Forensic Science International: Genetics 9 (2014) 93–101

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Forensic Population Genetics – Original Research

Exact computation of the distribution of likelihood ratios with forensic applications

Guro Dørum [a,*], Øyvind Bleka [b], Peter Gill [b,c], Hinda Haned [d], Lars Snipen [a], Solve Sæbø [a], Thore Egeland [a,b]

[a] Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences (UMB), Norway
[b] Department of Forensic Genetics, Norwegian Institute of Public Health, Oslo, Norway
[c] Rikshospitalet, University of Oslo, Oslo, Norway
[d] Netherlands Forensic Institute, Department of Human Biological Traces, The Hague, The Netherlands

ARTICLE INFO

Article history:
Received 20 December 2012
Received in revised form 25 October 2013
Accepted 30 November 2013

Keywords:
Likelihood ratios
Distribution of likelihood ratios
p-Values
Complex mixtures

ABSTRACT

If complex DNA profiles, conditioned on multiple individuals are evaluated, it may be difficult to assess the strength of the evidence based on the likelihood ratio. A likelihood ratio does not give information about the relative weights that are provided by separate contributors. Alternatively, the observed likelihood ratio can be evaluated with respect to the distribution of the likelihood ratio under the defense hypothesis. We present an efficient algorithm to compute an exact distribution of likelihood ratios that can be applied to any LR-based model. The distribution may have several applications, but is used here to compute a *p*-value that corresponds to the observed likelihood ratio. The *p*-value is the probability that a profile under the defense hypothesis, substituted for a questioned contributor e.g. suspect, would attain a likelihood ratio which is at least the same magnitude as that observed. The *p*-value can be thought of as a scaled version of the likelihood ratio, giving a quantitative measure of the strength of the evidence relative to the specified hypotheses and the model used for the analysis. The algorithm is demonstrated on examples based on real data. `R` code for the algorithm is freely available in the `R` package `euroMix`.

1. Introduction

DNA mixtures in forensics can be interpreted with the use of the random man not excluded (RMNE) approach or the likelihood ratio (LR) approach. The RMNE method estimates the probability that a random person in the population cannot be excluded as a possible contributor to the mixture evidence. It is easy to calculate and makes no assumptions about the number of contributors. It does not require the specification of any hypotheses. The LR approach, on the other hand, involves the calculation of a likelihood ratio between a prosecution hypothesis (H_p) and a defense hypothesis (H_d). More specifically, it is the ratio between the probability of the evidence given that H_p is true and the probability of the evidence given that H_d is true. It depends on the number of contributors specified in the hypotheses and the profiles of known and hypothesised contributors. RMNE does not have a formal framework for handling partially matching profiles, though an RMNE approach incorporating drop-out in a non-probabilistic way was presented in [1]. In comparison, likelihood ratios can include probabilities for drop-in and drop-out. The International Society for Forensic Genetics (ISFG) DNA commission [2] recommends the use of likelihood ratios instead of RMNE for the analysis of DNA mixtures. See e.g. [3] for a discussion of the pros and cons of RMNE and LR.

One of the main arguments against LR is that it may be difficult to explain the meaning of a large likelihood ratio in court [4]. The RMNE approach has a more intuitive understanding in the sense that it presents the probability of not excluding a random man as a contributor to the evidence, but the main criticism against its use is that it wastes information. The LR includes all the available information, and its value depends on a number of defined conditions. The LR is reported as supporting the prosecution hypothesis if it is >1, and if it is <1 then it supports the defence hypothesis. Often, likelihood ratios are only reported if the estimate is large and practice vary, but typically this critical 'number' is greater than 1 million.

The papers [5,6] propose to evaluate the robustness of the observed likelihood ratio by examining the distribution of the likelihood ratio under the defense hypothesis. We continue on this path by computing a *p*-value that works as a scaled version of the likelihood ratio. A *p*-value is the probability of obtaining a likelihood ratio larger or equally large to the one observed, conditioned on the defense hypothesis being true. Fig. 1 shows a

* Corresponding author at: Norwegian University of Life Sciences, Norway. Tel.: +47 64966247.
E-mail address: guro.dorum@nmbu.no (G. Dørum).

1872-4973/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fsigen.2013.11.008

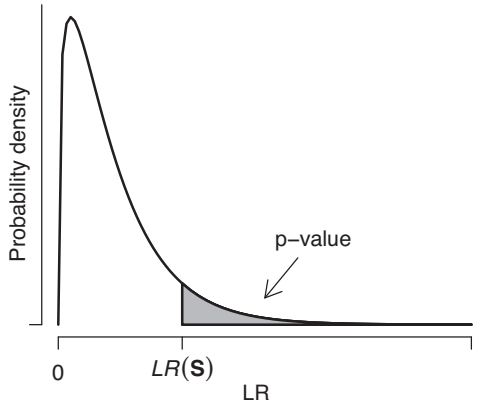

Fig. 1. A hypothetical probability distribution for likelihood ratios under H_d. The total area under the curve (i.e. the total probability) is equal to 1.0. The p-value is the grey area, which is the probability of observing a likelihood ratio larger or equally large as the observed $LR(\boldsymbol{S})$.

hypothetical probability distribution for likelihood ratios under the defense hypothesis when drop-in/drop-out is considered. Allowing for drop-in and drop-out in the evidence means that many genotype profiles could fit into the defense hypothesis, which again may result in a smooth distribution of likelihood ratios like the one in the figure. Note, in the example illustrated, the suspect is clearly indicated as a contributor since the majority of LR's are small, close to 0, whereas the observed LR, $LR(\boldsymbol{S})$, is large. The grey area under the curve gives the probability of observing a likelihood ratio at least as large as $LR(\boldsymbol{S})$. The p-value is based on likelihood ratios and therefore has the advantage of utilising all available information, and at the same time it has an intuitive interpretation as RMNE.

