profile). Due to the quality of biological material and its amount it is not always possible to investigate all of the polymorphisms and very often genetic profiles contain fewer loci than is necessary to uniquely identify them.

In the following text we will assume that we examine only one locus. Assuming independence of loci, generalization to a larger number of loci can be performed using product rule (i.e. multiplying the individual marginal probabilities).

## 2. Formalization

Denotation

- E - evidence or information about the crime (i.e. the circumstances, witness testimonies, crime scene evidence, etc.)
- G - an event at which the suspect is guilty
- I - an event at which the suspect is innocent
- $C_i$ - an event at which the culprit is a person $i$
- $\mathcal{I}$ - the population of alternative suspects.

Our goal is to determine the conditional probability of $\mathsf{P}(G|E)$ that, given circumstances $E$, the suspect is truly the culprit of the investigated crime. According to Bayes theorem

$$\mathsf{P}(G|E) = \frac{\mathsf{P}(E|G)\mathsf{P}(G)}{\mathsf{P}(E|G)\mathsf{P}(G) + \mathsf{P}(E|I)\mathsf{P}(I)}. \quad (1)$$

However, the expression $\mathsf{P}(E|I)$ cannot be counted directly. The suspect is innocent if and only if there exists an index $i \in \mathcal{I}$ in which the event $C_i$ occurs. Then the event $I$ is equivalent to the event $\cup_{i\in\mathcal{I}} C_i$ and thanks to the disjunction of events $C_i$ holds:

$$\mathsf{P}(I) = \mathsf{P}\left(\cup_{i\in\mathcal{I}} C_i\right) = \sum_{i\in\mathcal{I}} \mathsf{P}(C_i).$$

Thus

$$\begin{aligned}
\mathsf{P}(E|I)\mathsf{P}(I) &= \mathsf{P}\left(E|\cup_{i\in\mathcal{I}} C_i\right)\mathsf{P}\left(\cup_{i\in\mathcal{I}} C_i\right) = \\
&= \frac{\mathsf{P}\left(E\cap(\cup_{i\in\mathcal{I}} C_i)\right)}{\mathsf{P}\left(\cup_{i\in\mathcal{I}} C_i\right)}\mathsf{P}\left(\cup_{i\in\mathcal{I}} C_i\right) = \\
&= \mathsf{P}\left(\cup_{i\in\mathcal{I}}\left(E\cap C_i\right)\right) = \\
&= \sum_{i\in\mathcal{I}} \mathsf{P}\left(E\cap C_i\right) = \\
&= \sum_{i\in\mathcal{I}} \mathsf{P}(E|C_i)\mathsf{P}(C_i).
\end{aligned}$$

Define **likelihood ratio**

$$R_i = \frac{\mathsf{P}(E|C_i)}{\mathsf{P}(E|G)} \quad (2)$$

which expresses how many times the probability of evidence $E$ is greater under the condition that the culprit is a person $i$ than under the condition that the culprit is the suspect. Further define **likelihood weights**

$$w_i = \frac{\mathsf{P}(C_i)}{\mathsf{P}(G)}$$

which expresses how many times the prior probability of committing a crime by a person $i$ is greater than the prior probability of committing a crime by the suspect.

Then

$$\mathsf{P}(G|E) = \frac{1}{1+\sum_{i\in\mathcal{I}} w_i R_i}. \quad (3)$$

The formula (3) is usually called **the weight-of-evidence formula**.

### 3. The island problem

The simplest application of the previous part is the "island problem". This is a model where a crime is committed on an inaccessible island which contains $N$ people who are unrelated to each other. At the beginning, there is no information about the offender, so we assign to each of the islanders the same (prior) probability of committing a crime. Then the offender is found to possess a certain characteristic $\Upsilon$ and the suspect is also found to have that characteristic, $\Upsilon$. The question becomes, to what extent can we be sure that we have found the suspect who is truly the culprit?

