JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, January 2011, Vol. 56, No. 1
doi: 10.1111/j.1556-4029.2010.01550.x
Available online at: interscience.wiley.com

PAPER

CRIMINALISTICS

Hinda Haned,[1] *M.S.; Laurent Pène,*[2] *M.S.; Jean R. Lobry,*[1] *Ph.D.; Anne B. Dufour,*[1] *Ph.D.; and Dominique Pontier,*[1] *Ph.D.*

Estimating the Number of Contributors to Forensic DNA Mixtures: Does Maximum Likelihood Perform Better Than Maximum Allele Count?

ABSTRACT: Determining the number of contributors to a forensic DNA mixture using maximum allele count is a common practice in many forensic laboratories. In this paper, we compare this method to a maximum likelihood estimator, previously proposed by Egeland et al., that we extend to the cases of multiallelic loci and population subdivision. We compared both methods' efficiency for identifying mixtures of two to five individuals in the case of uncertainty about the population allele frequencies and partial profiles. The proportion of correctly resolved mixtures was >90% for both estimators for two- and three-person mixtures, while likelihood maximization yielded success rates 2- to 15-fold higher for four- and five-person mixtures. Comparable results were obtained in the cases of uncertain allele frequencies and partial profiles. Our results support the use of the maximum likelihood estimator to report the number of contributors when dealing with complex DNA mixtures.

KEYWORDS: forensic science, DNA typing, likelihood estimator, STR loci, DNA mixtures, population subdivision, allele count, partial profiles

Interpretation of forensic DNA mixtures is a challenging task in forensic casework. Mixtures arise when more than one individual contributes to the DNA stain. This is common in cases of sexual assault where the source of DNA evidence can include the victim, the perpetrator(s), and the consensual partner(s) of the victim.

The interpretation of DNA evidence is even more challenging when competing hypotheses are weighted using likelihood ratios because it is implicitly assumed that the number of contributors is known. As misclassified DNA mixtures can lead to dramatic effects on the result of a police investigation, several attempts have been made to assess this problem. Weir (1), Brenner et al. (2), Buckleton et al. (3), and Lauritzen and Mortera (4) have all suggested bounds on likelihood ratios. None of these authors considered the matter of inferring the number of contributors from the data although this is a prevalent line of questioning in court.

It is common laboratory practice to set the lower bound on the number of contributors to the minimum required to explain the observed set of alleles. This bound is based on the maximum allele count throughout the analyzed loci, i.e., the locus showing the maximum number of alleles determines the bound. This method is believed to be an unreliable predictor because of the effect of allele sharing between contributors to the mixture known as the masking effect (5,6). Setting a lower bound is obviously different from attempting to estimate the most supported number of contributors from the data alone. Egeland et al. (7) proposed to overcome this issue by making explicit use of the available allele frequencies of the target population. They suggested a likelihood-based estimator of the number of contributors using diallelic markers when conditions for Hardy–Weinberg equilibrium are met in the population. This method was shown to perform rather well for at least 200 diallelic markers and for mixtures of two and three contributors.

DNA stains from crime scenes are usually characterized through multiallelic short tandem repeat (STR) loci, so there is a need to investigate which approach is the most efficient in determining the number of individuals involved in a mixture. Moreover, several studies have shown that longer DNA fragment lengths carry a greater probability of lost information from allelic drop out (8), leading the forensic expert to conclude that the DNA evidence has partial profiles.

In this paper, we aim to (i) extend the work of Egeland et al. (2003) to an arbitrary number of alleles per locus and to dependencies between alleles because of population subdivision and (ii) investigate through simulations the performance of two methods for estimating the number of contributors to a DNA mixture from the genetic data alone and irrespective of background information that may affect this estimation: the maximum allele count and the maximum likelihood estimator.

