Estimating the probability of accidental mark locations on a shoe sole

Naomi Kaplan Damary∗, Micha Mandel†, Yoram Yekutieli‡, Sarena Wiesner§, Yaron Shor¶

December 19, 2019

Abstract
Footwear comparison is an important type of forensic evidence used to link between a suspect’s shoe and a footprint found at a crime scene. Besides the type and the size of the shoe, investigators compare the trace to the source using randomly acquired characteristics (RACs), such as scratches or holes, in order to distinguish between similar shoes. However, to date, the distribution of RAC characteristics has not been investigated thoroughly, and the evidential value of RACs is yet to be explored. A first important question concerns the distribution of the location of RACs on shoe soles in the general population, which can serve as a benchmark for comparison. In this paper, the location of RACs is modeled as a point process over the shoe sole and a data set of 386 independent shoes is used to estimate its rate function. As the shoes are differentiated by shape, level of wear and tear and contact surface, this process is complicated. This paper presents methods that take into account these challenges, either by

∗Department of Statistics, The Hebrew University of Jerusalem, 91905 Jerusalem, Israel. naomi.kaplan@mail.huji.ac.il
†Department of Statistics, The Hebrew University of Jerusalem, 91905 Jerusalem, Israel. micha.mandel@mail.huji.ac.il
‡Hadassah Academic College Dept. of Computer Science 37 Hanevi'im St. P.O.Box 1114, 91010 Jerusalem, Israel. yoramye@hac.ac.il
§Israel National Police Division of Identification and Forensic Science (DIFS) 1 Bar-Lev Road 91906 Jerusalem, Israel. sarenawiz@gmail.com
¶Israel National Police Division of Identification and Forensic Science (DIFS) 1 Bar-Lev Road 91906 Jerusalem, Israel. yaronshor@gmail.com
using natural cubic splines on high resolution data, or by using a piecewise-
constant model on larger regions defined by forensic experts. It is shown
that RACs are likely to appear at certain locations, corresponding to the
foot’s morphology. The results can guide investigators in determining the
evidential value of footprint comparison.

Keywords — Case-control sampling, Conditional maximum likeli-
hood, Natural cubic splines, Random effects model, Randomly acquired
characteristics (RACs), Shoeprints

1 Introduction

In recent years, forensic methods have been criticized for their shortcomings
in providing courts with objective and quantitative answers to the question of
whether a sample from a suspect matches a sample found at the crime scene.
Unlike DNA that is used routinely to link suspects to crime scenes because of its
scientific objectivity and accessible documentation, the evaluation of other types
of evidence such as shoeprints, hair, and even fingerprints has not reached this
gold standard. Both the 2009 NRC report, “Strengthening Forensic Science in
the United States: A Path Forward,” and the 2016 PCAST report to President
Obama, “Forensic Science in Criminal Courts: Ensuring Scientific Validity of
Feature-Comparison Methods,” have called for the strengthening of the scientific
basis of forensic procedures (NRC, 2009; PCAST, 2016).

Here shoeprint comparison is considered. The identification of footwear im-
pressions is based on the comparison of a print found at the crime scene with
a print made from a suspect’s shoe (Bodziak, 1999); see Supporting web ma-
terials 1 and a short film (Kaplan-Damary, 2014) for a detailed description of
the process of comparing shoeprints. The analysis of shoeprints by experts is
done in two broad stages. First, the pattern, size and wear of the shoe sole are
compared to the crime scene print. If these do not fit, the analysis is stopped
and the pair is classified as a non-match. In the second stage, the forensic expert
examines whether randomly acquired characteristics (RACs) on the shoe sole
match the RACs on the print from the crime scene. These RACs have various
characteristics, such as location, shape and orientation, that are used in the
expert’s evaluation.

The current study deals only with the location of RACs. Figure 1 presents
two lab prints taken from suspects’ shoes, with the location of RACs marked
by the examiner. The rarity of this set of RACs is of major interest, especially
in establishing a link between suspect’s shoes and the crime scene prints. Thus, the main focus of this study is understanding the spatial distribution of RACs’ locations, and specifically whether they are distributed uniformly as assumed by Yekutieli et al., (2016) and Stone, (2006) or concentrated in certain areas. This is an essential step in evaluating the degree of rarity of a given set of RACs, i.e. the probability that a random shoeprint has a pattern of RACs that is sufficiently similar. Marks at sparsely populated locations would be of much greater value in determining a match than marks at highly populated locations.

The RACs are modeled as a spatial process. Its intensity function is estimated in order to calculate the probability of observing RACs in different locations. This is done by using the Israeli Police Division of Identification and Forensic Science database of RACs, which includes 13,500 RACs from 386 lab shoeprints (Yekutieli et al., 2016). Estimation of intensity functions arises naturally in the framework of spatial statistics (Cressie, 1993), but there are several complications in the analysis of shoe impressions. First, spatial statistics typically deals with a single process that has many events, while in our case, there are many independent processes (shoe impressions), but each has only a small number of events (RACs); about 34 per shoe on average. A second complication is the variability of shoes: they differ in their types and sizes. Moreover, different shoes have different contact surfaces i.e, the part of the sole that actually
touches the floor or ground (see Figure 1, which presents an example of lab prints. The area in orange is the contact surface). This fact limits the area in which RACs appear, thus affecting the probability of observing them. On top of these difficulties, some shoes are scarred by many RACs while others have relatively few, apparently due to the level of wear and tear. This article models and estimates the intensity function while taking into account the challenges noted above.

The data used in this article are described in Sections 2 and the model is presented in Section 3. In Section 4, three estimators of the intensity function are introduced; a naive estimator, an estimator based on a random effects model and an estimator based on conditional maximum likelihood. These are presented in a model which employs pixels as areas, a logistic regression and natural cubic splines which create a smooth intensity function. In addition, sub-sampling techniques are described in order to deal with computational challenges. The section ends with an application of the estimated intensity function to the data described in Section 2. Using the exact location of RACs is problematic in practice due to characteristics of the data presented in Section 5. Instead, the shoe is divided into larger areas to which a piece-wise constant intensity function is fitted. Section 6 presents simulation results for the comparison of sub-sampling case-control techniques as well as a comparison of the piecewise constant estimators in different settings. Section 7 concludes the paper with a discussion.

2 Data

The Israeli Police Division of Identification and Forensic Science (DIFS) has amassed one of the most comprehensive RAC databases, including some 386 lab prints and 13,500 RACs (Yekutieli et al., 2016). An important initial pre-processing step was to normalize all shoe impressions to a standardized X-Y axis with identical length and orientation. This was done by first marking a shoe-aligned coordinate system on each print and then standardizing the shoe according to this system: for each lab print, the top and bottom of the shoeprint were marked to indicate the direction of the major axis and to determine the length of the shoe. The axes’ origin was set at the middle point between the two marked extremities. The minor axis was defined as the line perpendicular to the major axis that passes through the origin of axes. The standardization
was done by transforming all measurements from image coordinates to the shoe
aligned coordinate system as follows (More details of the normalization process
can be found in Yekutieli et al., 2016):

a) Translation of the marked origin of axes to (0,0).

b) Rotation by the direction of the shoe aligned coordinate system.

c) Scaling by the length of the shoeprint.

d) Multiplying the x-value of the points (x being the horizontal axis) by -1 or
   1, to mirror if needed such that all shoeprints will be turned to left shoes.

RACs are assumed to have a two-dimensional shape. This study focuses on
the location, measured as a point \((x, y) \in \mathbb{R}^2\), which is the center of gravity
calculated as the mean of all pixels included in the RAC.

The number of RACs per shoe varies between 1 and 190 with an average of
34 (for details see Figure 1 in Supporting web materials 2) except for one shoe
that has an unusually high number of RACs (309). The RACs were marked
by different examiners who were supervised by forensic experts. The guidelines
were to mark RACs that are larger than one square millimeter although smaller
marks could be included if visible without magnification. RACs can be observed
only where there is a contact surface, a feature which varies from one shoe to
the other. This should be taken into account in the analysis as described in
Section 3. The number of pixels with contact surface per shoe, varies between
3736 and 19500 (See Figure 2 in Supporting web materials 2). It is also shown
that the pad of the shoe and the 4 circles at the heal more frequently contain a
contact surface (see Figure 3 in Supporting web materials 2 for the cumulative
contact surface of all shoes). There is a weak correlation between the number
of pixels with contact surface per shoe and the number of RACs per shoe – the
Spearman correlation coefficient is equal to 0.115; see Figure 4 in Supporting
web materials 2.

