

## CHAPTER 4

# Catch them ... if you can

Ludolf Meester, Jaap Molenaar, Misja Nuyens, Yves Rozenholc, Koos van Winden

**ABSTRACT.** An important part of forensic science is dedicated to the evaluation of physical traces left at the crime scene like fingerprints, bullets, toolmarks etc. These traces are compared with traces from a suspect. The evaluation of physical traces can be interpreted as the comparison of two noisy signals. We introduce an evaluation of the matching of two noisy signals at diverse scales and localisations in space. In a multi-resolution way a "probability" of matching is computed. Furthermore, a description is given to evaluate the complexity of a shoemark. A likelihood ratio approach is used for comparing two shoemark traces.

**KEYWORDS:** adaptive testing, complexity measure, likelihood ratio

### 1. Introduction

The Netherlands Forensic Institute (NFI) is part of the Dutch Ministry of Justice. This large centre of expertise (about 300 staff members) has forensic casework as core activity. To maintain a high scientific standard, a considerable amount of the budget (about 5 million euro in total) is spent on Research and Development in order to develop new methods and procedures. The NFI also participates in training of crime scene investigators and public prosecutors and the institute maintains various collections and databases. The projects carried out by NFI cover a wide variety of topics such as: DNA-recognition, finger print recognition, drugs analysis, traffic accident analysis, hand writing analysis, paint recognition, etc.

**1.1. Forensic science.** One of the major areas of interest in forensic science is to develop methods to evaluate physical evidence found at scenes of crime. The type of physical evidence found can vary significantly, i.e. traces of blood, hair, fingerprints, bullets, shoemarks, toolmarks etc. But also the state of development of the methods of evaluation differ across different types of evidence. For DNA evidence for instance, scientists have developed models for evaluation which are accepted by the majority of scientists and judicial systems. For other types of evidence the evaluation model is not yet as developed. Much of the evidence is still evaluated by experts in a system that clearly should be classified as "statement of opinion based on experience". These methods are well accepted in most judicial systems and are still common place in most areas of expertise, see for instance [14] and [2].

The problem with most systems of evaluation that are based on "statement of opinion based on experience" is that the experts are more or less operating like a "black box" and it is very hard for other individuals involved in the judicial process to find out whether in a specific case the "black box" is working properly. To solve this problem forensic science has in recent years started to develop models to assist the experts with the evaluation of the evidence and at the same time give the judicial system the option to verify, to a certain extent, their conclusions.

The NFI knows the special function *statisticus* and nearly all activities involve a lot of statistics. The problems to be considered here are from the division for 'toolmarks, footwear and tyre impressions'. They concern recognition of toolmarks and recognition of footwear. We briefly describe them separately.

**1.2. Toolmark recognition.** The perpetration of crimes often involves the use of tools like knives and screwdrivers. Such a tool may leave a mark on the crime scene, e.g. on a window or door frame. On each tool a characteristic pattern is present, which can be observed on a microscale. Such a pattern initially stems from the production process, but it usually gets modified through use of the tool. The pattern makes the tool unique. It is to be expected that a mark stemming from a certain tool contains the same pattern, establishing a clear relationship between tool and mark. In the forensic practice the procedure concerning toolmarks is as follows. If at the crime scene a mark is detected, its pattern is carefully conserved, e.g. via photographs and/or special substances. Corresponding tools possessed by suspected persons are seized. With such a seized tool marks are made under various circumstances. For example, in case of a screwdriver marks are made at various angles of inclination between screwdriver and surface. The patterns of the resulting marks are compared to the mark found at the crime scene. From the result of this comparison a conclusion whether the crime mark corresponds to the tool or not is drawn. For this, probabilistic models are used and the probability that the mark is caused by the tool is expressed according to some classification with levels like "highly probable, probable, less probable, probably not, certainly not".

**1.3. Shoe print comparison.** For shoe print analysis the procedure is similar to the toolmark procedure. Now the tool is a shoe sole or a bare foot and the mark is a print of it in sand or soil. The difference with toolmark comparison is that in shoe print comparison (usually) more complex shapes have to be analysed.

**1.4. Concepts and organization of the report.** The conceptual questions are the same for both kinds of marks. We mention the main items. The accuracy with which the marks are recorded is not infinitely high. Measurements have of course always some inherent inaccuracy, but it also often happens that the mark at the crime scene has been degraded in the time interval between the crime and the spot investigation by the forensic experts. The patterns/marks produced in the laboratory by the tool of the suspect may differ from the mark at the crime scene, even if it is absolutely certain that the tool is indeed used during the crime, simply because the circumstances always slightly differ. So, the statement that a mark stems from a tool will always have a probabilistic character and be based on statistical considerations. The present project is meant to develop reliable models

to be used for these purposes.

In section 2 we focus on toolmark recognition. There the question to be addressed is the following. Given two one-dimensional patterns, i.e. functions on an interval, one representing the mark found at the crime scene and the other produced in the laboratory by a tool, what is the probability that these two marks stem from the same tool? Similarity between two functions on a finite interval may be by chance. How to incorporate the possibility of similarity by fortune into the eventual conclusion? In section 2 a highly promising method is proposed which turned out to be very effective in several bench mark tests.