We investigate the methods' properties in three distinct situations: in the first situation, all contributors to the mixture belong to the same population with known allele frequencies; in the second situation, we take into account the effect of not knowing with certainty the allele frequencies of the contributors' population, a situation that may arise from population subdivision; in the third

[1]CNRS UMR 5558—Laboratoire de Biométrie et Biologie Evolutive, Université Lyon 1, 43 bd du 11 novembre 1918, 69622 Villeurbanne, Cedex, France.
[2]Institut National de Police Scientifique, Laboratoire de Police Scientifique de Lyon, Lyon, France.

Received 1 June 2009; and in revised form 31 Oct. 2009; accepted 1 Nov. 2009.

situation, we seek to identify the effects of partial profiles on the estimation accuracy for both the maximum allele count and the likelihood-based estimators.

To facilitate reproducibility of our results and extension to other situations, our method is freely available in the package *forensim* for the R statistical software (9).

Methods

Extending the Likelihood Estimator to the Cases of Multiallelic Loci and Population Subdivision

Let A be a specific locus with alleles A_1,..., A_k with frequencies p_1,..., p_k in a given population. Let m be the set of observed alleles in a DNA stain from a crime scene. We are interested in estimating the probability of observing m knowing that there are x individuals contributing to the mixture. This is the likelihood of the data m conditional on x, denoted: $L_A(x)$.

Example—Suppose that a crime scene stain shows alleles A_1 and A_2 at locus A, the forensic expert wants to determine the likelihood that two contributors supply these alleles. Combining the observed alleles into two individual genotypes yields seven distinct pairs of possible genotypes for the two contributors: (A_1A_1, A_2A_2), (A_2A_2, A_1A_1), (A_1A_1, A_1A_2), (A_2A_2, A_1A_2), (A_1A_2, A_1A_1), (A_1A_2, A_1A_2), and (A_1A_2, A_2A_2).

Under the hypothesis that the contributors to the DNA stain are not related, the estimation of each genotype proportion can be obtained as a product of the allele frequencies using the Hardy–Weinberg formula. This assumes the independence of alleles between and within individuals. This simplifying hypothesis as a means to determine the genotype proportions from allele frequencies is termed the "product rule" (10).

The probability of observing the pair of genotypes (A_1A_1, A_1A_2), denoted Pr(A_1A_1, A_1A_2), corresponds to the probability of observing one homozygote for A_1 and one heterozygote A_1A_2, which is $p_1^2 \times 2p_1p_2$. By adding the probabilities for each possible genotype pair, we finally obtain:

$$L_A(x=2) = 4p_1^3p_2 + 6p_1^2p_2^2 + 4p_1p_2^3$$

These results could be derived analytically in a simple case (one locus and two hypothetic contributors), but the complexity of the likelihood computation increases dramatically with the numbers of loci and contributors; hence, there is a need for a general formulation of the likelihood function. To achieve this generalization, we follow the work of Curran et al. (11) who gave a general framework for interpreting DNA mixtures that can take population subdivision into account. In their paper, a general formula for mixture interpretation evaluation was given in the form: Pr($E|H$), where E is the DNA evidence and H is the hypothesis under which the data is being considered, for example, the prosecution hypothesis.

When only genetic data is considered, the evidence E is composed of the set of alleles observed in the mixture, denoted C. This set of alleles is composed of the following: (i) the set of alleles found in the typed individuals who are known to have contributed to the mixture, denoted T; (ii) the set of alleles found in the typed individuals known to be noncontributors to the mixture, denoted V; and (iii) the set of alleles carried by the unknown contributors, denoted U. For instance, in the case of a DNA stain from a rape case, T is the set of alleles carried by the victim, her consensual partner(s), and potentially the suspect(s); V is the set of alleles carried by cleared suspects; and U is the set of alleles carried by the unknown contributors to the mixture.

The general formula of the likelihood can thus be derived from the particular case where all contributors to the mixture are unknown and there are no typed individuals. This corresponds to $T = V = \varnothing$ and $C = U$. Note that the equality $C = U$ does not correspond to the degenerate case evoked in (11) where unknown contributors can have any genotypes in C. In our case, the x unknown contributors' genotypes must explain all alleles in C; thus, all possible genotypes attributable to the unknown individuals must explain the alleles present in the mixture, and they must all be taken into account in the likelihood calculation.