Using the formula (3) we get

$$\mathsf{P}(G|E) = \frac{1}{1+N\cdot p}, \quad (4)$$

where $p$ is the probability of the $\Upsilon$. For example if $p = 0.01$ and $N = 100$ then $\mathsf{P}(G|E) = 1/2$.

The previous result can be modified for more complex (and realistic) situations. Let's see where our simple model can fail:

- *Typing and handling errors*
  As the test may give erroneous results in a small percentage of cases, errors caused by human factor must also be considered: contamination or replacement of a sample from which the $\Upsilon$-status is investigated; incorrect evaluation of the results, or even intentional misrepresentation.

- *The population size*
  Often the population size $N$ is only estimated and furthermore, if there is migration in the population, then it is necessary to account for greater uncertainty within the population size.

- *The probability of occurrence $\Upsilon$ in the population*
  The value of $p$ is usually unknown and is therefore estimated on the basis of relative frequency of the $\Upsilon$ in a smaller sample or in a similar population, about which we have more information. However, these auxiliary data may be outdated or may only partially describe the ivestigated population.

- *Suspect searching*
  The suspect is not usually chosen randomly from the population but on the basis of other circumstantial evidence which increase the probability of guilt. Another possibility is choose the suspect by testing persons from the population for the presence of $\Upsilon$. In this way, people who are not $\Upsilon$-bearers can be excluded and thus the population size of alternative suspects is reduced.

- *Relatives and population subdivision*
  If the suspect (or other person being tested) is a $\Upsilon$-bearer and some of his relatives are included in the population too, then in the case of DNA profile

increases the probability of other persons having $\Upsilon$ due to inheritance. Similarly, unusually high relative frequency of a rare character usually occurs within the same subpopulation due to its shared evolution history.

- *The same prior probability of committing a crime* Although this requirement intuitively corresponds with the general presumption of innocence, we can asses varying prior probability (i.e. based on the distance from the scene, time availability, or a possible alibi).

We will analyze some of these cases in detail in the following sections.

## 4. Uncertainty about population size

The uncertainty in population size of possible alternative suspects affects the prior probability, $\mathsf{P}(G)$. Consider the population size $\tilde{N}$ as a random variable with mean $N$. Prior probability of guilt, conditional on value $\tilde{N}$, is

$$\mathsf{P}(G|\tilde{N}) = 1/(\tilde{N}+1).$$

However, since $\tilde{N}$ is not known, we use the expectation:

$$\mathsf{P}(G) = \mathbb{E}\left[G|\tilde{N}\right] = \mathbb{E}\left[\frac{1}{\tilde{N}+1}\right].$$

The function $1/(\tilde{N}+1)$ is not symmetric, but is convex on the interval $(0,\infty)$. Therefore Jensen's inequality for convex functions ($\mathbb{E}[f(x)] \geq f(\mathbb{E}[x])$) implies

$$\mathsf{P}(G) = \mathbb{E}\left[\frac{1}{\tilde{N}+1}\right] \geq \frac{1}{N+1}$$

because $\mathbb{E}[\tilde{N}] = N$.

Thus the uncertainty of the value $N$ tends to favor the defendant. This effect is usually very small. Let it be shown in a concrete example.

For $\varepsilon \in (0, 0.5)$ we put

$$\tilde{N} = \begin{cases} N-1 & \text{with probability } \varepsilon \\ N & \text{with probability } 1-2\varepsilon \\ N+1 & \text{with probability } \varepsilon. \end{cases}$$

Then

$$\begin{aligned} \mathsf{P}(G) &= \mathsf{E}\left[\frac{1}{\tilde{N}+1}\right] = \frac{\varepsilon}{N} + \frac{1-2\varepsilon}{N+1} + \frac{\varepsilon}{N+2} = \\ &= \frac{1}{N+1} + \frac{2\varepsilon}{N(N+1)(N+2)} \geq \frac{1}{N+1} \end{aligned}$$

and if we put $\varepsilon = 0.25$ with $N = 100$ then $\mathsf{P}(G)$ is greater than $1/(N+1)$ by only $0.000000485$.