3 Model

Consider \(m\) independent shoes with different levels of wear and tear. It is
assumed that the locations of RACs on each shoe follow a non-homogeneous
Poisson process. Specifically, let \(D \subset \mathbb{R}^2\) be the region representing the surface
of a generic shoe, and let \(B_i \subset D\) be the contact surface of shoe \(i, i = 1, \ldots, m\).
For any $A \subset D$, denote by $N_i(A)$ the number of RACs appearing on subset $A$ of shoe $i$. Note that $N_i(A)$ represents all of the RACs: those on the contact surface and those that are not, but only RACs on the contact surface are observed. $N_i$ is assumed to be a non-homogeneous Poisson point process on shoe $i$ with intensity function $\lambda(i)$ and corresponding cumulative intensity function $\Lambda(i)(A) = \int_{(x,y) \in A} \lambda(i)(x,y) \, dx \, dy$. Let $a_i \in \mathbb{R}^+$ be a random variable that indicates the degree of wear and tear of shoe $i$, such that $a_i$, $i = 1, \ldots, m$, are iid with $E(a_i) = 1$ and $\text{Var}(a_i) = \sigma^2$. Our basic assumption is that

$$\lambda(i)(x,y) = \lambda(0)(x,y) \cdot a_i, \quad (x,y) \in D,$$

where $\lambda(0)$ is joint to all shoes. Thus, it is assumed that all shoes have the same shape of the intensity function and only differ by the shoe-specific parameter $a_i$ that determines the height of the function. An equivalent assumption to (1) is

$$\Lambda(i)(A) = \Lambda(0)(A) \cdot a_i, \quad A \subset D, \quad \Lambda(0)(A) = \int_{(x,y) \in A} \lambda(0)(x,y) \, dx \, dy.$$

The main goal is to estimate the baseline intensity function $\lambda(0)$, which is, in general, a continuous function. To simplify the analysis while preserving the sole type for the observer, images were reduced in resolution from $7000 \times 3000$ to $397 \times 307$ pixels. Thus, the modeling assumption is that the function is constant within pixels in a high resolution grid. In Section 5 we consider a partition of larger areas. In general, region $D$ is partitioned into $J$ subsets $A_1, \ldots, A_J$ such that $\bigcup_j A_j = D$ and $A_j \cap A_k = \phi$, $\forall j \neq k$. The assumption is that $\lambda(0)(x,y) = \lambda_j$ for all $(x,y) \in A_j$. For each subset, two characteristics are of interest: the contact surface $A_j \cap B_i$ (considered fixed and not random) and the number of RACs.

As mentioned above, $N_i(A)$ is a point process of the RACs’ locations on the shoe sole and RACs are observed only where there is a contact surface. Let $N_{ij} = N_i(A_j \cap B_i)$ be the number of observable RACs on shoe $i$ and subset $j$, and $n_{ij}$ denote its realization. Also, denote by $n_i = \sum_j n_{ij}$, the total number of observed RACs on shoe $i$. All the modeling assumptions above reduce to

$$N_{ij} | a_i \sim \text{Poisson}(\lambda_j S_{ij} a_i),$$

where $S_{ij} = |B_i \cap A_j|$ is the area of the contact surface of shoe $i$ in subset $j$. Figure 5 in Supporting web materials 3 summarizes the model schematically on a lab print.
4 Estimation using maximum resolution

The maximum resolution is achieved when the regions are in fact pixels. In this case, $S_{ij} \in \{0, 1\}$ and $n_{ij} \in \{0, 1\}$. When RACs are created they may tear the shoe sole such that the location of the RAC appears to be on an area with no contact surface and thus the value of $S_{ij}$ is set to 1 in all cases where $n_{ij} = 1$. In addition, the value of $n_{ij}$ is set to 1 in 38 cases where $n_{ij} = 2$. Appearance of two RACs in the same pixel may be due to the way the data were pre-processed and the location was defined. More details regarding the complexities in defining the location and in using pixels are addressed in Section 5. Note that if there is no contact surface, RACs cannot be observed. Thus, $n_{ij} = 0$ whenever $S_{ij} = 0$.

The number of pixels is relatively large (an average of about 10,000 per shoe) compared to the number of RACs (average of approximately 34), meaning that there are many more pixels with $n_{ij} = 0$ than pixels with $n_{ij} = 1$.

4.1 A naive estimator for $\lambda^{(0)}$

By (2), if $S_{ij} > 0$ then

$$E \left( \frac{N_{ij}}{S_{ij}} \right) = E \left( \frac{E(N_{ij}|a_i)}{S_{ij}} \right) = E \left( \frac{\lambda_j S_{ij} a_i}{S_{ij}} \right) = \lambda_j E(a_i) = \lambda_j. $$

Therefore, a natural (unbiased) estimator for $\lambda_j$ is

$$\hat{\lambda}_j = \frac{1}{|m_j|} \sum_{i \in m_j} \frac{n_{ij}}{S_{ij}}, \quad (3)$$

where $m_j = \{i|S_{ij} > 0\}$. As noted above, in case of pixels, $S_{ij} \in \{0, 1\}$, $|m_j| = \sum_{i=1}^{m} S_{ij}$, and the estimator reduces to

$$\hat{\lambda}_j = \frac{1}{|m_j|} \sum_{i \in m_j} n_{ij} = \frac{\sum_{i=1}^{m} n_{ij}}{\sum_{i=1}^{m} S_{ij}},$$

where the last equality follows from $\{S_{ij} = 0\} \Rightarrow \{n_{ij} = 0\}$.

For the maximal resolution case, the estimator coincides with the one suggested by Yekutieli et al., (2016). In order to get a smooth estimator for the two-dimensional function $\lambda^{(0)}$, the set of estimators $\hat{\lambda}_j, j = 1, \ldots, J$ are smoothed using a Kernel smoother. A drawback of this estimation technique is the need to estimate a very large number of parameters separately while ignoring the
spatial structure. A possible alternative is to model this structure using smooth functions, as is done next.

4.2 Estimation of $\lambda^{(0)}$ using a random effects model

As noted in Section 3, a non-homogeneous Poisson process is assumed for the RACs locations. In the binary setting, the occurrence of RACs is approximated by a logistic regression model, which should work well as RACs are rare. Specifically, for $S_{ij} = 1$, it is assumed that

$$P(N_{ij} = 1 | a_i) = e^{g(\beta, x(A_j), y(A_j)) + a_i} / (1 + e^{g(\beta, x(A_j), y(A_j)) + a_i}) \tag{4}$$

where $(x(A_j), y(A_j))$ are the coordinates of pixel $j$ and $g$ is chosen here to be a product of natural cubic splines $g(x, y) = g_X(x)g_Y(y)$ (Hastie et al., 2009),

$$g_X(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \sum_{j=1}^{p} \beta_{3+j} (t - k_j)^3_+,$$

where $k_1, \ldots, k_p$ are the $p$ knots and $b_+ = \max(0, b)$; $g_Y$ is defined similarly.

A random effects model assumes that $a_i$ ($i = 1, \ldots, m$) are independent and identically distributed having a law $H_\theta$ (Myers et al., 2012 p.319) indexed by a parameter $\theta \in \Theta$. Here the standard assumption that $H_\theta$ is a normal distribution with zero mean and unknown variance is used. The likelihood reduces to

$$L(g(\beta, x(A_j), y(A_j)), \theta; n_{ij}) = \prod_{i=1}^{m} \int \prod_{j=1}^{J} \left( \frac{e^{n_{ij} \cdot g(\beta, x(A_j), y(A_j)) + a_i)}{1 + e^{g(\beta, x(A_j), y(A_j)) + a_i}} \right)^{S_{ij}} h_\theta(a) da,$$

where $h_\theta$ is the $N(0, \theta^2)$ density, and the estimators are obtained by maximizing the likelihood with respect to the parameters $\beta$ (of $g$) and $\theta$ (of $h$). Since the number of pixels is relatively large (there are millions of binary variables having $S_{ij} = 1$), computation is challenging and a sub-sampling technique is used, as described in Section 4.4

4.3 Estimation of $\lambda^{(0)}$ using conditional maximum likelihood

Instead of modelling the distribution of the degree of wear and tear, $a_i$, it can be treated as a nuisance parameter and be eliminated by conditioning on its
sufficient statistic (Bishop et al., 2007; Agresti, 2013). Maximizing the resulting likelihood is called conditional maximum likelihood (CML), and yields consistent estimators.

The sufficient statistics of \( x_i \) is \( n_i = n_{ij} \), and the resulting conditional likelihood is,

\[
\frac{e^{\sum_{j=1}^{J} n_{ij} \cdot g(\beta, x(A_j), y(A_j))}}{\Pi_{i=1}^{m} \sum_{u|n_i} e^{\sum_{j=1}^{J} u_j \cdot g(\beta, x(A_j), y(A_j))}}
\]

where \( n_{ij} = \sum_i n_{ij} \) and \( u|n_i \) indicates summation over all \( u = (u_1, \ldots, u_J) \) such that \( \sum_{j=1}^{J} u_j = n_i \). This sum includes \( \binom{|B_i|}{n_i} \approx \frac{|B_i|^3}{34} \) elements, where \( |B_i| \) is very large (about 10,000); see Figure 2 in Supporting web materials 2. This causes computational challenges and a case-control sub-sampling technique is used as described in the next section.