General Formulation of the Likelihood Function—Before giving the general formulation of the likelihood function, we first specify the notations used in this paper, following Curran et al. (11): x: The unknown number of contributors to the DNA mixture; c: The distinct number of alleles observed in the DNA stain; r: The number of unconstrained alleles, $r = 2x - c$; r_i: The unknown number of copies of allele A_i among the r unconstrained alleles of the stain; u_i: The unknown number of copies of allele A_i in the stain, with $\sum_{i=1}^{c} u_i = 2x$ and $u_i = r_i + 1$; θ: Wright's F_{ST} coefficient, which gives the probability of identity by descent of two alleles taken at random from a subpopulation in two distinct individuals.

In our case, all contributors are unknown. Consequently, the DNA evidence, E, is only composed of the alleles present in the stain, C, and all other quantities defined in (11) and related to the typed individuals, whether they are known to have contributed to the mixture, are set to zero. The likelihood of having x individuals giving the alleles observed at a locus A in the case of all individuals belonging to the same subpopulation is given by the general formula:

$$L_A(x) = \sum_{r_1=0}^{r} \sum_{r_2=0}^{r-r_1} \cdots \sum_{r_{c-1}=0}^{r-r_1-r_2-,\ldots-r_{c-2}} \frac{(2x)!}{\prod_{i=1}^{c} u_i!} \frac{\prod_{i=1}^{c} \prod_{j=0}^{u_i-1} [(1-\theta)p_i + j\theta]}{\prod_{j=0}^{2x-1} [(1-\theta) + j\theta]} \quad (1)$$

Equation (1) takes into account the variation in the subpopulation allele frequencies. When there is no need to consider population subdivision, the likelihood of the data is simply obtained by setting θ to zero.

The Likelihood Estimator—The maximum likelihood estimation of x, when a single marker A is considered, satisfies:

$$\max_{j=1,2,3,\ldots} L_A(x=j) \quad (2)$$

When multiple loci are considered simultaneously, the likelihood is calculated as the product of the likelihoods of each locus:

$$\max_{j=1,2,3,\ldots} \prod_{A} L_A(x=j) \quad (3)$$

The result in Eq. (3) is straightforward for the case of a homogeneous population, that is when $\theta = 0$ in Eq. (1). When there are allele dependencies in the general population because of subdivision, the overall loci likelihood (in the subpopulation) is still, to a close approximation, the product of the single locus probabilities, because the dependencies between alleles at different loci are corrected through θ (12).

In fact, the likelihood estimator defined by Eqs. (2) and (3) extends the likelihood-based estimator derived by Egeland et al. (7) to the case of multiallelic loci and allows population subdivision to

be taken into account through θ. Thus, the value for θ must be chosen according to the level of subdivision of the population. Typically, θ is chosen in the interval [0,0.03] when dealing with human populations (13).

Most forensic DNA mixtures consist of two-person mixtures (14); thus, for the estimator to be biologically meaningful, estimates were searched in the discrete interval [1,6]. This is a sensible upper limit for the number of contributors that can be analyzed in practice.

Evaluation of the Methods' Performance

Known Allele Frequencies Case—We used a published data set of allele frequencies in three U.S. populations (15): African-Americans, Caucasians, and Hispanics. These populations were characterized by 15 STR loci, of which 13 correspond to the core CODIS loci.

Genotypes were simulated by drawing alleles independently at their relative frequencies from each population data base. Mixtures were then simulated by randomly drawing genotypes at each locus. The performances of the likelihood-based estimator and maximum allele count were compared on 1000 simulated mixtures comprising two to five contributors.

Uncertain Allele Frequencies Case—Generally, in the case of population subdivision, allele frequencies of the subpopulations are not known with certainty. This is because of the difficulty of defining the subpopulation of an individual (16). In this paper, we analyze the effect of uncertainty on allele frequencies by modeling the differences in allele frequencies between the global population and a subpopulation through a Dirichlet model. The term "subpopulation" means that the allele frequencies in the target population are not known with certainty and does not imply allele dependencies between and within loci.

