Science and Justice 56 (2016) 380–382

Contents lists available at ScienceDirect

Science and Justice

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

Position paper

Admitting to uncertainty in the LR☆

James M. Curran

University of Auckland, Department of Statistics, Private Bag 92019, Auckland 1142, New Zealand

ARTICLE INFO

Article history:
Received 9 May 2016
Accepted 11 May 2016

Keywords:
Statistics
Bayesian
Likelihood ratio
Bayes factor
Sampling uncertainty

ABSTRACT

In this paper I argue that, given our current state of knowledge, reporting uncertainty in the likelihood ratio is best practice. This may in time be replaced by reporting a Bayes factor, but we are currently unable to do this in all but the simplest of examples.

1. Introduction

In late 1997 I began to consider the problem of addressing sampling uncertainty when assessing random match probabilities, or likelihood ratios for DNA evidence, especially in complex mixtures. I was not the first person to consider this problem. It was a feature of both the first and second NRC reports [1,2] as well as publications from Chakraborty et al. [3], Balding [4], and Weir [5]. These earlier publications were primarily focused on assessment of random match probabilities for single contributor stains. However under the simplest assumptions this is numerically equivalent to the reciprocal of the LR. The issue of concern for these researchers and me was that the allele frequencies used in calculating genotype probabilities typically arose from very small samples of individuals, and therefore might have moderate sampling uncertainty associated with them. In the United States, many jurisdictions forbade the use of convicted offender databases for evidence assessment. As a result, many DNA laboratories relied directly or indirectly through the use of the program PopStats, on a set of allele frequencies produced by the FBI. This set of allele frequencies covered the major population groups of the United States: African American, Caucasian and (South Western) Hispanics, and consisted of approximately 200 individuals in each of these three groups. There were similar rules in the United Kingdom where, for a considerable period, statistical assessment was performed using a database which consisted of 602 Caucasians, 190 Afro-Caribbeans and 257 Asians of Indo/Pakistani descent [6]. DNA experts testifying in the 1990s and 2000s were routinely asked to justify how they could give a random match probability in the order of one in a billion (1×10^{-9}) from such a small sample of people. An astute lawyer would also ask "If I took another sample of size 200, would this figure change?" The single most effective response to this question is "Yes, and my method for assessing this probability has already taken this into account."

I believe that an expert witness who has used a statistically justifiable method for quantifying and adjusting for sampling uncertainty in his or her evaluation will be well-equipped to respond to the sample size question. This is and always has been our motivation for the incorporation of sampling uncertainty into the evaluation of DNA evidence. This recommendation extends beyond DNA evidence type to the evaluation of any evidence where small experiments or small databases introduce sampling variation.

2. What does uncertainty about a probability mean?

Let us consider a very simple DNA case. A stain is recovered from a crime scene and is genotyped. The alleles present in the scene stain are *a* and *b*. Further investigation uncovers a suspect who gives a DNA sample which is also genotyped and found to be of type *ab*. In the absence of all other information and ignoring what we know about the PCR process, we would say the suspect matches the scene. If the hypotheses of interest are:

Hp. the suspect is the only contributor to this stain

Hd. someone unrelated to the suspect is the only contributor to this stain, then the likelihood ratio for this case would be:

$$\mathrm{LR} = \frac{1}{P_{ab}}$$

☆ This paper is part of the Virtual Special Issue entitled: Measuring and Reporting the Precision of Forensic Likelihood Ratios, [http://www.sciencedirect.com/science/journal/13550306/vsi], Guest Edited by G. S. Morrison.