4.4 Sub-sampling techniques

Estimating the intensity function at a high resolution is computationally challenging since the average number of pixels with contact surface per shoe is about 10,650, while the average number of RACs per shoe is around 34. A possible approach is to use random sub-sampling for inference and specifically to employ case-control sub-sampling technique as recently suggested by Wright et al., (2017); (see also Fithian and Hastie, 2014).

In logistic regression, the estimated effect is consistent when using case-control sampling, but the estimated intercept may not be valid (Agresti, 2013). As we are mainly interested in \( \lambda_0 \), the intercepts are of a secondary importance, thought they can be readily estimated as described by Wright et al., (2017). The implication of random sub-sampling from the original data is as follows. Let \( Z_{ij} \) indicate whether pixel \( A_j \) of shoe \( i \) is sampled (1 = yes, 0 = no) and let \( \rho_i^{(1)} = P(Z_{ij} = 1|N_{ij} = 1) \) and \( \rho_i^{(0)} = P(Z_{ij} = 1|N_{ij} = 0) \) be the (possibly shoe-dependent) sub-sampling probabilities of cases and controls respectively. It follows from Bayes’ theorem that,

\[
P(N_{ij} = 1|Z_{ij} = 1) = \frac{\rho_i^{(1)}P(N_{ij} = 1)}{\sum_{k=0}^{k} \rho_i^{(k)}P(N_{ij} = k)}.
\]

4.4.1 Case-control sub-sampling in a random effects model

Using (4), Equation (5) simplifies to
\[ P(N_{ij} = 1 | Z_{ij} = 1) = \frac{\rho_i^{(1)} \exp(g(\beta, x(A_j), y(A_j))) + a_i)}{\rho_i^{(0)} + \rho_i^{(1)}(\exp(g(\beta, x(A_j), y(A_j))) + a_i)} = \frac{\exp(g^*(\beta, x(A_j), y(A_j))) + a_i)}{1 + \exp(g^*(\beta, x(A_j), y(A_j))) + a_i}, \]

where \( g^*(\beta, x(A_j), y(A_j)) = \log(\rho_i^{(1)}/\rho_i^{(0)}) + g(\beta, x(A_j), y(A_j)) \).

Thus, the underlying model of the random sub-sample is identical to the original model except for the intercept. Specifically, the location effect parameters are the same. This means that the estimators of the location parameters using the sub-sample instead of the original sample are consistent while the intercept estimator is biased by a factor of \( \log(\rho_i^{(1)/\rho_i^{(0)}}) \). The latter can be adjusted with the inclusion of simple offset terms reflecting shoe-specific sampling probabilities (see Wright et al., 2017). Note that in the case of identical sampling probabilities for all shoes, the intercept can be adjusted after the estimation.

### 4.4.2 Case-control sub-sampling in a CML estimation

Following similar steps as in the random effects case (see also Agresti, 2013, p.168) it can be shown that the likelihood under case-control sub-sampling and the original likelihood differ only by the intercept. Since the intercept term does not appear in the conditional likelihood, the CML under both scenarios yields estimators for the same location effect parameters.

In Section 6, simulations are carried out to compare different types of within cluster sub-sampling to sub-sampling across the whole data frame.

### 4.5 Data analysis

#### 4.5.1 Estimation of the baseline intensity function

Figure 2 presents the three estimators applied to the shoe data. The naive estimator of Section 4.1 was smoothed using the \texttt{kernel2dsmooth} function in the \texttt{smoothie} package (Gilleland, 2003) in the R program (R Core Team, 2014). A uniform kernel is used where each entry of the smoothed matrix is calculated as the average of its \( 21^2 \) neighbor entries in the original matrix.

The random effects and the CML estimates were calculated using a product of natural cubic splines. Three knots for the X-axis and five knots for the Y-axis were used and their positions were set according to equal quantiles. These numbers of knots enabled flexibility and still avoided computational problems. The \texttt{ns} function under the splines package (R Core Team, 2014) in the R program
was employed. The calculations were based on within-cluster case-control sub-
sampling, which includes all cases (pixels with RACs, $n_{ij} = 1$) and 20 random
controls (pixels without RACs, $n_{ij} = 0$) from each shoe, since, as shown in section 6.1, this sampling method performed the best. The \textit{glmer} function under
the \textit{lmer4} package (Bates et. al., 2015) and the \textit{clogit} function under the survival
package (Therneau, 2015), both in R, were used to calculate the random effects
and the CML estimators respectively. In the random effects case it was assumed
that $a_i \sim N(0, \theta^2)$.

As is evident from Figure 2, the CML and random effects estimators brought
about similar results which were relatively close to the naive estimator (but much
smoother). The estimated intensity function is highest at the ball and heel of
the foot. In addition, using the random effects model, the global test of all
coefficients equal to zero (with the exception of the intercept) was conducted in
order to check the constant intensity function assumption. Although the test was
highly significant in rejecting the constant intensity assumption ($p-value \approx 0$),
the maximum estimated intensity value is about twice that of the minimum
value; meaning that it is not far from a uniform intensity function. Thus, the
probability of finding a RAC is relatively similar across the entire shoe sole, and
there is no area in which observing a RAC increases dramatically the evidential
value against a suspect.

4.5.2 Confidence intervals

To measure the uncertainty in the estimated intensity function, pointwise con-
fidence intervals were calculated under the postulated model. A moment type
estimator for the variance under the naive approach is given in Web appendix 4.
In the CML case, the asymptotic covariance matrix of $\hat{\beta}$ can be estimated using
the observed information matrix based on conditional likelihood evaluated at
$\hat{\beta}$ (Sartori and Severini, 2004). The covariance matrix was estimated using the
\textit{clogit} function under the survival package (Therneau, 2015) in R. In the random
effects case, the same estimation approach of employing the observed information
evaluated at $\hat{\beta}$ is performed using the random effects likelihood. The \textit{glmer}
function under the \textit{lmer4} package (Bates et. al., 2015) in R was employed. Using
these covariance matrices, confidence intervals of the intensity function were cal-
culated. See Figure 3 for the resulting 95% pointwise confidence intervals of the
estimators using random effects and CML models, in three chosen locations on
the Y axis. The confidence intervals based on the naive estimator (not shown)
Figure 2: Comparison of the three estimates of the intensity function are much wider as the result of the local estimation approach. One can see that random and CML confidence intervals are relatively close and in most locations the CML is slightly wider.

5 Estimation using larger areas

The estimates presented in the previous section are local and therefore heavily rely on valid location data. However, the definition of location is problematic for at least two reasons. First, a RAC is not a point in two dimensions but a set of points, which is marked by trained experts. The marking process is somewhat subjective and is exposed to marking errors, therefore the RACs’ centers are prone to inaccuracies. Second, different shoes have different shapes, and it is not clear if they can be appropriately normalized. Here the shoes were normalized according to the Y axis, which is the standard measure of a size of a shoe, but the X axes of different shoes vary. This means that a RAC having a certain X-coordinate can appear near the middle in one shoe and near the edge in another.

These suggest that “locations” of RAC’s are more regional than local, and in order to overcome this, pixels should be grouped to larger subsets. The question
of how to divide the shoe and determine these subsets remains. As noted, these sets should be large enough in order to minimize the errors resulting from the normalization problem especially on the X axis, but also should reflect the differences in the intensity function in different areas. It is assumed that the relative frequencies of RACs in different areas of the sole are mostly attributed to walking patterns. Based on the expertise of the authors from the police laboratory, it was decided that the Y axis of the shoe sole would be divided into 5 layers, the X axis would be divided into 2 layers, and the upper part of the shoe sole that comes in contact with the pad of the foot would be divided into an outer and inner part as they are expected to behave differently. Figure 4 presents the resulting 14 areas that are believed to be quite homogeneous, but have different probabilities of observing RACs. This partition is used in the following analysis. The number of RACs in each area follows the Poisson distribution given in (2).

The naive approach results in the estimator (3). However, since the data used here are not restricted to have 0 or 1 RACs per region, this estimator does
Figure 4: Subsets of the shoe obtained according to expert knowledge

not coincide with the estimator presented in Yekutieli et al., (2016).

For the random effects approach, the likelihood is

$$L(\lambda_1, \ldots, \lambda_J; n_{ij}) = \prod_{i=1}^{m} \int \prod_{j=1}^{J} \frac{e^{-\lambda_j a_i S_{ij} (\lambda_j a_i S_{ij})^{n_{ij}}}}{n_{ij}!} h_\theta(a) da$$

(6)

The random effects estimators are obtained by maximizing the likelihood with respect to $\lambda_1, \ldots, \lambda_J$ and the parameters $\theta$ of $h_\theta(a)$.