The allele frequencies for a given locus in a given subpopulation are generated as random deviates from a Dirichlet distribution (17,18). Each allele frequency is a random variable with a parameter $\alpha_i = p_i(1 - \theta)/\theta$, where θ is the F_{ST} coefficient. Denoting p'_i the frequency of allele A_i in the subpopulation, the allele frequencies are modeled as:

$$p'_1, \ldots, p'_k \rightarrow \text{Dirichlet}(\alpha_1, \ldots, \alpha_k)$$

The global allele frequencies were taken from the African-American population (15).

We chose to set $\theta = 0.03$ in the variance parameter α_i. This value corresponds to the correction factor suggested by the National Research Council (19) for dealing with highly subdivided human populations. Because we were only interested in studying the effect of uncertainty on the subpopulation allele frequencies, all loci were simulated independently within the subpopulation.

We compared the results of the maximum allele count to the likelihood-based estimator on 1000 simulated mixtures of two to five contributors. We investigated the differences between results when the uncorrected form of the likelihood-based estimator is used ($\theta = 0$) and compared them to the results obtained using the corrected form by setting $\theta = 0.03$.

Evaluation of the Methods' Robustness to Partial Profiles—We analyzed the effect of successively removing loci while estimating the number of contributors on 1000 simulated mixtures of two to five individuals. The markers were successively removed according to their alleles' expected median length (20). This corresponds to what happens in the case of a degraded DNA sample: Longer DNA fragments drop out first (8).

All programs used for the simulations were implemented in the *forensim* package for the R statistical software, available at http://forensim.r-forge.r-project.org/.

Results

Known Allele Frequencies Case

The accuracy of estimations decreased with the number of contributors for both the maximum allele count and the maximum likelihood estimators (Table 1). The probability of a correct estimation was always >90% for mixtures of two or three individuals. Maximum allele count produced better estimates for three-person mixtures, but the efficiency of this method decreased dramatically for complex mixtures of four or five individuals, while maximum likelihood gave a correct classification rate ranging from 64% to 79% in the three populations.

Uncertain Allele Frequencies Case

The effect of uncertainty on allele frequencies was investigated for the case where the real allele frequencies deviate greatly from those used in the estimator ($F_{ST} = 0.03$, Table 2). Accurate estimates were obtained with the maximum allele count for mixtures with two or three contributors (success rate >90%). The percentage

TABLE 1—*Percentages of correctly identified mixtures for all three studied populations. The first column gives the true number of contributors,* x. *The second and third columns give the percentages of mixtures correctly identified by the two methods: the maximum allele count and the maximum likelihood estimator.*

x	Maximum Allele Count (%)	Likelihood Estimator (%)
African-Americans		
2	100	100
3	99	94
4	45	79
5	5	67
Caucasians		
2	100	99
3	97	92
4	34	77
5	2	64
Hispanics		
2	100	100
3	98	93
4	45	79
5	2	67

TABLE 2—*Percentages of correctly identified mixtures in the uncertain allele frequencies case. The first column gives the true number of contributors,* x. *The next two columns give the percentages of accurate estimation for the maximum allele count and the maximum likelihood methods. For the latter, two estimates are displayed corresponding to the form used in the estimator: the uncorrected form ($\theta = 0$) and the corrected form ($\theta = 0.03$).*

x	Maximum Allele Count (%)	Likelihood Estimator (%)	
		Uncorrected Form	Corrected Form
2	100	99	99
3	94	95	91
4	21	56	76
5	0.7	27	60

of correctly identified stains was lower when dealing with four or five contributors. For instance, only 21% of five-person mixtures were correctly identified.

The corrected ($\theta = 0.03$) and uncorrected forms ($\theta = 0$) of the likelihood-based estimator produced similar results for mixtures of two or three individuals. The corrected form was more efficient in cases of a greater number of contributors: 60% of five-person mixtures were correctly identified, which was more than twofold the maximum allele count success rate.