E-mail address: curran@stat.auckland.ac.nz.

http://dx.doi.org/10.1016/j.scijus.2016.05.005
1355-0306/

where P_{ab} is the probability of finding another person in the population of interest who is of type *ab*. If there is an (unrealistic) assumption independence between individuals in the same population, then this probability reduces to the probability that a randomly selected individual from the population of interest is of type *ab*. How do we evaluate this probability? The statistical idealization would be that we would take a sample of *N* individuals at random from our population of interest and record *x*, the number of people who have genotype *ab*. In reality, this is rare, if ever possible, and so we perform the same "experiment" with a convenience sample. This sample might be an offender database, or it might be from a blood donor bank. The justification for using such samples is that the mechanism for being selected in the sample is unlikely to be related to the individual's genotype, and therefore the selection mechanism can be regarded as random. Therefore, our natural (and also the maximum likelihood) estimate of the genotype probability would be:

$$\hat{P}_{ab} = \frac{x}{N}.$$

In making this calculation we have "estimated a probability." Some Bayesians argue vehemently against this idea. For example Jaynes [7], p. 292, makes the following statement:

> A frequency is a factual property of the real world that we measure or estimate. The phrase "estimating a probability" is just as much an incongruity as "assigning a frequency." The fundamental, inescapable distinction between probability and frequency lies in this relativity principle: probabilities change when we change our state of knowledge.

Eminent Bayesian Christian Robert discusses (the over-interpretation of) this quote in his blog [8], and notes how Jaynes later appears to contradict himself. Robert's summary provides some direction on how we might proceed.

> I see no problem with estimating a parameter driving the probability distribution assumed on the data as long as point estimates are not the final answers. Estimation is one incomplete if useful summary of the posterior distribution.

3. Bayes factors

Robert's statement, I believe, justifies the solution that Balding [4] employed and that my coauthors and I extended [9,10]. It also has a consequence. If we regard the likelihood ratio as function of parameters which have a distribution, then this means that the LR will also have a distribution. David Balding (pers. comm.) has said that calculation of the LR is the issue and that realistically the allele probabilities are just nuisance parameters. In statistics, the way to deal with nuisance parameters is to integrate them out. The resulting quantity then is a Bayes factor [11]. Bayes factors are what we ultimately want, and Ian Evett (pers. comm.) has remarked that he "wished he had used the term 'Bayes Factor' instead of 'Likelihood Ratio' from day one." However, consider the DNA example:

Under the assumptions of Hardy–Weinberg equilibrium, the genotype probability may be written as a product of the allele frequencies, i.e. $P_{ab} = 2p_a p_b$. The factor of 2 arises from the fact that humans are a diploid species with two copies of each chromosome. Generally we cannot distinguish between someone who has genotype *ab* and someone who has genotype *ba*. We proposed a Dirichlet prior for all *k* possible alleles at a specific locus. The Dirichlet distribution is a conjugate prior for the probabilities in a multinomial sampling situation. This means that the posterior distribution of the probabilities will also be a Dirichlet distribution. The parameters of the posterior distribution are updated using the alleles counts observed in a database of size *N* individuals, and hence 2 *N* alleles. The posterior distribution can be collapsed to the marginal distribution consisting only of the alleles of interest, so our Bayes factor for this simple case would be:

$$\mathrm{BF} = \frac{\int\int 1 \times g(p_a, p_b | x_a, x_b, N) dp_a p_b}{\int\int 2p_a p_b \times g(p_a, p_b | x_a, x_b, N) dp_a p_b}$$

The numerator simply integrates to one in this case. The denominator in this case is reasonably simple — it is equal to the posterior mean. However, a standard DNA case does not consist of just one locus as we have in our example. A modern DNA laboratory uses somewhere between 17 and 29 loci to type evidential stains and donor samples. That means that our Bayes factor involves the integration, for a single contributor case, over 17 to 58 dimensions. This is aggravated further when we consider mixtures, and added to the additional computational burden of a fully continuous model for DNA interpretation is not currently feasible. I understand that Chris Saunders and Cedric Neumann (pers. comm.) are currently evaluating methods that make it simpler to compute a Bayes factor, and I look forward to the outcomes of their research.