The conditional maximum likelihood is obtained by conditioning on the $a_i$’s sufficient statistic, $N_i$. Since $N_{i1}, \ldots, N_{iJ}$ are independent, $N_{ij} | a_i \sim \text{Poisson}(\lambda_j a_i S_{ij})$ and $N_i | a_i \sim \text{Poisson}(\sum_j \lambda_j a_i S_{ij})$.

$$N_{i1}, \ldots, N_{iJ} | N_i = n_i \sim \text{Multinomial} \left( n_i, \frac{e^{\log(\lambda_j) + \log(S_{ij})}}{\sum_j e^{\log(\lambda_j) + \log(S_{ij})}} \right).$$

Thus, the log likelihood (up to a constant) is,

$$\ell(\lambda_1, \ldots, \lambda_J) = \sum_{i=1}^{m} \sum_{j=1}^{J} n_{ij} \left[ \log(\lambda_j) + \log(S_{ij}) - \log \left( \sum_{j'} e^{\log(\lambda_{j'}) + \log(S_{ij'})} \right) \right].$$

(7)

Since $\ell(c\lambda_1, \ldots, c\lambda_J) = \ell(\lambda_1, \ldots, \lambda_J)$ for all $c > 0$, the vector of parameters $(\lambda_1, \ldots, \lambda_J)$ can be estimated only up to a multiplicative constant. We therefore restrict $\lambda_1 = 1$, and the CML approach reduces to solving the following simple
set of equations for \( \lambda_2, \ldots, \lambda_J \),

\[
\frac{\partial \ell}{\partial \lambda_k} = \sum_{i=1}^{m} \left[ \frac{n_{ik}}{\lambda_k} - \frac{n_i}{\sum_{j'} S_{ij'} \lambda_{j'}} S_{ik} \right] = 0, \tag{8}
\]

which can be carried out numerically. This method is restricted to a relatively small number of regions, as the number of parameters \( J \) must be small relative to the number of shoes \( m \).

Remark Due to the identifiability issue discussed above, \( \lambda_1, \ldots, \lambda_J \) cannot be fully estimated by using the conditional likelihood. A possible approach is to scale the CML estimates by equating their average to that of the estimated parameters under the naive approach \([3]\). In this case, there is an added source of variance due to the variance of the naive estimators. This issue requires further research and is not addressed in the current study.

5.1 The case of a single shoe model

Proposition 5.1 If \( S_{ij} = S_j, \forall i, j \), then

\[
\frac{\hat{\lambda}_j}{\hat{\lambda}_k} = \frac{n_{.j}}{n_{.k}} \cdot \frac{S_k}{S_j}
\]

for any \( k, j \) in all three estimators.

Proposition 5.1 refers to the simple case of all shoes having the same model, meaning that they have the same contact surface. Actually, it is true even if shoes have different models but have the same amount of contact surface in the different areas. The Proposition states that the three methods differ only when the contact surface differs, and it holds for any partition of the shoe sole. The proof is deferred to Supporting web materials 5.

5.2 Data analysis

The results of the three estimators (naive, random and CML) applied to large areas are presented in Figure 5. In order to calculate the estimator in the random effects case, the \textit{hglm} function under the \textit{hglm} package (Rönnegård, 2010) in R is used, where \( a_i \) are iid, with Gamma distribution having mean and variance: \( E(a_i) = 1, \text{Var}(a_i) = \sigma^2 \). The two other estimators were implemented using a self written code. All three estimates agree on the areas with high and low
intensity. The results are consistent with the previous analysis as the intensity function is the highest at the ball and heel of the foot and the differences between the maximum and minimal estimated intensity function is by a factor of 2.

In addition, 95% confidence intervals based on the three approaches were calculated. The interval of the naive estimator was calculated using the normal approximation with variance estimated as described in Section 4 of the Supporting web materials. For the random effects estimator, the \texttt{hglm} function under the \texttt{hglm} package (Rönnegård, 2010) was used to calculate the variance. The variance of the CML estimator is based on the observed information matrix. The confidence intervals are presented in Table 1. The number of the sub-area indicates the area marked in Figure 5. The confidence intervals of the three approaches are relatively close. The confidence interval of the CML approach is narrower as a result of the estimators’ lower variance due to conditioning on the $n_i$’s and treating the scaling factor as a constant (see the discussion in Remark 5). The estimators agree on the areas with relatively wide and narrow intervals. The widest interval is of area 3 of the shoe, which is characterized by a low amount of contact surface (see Figure 3 in Supporting web materials 2). In addition, using the random effects model, the hypothesis that the $\lambda_j$ parameters are equal for all $j$, meaning that
Table 1: Table of the confidence intervals for the estimators based on the piece-wise constant model. The number of the sub-area indicates the area marked in Figure 5.

| Sub area | Naive         | Random        | CML           |
|----------|---------------|---------------|---------------|
| 1        | (0.0020, 0.0026) | (0.0020, 0.0026) | (0.0021, 0.0024) |
| 2        | (0.0036, 0.0046) | (0.0036, 0.0045) | (0.0039, 0.0043) |
| 3        | (0.0033, 0.0047) | (0.0031, 0.0042) | (0.0034, 0.0040) |
| 4        | (0.0036, 0.0047) | (0.0035, 0.0046) | (0.0038, 0.0044) |
| 5        | (0.0028, 0.0035) | (0.0027, 0.0034) | (0.0029, 0.0033) |
| 6        | (0.0030, 0.0037) | (0.0028, 0.0035) | (0.0031, 0.0034) |
| 7        | (0.0026, 0.0038) | (0.0028, 0.0039) | (0.0030, 0.0037) |
| 8        | (0.0030, 0.0057) | (0.0029, 0.0043) | (0.0033, 0.0043) |
| 9        | (0.0038, 0.0048) | (0.0038, 0.0048) | (0.0042, 0.0046) |
| 10       | (0.0023, 0.0040) | (0.0022, 0.0029) | (0.0024, 0.0028) |
| 11       | (0.0036, 0.0046) | (0.0035, 0.0045) | (0.0039, 0.0044) |
| 12       | (0.0028, 0.0036) | (0.0027, 0.0034) | (0.0030, 0.0033) |
| 13       | (0.0035, 0.0045) | (0.0034, 0.0044) | (0.0038, 0.0042) |
| 14       | (0.0037, 0.0049) | (0.0035, 0.0047) | (0.0039, 0.0046) |

the intensity is uniform over the whole shoe, is rejected with a $p$-value $\approx 0$.

6 Simulation

6.1 Comparison of sub-sampling case-control techniques

Simulations are carried out to compare the random effects and CML estimators using different types of within-cluster case-control sub-sampling and sub-sampling across the whole data frame. Specifically, results based on the entire sample are compared to (i) random sub-sampling, (ii) case-control sub-sampling without taking into account the clusters, (iii) within-cluster case-control sub-sampling where controls are sampled from each cluster proportional to the cluster size and (iv) within-cluster case-control sub-sampling where controls are sampled from each cluster proportional to the number of cases in the cluster.

In all scenarios, all cases are sampled in the case-control sampling.

The full data includes 500 clusters with 500 observations in each cluster. The probability that a RAC appears is according to the logistic model:

$$P(N_{ij} = 1 | a_i) = \frac{e^{\beta_0 + \beta_1 x + \beta_2 x^2 + a_i}}{1 + e^{\beta_0 + \beta_1 x + \beta_2 x^2 + a_i}},$$

where $x$ represents a simple one-dimensional location and the parameters used are: $\beta_0 = -3$, $\beta_1 = 2$, $\beta_2 = -2$. The distribution of $a_i$ is $N(0, 0.75^2)$.

The comparison was made based on 300 replications for each sampling technique. Figures 6 and 7 show the bias and MSE of $\beta_1$ and $\beta_2$, where each sampling
technique is marked as written above. In addition, “r” indicates that the results are based on the random effects estimator and “c” indicates that they are based on the “CML” estimator.

Method 4 of within-cluster case-control sub-sampling where controls are sampled from each cluster proportional to the number of cases in the cluster, has a relatively large bias, but still performs best in terms of MSE, which is almost identical to the MSE of the results based on the entire sample. This is true using both the random effects and CML estimators for $\beta_1$ and $\beta_2$. In addition it can be seen that although differences exist in the bias of the random effects and CML estimators, they are relatively close in their MSE. The initial intention was to find the best sampling technique (method 4), and then repeat it
multiple times and averaging the resulting estimates in order to improve the performance. However, since the MSE using method 4 is very close to the one using the entire sample, the expected contribution of such a re-sampling method is not worth the effort.