Method Robustness to Partial Profiles

The effects of partial profiles on the estimators' accuracy are shown in Fig. 1. Only mixtures simulated from African-American allele frequencies are shown here in the known allele frequencies case. Similar results were obtained for the other two populations (Caucasians and Hispanics) as well as in the uncertain allele frequencies case for all three populations (results not shown). Consistent with previous results (Tables 1 and 2), the accuracy of both methods decreased with the number of contributors. The relative performance of both methods changed with the number of contributors in the mixture. The maximum allele count was revealed to be more efficient for mixtures of two or three persons, while the likelihood-based estimator performed better for mixtures of more than three individuals (see Fig. 1). A 90% success rate was reached using the maximum allele count for a two-person mixture when exploring only two loci, while five were needed for the maximum likelihood estimator. For three-person mixtures, the loci number increased to 10 and 14, respectively. For complex mixtures of four or five contributors, the success rates fell to 63% for the likelihood-based estimator and to 0.042% for the maximum allele count using all 15 loci.

Finally, to further our understanding of the aforementioned results, we looked at the characteristics of the profiles responsible for the biased estimations with the maximum likelihood estimator (Tables 1 and 2, Fig. 1). We analyzed the sensitivity of the estimator to allele frequencies. An illustration of our results is shown in Fig. 2 for a three-person mixture characterized by one locus. The