In sections 3 and 4 we deal with the shoe print problem. Here, a new type of question arises. If in two marks small two-dimensional subpatterns are found that agree to a high degree, what weight should this get in the whole procedure? Similarity in highly complex subpatterns must clearly be rated higher than when two simple shapes agree. This leads to the problem how to define the ‘complexity’ of geometric patterns in the plane. This issue is dealt with in section 3, while section 4 is devoted to the development of a statistical model to draw conclusions about similarity of complete shoe print patterns.

## 2. Toolmark recognition: testing nullity in a regression framework

In this section <sup>1</sup> we develop a model for the efficient comparison of striation marks as found when comparing for instance toolmarks or bullets. We here focus on toolmarks. The classical problem description is as follows: one mark - made by a tool - is found on the crime scene, the other is made in laboratory after having found a potential associated object by a suspect. The question then is ”To what extent do these marks match?” and further ”what is the evidential value of such a match?”.

One way to evaluate this evidential value is to translate the difference between the probability of declaring a match if the two marks are made by the same tool and the probability of declaring the match when the marks are made by different tools into the *likelihood ratio*

$$LR = \frac{P(\text{match} \mid \text{marks made by the same tool})}{P(\text{match} \mid \text{marks are made by different tools})}$$

The problem that rises immediately when one tries to operationalize this likelihood ratio approach is that in practice it is very hard to estimate the denominator. In section 4 we will calculate this likelihood ratio under simplifying assumptions.

In forensic science the physical marks are considered as ”fingerprints” of the associated tool or gun. It is assumed that if the same tool makes several marks in a row, the marks will be very similar and distinguishable from marks made with another tool or gun. The assumption is that this holds for ”good” or ”perfect and complete” marks. However, in practice the marks are unfortunately not perfect. Sometimes marks found display only a partial mark, the crime scene mark and comparison mark are made under different circumstances, tools are used between the time of the crime and time of making the comparison mark etc. Therefore,

---

<sup>1</sup>This section is essentially an abbreviated version of [4]

noise is added to the "fingerprint". Another source of noise is due to the fact that the marks have to be recorded, photographed and normalized to make a comparison possible. In the method described below we assume that this preprocessing has already taken place. A good example of a possible normalization can be found in [3].

In this section these "fingerprints" are modeled as a noisy one-dimensional or two-dimensional signal depending on whether it is a striation comparison (reduced to gray levels along the perpendicular direction of the axis of striation) or a shape print comparison (reduced to gray levels on a surface). Hence, we have at our disposal two noisy signals

$$\check{Y} = \check{s} + \check{\varepsilon} \text{ and } \tilde{Y} = \tilde{s} + \tilde{\varepsilon},$$

where  $\check{s}$  and  $\tilde{s}$ , are two vectors in  $\mathbb{R}^n$  (for one-dimensional signals) or two  $n_1$  by  $n_2$  matrices (for two-dimensional signals) and where  $\check{\varepsilon}$  and  $\tilde{\varepsilon}$  are vectors (resp. matrices) of i.i.d. random errors of corresponding sizes.

To answer the question "Do these marks match?" and to evaluate the strength of the evidence, we propose to use the difference of the noisy signals  $Y = \tilde{Y} - \check{Y}$  and to test whether  $s = \tilde{s} - \check{s}$  equals 0 or not. This test is done not only globally but also locally: two "fingerprints" may not fit globally because additional use of the tool between the making of the two marks might have lead to a change of part of the "fingerprint". Instead of estimating the likelihood ratio of a match based on the probability of matching when the marks are made by the same tool and the probability of a "random match" when the marks are made by different tools, we use tests of the nullity of  $s$  in a specific piecewise way at different resolution levels to compute a *measure of matching*. To each piece (i.e. each test of nullity) is associated an elementary measure of matching. We compute the global measure of matching as a weighted sum of these elementary measures. The knowledge of the expert then should be used to convert this measure of matching into a real probability of matching.

**2.1. The statistical problem.** Let us consider the regression model

$$Y = s + \varepsilon,$$

where  $Y$  is a vector in  $\mathbb{R}^n$  of  $n$  observations,  $s$  is an unknown vector in  $\mathbb{R}^n$  and the  $\varepsilon$  is an unknown random centered vector with i.i.d. components. We consider a test of  $s = 0$  against  $s \neq 0$ . More precisely, we test the hypothesis

$\mathbf{H}_0$  : "Y is a random centered vector with i.i.d. components"  
against

$\mathbf{H}_s$  : "there exists a non-zero vector  $s$  in  $\mathbb{R}^n$  and a random centered vector with i.i.d. components  $\varepsilon$  such that  $Y = s + \varepsilon$ "

Here, various tests of nullity can be considered, associated to specific assumptions under the non zero hypothesis. Most are based on Gaussian assumptions on the vector of errors  $\varepsilon$  and/or an asymptotic construction for  $n$  going to infinity and/or the variance of the signal going to zero. Several tests of nullity in regression or related frameworks have been studied by many authors. We cite here the work of [1] in a classical non asymptotic Gaussian framework.