6.2 Comparison of the three estimators

A second simulation study compares the three estimators based on parameters from the shoe data. As shown in Section 5 using this model, it is assumed that \( N_{ij} | a_i \sim \text{Poisson}(a_i \lambda_j S_{ij}) \). The first simulation uses the estimates for the \( \lambda \)'s obtained from the RACs’ database by applying the naive approach and the observed contact surfaces, \( S_{ij} \), in the \( J = 14 \) sub areas. The number of RACs in shoe \( i \), \( n_i \), is used as \( a_i \). In addition, these \( n_i \)'s were divided by their mean in order to apply a constraint of \( \bar{a} = 1 \). The results are based on 500 replications.

Figures 8 and 9 present the bias (in absolute value) and MSE of the estimators in the different sub areas relative to the naive estimator. The black line represents \( y = 1 \) which is the bias or MSE of the naive estimator divided by itself. The random effects estimator has the lowest mean absolute value bias and mean MSE (the mean is calculated over all the sub areas), and the naive estimator has the highest bias and MSE (see Table 1, row 1 in Supporting web materials 7).

The amount of contact surface in the different sub areas may affect the resulting estimators. See Supporting web materials 6 for histograms of the contact surface in the different sub areas.
Additional simulations that investigate the effect of \( \lambda \), the effect of the number of \( \lambda \)'s (the number of sub areas), the effect of the sample size, the number of shoes and the \( a_i \) are conducted. The specific settings and the results are given in Supporting web materials 7. In summary, the random effects estimator is found to be the best among the estimators in most settings, with good performance in all. The CML estimator is very close to it.

7 Discussion

This paper suggests three estimators for the intensity function of RAC locations, which facilitates the calculation of the probability that a RAC will appear in a certain place. Results based on real data were presented. The naive, random and CML estimators were compared in two analyses; the first using pixels as areas and achieving maximum resolution and the second using large areas and a piece-wise constant model that was meant to reduce the sensitivity to the challenges caused by the process of normalization.

In these two analyses, the CML and random effects estimators produced very similar results which were moreover relatively close to the naive estimator. The simulations show that the random effects estimator is found to
be the best among the three estimators and thus may be proffered. Although the hypothesis of a uniform intensity function is rejected, the maximum estimated intensity value is approximately twice that of the minimum value. In other words, it is not far from a uniform intensity function. This means that the probability of finding a RAC is relatively similar across the entire shoe sole. The estimated intensity function is highest at the ball and heel of the foot. It seems reasonable that the deviation from uniformity is a result of the morphology of the foot and the areas of the foot that cause pressure on the shoe. This assumption fits the shape of the estimated intensity function presented here.

The estimation of the intensity function has been made under the assumption of independence among RACs (using the logistic model and assuming independence among pixels or the Poisson process assumption which implies independence among RACs). However, as shown in Kaplan Damary et al. (2018), the assumption of independence among RACs is unjustified. It is well known that using an independence working assumption even when dependence exists results in consistent estimators. Thus, the estimators of the intensity function are still valid but their variance may not be.

In addition, the assumption that all shoes have the same shape of the intensity function and only differ by the shoe-specific parameter which determines the height of the function, presented in [1], may not be realistic. The reason is that people have different walking patterns which can affect their baseline intensity function. Using larger areas as building blocks for the model may solve part of this problem, but further study is needed to understand the importance of the assumption.

A few questions remain. As noted above, normalization of the shoe sole coordinates caused the location of a pixel, after the stretching of the shoe along the Y axis, to move from its original position, and it is not clear how the shoes should be compared. It is clear that the normalization affects the results presented here and other types of normalizations should be considered. Further investigation is needed in this respect.

Other characteristics of the RAC such as size and shape should be included in the model for the intensity function. This can be done by calculating a separate intensity function for each characteristic or by including these characteristics using a parametric model. In addition, the intensity function may be modelled by properties of the shoe such as the shoe type (sport vs. army), the manufacture, the material of the shoe etc.
Most importantly, the probability of a RAC at a given point on the shoe sole may have a serious effect on the evidential value of the match between the crime scene print and the suspect’s shoe. As mentioned above, it was found that in certain areas of the shoe the estimated intensity value is approximately twice that of other areas. This factor of 2 is not of great significance when finding only a single RAC, but may be much more important when finding multiple RACs, which is often the case.

The findings of this study take us a step forward in assessing the evidential value of shoeprint comparison. The use of prints taken off suspect shoes provides a best case scenario. Additional challenges confront the researcher when analyzing crime scene prints, which are complicated by noise of various forms. Further research on actual crime scene RACs is therefore of vital importance.

References

[1] A. Agresti. *Categorical data analysis*. John Wiley & Sons, Hoboken, N.J., third edition, 2013.

[2] Y. M. Bishop, S. E. Fienberg, and P. W. Holland. Estimating the size of a closed population. *Discrete Multivariate Analysis: Theory and Practice*, pages 229–256, 2007.

[3] W. J. Bodziak. *Footwear impression evidence: detection, recovery and examination*. CRC Press, 1999.

[4] A. Cohen, S. Wiesner, A. Grafit, and Y. Shor. A new method for casting three dimensional shoe prints and tire marks with dental stone. *Journal of forensic sciences*, 56(s1), 2011.

[5] N. Cressie. *Statistics for Spatial Data: Wiley Series in Probability and Statistics*. Wiley-Interscience New York, 1993.

[6] ENFSI Expert Working Group Marks Conclusion Scale Committee and others. Conclusion scale for shoeprint and toolmarks examination. *Journal of Forensic Identification*, 56(2):255–280, 2006.

[7] Executive Office of the President President’s Council of Advisors on Science and Technology. *Forensic Science in Criminal Courts: Ensuring
Scientific Validity of Feature-Comparison Methods. PCAST, Washington, D.C., 2016.

[8] W. Fithian and T. Hastie. Local case-control sampling: Efficient subsampling in imbalanced data sets. Annals of statistics, 42(5):1693, 2014.

[9] E. Gilleland. Two-dimensional kernel smoothing: Using the r package smoothie. 2013.

[10] B. Glattstein, Y. Shor, N. Levin, and A. Zeichner. Ph indicators as chemical reagents for the enhancement of footwear marks. Journal of Forensic Science, 41(1):23–26, 1996.

[11] T. Hastie, R. Tibshirani, and J. Friedman. The elements of statistical learning data mining, inference, and prediction. Springer, New York, N.Y., second edition, 2009.

[12] D. S. Hilderbrand. Footwear: the missed evidence. Staggs Pub., second edition, 2007.

[13] D. W. Hosmer, S. Lemeshow, and S. May. Applied survival analysis. John Wiley & Sons, Hoboken, N.J., second edition, 2008.

[14] N.H Kaplan-Damary. The Process of Documenting Shoe Prints, http://www.youtube.com/watch?v=sSLQ8G8qGE&feature=youtu.be, 2014.

[15] R. H. Myers, D. C. Montgomery, G. G. Vining, and T. J. Robinson. Generalized linear models: with applications in engineering and the sciences, volume 791. John Wiley & Sons, Hoboken, N.J., 2012.

[16] National Research Council. Strengthening forensic science in the United States: a path forward. The National Academies Press, Washington, D.C., 2009.

[17] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2014.

[18] N. Sartori and T. A. Severini. Conditional likelihood inference in generalized linear mixed models. Statistica Sinica, pages 349–360, 2004.
[19] Y. Shor, T. Tsach, A. Vinokurov, B. Glattstein, E. Landau, and N. Levin. Lifting shoeprints using gelatin lifters and a hydraulic press. *Journal of forensic sciences*, 48(2):368–372, 2003.

[20] Y. Shor, A. Vinokurov, and B. Glattstein. The use of an adhesive lifter and pH indicator for the removal and enhancement of shoeprints in dust. *Journal of Forensic Science*, 43(1):182–184, 1998.

[21] J. A. Speir, N. Richetelli, M. Fagert, M. Hite, and W. J. Bodziak. Quantifying randomly acquired characteristics on outsoles in terms of shape and position. *Forensic Science International*, 266:399–411, 2016.

[22] R. S. Stone. Footwear examinations: mathematical probabilities of theoretical individual characteristics. *Journal of Forensic Identification*, 56(4):577, 2006.

[23] SWGTREAD. Guide for the detection of footwear and tire impressions in the field. *J Forensic Ident*, 55(6):766–769, 2005a.

[24] SWGTREAD. Guide for the collection of footwear and tire impressions in the field. *J Forensic Ident*, 55(6):770–773, 2005b.