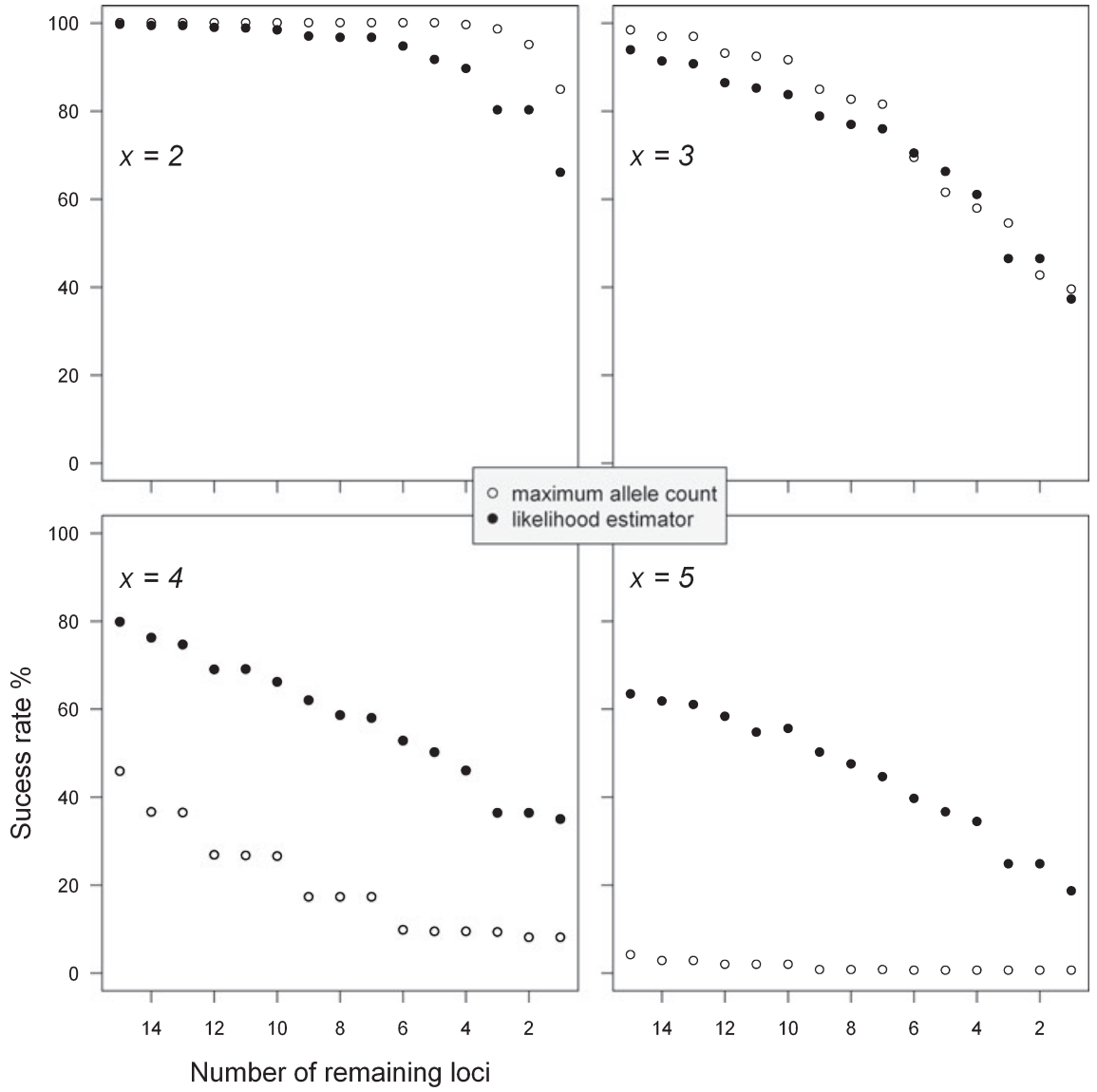

FIG. 1—*Percentages of correctly identified mixtures for x contributors, where x ranges from 2 to 5 in the case of partial profiles, for the maximum allele count and the maximum likelihood methods.*

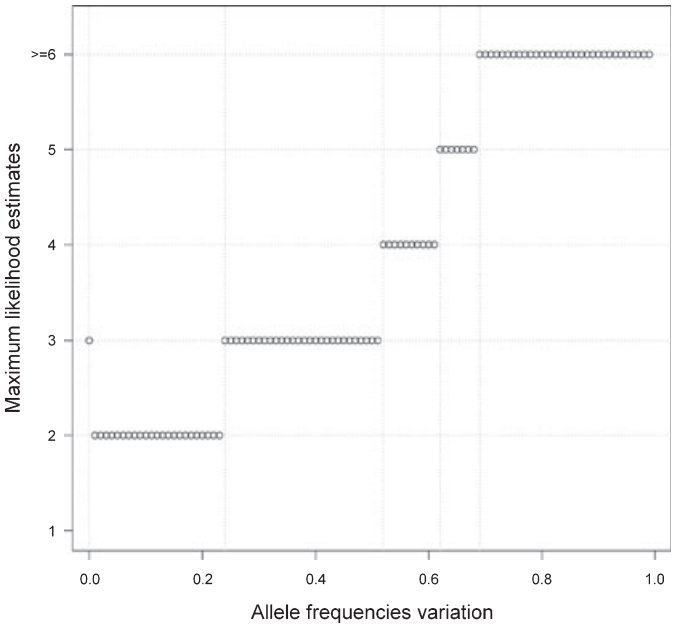

FIG. 2—*Sensitivity of the maximum likelihood estimations of the number of contributors to variations in allele frequencies for a simulated three-person mixture. A single locus, "vWA," was considered. At this locus, the mixture included alleles "16," "17," "18," and "19," with initial allele frequencies taken as 0.25, 0.24, 0.15, and 0.06 from the African American population. We varied the frequency of the less frequent allele "19" from 0 to 1 (x-axis), values of the three other alleles being also varied by keeping their relative frequencies constant. Each point on the plot represents the estimation yielded by the maximum likelihood estimator (yaxis). Correct estimates are obtained with the original allele frequencies (origin of the x-axis), and when the frequency of allele "19" varies between 0.24 and 0.52. Underestimation of the number of contributors occurs when frequency of allele "19" is under 0.24, while overestimations occur when its frequency is greater than 0.52.*

maximum allele count can only give a lower bound to the real number of people involved in the mixture; thus, it cannot give overestimates. In contrast, maximizing the likelihood can lead to either underestimation or overestimation. Underestimation occurred when there are rare alleles in the mixture, while mixtures with frequent alleles also tended to be misclassified.

Discussion

We compared the efficiency of the commonly used maximum allele count and an estimator based on likelihood maximization in inferring the number of contributors to forensic DNA mixtures.