In other classical frameworks related to this problem, such as the Gaussian sequence model (Gaussian errors and a variance assumed to decrease like  $1/n$ ) or functional regression ( $s = (S(x_1), \dots, S(x_n))$  for  $S$  a real function on  $[0, 1]$  or  $\mathbb{R}$ ), the authors consider on one hand stronger assumptions on the noise structure such as Gaussian noise and/or suppose a known variance, on the other hand regularity assumptions on  $S$  (supposed to be in a Sobolev or a Besov ball for example) and prove (except for the previously cited work of Baraud & *al.*) asymptotic results with respect to the size  $n$ . We refer to the recent book [9] of Ingster and Suslina for a complete study on these asymptotic approaches in Gaussian model for nonparametric testing and point out the papers of [6], [7], [8], [11], [10], [15], [5].

Let us consider a sequence of subintervals of the integers in  $\{1, \dots, n\}$  denoted  $(I_m)_m$  where  $m$  represents an index belonging to a set  $\mathcal{M}_n$  which can vary with  $n$ . Later, we will state precisely the possible choices of this sequence of intervals. We denote by  $\ell_m$  the number of integers in the interval  $I_m$ . For each interval  $I_m$ , we consider the vector  $Y_m$  of  $\mathbb{R}^{\ell_m}$  which are those coordinates of  $Y$  which are indexed by  $I_m$ .

Let us consider a real number  $\alpha$  in  $[0, 1]$  and the sequence of levels

$$\alpha_m = \alpha \frac{[\ell_m]^c}{\#\mathcal{M}_n}, \text{ for } m \in \mathcal{M}_n$$

where  $c$  is a fixed parameter - used to control the significance of the tests with respect to the length of the intervals. For each  $m$  in  $\mathcal{M}_n$ , we test at the level  $\alpha_m$  whether the sub-vector  $Y_m$  is 0 or not. Let us note  $\Phi_m$  the result of this test:  $\Phi_m = 0$  if  $\mathbf{H}_0$  is not rejected and 1 otherwise.

One can remark that the global test of nullity  $\Phi$ , defined by  $\Phi = 0$  if all  $\Phi_m$  are null and 1 otherwise, is of level less or equal to  $\alpha \sum_m \ell_m^c / \#\mathcal{M}_n$ . Indeed

$$\mathbb{P}_{\mathbf{H}_0}(\Phi = 1) = \mathbb{P}_{\mathbf{H}_0}(\exists m \in \mathcal{M}_n, \Phi_m = 1) \leq \sum_{m \in \mathcal{M}_n} \mathbb{P}_{\mathbf{H}_0}(\Phi_m = 1) = \frac{\alpha}{\#\mathcal{M}_n} \sum_{m \in \mathcal{M}_n} \ell_m^c.$$

Under Gaussian assumptions and following [1], it is possible to derive the power of the test  $\Phi$  for our family of sub-tests  $(\Phi_m)_m$ .

**2.2. A model for evidential testing.** In order to construct a model for evidential testing, one has to keep in mind the two characteristics of this problem. The time interval between acquiring the two noisy signals to be compared can be large. Then, it is natural to imagine that the signals will always be globally different and a global test of nullity will always reject the hypothesis  $\mathbf{H}_0$ , hence such a test is not relevant in practice. For example, the shoe of a criminal may have been used after a crime and the shape of its surface may have changed. It is important to understand that the experts usually look only for small features as evidence as they are usually considered as strongly individualizing. In one-dimensional comparison

for tool-marks, there are experts that consider five matching lines in a row on two toolmarks as sufficient evidence. In our framework, this means that  $s$  has five coordinates in a row equal to 0.

The model introduced here is based on a dyadic multi-resolution evaluation of the *measure of matching*, or *measure of similarity*, using a dyadic sequence of intervals. The measure of matching  $\mu$  has to be all the greater so that the signals match (i.e.  $s = 0$ ) on larger interval. It has to be additive with respect to disjoint intervals, i.e.

$$\mu(\text{match on } A \cup B) = \mu(\text{match on } A) + \mu(\text{match on } B),$$

for any  $A, B$  s.t.  $A \cap B = \emptyset$ . Finally it has to be additive with respect to the resolution, i.e. for any  $B \subset A$

$$\mu(\text{match on } A, B) = \mu(\text{match on } A) + \mu(\text{match on } B).$$

We make precise our construction for one-dimensional signals (tool-mark evidence). Similar constructions can be made for higher dimensions. Let us assume that  $n = 2^J$ . We consider the set of index  $\mathcal{M}_n = \{m = (j, k), j = 0, \dots, J-1, k = 1, \dots, 2^j\}$  and the integer intervals of the form

$$I_{j,k} = \left\{ i \in \mathbb{N}, \frac{k-1}{2^j} < \frac{i}{n} \leq \frac{k}{2^j} \right\}.$$

The tests of nullity associated with these intervals are denoted by  $\Phi_{j,k}$ . Let  $(\pi_j, j \in \mathbb{N})$  denote a sequence of non negative reals such that  $\sum_{j=0}^{\infty} \pi_j = 1$ . The measure of matching associated to the interval  $I_{j,k}$  is  $2^{-j} \pi_j (1 - \Phi_{j,k})$ . The global measure of matching is equal to

$$\mu(\text{matching}) = \sum_{j=0}^{J-1} 2^{-j} \pi_j \sum_{k=1}^{2^j} (1 - \Phi_{j,k}) = 1 - \sum_{j=0}^{J-1} 2^{-j} \pi_j \sum_{k=1}^{2^j} \Phi_{j,k}.$$