[25] SWGTREAD. Guide for the preparation of test impressions from footwear and tires. 2005c. Available at [http://www.swgtread.org/images/documents/standards/published/swgtread_05_test Impressions_200503.pdf](http://www.swgtread.org/images/documents/standards/published/swgtread_05_test Impressions_200503.pdf)

[26] SWGTREAD. Guide for the forensic documentation and photography of footwear and tire impressions at the crime scene. 2006a. Available at [http://www.swgtread.org/images/documents/standards/published/swgtread_09_documentation_and_photography_200603.pdf](http://www.swgtread.org/images/documents/standards/published/swgtread_09_documentation_and_photography_200603.pdf)

[27] SWGTREAD. Guide for the examination of footwear and tire impression evidence. 2006b. Available at [http://www.swgtread.org/images/documents/standards/published/swgtread_08_test_examination_200603.pdf](http://www.swgtread.org/images/documents/standards/published/swgtread_08_test_examination_200603.pdf)

[28] SWGTREAD. Guide for casting footwear and tire impression evidence. 2007a. Available at [http://www.swgtread.org/images/documents/standards/published/swgtread_11_CastingFootwearandTireImpressionEvidence200703.pdf](http://www.swgtread.org/images/documents/standards/published/swgtread_11_CastingFootwearandTireImpressionEvidence200703.pdf)
[29] SWGTREAD. Guide for lifting footwear and tire impression evidence. 2007b. Available at http://www.swgtread.org/images/documents/standards/published/swgtread_12_LiftingFootwearandTireImpressionEvidence200703.pdf.

[30] SWGTREAD. Guide for casework documentation. 2008. Available at http://www.swgtread.org/images/documents/standards/published/swgtread_13_CaseworkDocumentation200803.pdf.

[31] T.M. Therneau. A Package for Survival Analysis in S, 2015. version 2.38.

[32] S.T. Wright, L.M. Ryan, and T. Pham. A novel case-control subsampling approach for rapid model exploration of large clustered binary data. Statistics in medicine, 2017.

[33] Y. Yekutieli, Y. Shor, S. Wiesner, and T. Tsach. Expert Assisting Computerized System for Evaluating the Degree of Certainty in 2D Shoeprints. Final report, NIJ Task plan 3211, 2016. Available at https://www.ncjrs.gov/pdffiles1/ijj/grants/250336.pdf.
Supporting Web Material

1 The process of evaluating shoeprints

Here the process of documenting and evaluating shoeprints used in the Israel National Police Division of Identification and Forensic Science (DIFS) is described.

1. CSIs (Crime Scene Investigators) who arrive at the crime scene search for locations where the perpetrators have likely left their shoe prints, preferably an area left untouched by other inhabitants. Visible shoe prints are identified and the area is then darkened while investigators search for shoe prints using oblique light (SWGTREAD,2005a).

2. After locating a shoeprint, the CSI places an L-shaped photography scale next to it, illuminates it at an angle that reveals as much detail as possible and photographs it with the camera positioned on a tripod directly over the print (SWGTREAD,2006).

3. If possible, the surface on which the print appears is wrapped in paper and sent to the crime lab for further processing (SWGTREAD,2005b).

4. If this is not possible, prints are lifted using one of the following methods:
   a. Two dimensional print methods (SWGTREAD,2007b and Manual for BVDA lifters); white adhesive lifter (most common method used in this lab for 2D shoeprints), black gelatin lifter and electrostatic lifter.
   b. Three dimensional Prints are casted using dental stone (SWGTREAD,2007a and Cohen et al., 2011).

5. Once a suspect is apprehended, the suspect’s shoes are sent to the crime lab for comparison with the shoeprints. Usually the shoes will be removed from the suspect’s feet, but occasionally, the suspect’s house will be searched for shoes with soles that resemble the shoeprints from the crime scene.

6. All exhibits are registered by the investigating unit and are then sent to the crime lab through the “evidence office” which is responsible for assigning file numbers and passing on the exhibits to the relevant lab.
The investigators add a letter that contains details about the crime and the collection of the relevant evidence.

7. All the exhibits and their packaging are documented and marked at the crime lab. The examiner documents all of the information concerning the exhibits as they were received by the lab (date of reception, description of packaging and exhibit etc.) (SWGTREAD, 2008). This is known as the “chain of custody.” Keeping the chain of custody is important to ensure that the exhibits examined by the lab are the same exhibits confiscated in this case.

8. All the shoeprints (exhibits, white adhesive lifters, black gelatin lifters and casts) and the shoes are photographed at high resolution (1000 dpi). The photographs taken at the crime scene are calibrated to 1000 dpi as well. Minimal image processing is carried out if necessary (conversion to gray scale, contrast and brightness adjustment etc.).

9. If applicable, the shoeprints collected at the crime scene go through a process of enhancement. White adhesive lifters are sprayed with Bromophenol Blue which has a yellow color that turns blue when reacting with the dust, thus creating the shoeprint (Glattstein et al., 1996 and Shor et al., 1998). Shoeprints on items collected at the crime scene are lifted using the most appropriate method, black gelatin with a press (Shor et al., 2003) or white adhesive lifter.

10. If the patterns are similar, as explained in the next paragraph, two dimensional lab prints are made from the suspect’s shoes. The shoe soles are dusted with fingerprint powder and then impressions are made, while wearing the shoe, by stepping onto a clear adhesive film. At least two lab prints are made from each shoe (Hilderbrand, 2007), first by walking and again by pressing the adhesive to the shoe sole while it is in the air. If necessary, for the comparison stage additional three dimension lab prints are made using Biofoam © (SWGTREAD, 2005c).

11. Comparison and evaluation Stages (SWGTREAD, 2006b): The first step is visual examination of the crime scene prints and the shoe soles.

   a. Pattern - If the exact shoe pattern does not match, the result of the comparison is exclusion.
b. Size - If the shoe pattern matches, but the physical size of shoeprint differs from the corresponding area of the shoe sole this might be explained by incorrect scaling, photography at an angle, movement of the shoe while producing the shoeprint etc. If no explanation satisfies the examiner as a logical explanation for the dissimilarities, exclusion is determined.

c. Wear - If exclusion is not determined by now, the degree of general wear and local wear are compared. If the wear areas differ or crime scene print is worn more than the shoe, exclusion is determined. If the shoe is worn more than the crime scene print, the time elapsed between creation of the shoeprint and confiscation of the shoes is considered.

d. The crime scene print is searched for locations where the pattern isn’t complete and the reason might be the presence of RACs. If they indeed appear on both lab and crime scene prints, they are marked and their clarity, complexity and rarity are evaluated based on the examiner’s experience.

12. If the patterns are similar, lab prints are made from the suspect’s shoes as described above and these are compared to the shoeprints found at the crime scene. The two common comparison methods are:

i. Overlay - a transparency of the suspect’s shoe lab prints is positioned over a photograph of the print from the crime scene.

ii. Side by Side - the lab prints and the photograph of the print from the crime scene are laid out side by side in order to compare the similarities between them. The method used in the lab is on screen overlay (using Lucia TrasoScan, by LIM ©) with 1000 ppi (pixel per inch) images. The examination process includes comparison of class characteristics (pattern, size and wear) and the identification of accidentals (RACs and unique wear)

13. The examiner determines the level of certainty and writes a report. An ordinal scale is used to describe the degree of the match (ENFSI, 2006). The scale used in the Israel National Police Division of Identification and Forensic Science (DIFS) is presented next.

The conclusion scale:
a. Negative - the shoe under investigation is significantly different from the crime scene print, and therefore the suspect’s shoe couldn’t have left the crime scene print.

b. Indication of non-association - Differences were found during the comparison but the quality of the print or the essences of the differences are not sufficient for total exclusion.

c. Lacks sufficient details - The crime scene shoeprint lacks information that would enable significant comparison.

d. Cannot be eliminated - Similarities in class characteristics were found, but the details on the crime scene print were limited and therefore, specific association is not possible.

e. Possible - there is a match regarding class characteristics between the crime scene print and the suspect shoe.

f. Probable - Matching class and identifying characteristics were both found, but the amount of information in the identifying characteristics is not sufficient to determine Identification.

g. Highly probable – Similar to Probable, but a higher degree of certainty.

h. Identification - Besides the match in class characteristics, the match in identifying characteristics is of sufficient quality and quantity.

Based on the conclusion distribution of the lab, in approximately 40% of all cases the suspects’ shoes sent to the lab result in a non-match to the crime scene prints. A match of class characteristics (“Possible”) is found in approximately a third of the cases and accidentals are found in nearly 25% of all cases. Less than 2% of the crime scene prints lack sufficient details for comparison.

2 Descriptive statistics
Figure 1: A Histogram of the number of RACs per shoe

Figure 2: A Histogram of the number of pixels with contact surface per shoe. An image of a shoe contains $395 \times 307 = 121,265$ pixels.