Let's see what uncertainty in population size causes by using formula (4):

$$\begin{aligned} \mathsf{P}(G|E) &= \frac{1}{1+\sum_i R_i \frac{\mathsf{P}(C_i)}{\mathsf{P}(G)}} = \\ &= \frac{1}{1+p\frac{1}{\mathsf{P}(G)}\underbrace{\sum_i \mathsf{P}(C_i)}_{=1-\mathsf{P}(G)}} = \\ &= \frac{1}{1+p\frac{N(N+1)(N+2)}{N^2+2N+2\varepsilon}(1-\frac{N^2+2N+2\varepsilon}{N(N+1)(N+2)})} = \\ &= \frac{1}{1+Np\frac{N^3+2N^2-2\varepsilon}{N^3+2N^2+2N\varepsilon}} = \\ &= \frac{1}{1+Np\left(1-2\varepsilon\frac{N+1}{N^3+2N^2+2N\varepsilon}\right)}. \end{aligned}$$

Again substituting $\varepsilon = 0.25$ and $N = 100$ we conclude that $\mathsf{P}(G|E) = 0.5000124$ which, despite the high value of $\varepsilon$, differs from the original result of $50\ \%$, which was calculated with a fixed $N$, by just one thousandth of a percent. Therefore, continuing with uncertainty about $N$,

$$\mathsf{P}(G|E) \approx \frac{1}{1+Np\,(1-2\varepsilon/N^2)}$$

is very good approximation to take. In this example the approximation gives $\mathsf{P}(G|E) = 0.5000125$, which is $50.00125\ \%$.

Balding [1] uses an approximation order of worse magnitude

$$\mathsf{P}(G|E) \approx \frac{1}{1+Np\,(1-4\varepsilon/N^3)}$$

which gives our example the value $\mathsf{P}(G|E) = 0.5000003$, or $50.00003\ \%$.

## 5. Relatives and population structure

Alleles, which are identical and come from a common ancestor, are called identical by descent (*ibd*). The commonality of recent evolution history between two persons, whether relatives or members of the same subpopulation, increases the probability of ibd alleles occurrence. Therefore, the coancestry coefficient $\theta$, indicating the probability that two randomly selected alleles

on fixed locus are ibd, is used as the measure of relatedness within subpopulations. Neglecting the influence of kinship and population structure leads to an overestimation of posterior probability of the suspect's guilt, and therefore ignoring this influence tends to cause disfavor for defendant. Thus, this topic is given considerable attention.

Consider a given locus with $J$ alleles $A_1, \ldots, A_J$ whose probability of occurrence in the population is $p_1, \ldots, p_J$, $\sum_{i=1}^{J} p_i = 1$. Allele proportions in the subpopulation can be modeled by the Dirichlet distribution ( [5]) with parameters $\lambda p_i, \lambda = \frac{1-\theta}{\theta(1-k)}$ where $\theta$ is the coancestry coefficient characterizing the subpopulation and $k$ is the proportion of the subpopulation within the general population. Thus the probability of drawing $m_i$ alleles $A_i$ ($\sum_i m_i = n$) is given by

$$\mathsf{P}(m_1, \ldots, m_J) = \frac{\Gamma(\lambda)}{\Gamma(\lambda + n)} \prod_{i=1}^{J} \frac{\Gamma(\lambda p_i + m_i)}{\Gamma(\lambda p_i)}. \quad (5)$$

Putting $m = (m_1, \ldots, m_J)$ we can adjust formula (5) to

$$\mathsf{P}(m) = \frac{\prod_{j=1}^{J} \prod_{i=0}^{m_j - 1} [(1-\theta) p_j + \theta i (1-k)]}{\prod_{i=0}^{n-1} [1 - \theta + \theta i (1-k)]}. \quad (6)$$

The formula (6) is usually called **the beta-binomial sampling formula** and applies to ordered samples. If we want to use unordered samples, it is necessary to multiply the result by $\frac{n!}{m_1! \cdots m_J!}$.