Hence,  $2^{-j} \sum_{k=1}^{2^j} (1 - \Phi_{j,k})$  appears as the measure of matching at resolution level  $j$ . Our definition of the measure of matching can be interpreted using conditional expectations as the expectation of matching at any levels using a prior distribution  $\pi$  on the resolution levels. More simply, if all tests of the level  $j$  are equal to 0 ( $s = 0$  on  $I_{j,k}$  for all  $k = 1, \dots, 2^j$ ) the contribution to the measure of matching is equal to  $\pi_j$  and the previous formula defines a number between 0 and  $1 - \sum_{j=J}^{\infty} \pi_j$ .

To ensure that the measure of matching is invariant by translation a natural sequence of  $\pi_j$  is the uniform distribution by level:  $\pi_j = 1/(2^{\log n})$  for  $j = 0, \dots, J-1$  and 0 otherwise.

By construction, on each interval  $I$ , the test associated with  $I$  makes an error of first kind (i.e. does not reject  $\mathbf{H}_0$  when  $s$  is non zero) if  $s$  is too close to 0 on  $I$  - the precise sense of "close" depend on the choice of the test of nullity considered and is not discussed here as this has to be studied in practice using real data. Using a uniform distribution by level ( $\pi_j = 1/(2^{\log n})$ ), we give to the errors of first kind a weight which is only proportional to the level of each elementary test and hence which is only proportional to the length of the interval up to the power constant  $c$ .

An other way to compute a measure of matching is to consider the  $p$ -value of each test  $\Phi_{j,k}$  denoted  $\rho_{j,k}$  and to define

$$\mu(\text{matching}) = \sum_{j=0}^{J-1} 2^{-j} \pi_j \sum_{k=1}^{2^j} \rho_{j,k}.$$

In order to keep translation invariance here again a natural distribution for the  $\pi_j$  is the uniform by level previously defined.

### 3. Shoe print comparison: complexity of patterns

Shoe print patterns usually contain a lot of details, but for shoe print comparison only those subpatterns are relevant that are not general and found on all shoes of a specific type. In the forensic practice one is interested in subpatterns that are supposed to be characteristic for the shoe under consideration. When comparing the patterns of two shoe prints it may happen that the prints match at some subpatterns but differ at others. Matching of subpatterns means that both prints have similar subpatterns at similar positions. The matching is never perfect due to circumstances. In Figure 3 examples of realistic subpatterns are shown.

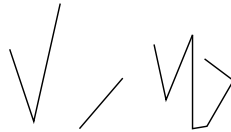


FIGURE 1. Some realistic subpatterns.

The probability that the two prints stem from the same shoe has to be determined via some probabilistic model. Such a model will be proposed in the next section. A necessary ingredient of such a model is the rate of complexity of a subpattern. If two prints match with respect to a simply shaped pattern, this agreement has to be weighed differently from a match at a highly intricate pattern, simply because a shape with high complexity characterizes the shoe to a higher degree. The underlying idea for this is: the more complex a feature, the more rare it is. The question here is how to define the complexity of a pattern. We shall deal separately with one and two-dimensional features. In fact it could suffice to deal only with two-dimensional patterns, since the one-dimensional case is a special case then. However, we pay rather extensive attention to the one-dimensional case, since it helps in developing and clarifying ideas. Moreover, in the forensic practice one-dimensional features on shoe prints are treated as a special class. For this class tables are available in which the complexity of these patterns is specified according to common sense and expertise. Comparison of the numbers of these tables and the present results is one of the intriguing aspects of this project. This comparison can not yet be reported here, since this is part of future research.

The complexity of shapes is hard to define in general. It is to be expected that

any definition involves a relative measure with which two shapes can be compared. The absolute value of the complexity of shapes is in practice irrelevant. A particular nice starting point is the following definition of complexity, which is from [12] and [13].

DEFINITION 3.1 (Complexity of an object). The complexity of an object is the length of its description in some language.

Of course, one has to choose the language suitable, depending on the nature of the object. To illustrate this definition, consider the following examples.

*What is the complexity of a natural number?*

The numbers 0, 1, 2, ..., 9 are described with one digit, the numbers 10, 11, 12, ..., 99 are described with two digits, etcetera. In general, the numbers  $10^n, \dots, 10^{n+1} - 1$  are described with  $n$  digits. In this simple case it is clear that the parameter  $n$  is a good measure of the complexity  $c$ . So for a natural number  $N$ :

$$(1) \quad c(N) \sim n \sim \log(N) .$$

Here, the symbol  $\sim$  stands for 'scales with'; it implies that possible multiplicative and/or additive constants are left out.