3 Representation of the model parameters

Figure 5 presents the notation. $\lambda_j$ is the intensity within pixel $j$, $a_i$ is the wear and tear parameter of shoe $i$, $S_{ij}$ is the area of the contact surface of shoe $i$ and pixel $j$ and $n_{ij}$ is the observed number of RACs on shoe $i$ and pixel $j$. 
4 Variance of estimators under the naive approach

Under the naive approach, $\hat{\lambda}_j = \frac{\sum_i n_{ij}}{\sum_i S_{ij}}$. Since $N_j|a_i \sim \text{Poisson}(\sum_i \lambda_j a_i S_{ij})$, where $\sum N_j = \sum_i N_{ij}$ and it is assumed that
Shoe $i$, $\lambda_j = ?$

$n_{ir} = 1$

$S_{ik} = 0$

$S_{i\ell} = 0.5$

Figure 5: Representation of the model parameters on the lab print

$E(a_i) = 1$, the variance of the estimator is

$$\text{Var} \left( \frac{1}{|m_j|} \sum_{i \in m_j} \frac{N_{ij}}{S_{ij}} \right) = \frac{1}{|m_j|^2} \sum_{i \in m_j} \text{Var}(N_{ij}) = \frac{1}{|m_j|^2} \sum_{i \in m_j} \lambda_j S_{ij}^2 \text{Var}(a_i) + \lambda_j \sum_{i \in m_j} \frac{1}{S_{ij}},$$

which under maximal resolution ($S_{ij} = 0, 1$) reduces to

$$\text{Var}(\hat{\lambda}_j) = \frac{\lambda_j^2 \text{Var}(a_i) + \lambda_j}{|m_j|}.$$  \hspace{1cm} (9)

Let $U_i = \frac{N_i^2 - N_i}{(\sum_j \lambda_j S_{ij})^2}$, then simple calculations show

$$E(U_i) = \frac{E \left( E^2(N_i|a_i) + \text{Var}(N_i|a_i) - E(N_i|a_i) \right)}{(\sum_j \lambda_j S_{ij})^2} = \frac{E \left( (a_i \sum_j \lambda_j S_{ij})^2 \right)}{(\sum_j \lambda_j S_{ij})^2} = \text{Var}(a_i) + E^2(a_i) = \text{Var}(a_i) + 1.$$

Thus, by defining $\hat{U}_i = \frac{N_i^2 - N_i}{(\sum_j \lambda_j S_{ij})^2}$ with $\hat{\lambda}_j$ being the naive estimator, $\text{Var}(\hat{\lambda}_j)$ is readily estimated by plugging $\hat{\lambda}_j$ and $\text{Var}(a) = m^{-1} \sum_i \hat{U}_i - 1$ in (9).

32
5 The case of a single shoe model

Proposition 5.1 If $S_{ij} = S_j, \forall i, j$, then

$$\frac{\hat{\lambda}_j}{\hat{\lambda}_k} = \frac{n_{j\cdot}}{n_{k\cdot}} \cdot \frac{S_k}{S_j}$$

for any $k, j$ in all three estimators.

Proof The naive estimator presented in Section 4.1 reduces to:

$$\hat{\lambda}_j = \frac{1}{m} \sum_i n_{ij} S_j = \frac{n_{j\cdot}}{mS_j}$$

Note that in this case where the shoes are of the same shoe model, $m_j = m$.

As mentioned in Section 5, the CML estimator solves equation (9) and in this case solves:

$$\sum_{i=1}^{m} \left[ \frac{n_{ik}}{\hat{\lambda}_k} - \frac{n_i}{\sum_{j'} S_{j'} \hat{\lambda}_{j'}} S_k \right] = 0.$$  

It follows that

$$\hat{\lambda}_k = \frac{n_{k\cdot}}{mS_k} \cdot \frac{m \sum_{j'} S_{j'} \hat{\lambda}_{j'}}{n_{..}}$$

where $n_{..} = \sum_{i=1}^{m} \sum_{j=1}^{J} n_{ij}$. The right fraction does not depend on $k$ and thus the ratio $\frac{\hat{\lambda}_j}{\hat{\lambda}_k}$ is equal to the ratio using the naive estimator.

Using equation (7) in Section 5 the random effects log likelihood is proportional to the following equation (the $n_{ij}$! is eliminated in the denominator)

$$\ell(\lambda_1, \ldots, \lambda_J, \theta; n_{ij}) = \sum_{i=1}^{m} \log \int e^{-a \sum_{j=1}^{J} \lambda_j S_j \cdot a \sum_{j=1}^{J} n_{ij} \Pi_{j=1}^{J} (S_j \lambda_j)^{n_{ij}} h_\theta(a) da}$$
and the derivatives are

\[
\frac{\partial \ell}{\partial \lambda_k} = \sum_{i=1}^{m} \frac{\partial}{\partial \lambda_k} \left[ \int e^{-a \sum_{j=1}^{j'} S_j \lambda_j} \cdot a \sum_{j'=1}^{j'} n_{ij} \prod_{j=1}^{j} (S_j \lambda_j)^{n_{ij}} h_\theta(a) da \right] \\
= -\sum_{i=1}^{m} \int a \cdot S_k \cdot e^{-a \sum_{j=1}^{j'} S_j \lambda_j} \cdot a \sum_{j'=1}^{j'} n_{ij} \prod_{j=1}^{j} (S_j \lambda_j)^{n_{ij}} h_\theta(a) da \\
+ \sum_{i=1}^{m} \frac{n_{ik}}{\lambda_k} \int e^{-a \sum_{j=1}^{j'} S_j \lambda_j} \cdot a \sum_{j'=1}^{j'} n_{ij} \prod_{j=1}^{j} (S_j \lambda_j)^{n_{ij}} h_\theta(a) da \\
= -\sum_{i=1}^{m} \frac{E(a \cdot S_i \cdot f_i(a))}{E(f_i(a))} + \sum_{i=1}^{m} \frac{n_{ik}}{\lambda_k} \\
= -S_k \sum_{i=1}^{m} \frac{E(a \cdot f_i(a))}{E(f_i(a))} + \sum_{i=1}^{m} \frac{n_{ik}}{\lambda_k}
\]

(10)

where \( f_i(a) = e^{-a \sum_{j=1}^{j'} S_j \lambda_j} \cdot a \sum_{j'=1}^{j'} n_{ij} \prod_{j=1}^{j} (S_j \lambda_j)^{n_{ij}} \).

Note that \( \frac{E(a \cdot f_i(a))}{E(f_i(a))} \) does not depend on \( k \), i.e. it is equal for all \( \lambda_k \) for any \( h_\theta(a) \). The random effects estimator is the solution of equation (10) to 0.

Denote the constant \( \frac{E(a \cdot f_i(a))}{E(f_i(a))} \) by \( c_i \) and \( \sum_{i=1}^{m} n_{ik} \) by \( n_k \). Then, the estimator solves the following equation:

\[-S_k \sum_{i=1}^{m} c_i + \frac{n_k}{\lambda_k} = 0,\]

and thus, the random effects estimator is equal to:

\[\hat{\lambda}_k = \frac{n_k}{S_k \sum_{i=1}^{m} c_i}.\]

(11)

Thus, the ratio \( \frac{\lambda_j}{\lambda_k} \) is equal for any \( k, j \) in all three estimators.

If \( h_\theta(a) \) satisfies \( c_i = \frac{E(a \cdot f_i(a))}{E(f_i(a))} = m \), all three estimators are identical. An example of this is the case of \( h_\gamma(a) = \Gamma(a, \gamma) \) where \( E(a) = 1 \) and \( \text{Var}(a) = \frac{1}{\gamma} \). Here,

\[c_i = \frac{E(a \cdot f_i(a))}{E(f_i(a))} = \frac{n_i + \gamma}{\sum_{j=1}^{j} \lambda_j + \gamma}.
\]

(12)
Summing Equation (11) and placing $c_i$ of Equation (12) results in:

$$\sum_{j=1}^{J} \hat{\lambda}_j = \frac{n_i + m\gamma}{\sum_{j=1}^{J} \hat{\lambda}_j + \gamma}.$$ 

Thus, in this case:

$$\sum_{j=1}^{J} \hat{\lambda}_j = \frac{n_i}{m}, \quad (13)$$

Now, summing Equation (12) and placing Equation (13) inside results in $\sum_{i=1}^{m} c_i = m$ and thus $\hat{\lambda}_j = \frac{n_j}{m}$ and all three estimators are identical.
6 Histograms of contact surface in 14 areas

Figure 6: The contact surface in sub areas 1 to 14
7 Additional comparisons of the piece-wise constant estimators

Here the three estimators using the piece-wise constant intensity function are compared through simulations that test the effect of \( \lambda \), the effect of the number of sub areas, the effect of the sample size, the number of shoes and the \( a_i \). The first simulation (presented in the article) uses the estimates for the \( \lambda \)'s obtained from the RACs' database by applying the naive approach. Instead of sampling contact surfaces, \( S_{ij} \) in \( J = 14 \) sub areas, the observed ones included in the data set are used. The number of RACs in shoe \( i \), \( n_i \), is used as \( a_i \). In addition, these \( n_i \)'s were divided by their mean in order to apply a constraint of \( \bar{a} = 1 \). The results are based on 500 replications. Figures 7-28 present the bias (in absolute value) and MSE of the estimators in the different sub areas relative to the naive estimator. The black line represents \( y = 1 \) which is the bias or MSE of the naive estimator divided by itself. Table 2 summarizes the mean MSE, \( \overline{MSE} = J^{-1} \sum_{j=1}^{J} \{ (\hat{\lambda}_j - \lambda_j)^2 + \text{Var}(\hat{\lambda}_j) \} \), and the bias absolute value over all estimates, \( \overline{\text{bias}} = J^{-1} \sum_{j=1}^{J} |\hat{\lambda}_j - \lambda_j| \).