4. What do I do with zeros?

Occasionally a DNA case may arise where one or more of the alleles typed in the crime scene stain has either not been observed in the population of interest or is a rare genetic variant. This situation is problematic for the analyst, as the probability of interest estimated as zero. This contradicts the analyst's belief and knowledge. It contradicts belief because the allele may have been observed in another population, and it contradicts knowledge because usually the allele has been observed in both the crime scene sample and a suspect. If the hypothesis H_p is true, then perhaps a better estimate would be $1/(2N + 1)$ which is equivalent to adding the suspect's copy to the database. If the hypothesis H_d is true, then this means that the suspect and at least one other individual in the population have this allele, so an estimate would be $2/(2N + 2)$. Regardless of whichever estimate is chosen, no one actually believes that the true probability is zero. Therefore any method of assessment should reflect both this and some degree of uncertainty about the true value of the allele probability. Such practice is clearly the purview of Bayesian statistics. This problem is not limited to DNA. Many evidence types are quantified with elemental concentration techniques such as laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS). In many cases, the relative concentration of the elements in the questioned and known samples may not occur in any other sample contained in the forensic scientists database. Does that scientist believe that this substance is so rare that it will never be observed ever again? The answer of course is no. Any technique used to interpret this evidence type needs to be able to make some estimate of relative rarity and to indicate any uncertainty in that estimate.

5. Conclusion

There is no argument that the sources of uncertainty inherent in obtaining estimates of parameters in LR calculations are nuisance parameters. That is, we wish to reflect that we are uncertain about their true value, but ultimately we are focused on calculating the LR. Equally there is no argument that the solution to this problem is the Bayes factor. However, we do not currently possess the techniques to compute the Bayes factor efficiently, nor do we currently know how accurate approximation techniques are. We know that in Bayesian statistics we should evaluate competing models by computing the Bayes factor. However the difficulty in computation means that it is almost never used. In the absence of a simple way to compute the Bayes factor, assessing and reporting uncertainty in the LR is our current best

approach to dealing with imperfect data sources and imperfect knowledge.

Acknowledgments

This work was supported in part by grant 2014-DN-BX-K028 from the US National Institute of Justice. Points of view in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice.

References

[1] National Research Council (U.S.), Committee on DNA Technology in Forensic Science, DNA Technology in Forensic Science, National Academy Press, 1992.
[2] National Research Council (U.S.), Committee on DNA Forensic Science: An Update, The Evaluation of Forensic DNA Evidence, National Academy Press, 1996.
[3] R. Chakraborty, M.R. Srinivasan, S.F. Daiger, Evaluation of standard error and confidence intervals of estimated multilocus genotype probabilities and their implications in DNA, Am. J. Hum. Genet. 52 (1993) 60–70.
[4] D.J. Balding, Estimating products in forensic identification using DNA profiles, J. Am. Stat. Assoc. 90 (1995) 839–844.
[5] B.S. Weir, Genetic Data Analysis II, Sinauer, 1996.
[6] I.W. Evett, P.D. Gill, J.A. Lambert, N. Oldroyd, R. Frazier, S. Watson, S. Panchal, A. Connolly, C. Kimpton, Statistical analysis of data for three British ethnic groups from a new STR multiplex, Int. J. Legal Med. 110 (1997) 5–9.
[7] E.T. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, 2003.
[8] C.P. Robert, Frequency vs. probability, https://xianblog.wordpress.com/2011/05/06/frequency-vs-probability/ (Last Accessed: 2016–05-01).
[9] J.M. Curran, J.S. Buckleton, C.M. Triggs, B.S. Weir, Assessing uncertainty in DNA evidence caused by sampling effects, Sci. Justice 42 (2002) 29–37.
[10] J.M. Curran, An introduction to Bayesian credible intervals for sampling error in DNA profiles, Law Probab. Risk 4 (2005) 115–126.
[11] R.E. Kass, A.E. Raftery, Bayes factors, J. Am. Stat. Assoc. 90 (430) (1995) 773–795.