From the formula (6) we can also deduce the probability of certain allele withdrawal by using our knowledge of previous allele's withdrawal:

$$\mathsf{P}(m_j + 1 | m_1, \ldots, m_j, \ldots, m_J) = \\ = \frac{(1-\theta) p_j + m_j \theta (1-k)}{1 - \theta + n\theta (1-k)}. \quad (7)$$

### 5.1. Aplication of beta-binomial formula

Denote $G_C$ and $G_S$ as culprit and suspect genotypes, respectively, and denote $G_i$ as the genotype of a general person $i$. Then the likelihood ratio (2) can be rewritten as

$$R_i = \frac{\mathsf{P}(G_C = G_S = D | C_i)}{\mathsf{P}(G_C = G_S = D | G)} = \\ = \frac{\mathsf{P}(G_i = G_S = D)}{\mathsf{P}(G_S = D)} = \mathsf{P}(G_i = D | G_S = D).$$

Suppose first that the culprit has a homozygous profile $A_j A_j$. Then calculate the probability that the suspect has the same homozygous profile:

$$R_i = \mathsf{P}(G_i = A_j A_j | G_S = A_j A_j) \equiv \mathsf{P}(A_j^2 | A_j^2) = \\ = \mathsf{P}(A_j | A_j^3) \cdot \mathsf{P}(A_j | A_j^2)$$

We know to calculate these conditional probabilities using (7). First we put $m_j = n = 2$ and then $m_j = n = 3$. Therefore

$$R_i = \frac{[(1-\theta) p_j + 2\theta (1-k)]}{[1 - \theta + 2\theta (1-k)]} \times \\ \times \frac{[(1-\theta) p_j + 3\theta (1-k)]}{[1 - \theta + 3\theta (1-k)]}. \quad (8)$$

Similarly, we proceed for culprit with a heterozygous profile $A_j A_k$:

$$R_i = \mathsf{P}(G_i = A_j A_k | G_S = A_j A_k) \equiv \\ \equiv \mathsf{P}(A_j A_k | A_j A_k) = \\ = \mathsf{P}(A_k | A_j^2 A_k^1) \mathsf{P}(A_j | A_j^1 A_k^1) + \\ + \mathsf{P}(A_j | A_j^1 A_k^2) \mathsf{P}(A_k | A_j^1 A_k^1). \quad (9)$$

To quantify both expressions on the bottom line we put $m_j = 1, n = 2$ and $m_k = 1, n = 3$; $m_k = 1, n = 2$ and $m_j = 1, n = 3$ respectively. In total

$$R_i = 2 \frac{[(1-\theta) p_j + \theta (1-k)]}{[1 - \theta + 2\theta (1-k)]} \times \\ \times \frac{[(1-\theta) p_k + \theta (1-k)]}{[1 - \theta + 3\theta (1-k)]}. \quad (10)$$

## 6. DNA mixtures

If the DNA sample is found to have more than two alleles at one locus, then it is defined as a mixture. The number of contributors to the mixture can be known or estimated, usually as $\lceil \frac{n}{2} \rceil$ where $n$ is the maximum number of alleles detected. Due to the large number of situations which may arise we show for illustration only the case in which the victim ($V$) and one other person contribute to the mixture.

Thus the likelihood ratio $R_i$, defined by formula (2), can be rewritten as

$$R_i = \frac{\mathsf{P}(E_C, G_S, G_V | C_i)}{\mathsf{P}(E_C, G_S, G_V | G)} = \\ = \frac{\mathsf{P}(E_C | G_S, G_V, C_i)}{\mathsf{P}(E_C | G_S, G_V, G)} \cdot \frac{\mathsf{P}(G_S, G_V | C_i)}{\mathsf{P}(G_S, G_V | G)} = \\ = \frac{\mathsf{P}(E_C | G_S, G_V, C_i)}{\mathsf{P}(E_C | G_S, G_V, G)} = \frac{\mathsf{P}(E_C | G_V, C_i)}{\mathsf{P}(E_C | G_S, G_V, G)}. \quad (11)$$

### 6.1. Four alleles mixture

First we look at the case where the mixture consists of four alleles.