*What is the complexity of the repeating sequence of finite length*

$$S = 104104104104104?$$

Since the sequence is repeating, it can be described by mentioning the periodic part (104) and the number of times it is repeated (5). So, the parameters are period and repetition rate. As seen above, the length of a natural number grows with the log of its value. It is not directly clear how to combine the complexities of period and repetition rate. At first we tend to add them. For short series this is appropriate. However, in the limit of a very long series, the contribution from the repetition rate will then strongly dominate the contribution from the period and the information in the period gets lost. For long series we prefer to multiply both contributions, so

$$(2) \quad c(S) \sim c(\text{period}) * c(\text{repetition rate}) .$$

*What is the complexity of an arbitrary real number?*

An arbitrary real number needs infinitely many digits to describe it, so it has complexity  $c = \infty$ . Note that the real number  $\pi$  has also infinitely many digits. However, it can be described alternatively by a few words, i.e. as the ratio of circumference and diameter of a sphere. The number  $\pi$  is certainly not 'arbitrary'.

**3.1. The complexity of one-dimensional patterns.** Here, we restrict ourselves to shapes that consist of one-dimensional components such as line and arc segments.



*Line segment L*

From the very general definition above we may deduce a definition for a line segment. Such a segment can be considered as a sequence of identical pixels, so with a simple internal structure, and with a certain length  $D$ . We assign to a single pixel the complexity one. Formula (2) then yields

$$(3) \quad c(L) = \log(D/D_0) .$$

Here,  $D_0$  is the minimal length that is relevant for the shoe print under consideration. This is dependent on the signal to noise level of the print. A typical value is  $D_0 \approx 1.5$  mm. For  $D = D_0$  we have  $c(L) = 0$ .

*Two line segments  $L_{2,V}$  forming a 'V'-shape.*

In practice these shapes are considered a relevant separate class. The total size of this pattern is measured by the diameter  $D$  of the smallest circumscribing sphere. The complexity of the internal pattern is measured by counting the number of parameters characterizing it. We find two lengths and an angle. Only the ratio of the lengths is relevant, since the pattern may be freely scaled: its size is accounted for via  $D$ . So, we have two dimensionless parameters and this leads to

$$(4) \quad c(L_{2,V}) = 2 \log(D/D_0) .$$

*Two crossing line segments  $L_{2,C}$ .*

We proceed in an analogous way. The internal structure of  $L_{2,C}$  has 4 lengths and one angle. So we have 4 dimensionless parameters and

$$(5) \quad c(L_{2,C}) = 4 \log(D/D_0) .$$

*Triangle  $L_{3,T}$ .*

The internal structure of  $L_{3,T}$  is characterized by 3 lengths. So we have 2 dimensionless parameters and

$$(6) \quad c(L_{3,T}) = 2 \log(D/D_0) .$$

*Three crossing line segments  $L_{3,C}$ .*

The internal structure of  $L_{3,C}$  is characterized by 7 lengths and two angles. So we have 8 dimensionless parameters and

$$(7) \quad c(L_{3,C}) = 8 \log(D/D_0) .$$

*Arc (circle segment) A.*

The internal structure of  $A$  is characterized by its length, the radius of the circle, and the angle. We remark that the length of the segment is different from the diameter  $D$  of the circumscribing sphere. For a full circle,  $D$  equals the diameter of the circle, but the length of  $A$  equals  $2\pi D$ . We have 2 dimensionless parameters and

$$(8) \quad c(A) = 2 \log(D/D_0) .$$

*Arbitrarily curved line segment*

An arbitrarily curved line segment can in good order be approximated by series of arcs of different radii. This introduces an error and the expert has to specify which error is still acceptable. The complexity of the curve is then the sum of the complexities of the circle segments. So, if  $N$  segments are used, the internal

structure of the curve is  $2N$ . An alternative approach is to construct a polynomial that fits the curve reasonably well. The number of coefficients of the polynomial then yields the internal complexity of the curve. So, if the polynomial is of order  $N$ , the complexity is  $N + 1$ .

**3.2. The complexity of two-dimensional features.** Figure 3.2 gives some idea how two-dimensional features look like in the shoe print practice: from a completely filled circle to a more or less randomly distributed pattern. Here, black and

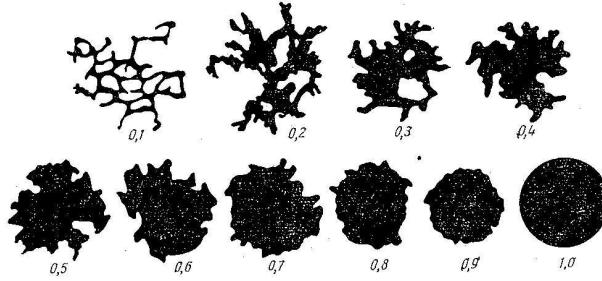


FIGURE 2. Some two-dimensional shapes.

white shapes are studied that are specified via pixels on a grid. We start from the intuitive observation that the complexity of the shape is low if it contains a lot of symmetry and is high when we can not discover any regular pattern in it. It is clear that if the pattern is completely random we have to specify it pixel by pixel. In the latter case we expect that the white and black pixels are distributed more or less uniformly over the shape. This suggests to introduce for a two-dimensional feature the concept of entropy. We define it in this context as follows.

- Put a square grid with boxes over the object. Take the grid size equal to  $n$  pixels. The resulting number of boxes is  $N_n$ , say.
- The number of pixels in each box is  $n^2$ . In box  $i \in \{1, \dots, N_n\}$  we count the number  $n_b$  of black pixels; so box  $i$  has  $n^2 - n_b$  white pixels.