Globally, maximizing the likelihood did not perform better than the maximum allele count for mixtures of two or three individuals. When all loci were documented and all mixture contributors belonged to the same population with known allele frequencies, the maximum allele count gave lower misclassification rates (varying from 1% to 3%) than the likelihood-based estimator (varying from 6% to 8%). These results corroborate previous findings for the former estimator (5).

Maximum allele count gives correct estimates for mixtures comprising x individuals when there are at least $2x - 1$ alleles at one of the considered loci in the stain. While this condition is often met in two- or three-person mixtures, it is unlikely to find as many distinct alleles in mixtures of high order because of allele sharing (6). For instance, five-person mixtures are unlikely to show nine distinct alleles at any of the considered loci, even if very polymorphic markers are used. Consequently, the maximum allele count method, which tends to underestimate the real number of contributors in mixtures of high order ($x > 3$), still gives satisfactory results for two- and three-person mixtures. Maximum likelihood estimator can either over- or underestimate the real number of contributors for all mixture types.

As expected, the uncertainty of estimations increased with the number of contributors for both methods, while four- and five-person mixtures were more accurately identified by maximizing the likelihood. This is owing to allele sharing between contributors. As maximum allele count relies only on the number of distinct alleles, mixtures with greater numbers of contributors have greater amounts of allele sharing, which leads to the underestimation of the number of contributors.

Previous studies showed that using maximum allele count in the case of substantial allele sharing leads to biased estimates (5). The bias is likely to increase in cases of population subdivision. Here, we were more interested in one of the consequences of subdivision on the likelihood-based estimator, namely, the uncertainty on allele frequencies of the subpopulation, because the estimator explicitly makes use of the allele frequencies. In the case of uncertain allele frequencies, we observed that the corrected form of our estimator performed better than the uncorrected one only for mixtures consisting of four or five contributors. Mixtures involving two or three individuals were more accurately classified with the uncorrected form of the estimator. The correction for subdivision was thus efficient in the uncertain allele frequencies case only for complex mixtures, but this might not be the case in highly subdivided populations, where the independence of individual genotypes might not be realized.

In the case of partial profiles, both of the estimators showed a similar decrease in precision for two- and three-person mixtures, while the likelihood-based estimator was clearly more robust to partial profiles when dealing with four- and five-person mixtures. The lack of robustness of maximum allele count is explained by the fact that decreasing the number of loci decreases the chance of encountering in the mixture a locus that shows enough distinct alleles to allow a correct estimation using only the maximum allele count. This effect is likely to be increased when dealing with complex mixtures of more than three contributors.

Overall, it is difficult to specify the minimum number of loci needed to accurately resolve a mixture because this number depends on the tolerated error rate that relies on the forensic expert's experience; however, even with all 15 STR loci, five-person mixtures could not be resolved satisfactorily: The maximum allele count yielded an error rate of more than 95%, while maximizing the likelihood misclassified more than 30% of the mixtures.

The bias in estimations is due in part to profiles with multiple masked alleles. This problem could be circumvented using quantitative data given by the mixture profiles' peak heights or areas (21). In fact, our estimator only takes into account qualitative information consisting of the allele types present in the stain. We assumed that the forensic expert had already determined the alleles present in the mixture and that there was no ambiguity during this stage of the evidence analysis. Further work could thus include the use of quantitative information to help in revealing masked alleles.

Most forensic laboratories use the maximum allele count method to specify the number of contributors to mixed stains. Complex mixtures comprising multiple masked alleles are likely to be misclassified by this method. This issue could have dramatic consequences especially when the number of contributors is determined solely on genetic data. This might be the case when dealing with DNA casework. Very often no suspect is available in such

stains. Consequently, having an estimate of the number of contributors could help investigators when new elements emerge in the case. Therefore, it appeared to us that in case the number of contributors is determined on genetic data, maximizing the likelihood should be preferred to maximum allele count especially when dealing with stains suspected to be mixtures of three or more individuals.