Suppose the following conditions apply:

1. None of the persons are considered relatives to each other.
2. The population is homogeneous (i.e. $\theta = 0$).
3. The population follows Hardy-Weinberg equilibrium.

Let the mixture be made up of alleles $A, B, C$, and $D$, with known probabilities of occurrence in the total population $p_A, p_B, p_C$, and $p_D$. Also let the suspect have alleles $A$ and $B$ and let the victim have alleles $C$ and $D$. Then the denominator in the formula (11) is equal to one, the numerator is equal to the probability of observing the person with alleles $A$ and $B$ (which using the information above assumes the probability of occurance $2p_Ap_B$), and therefore, the likelihood ratio is

$$R_i = 2p_Ap_B.$$

Suppose now that all considered persons have the same degree of relatedness to each other as expressed by the coancestry coefficient $\theta$. Then according to (7)

$$R_i = \mathsf{P}\left(AB|ABCD\right) =$$

$$= \frac{2\left[(1-\theta)p_A + \theta\left(1-k\right)\right]\left[(1-\theta)p_B + \theta(1-k)\right]}{\left[1-\theta+4\theta(1-k)\right]\left[1-\theta+5\theta\left(1-k\right)\right]}.$$

### 6.2. Three alleles mixture

In the case of three alleles in the sample it is necessary to assume at least two contributors to the mixture. Consider alleles $A, B$, and $C$ with probabilities $p_A, p_B$, and $p_C$. If the victim is homozygous for allele $C$, we get the same results as in the four allele's mixture.

Assume that the victim is heterozygous with alleles $A$ and $B$ and that the suspect is homozygous for allele $C$. Furthermore, assume that conditions 1 to 3 are fulfilled. Then the denominator of the formula (11) is again equal to one, the numerator is equal to the probability of observing a person who has the allele $C$ and does not have a different allele other than $A, B$, or $C$, and

$$\begin{aligned} R_i &= \mathsf{P}(AC) + \mathsf{P}(BC) + \mathsf{P}(CC) = \\ &= 2p_Ap_C + 2p_Bp_C + p_C^2\,. \end{aligned} \tag{12}$$

To include the population structure we use the formula (7) again:

$$\begin{aligned} R_i \quad &= \quad \mathsf{P}\left(AC|ABCC\right) + \mathsf{P}\left(BC|ABCC\right) + \\ &+ \quad \mathsf{P}\left(CC|ABCC\right) = \\ &= \frac{2\left[(1-\theta)\,p_A + \theta\left(1-k\right)\right]\left[(1-\theta)\,p_C + 2\theta\left(1-k\right)\right]}{\left[1-\theta+4\theta\left(1-k\right)\right]\left[1-\theta+5\theta\left(1-k\right)\right]} \\ &+ \frac{2\left[(1-\theta)\,p_B + \theta\left(1-k\right)\right]\left[(1-\theta)\,p_C + 2\theta\left(1-k\right)\right]}{\left[1-\theta+4\theta\left(1-k\right)\right]\left[1-\theta+5\theta\left(1-k\right)\right]} \\ &+ \frac{\left[(1-\theta)\,p_C + 3\theta\left(1-k\right)\right]\left[(1-\theta)\,p_C + 2\theta\left(1-k\right)\right]}{\left[1-\theta+4\theta\left(1-k\right)\right]\left[1-\theta+5\theta\left(1-k\right)\right]} \\ &= \quad \frac{\left[(1-\theta)\,p_C + 2\theta\left(1-k\right)\right]}{\left[1-\theta+4\theta\left(1-k\right)\right]} \times \\ &\quad \times \frac{\left[(1-\theta)\left(2p_A + 2p_B + p_C\right) + 7\theta\left(1-k\right)\right]}{\left[1-\theta+5\theta\left(1-k\right)\right]}. \end{aligned}$$

