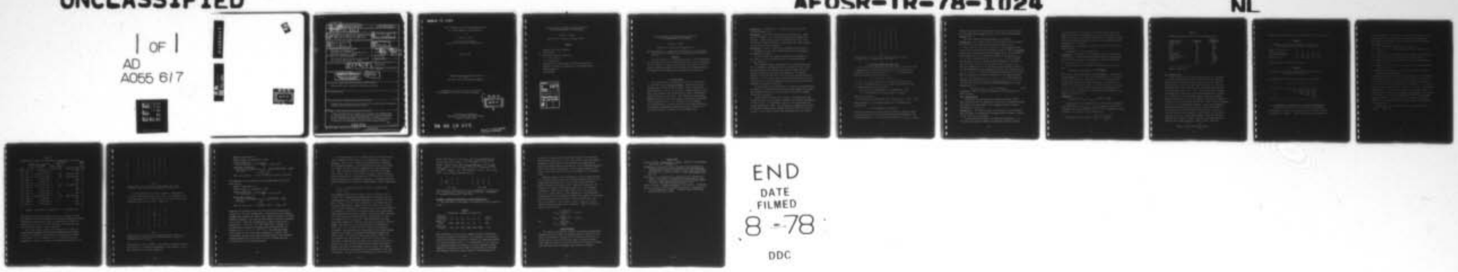
AD-A055 617 ILLINOIS UNIV AT CHICAGO CIRCLE DEPT OF MATHEMATICS F/G 5/4
THE OCCURRENCE OF FINGERPRINT CHARACTERISTICS AS A TWO-DIMENSIO--ETC(U)
MAY 78 S L SCLOVE AFOSR-77-3454
UNCLASSIFIED AFOSR-TR-78-1024 NL

1 OF 1
AD
A055 617



END
DATE
FILMED
8-78
DDC

②

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR-78-1024	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) THE OCCURRENCE OF FINGERPRINT CHARACTERISTICS AS A TWO-DIMENSIONAL POISSON PROCESS		5. TYPE OF REPORT & PERIOD COVERED Interim rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Stanley L. Sclove		8. CONTRACT OR GRANT NUMBER(s) AFOSR-77-3454
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Mathematics University of Illinois at Chicago Circle Box 4348, Chicago, IL 60680		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5 A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, Washington, DC 20332		12. REPORT DATE May 15, 1978
		13. NUMBER OF PAGES 16
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) 15 May 78		15. SECURITY CLASS (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

19p.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Approved for public release; distribution unlimited

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Fingerprints; Identification; Criminalistics; Two-dimensional stochastic process; Markov process; Poisson process.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The individuality of a fingerprint is based on the configuration of occurrences of the ten Galton characteristics. The existing model for the occurrence of these characteristics, in terms of a grid of cells, is further developed. The occurrence of the characteristics is modelled as a two-dimensional multivariate Poisson process.

409 950

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AFOSR-TR- 78-1024

THE OCCURRENCE OF FINGERPRINT CHARACTERISTICS AS A TWO-DIMENSIONAL POISSON PROCESS

by

Stanley L. Sclove

Department of Mathematics
University of Illinois at Chicago Circle

May 15, 1978

PREPARED UNDER GRANT AFOSR 77-3454
FOR
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS AT CHICAGO CIRCLE
CHICAGO, ILLINOIS 60680

THE OCCURRENCE OF FINGERPRINT CHARACTERISTICS AS A TWO-DIMENSIONAL POISSON PROCESS

Stanley L. Sclove

University of Illinois at Chicago Circle

OUTLINE

THE OCCURRENCE OF FINGERPRINT CHARACTERISTICS AS A TWO-DIMENSIONAL POISSON PROCESS

Stanley L. Sclove

University of Illinois at Chicago Circle

Key Words & Phrases: fingerprints; identification; criminalistics; two-dimensional stochastic process; Markov process; Poisson process.

ABSTRACT

The individuality of a fingerprint is based on the configuration of occurrences of the ten Galton characteristics. The existing model for the occurrence of these characteristics, in terms of a grid of cells, is further developed. The occurrence of the characteristics is modelled as a two-dimensional multivariate Poisson process.

1. INTROUDCTION

The individuality of a fingerprint is based on the pattern of occurrence of the ridge-line details. These details, first systematically studied by Galton (1892), are now called Galton characteristics. They are of ten types: islands, bridges, spurs, dots, ridge endings, forks (bifurcations), lakes, trifurcations, double bifurcations, and deltas. [See Osterburg, Parthasarathy, Raghavan and Sclove (1977) for diagrams and detailed descriptions of the characteristics.] It is desired to develop formulas for the probability of partial prints, such as crime-scene prints. In Osterburg, Parthasarathy, Raghavan and Sclove (1977) such a development was made, according to the following model.