Specifically, the following scenarios are compared. In each scenario the setting is as presented in the original simulation except for a change of one parameter presented as follows. The corresponding row of Table 2 and the specific figures are indicated.

1. Using \( \lambda \)'s of equal value, \( \lambda_j = 32, j = 1, \ldots, 14 \) (row 2 and Figures 7 and 8).
2. Using \( \lambda \)'s, half of which are relatively small - equal to 16 and half of which are relatively large - equal to 48 (row 3 and Figures 9 and 10).
3. Using relatively small \( \lambda \)'s - equal 2.5 except for one relatively large \( \lambda \) - equals 416 (row 4 and Figures 11 and 12).
4. Using a large number of \( \lambda \)'s (36 instead of 14). This means that the shoe is divided to more sub areas. The \( \lambda \)'s used here are the naive estimators calculated from the RACs data base (row 5 and Figures 13 and 14).
5. Using 500 shoes instead of 386 as in the original setting. These are sampled with replacement from the database of 386 shoes. The \( a_i \)'s are sampled as well, from the \( n_i \) vector of the number of RACs in the shoe. The \( a_i \) and the contact surface are sampled independently, in contrast to
the original case where each contact surface goes along with its corresponding \( n_i \) (row 6 and Figures 15 and 16).

6. As in the previous scenario, but using 1000 shoes (row 7 and Figures 17 and 18).

7. As in the previous scenario, but using 386 shoes. This is conducted to compare it to the last two scenarios (row 8 and Figures 19 and 20).

8. Using \( a_i \sim \text{Gamma}(1/3, 3) \) (row 9 and Figures 21 and 22).

9. Using equal \( a_i, a_i = 1 \) \( i = 1, \ldots, m \) (row 10 and Figures 23 and 24).

10. Using \( a_i \sim U(0, 2) \) (row 11 and Figures 25 and 26).

11. Using \( a_i \) from a Bernoulli distribution which is shifted by 0.5, i.e. \( a_i \) are equal to 0.5 with probability of 0.5 and equal to 1.5 otherwise (row 12 and Figures 27 and 28).

Cases 1-3 investigate the use of different sets of \( \lambda \)'s. In all cases the random effects estimator has the lowest mean bias absolute value and mean MSE (though relatively comparable to the CMLs’ MSE), and the naive estimator has the highest bias and MSE, except for case 2 (half of the \( \lambda \)'s are relatively small and half are relatively large), presented in row 3 of Table 2 where the CML has the highest bias.

Case 4 investigates the effect of the number of \( \lambda \)'s (meaning the number of sub areas used) presented in row 5 of Table 2. The random effects estimator has the lowest mean bias absolute value and mean MSE, the naive estimator has the highest bias and the CML has the highest mean bias absolute value. In addition the MSE of all estimators is higher than the one obtained using 14 sub areas.

Cases 5-7 presented in rows 6-8 of Table 2 investigate the effect of the number of shoes. The bias and the MSE of the naive, CML and random effects estimators are compared using 500 and 1000 shoes (contact surfaces) sampled with replacement from the database of 386 shoes (rows 6 and 7 respectively). As noted, the \( a_i \)'s are sampled with replacement as well, from the \( n_i \) vector of the number of RACs in a shoe. In these analyses, the \( a_i \) and the contact surfaces are sampled independently in contrast to the original case where each contact surface goes along with its matching \( n_i \). Thus, they should be compared to the analysis where the original 386 shoes are used together with
sampled 386 $a_i$ with replacement from the $n_i$ vector (row 8). As expected, the MSE of all estimators decreases as the number of shoes increases. In the 500 shoe analysis, the random effects estimator has the lowest mean bias absolute value and mean MSE. The naive estimator has the highest MSE and the CML has the highest bias. In the 1000 shoe analysis, the random effects estimator and the CML have the lowest mean bias absolute value.

Cases 8-11 presented in rows 9-12 of Table 2 investigate the effect of the $a_i$’s. In all cases the random effects estimator has the lowest mean MSE (though relatively close to the CML), and the naive estimator has the highest MSE. The random effects estimator is found to be the best among the estimators in most settings, with good performance in all. The CML estimator is very close to it.

| Sim | mean MSE | mean | | bias | | bias | |
|-----|----------|------|-----|------|-----|-----|-----|
| 1   | 1.54     | 3.00 | 1.66 | 0.06 | 0.08 | 0.08 | 0.062 |
| 2   | 1.58     | 3.00 | 1.72 | 0.05 | 0.09 | 0.07 | 0.07 |
| 3   | 1.57     | 3.83 | 1.93 | 0.06 | 0.07 | 0.08 | 0.08 |
| 4   | 2.60     | 2.79 | 2.69 | 0.01 | 0.04 | 0.03 | 0.03 |
| 5   | 4.54     | 11.08| 5.04 | 0.09 | 0.17 | 0.26 | 0.26 |
| 6   | 1.20     | 3.33 | 1.39 | 0.05 | 0.07 | 0.10 | 0.10 |
| 7   | 0.59     | 1.66 | 0.65 | 0.02 | 0.03 | 0.02 | 0.02 |
| 8   | 1.59     | 4.58 | 1.96 | 0.03 | 0.06 | 0.07 | 0.07 |
| 9   | 4.07     | 6.31 | 4.15 | 1.59 | 1.62 | 1.56 | 1.56 |
| 10  | 1.56     | 4.18 | 1.82 | 0.04 | 0.06 | 0.08 | 0.08 |
| 11  | 2.37     | 5.08 | 2.83 | 0.86 | 0.91 | 0.97 | 0.97 |
| 12  | 1.98     | 4.91 | 2.44 | 0.57 | 0.64 | 0.70 | 0.70 |

Table 2: The mean MSE and the mean bias absolute value of the estimators.
Equal $\lambda$'s

The bias of the estimators

{Figure 7: The bias of the estimators}

The MSE of the estimators

{Figure 8: The MSE of the estimators}

$\lambda$ - small and large

The bias of the estimators

{Figure 9: The bias of the estimators}

The MSE of the estimators

{Figure 10: The MSE of the estimators}
One large $\lambda$

The bias of the estimators

| Bias | CML/Naive | Random/Naive |
|------|-----------|--------------|
| 0    |           |              |
| 1    |           |              |
| 2    |           |              |
| 3    |           |              |
| 4    |           |              |

The MSE of the estimators

| MSE  | CML/Naive | Random/Naive |
|------|-----------|--------------|
| 0.0  |           |              |
| 0.2  |           |              |
| 0.4  |           |              |
| 0.6  |           |              |
| 0.8  |           |              |
| 1.0  |           |              |

Figure 11: The bias of the estimators

Figure 12: The MSE of the estimators

Large number of $\lambda$'s

The bias of the estimators

The MSE of the estimators

Figure 13: The bias of the estimators

Figure 14: The MSE of the estimators

500 shoes
The bias of the estimators

Lambda values

Bias

CML/Naive
Random/Naive

Figure 15: The bias of the estimators

The MSE of the estimators

Lambda values

MSE

CML/Naive
Random/Naive

Figure 16: The MSE of the estimators

1000 shoes

The bias of the estimators

Lambda values

Bias

CML/Naive
Random/Naive

Figure 17: The bias of the estimators

The MSE of the estimators

Lambda values

MSE

CML/Naive
Random/Naive

Figure 18: The MSE of the estimators

386 shoes with sampled $a_i$
Figure 19: The bias of the estimators
Figure 20: The MSE of the estimators

Using $a_i$ sampled from gamma distribution

Figure 21: The bias of the estimators
Figure 22: The MSE of the estimators
Equal $a_i$

Figure 23: The bias of the estimators

Figure 24: The MSE of the estimators

Uniform $a_i$

Figure 25: The bias of the estimators

Figure 26: The MSE of the estimators
Shifted Bernoulli $a_i$

The bias of the estimators

The MSE of the estimators

Figure 27: The bias of the estimators

Figure 28: The MSE of the estimators