We assumed in the previous calculation that the suspect is homozygous with alleles $C$. If he is heterozygote with alleles $A$ and $C$, or $B$ and $C$ respectively, formula (12) remains unchanged under conditions 1 to 3. If population structure is included the likelihood ratio is

$$\begin{aligned} R_i \quad &= \quad \frac{\left[(1-\theta)\,p_C + \theta\left(1-k\right)\right]}{\left[1-\theta+4\theta\left(1-k\right)\right]} \times \\ &\quad \times \quad \frac{\left[(1-\theta)\left(2p_A + 2p_B + p_C\right) + 8\theta\left(1-k\right)\right]}{\left[1-\theta+5\theta\left(1-k\right)\right]}. \end{aligned}$$

## 7. DNA database

DNA profiles, as sequences of alphanumeric data, allows relatively easy storage in the database. Therefore national databases began being created in the late 1990's and have continued to function since then. Currently there are three major forensic DNA databases: the Combined DNA Indexing System (CODIS), which is maintained by the United States FBI; the European Network of Forensic Science Institutes (ENFSI) DNA database; and the Interpol Standard Set of Loci (ISSOL) database maintained by Interpol.

All of these systems divide DNA database into two subdatabases. In *the crime scene database* the biological samples which are collected at the scene are stored and in *the convicted offender database* genetic profiles of persons convicted in the past are stored. These two databases are compared with one another and eventual agreement of profiles is examined by qualified professionals.

The type of offenses for which DNA is stored differs among countries and states. Initially, these databases contained only samples from violent offenders, such as those convicted of aggravated assault, rape, or murder.

However, the value of obtaining DNA from offenders of less severe crimes has been recognized more in recent times, as it has been discovered that many small time criminals often become repeat offenders, and in some cases more violent future offenders. However, the power of a large bank of DNA samples can sometimes serve as a deterrent. A match of DNA evidence from a crime scene (which would then be logged in the crime scene database) to one in the convicted offender database rapidly solves the crime rapidly and efficiently, saving time, effort, and money. Conversely, the use of DNA evidence can also immediately prove a suspect's innocence ( [6]).

According to data from the United States in August of 2006, the crime scene database included approximately $150\ 000$ profiles and the convicted offender database more than $3\ 500\ 000$ profiles ( [2]). The national database of United Kingdom currently consists of over four million profiles, and increases monthly by forty to fifty thousand. The success of this approach has been confirmed by the increase in the number of solved crimes from twenty-four to forty-three percent within the United Kingdom, since the creation of the DNA databases.

Therefore, the database system has the support of public. From a negative standpoint, the DNA often reveals very sensitive, personal information and therefore it is necessary that databases are kept confidential and are thoroughly protected from abuse.

The Czech national database was created in 2002. After rapid development, the database now contains approximately ninety thousand genetic profiles.

**References**

[1] D.J. Balding, *Weight-of-evidence for forensic DNA profiles*, John Wiley & Sons, Ltd, pp.15-63, 2005.

[2] www.uoou.cz/uoou.aspx?menu=287&submenu =288, (available August 25, 2011).

[3] H. Kubátová and J. Zvárová (supervisor), *Statistical methods for interpreting forensic DNA mixtures*, MFF UK, Praha, p. 20, 2010.

[4] D. Slovák and J. Zvárová (supervisor), *Statistické metody stanovení váhy evidence v procesu identifikace jedince*, MFF UK, Praha, 2009.

[5] S. Wright, *The genetical structure of populations*, Ann. Eugen. 15, pp. 323-354, 1951.

[6] www.enotes.com/forensic-science/dna-evidence-social-issues, (available August 25, 2011).