Assumption 1. A fingerprint is considered in terms of a grid of one millimeter cells.

Assumption 2. For each cell there are 13 possibilities: either the cell is empty, or one of the following 12 possibilities has occurred: island, bridge, spur, dot, ending ridge, fork, lake, trifurcation, double bifurcation, broken ridge (two ridge endings), or some other multiple occurrence.

Assumption 3. There is statistical independence between cells.

Assumptions 2 and 3 bear further study, with a view toward developing better models.

In Sclove (1978) the extent of departure from Assumption 3 was studied, and the multinomial model was refined according to that departure.

The categories defined in Assumption 2 are somewhat arbitrary. The ten categories corresponding to the occurrence of each of the ten characteristics as singletons are natural enough; it is the special treatment accorded two ridge endings and the lumping together of all other multiple occurrences which warrant alternative treatment. In the present paper a method is presented which not only takes account of inter-cell dependence but also provides a different treatment of multiple occurrences.

In the present paper the occurrence of fingerprint characteristics is modelled as a two-dimensional Poisson process, taking into account the dependence between cells and providing alternative treatment of multiple occurrences.

A configuration such as that of Figure 1 represents the result of placing a grid of one-millimeter squares over a partial fingerprint. It has 43 cells, 37 of them empty, the other six being occupied by 4 ridge endings and 2 forks. Its probability, estimated by the method of Osterburg, Parthasarathy, Raghavan and Sclove (1977) is given by $-\log_{10}\hat{P} = 11.4$, where P denotes the probability and $\hat{P}$ its estimate.

	1	2	3	4	5	6
a	0	0	0	0	0	0
b	E	0	0	0	E	0
c	0	0	F	0	0	0
d	0	0	0	0	0	0
e	0	0	0	E	0	0
f	0	0	0	0	F	0
g	0	E	0	0	0	0
h			0			

FIG. 1

Configuration of 43 cells with 4 ending ridges and 2 forks.
0 = empty cell, E = ending ridge, F = fork.

2. MODELLING DEPENDENCE AMONG CELLS

Let the cells be numbered in some fixed order, say, as one reads a language such as English, starting with the top row and moving from left to right within each row. Let Y_c be a random vector giving the outcome in the c-th cell,

$$Y_c = (Y_{1c}, Y_{2c}, \ldots, Y_{10,c}),$$

$c = 1,2,\ldots,t$ = total number of cells in the print, where, for $v = 1,2,\ldots,10$ characteristics, Y_{vc} = number of occurrences of the v-th characteristic in the c-th cell.

Let P denote the probability of a configuration. Then

$$P = P(\text{configuration}) = P(Y_1=y_1, Y_2=y_2, \ldots, Y_t=y_t).$$

This can be written as

$$P = P(Y_1=y_1)P(Y_2=y_2|Y_1)P(Y_3=y_3|Y_2,Y_1)\cdots P(Y_t=y_t|Y_{t-1},\ldots,Y_1). \quad (2.1)$$

Under Assumption 3, independence among cells, expression (2.1) simplifies to

$$P = P(Y_1=y_1)P(Y_2=y_2)\cdots P(Y_t=y_t).$$

Under this assumption, all that would be required to obtain a model to give P would be to describe the within-cell distributions of the ten characteristics, i.e., the joint distribution of $Y_{1c}, Y_{2c}, \ldots, Y_{10,c}$.

However, any departure from Assumption 3 means that we need to model not only what happens within cells but also to model the dependence between cells.

As a step toward modelling dependence, we introduce

Assumption 3'. The outcome in the c-th cell depends on the outcomes in the other cells only through the outcomes in adjacent cells.

Due to the fact that (2.1) forces one to use a linear ordering of the cells, one must develop the model in terms of the four preceding adjacent cells rather than all eight adjacent cells. More precisely, under Assumption 3', the conditional probability $P(Y_c=y_c|Y_{c-1},Y_{c-2},\ldots,Y_1)$ will not depend upon all of $Y_{c-1},Y_{c-2},\ldots,Y_1$ but only upon four of these variables, namely, those corresponding to the cell to the left (west) of cell c, the cell above (north of) cell c, the cell just northwest of cell c, and the cell just northeast of cell c. [If the configuration were rectangular and cells were indexed as (i,j), $i = 1,\ldots,I$, $j=1,\ldots,J$, then the conditional distribution of the variable $Y_{i,j}$, given the variables preceding it in the ordering, namely $Y_{i,j-1},\ldots,Y_{i,1},Y_{i-1,J},Y_{i-1,J-1},\ldots,Y_{i-1,1},\ldots,Y_{11}$, would depend only upon the four variables $Y_{i,j-1},Y_{i-1,j+1},Y_{i-1,j},Y_{i-1,j-1}$.] If W_c is a matrix whose columns are the four preceding neighbors of Y_c, then Assumption 3' is

$$P(Y_c=y_c|Y_{c-1},Y_{c-2},\ldots,Y_1) = P(Y_c=y_c|W_c) \ . \qquad (2.2)$$

Assumption 3' may be viewed as an assumption that the process is a Markov process.