The entropy in box  $i$  is then defined as:

DEFINITION 3.2 (Entropy of box  $i$ ).

$$s_i = \frac{1}{a} \log \binom{n^2}{n_b},$$

with  $\binom{N}{M}$  for arbitrary  $N$  and  $M$  defined as usual by

$$\binom{N}{M} = \frac{N!}{M!(N-M)!}.$$

The normalization factor  $a$  is defined as  $a = \log \binom{n^2}{n^2/2}$ . This factor is chosen such that the entropy in box  $i$  is maximal and equal to 1 if there are as many black pixels as white pixels, i.e.  $n_b = \frac{1}{2}n^2$ .

Summing the entropies over all the boxes we get the total entropy  $s_n$  of the grid under consideration:

$$(9) \quad s_n = \sum_{i=1}^{N_n} s_i.$$

The entropy of course depends on the grid size  $n$ . By varying  $n$ , it is measured on which length scale the object has most structure. The complexity is therefore defined as the maximum entropy under variations of the grid size. So

DEFINITION 3.3 (Complexity of two-dimensional objects (not normalized)).

$$\bar{c} = \max_n s_n .$$

Here,  $n$  is varied continuously. In practice, however, one usually will halve the grid size again and again. It is clear that boxes with no variation in colour have vanishing entropy. So, if an object consists of subpatterns of one colour, the inner part of such a subpattern will not contribute as soon as the grid size becomes smaller than the size of the pattern. Only boxes containing parts of the boundaries have variation in colour and contribute to the entropy. We conclude that this complexity  $\bar{c}$  determines the total length of the boundaries in and around the object. This agrees with the intuitive notion of complexity for objects of this kind. The important point here is that the definition also applies to objects that do not consist of well-defined subshapes of one colour, but have a more or less fractionated appearance or even a fractal dimension.

The definition of  $\bar{c}$  above has the disadvantage that it does not yet reduce to the definition for complexity in one dimension in case of a line segment. This is because  $\bar{c}$  is not yet normalized by the size of the object as a whole.

**Example: reduction to a line segment**

As an example we take a line segment consisting of 8 horizontal pixels. According to the procedure described above, it has to be covered by an  $8 * 8$  grid. If we take  $n = 8$ , this implies that the segment is covered with only one box of size  $8 * 8$ . Applying the formula we find  $s_8 \approx 0.5$ . Halving the grid size and thus taking  $n = 4$ , we obtain that  $s_4 \approx 1.5$ . Two grid boxes have vanishing entropy since they are uniform in colour. The other two contribute equally. Halving the grid size again we find that now 4 grid boxes contribute and  $s_2 \approx 4.0$ . Of course,  $s_1 = 0$ , since grid boxes consisting of one pixels have vanishing entropy. The maximum is thus found for  $n = 2$  and we denote this  $n$ -value by  $n_{max}$ .

To manage that the two-dimensional complexity definition reduces to the one-dimensional one if applied to a line segment, we adjust the  $\bar{c}$  a bit. Let, as usual,  $D$  be the diameter of the smallest circle around the object. Then, the following definition has the required properties:

DEFINITION 3.4 (Complexity of two-dimensional objects (normalized)).

$$c = \bar{c} \frac{n_{max}}{D} \log(D/D_0).$$

This complexity could be conveniently used as weight function when two shoe print patterns agree on one or more shapes.

#### 4. Shoe prints comparison: likelihood ratios

We are given two prints. One is a trace found on the crime scene. The second is a print of the shoe of the (a) suspect. What can we say, on basis of these two prints, about the two statements '*The shoe of the suspect made the print on the crime scene*' and '*The shoe of the suspect did not make the print on the crime scene*'? We will measure the strength of the evidence by a *likelihood ratio*.

The characteristics of a shoe (print) are the marks (damages) on it. Each mark is given by a vector  $(x, y, s, \alpha)$ , where  $(x, y)$  are the coordinates of the mark on the shoe,  $s$  is the complexity of the shapes of the mark and  $\alpha$  is the orientation of the mark. For reasons of simplicity, we describe the shape of the mark by its complexity and not shape itself. Of course this results in the loss of some information.

Suppose the shoe of the suspect contains  $K$  marks and the crime scene print contains  $K'$  marks. We then define

DEFINITION 4.1. The *suspect evidence*  $E$  is given by the matrix

$$E = \{(x_i, y_i, s_i, \alpha_i), i = 1..K\}.$$

The *crime scene evidence*  $E'$  is given by

$$E' = \{(x'_i, y'_i, s'_i, \alpha'_i), i = 1..K'\}.$$

Note that the order of the rows of  $E$  is not important. Furthermore we define  $(x, y) = (x_i, y_i)_{i=1..K}$ ,  $s = (s_i)_{i=1..K}$  and  $\alpha = (\alpha_i)_{i=1..K}$  and likewise for the columns of the crime scene evidence matrix.

There are two possible worlds. We call them  $H_p$  and  $H_d$ , where the  $p$  stands for prosecutor and the  $d$  for the *defender*. In the world  $H_p$ , the shoe of the suspect made the print. The probability measure in this world is  $P_p$ . In  $H_d$ , the shoe of the suspect did not make the print. In this second world the probability measure is  $P_d$ .