To conclude, we would like to point out that we do not recommend one method over the other. Our work is intended to provide insight into forensic practitioners on the differences in efficiency between the two estimators with respect to situations frequently encountered in forensic casework, namely, uncertainty about the population allele frequencies and partial profiles. Our methodology is freely available in the package *forensim* for the R statistical software to allow investigations in contexts not explored here.

Acknowledgments

We thank two referees for a thorough review and constructive comments. We are grateful to Frank Sauvage for his helpful comments.

References

1. Weir BS. DNA statistics in the Simpson matter. Nat Genet 1995;11(4): 365–8.
2. Brenner CH, Fimmers R, Baur MP. Likelihood ratios for mixed stains when the number of donors cannot be agreed. Int J Legal Med 1996;109(4):218–9.
3. Buckleton JS, Evett IW, Weir BS. Setting bounds for the likelihood ratio when multiple hypotheses are postulated. Sci Justice 1998;38(1):23–6.
4. Lauritzen SL, Mortera J. Bounding the number of contributors to mixed DNA stains. Forensic Sci Int 2002;130(2–3):125–6.
5. Paoletti DR, Doom TE, Krane CM, Raymer ML, Krane DE. Empirical analysis of the STR profiles resulting from conceptual mixtures. J Forensic Sci 2005;50(6):1361–6.
6. Buckleton JS, Curran JM, Gill P. Towards understanding the effect of uncertainty in the number of contributors to DNA stains. Forensic Sci Int Genet 2007;1(1):20–8.
7. Egeland T, Dalen I, Mostad PF. Estimating the number of contributors to a DNA profile. Int J Legal Med 2003;117(5):271–5.
8. Butler JM. Forensic DNA typing. Burlington, MA: Elsevier Academic Press, 2001.
9. R Development Core Team 2009. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna: Austria. ISBN 3-900051-07-0, http://www.R-project.org.
10. Evett IW, Weir BS. Interpreting DNA evidence. Sunderland, MA: Sinauer, 1998.
11. Curran JM, Triggs CM, Buckleton JS, Weir BS. Interpreting DNA mixtures in structured populations. J Forensic Sci 1999;44(5):987–95.
12. Buckleton JS, Triggs CM, Walsh SJ. Forensic DNA evidence interpretation. Boca Raton, FL: CRC Press, 2005.
13. Curran J, Buckleton JS, Triggs CM. What is the magnitude of the subpopulation effect? Forensic Sci Int 2003;135(1):1–8.
14. Torres Y, Flores I, Prieto V, López-Soto M, Farfán MJ, Carracedo A, et al. DNA mixtures in forensic casework: a 4-year retrospective study. Forensic Sci Int 2003;134(2–3):180–6.
15. Butler JM, Schoske R, Vallone PM, Redman JW, Kline MC. Allele frequencies for 15 Autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. J Forensic Sci 2003;48(4):908–11.
16. Balding DJ, Nichols RA. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 1994;64(2–3):125–40.
17. Nicholson G, Smith AV, Jónsson F, Gústafsson O, Stefánsson K, Donnelly P. Assessing population differentiation and isolation from single-nucleotide polymorphism data. J R Stat Soc B 2002;64:695–716.
18. Marchini JL, Cardon LR. Discussion of Nicholson et al. J R Stat Soc B 2002;64:740–1.
19. NRC II—National Research Council Committee on DNA Forensic Science. The evaluation of forensic DNA evidence. Washington, DC: National Academy Press, 1996.
20. Applied Biosystems. AmpFlSTR® Identifiler™ PCR Amplification Kit user's manual. Foster City, CA: Applied Biosystems, 2001;P/N 4323291.
21. Evett IW, Gill PD, Lambert JA. Taking account of peak areas when interpreting mixed DNA profiles. J Forensic Sci 1998;43(1):62–9.

Additional information and reprint requests:
Hinda Haned, M.S.
CNRS UMR 5558—Laboratoire de Biométrie et Biologie Evolutive
Université Lyon 1
43 bd du 11 novembre 1918
69622 Villeurbanne, Cedex
France
E-mail: Hinda.Haned@univ-lyon1.fr