1.1. Hypothesis testing and p-values

The term p-value comes from the statistical hypothesis testing framework, where the purpose is to draw inferences regarding some hypotheses based on the data. A null hypothesis H_0 and an alternative hypothesis H_1 are formulated. A test is performed on the basis of observed data to decide whether H_0 can be rejected. Typically, H_0 is chosen such that incorrectly rejecting it is considered more serious than incorrectly accepting it, hence the evidence against H_0 in the data should be persuasive before H_0 is rejected. The p-value of the test is the probability of observing the data when H_0 is true, and H_0 is rejected if the p-value is smaller than some significance level. This cut-off level is chosen rather arbitrary, but typically values such as 0.05 or 0.01 are used.

The p-value in statistical hypothesis testing is used to decide whether the null hypothesis should be rejected or not. The purpose of the p-value proposed in this paper is not to introduce any formal tests of hypotheses, but rather to give a quantitative measure of the strength of the evidence relative to the profiles of known contributors, the number of unknown contributors, and the probabilities of drop-in and drop-out. The p-value can thus be seen as a scaled version of the likelihood ratio.

1.2. Simulations

The LRmix module in the `R` package `Forensim` [7,8] has a performance test that evaluates the strength of the observed LR by simulations. The method is based on Tippet plots [5], but simulations are only carried out under the defense hypothesis. The suspect's profile is replaced with a random profile drawn from the population allele frequencies, and a new likelihood ratio is computed. This is repeated for a large number of random profiles. A performance plot shows the cumulative distribution of likelihood ratios for the random profiles, and gives an idea of how well the suspect can be distinguished from a random person with the given model.

The approach of the performance test is similar to Monte Carlo simulations in statistical hypothesis testing, where repeated sampling is used to estimate the p-value (see e.g. [9]). In theory this type of simulation could have been used to compute a p-value. The approach is however limited by the number of simulations that can be physically carried out. If the defense hypothesis proposes an unknown, unrelated person as the contributor, and we account for drop-in and drop-out in the evidence, the distribution of likelihood ratios under the defense hypothesis includes all possible profiles that can arise from the alleles in the relevant population database. Let A be the total number of alleles found in a database for a given locus. The total number of genotypes the locus may have is:

$$\text{Number of genotypes} = \frac{A(A+1)}{2} \tag{1}$$

Let M be the number of marker loci. To simplify the formula we assume that all loci have the same number of alleles (note that we do not make this assumption in the remainder of the paper). The total number of profiles that may occur is then found as:

$$\text{Number of profiles} = \left[\frac{A(A+1)}{2}\right]^M \tag{2}$$

With e.g. 16 loci with 10 alleles each, the number of profiles is:

$$\left[\frac{10(10+1)}{2}\right]^{16} \approx 7.01 \times 10^{27}$$

In practice maybe 1000 profiles would be simulated, and in most cases would only allow a p-value <0.001 to be reported. The problem is to calculate the probabilities in the upper tail of the LR distribution. If the suspect's profile matches the evidence in all or most loci, relatively few of the 7.01×10^{27} profiles will have a likelihood ratio at least as large as the suspect's profile. Even if we do billions of simulations, it is very unlikely that any of these rare profiles will be sampled.

Note that if we do not account for drop-in and drop-out, or if the defense hypothesis specifies a relative of the suspect as the contributor, the number of profiles under H_d with a non-zero likelihood ratio may be drastically reduced. Subsequently, the problem with a simulation based approach may be reduced by sampling from this limited population.

In the next section we present an algorithm for fast computation of the distribution of likelihood ratios that considers all possible profiles that may occur under the defense hypothesis. Calculating the LR for all possible profiles in a population allele frequency database, even for a modest number of marker loci, is beyond the computing power available today. However, we are only interested in the profiles with a likelihood ratio equally large or larger than the observed $LR(\boldsymbol{S})$, and it is unnecessary to evaluate all profiles. Let $LR(\boldsymbol{R})$ denote the likelihood ratio of a random profile under H_d. The algorithm identifies the profiles that meet the criterion $LR(\boldsymbol{R}) \geq LR(\boldsymbol{S})$, and calculates the p-value as $Pr(LR(\boldsymbol{R}) \geq LR(\boldsymbol{S}))$. Note that computing p-values is only one application of the algorithm. Generally speaking, the algorithm can be used to find all probabilities of the form $Pr(LR(\boldsymbol{R}) \geq T)$, where T is some chosen threshold. The algorithm is implemented in the freely available `R` package `euroMix` (http://euromix.r-forge.r-project.org).