**4.1. A likelihood ratio.** A measure to indicate the strength of the evidence supplied by the two marks is the so-called *likelihood ratio*. The likelihood ratio of an event  $A$  in the two worlds  $H_p$  and  $H_d$  is defined by the ratio

$$(10) \quad LR(A) = \frac{P_p(A)}{P_d(A)}$$

if the two probabilities are well-defined and the denominator is non-zero. In case the event  $A$  is defined in terms of continuous random variables, the likelihood ratio is defined by the quotient of the densities  $f_p$  and  $f_d$  i.e.  $LR(A) = f_p(A)/f_d(A)$ . In the following we will encounter events defined in terms of discrete and continuous random variables. For reasons of simplicity, we will, with abuse of notation, use the notation of (10) for the continuous case as well.

The event  $\{E, E'\}$  consists of those outcomes in the probability space that result in the suspect evidence  $E$  and crime evidence  $E'$ . To measure the strength of the evidence supplied by the two prints, we calculate the likelihood ratio  $LR$  of  $\{E, E'\}$  in the two worlds  $H_p$  and  $H_d$ , i.e.

$$LR = \frac{P_p(E, E')}{P_d(E, E')}.$$

The numerator of  $LR$  is the probability that in the world where the shoe of the suspect is the shoe that made the crime scene mark, the event  $\{E, E'\}$  happens. The denominator is the probability that  $\{E, E'\}$  happens in the world where the suspect is innocent. The quotient of these two terms can be used to measure how likely it is that the suspect is guilty: a high value of  $LR$  pleads against him, a low value strengthens his defense.

In the following we assume that the position of the mark(s), their shape and their orientation are independent. This assumption is not clearly untrue, but whether this independence is valid should be checked with (the) data. Conditioning on  $K$  and  $K'$  then yields

$$\begin{aligned} LR &= \frac{P_p(K, K') P_p((x, y), (x', y') | K, K') P_p(s, s' | K, K') P_p(\alpha, \alpha' | K, K')}{P_d(K, K') P_d((x, y), (x', y') | K, K') P_d(s, s' | K, K') P_d(\alpha, \alpha' | K, K')} \\ (11) \quad &= LR_{number} \times LR_{position} \times LR_{shape} \times LR_{orientation}. \end{aligned}$$

**4.2. One mark.** As an indication how to obtain (11), we now calculate each of the four factors of the likelihood ratio  $LR$  in the case that both the crime scene print and the shoe of the suspect show one feature, i.e.  $K = K' = 1$ .

The calculation of the four likelihood ratios in (11) follows the same pattern. In the numerator we calculate the probability under the assumption that the crime scene print was made by the suspect's shoe. Hence there is in fact one shoe! The probability is then a product of two factors: the probability (density) that an arbitrary shoe has the given property and the probability (density) that two different prints from the same shoe differ a certain (given) way.

For the denominator, the assumption is that the two prints are made by two different shoes. The two shoes have nothing to do with each other and are independent. The probability (density) that they both contain certain features is then the product of the probability (density) that an arbitrary shoe contains certain feature(s) and the probability that a print from an arbitrary shoe contains certain (other) feature(s). Note that these two probability (densities) are not (necessarily) the same: in the shoe print there is less detail and much more noise.

• $LR_{position}$ . Let us first look at the numerator. In the world  $H_p$  the shoe of the suspect made the crime scene print. The positions of the two marks however do not have to coincide: both in the making of the print by the suspect and in the measuring of the positions  $(x, y)$  and  $(x', y')$  (little) errors are made. These errors depend on the surface on which the crime scene print is found as well. It seems reasonable to assume that  $(x - x', y - y')$  has approximately a bivariate normal distribution with mean zero, covariance matrix  $\Sigma_{surface}$ . and density  $f_{surface}$ . Let the

surface area of the shoe be  $B$ . Assuming that the position of the mark is uniformly distributed over the surface of the shoe, we have

$$P_p((x, y), (x', y') \mid K = 1, K' = 1) = B^{-1} f_{surface}(x - x', y - y').$$

For the denominator we move to the world  $H_d$ , where the shoe of the suspect did not make the print. Hence the numerator is the probability (density) that the marks on two different shoes are situated in positions  $(x, y)$  and  $(x', y')$  respectively. By the uniformity assumption above and the independence, this density is  $1/B^2$ . Hence

$$\begin{aligned} LR_{position} &= \frac{P_p((x, y), (x', y') \mid K = 1, K' = 1)}{P_d((x, y), (x', y') \mid K = 1, K' = 1)} \\ (12) \quad &= B^{-1} f_{surface}(x - x', y - y') / B^{-2} = B f_{surface}(x - x', y - y'). \end{aligned}$$

Note that (12) does not depend on the unit in which  $x, y$  and  $B$  are measured. Indeed, changing the unit of measurement by a factor  $c$  results in  $c^{-2} B c^2 f_{surface}$  and  $c$  nicely drops out.

•  $LR_{angle}$ . Assuming the distribution of the angle of the mark is uniform, again, the numerator consists of a 'uniform' term  $1/(2\pi)$  and Gaussian 'error' term  $g_{surface}(\alpha - \alpha')$ , again depending on the surface where the crime scene print was found. In the denominator the uniform term  $1/(2\pi)$  is squared. Then  $LR_{angle}$  is given by

$$LR_{angle} = 2\pi g_{surface}(\alpha - \alpha').$$

•  $LR_{shape}$ . For calculating the likelihood ratio of (complexity of) the shapes, we use the theory of section 3. The complexity of a mark is an element of a countable set. Its discrete distribution  $P_{compl}$  should be obtained from the data, possibly in combination with the theory of section 23. Let  $P_{err}(s, s')$  be the probability that a shoe with a mark of complexity  $s$  produces a print with a mark of complexity  $s'$ . As before, this is term caused by a 'reproduction error'. Then

$$(13) \quad LR_{shape} = \frac{P_{compl}(s)P_{err}(s, s')}{P_{compl}(s)P_{compl}(s')} = \frac{P_{err}(s, s')}{P_{compl}(s')}.$$

Note that  $P_{err}(s, s')$  is not the same as the probability that a print with a mark of complexity  $s'$  was produced by a shoe with a mark of complexity  $s$ , nor is it equal to  $P_{err}(s', s)$ .

•  $LR_{number}$ . In the numerator we have  $P_p(K, K') = P(K' = 1)P_p(K' = K \mid K' = 1)$ . The probability that the shoe of the suspect contains one mark given that the crime scene print (made by the same shoe!) contains one mark,  $P_p(K' = 1 = K \mid K' = 1)$ , depends on a number of things:

- the time the shoe has been used between the crime time and the moment the shoe was secured
- the probability that a mark on a shoe actually is found in a print the shoe leaves behind - this in its turn depends on the surface
- the probability that a mark in the crime scene print is caused by something temporarily attached to the shoe, for instance a leaf, a twig, etcetera.

Therefore this probability is perhaps the most complex of all the probabilities in this section. Further study (of data) could reveal its nature. Let us conclude by calculating the denominator:  $P_d(K, K') = P(K = 1)P(K' = 1)$ . The likelihood ratio then satisfies

$$LR_{number} = \frac{P_p(K = 1 | K' = 1)}{P(K = 1)}.$$

If the marks on the suspect's shoe are similar to those on the crime scene, the likelihoods described above will in general be larger than one. Hence, when it is impossible to compute all likelihoods, we can find a lower bound for  $LR$  by computing some of them. Such a lower bound might already be powerful enough for a clear conclusion about strength of the evidence.

**Acknowledgements** We thank the members of the NFI that were part of our study group. Huub Hardy, Andre Hoogstrate, Ies Keereweer and Marjan Sjerps were almost continuously present, usually arrived even before the mathematicians did and created a very pleasant atmosphere. Some of the ideas brought forward by them have found their way into this report. We thank Sergei Anisov and Dinard van der Laan for interesting discussions and for developing some of the ideas written down in these proceedings. Finally, we thank the NFI for organizing a symposium in May 2003 where we have explained our ideas and have had the opportunity to 'match' them with the views of the NFI experts.

### Bibliography

- [1] Y. Baraud, S. Huet, and B. Laurent. Adaptive tests of linear hypotheses by model selection. *Annals of Statistics*, 31(1), 2003.
- [2] A. P. A. Broeders. Daubert en Saks: het einde van de klassieke identificatiecriminalistiek? *Nederlands Juristenblad*, 36:1786–98, October 2002.
- [3] M. Heizman. Strategies for the automated recognition of marks in forensic science. *In Proceedings SPIE 4709*, April 2002.
- [4] A. Hoogstrate and Y. Rozenholc. Testing nullity in a regression framework to evaluate forensic evidence. *To appear*.
- [5] Joel L. Horowitz and Vladimir G. Spokoiny. An adaptive, rate-optimal test of linearity for median regression models. *J. Amer. Statist. Assoc.*, 97(459):822–835, 2002.
- [6] Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Statist.*, 2(2):85–114, 1993.
- [7] Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. II. *Math. Methods Statist.*, 2(3):171–189, 1993.
- [8] Yu. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. III. *Math. Methods Statist.*, 2(4):249–268, 1993.
- [9] Yu. I. Ingster and Irina A. Suslina. *Nonparametric goodness-of-fit testing under Gaussian models*, volume 169. Lecture notes in statistics (Springer-Verlag), 2003.
- [10] O. V. Lepski and A. B. Tsybakov. Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Related Fields*, 117(1):17–48, 2000.

- [11] Oleg V. Lepski and Vladimir G. Spokoiny. Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*, 5(2):333–358, 1999.
- [12] Li, Ming & Vitányi, Paul. An introduction to Kolmogorov complexity and its applications. Second edition. Graduate Texts in Computer Science. Springer-Verlag, 1997.
- [13] Li, Ming & Vitányi, Paul. An introduction to Kolmogorov complexity and its applications. Texts and Monographs in Computer Science. Springer-Verlag, 1993.
- [14] M. J. Saks. Merlin and Solomon: Lessons from the laws formative encounters with forensic identification science. *Hastings Law Journal*, 49:1069–1141, 1998.
- [15] V. Spokoiny. Data-driven testing the fit of linear models. *Math. Methods Statist.*, 10(4), 2002.