2.1. Data Analysis

A between-cells data analysis is discussed in Sclove (1978). The results were that the probability that a cell is occupied increases monotonically with the number of its neighbors that are occupied. Accordingly, we introduce

Assumption 4'. The expected number of occurrences in a cell depends upon the number of neighboring cells depends upon the number of neighboring cells that are occupied.

A within-cells data analysis is discussed in Appendix A. It was found that negative binomial distributions provided a good fit to the distribution of the number of characteristics, and to the numbers of different characteristics. This is consistent with a model of a mixture of Poisson distributions, for a negative binomial distribution can be obtained as a mixture of Poisson distributions. Accordingly, we shall make

Assumption 4''. The number of occurrences in a cell is a Poisson random variable.

Assumptions 4' and 4'' combine into

Assumption 4. The number of occurrences in a cell is distributed according to a Poisson distribution with parameter λ which depends upon the random variable a, the number of adjacent cells which are occupied: $\lambda = \lambda(a)$, $a = 0,1,2,3$, or 4.

Thus, letting N_c be the number of occurrences in cell c, we have

$$P(N_c=n|W_c) = P(N_c=n|A_c=a) = e^{-\lambda(a)}\frac{[\lambda(a)]^n}{n!} .$$

Now, this is the distribution of the total number of occurrences of all ten types in a cell. The relative frequencies p_v of the types, $v = 1,2,\ldots,10$ types, are given in Table I. It is reasonable to assume that the expected value of the number of occurrences Y_{vc} of the v-th characteristic in the c-th cell, given $A_c=a$, is $\lambda(a)p_v$, $v = 1,2,\ldots,10$. We shall assume that the joint distribution of $Y_{1c},Y_{2c},\ldots,Y_{10,c}$, given $N_c=n$, is multinomial with parameters n and $p_1,p_2,\ldots,p_{10}$. This gives

Assumption 5.

$$P(Y_c=y_c|N_c=n) = \binom{n}{y_{1c},\ldots,y_{10,c}} p_1^{y_{1c}} p_2^{y_{2c}} \cdots p_{10}^{y_{10,c}} .$$

Assumptions 4 and 5 combined give the result that the marginal joint distribution of $Y_{1c},Y_{2c},\ldots,Y_{10,c}$ (marginal in the sense of averaging over the distribution of N_c) is that of independent Poisson variables, the parameter of Y_{vc} being $\lambda(a_c)p_v$:

$$P(Y_c=y_c|W_c=w_c) = P(Y_c=y_c|A_c=a) = \prod_{i=1}^{10} e^{-\lambda(a)} \frac{[\lambda(a)p_i]^{y_{ic}}}{y_{ic}!} .$$

TABLE I

Relative Frequencies of the Ten Galton Characteristics

Characteristic	Frequency*	Relative Frequency*
Ending ridge	1247	.497
Fork	399	.159
Delta	34	.0135
Spur	88	.0350
Dot	255	.102
Island ridge	259	.103
Eye (lake)	66	.0263
Double bifurcation	16	.00637
Trifurcation	7	.00279
Bridge	140	.0558
	2511	1.00

*Based on 8591 cells

2.2. Border Cells

Cells at the border, not being touched by the full complement of four preceding adjacent cells, require some special treatment. One could take the results in border cells as given and take the probabilities for the other cells conditionally on the outcomes in the border cells, but this would result in considerable reduction in the effective sample size. (E.g., 18 of the 43 cells in Figure 1 are border cells.) We wish to use such information as is present. I.e., some but not all 4 preceding adjacent cells are sometimes present; we make use of this. For example, if there are 3 adjacent cells and 2 are occupied, then we know that if all 4 cells were present, then 2, 3 or 4 of them would be occupied. To use such information we need to know the conditional expected value of the number of occurrences, given that the number of adjacencies, A, is in the interval $[a_1,a_2]$, where $0 \le a_1 \le a_2 \le 4$. Recall that N_c is the number of occurrences in cell c and A_c is the number of the four preceding adjacent cells that are occupied. Then the conditional expected value we need is $E[N_c | a_1 \le A_c \le a_2] = \lambda(a_1,a_2)$, say. Note that

$$\lambda(a_1,a_2) = \sum_{a=a_1}^{a_2} \lambda(a)P(A_c=a) / \sum_{a=a_1}^{a_2} P(A_c=a).$$

Table II gives the estimates of $P(A_c=a)$ and of $\lambda(a)$, a = 0,1,2,3,4.

TABLE II

Estimates of Distribution of Number of Adjacencies and Conditional Expected Number of Occurrences

Number of adjacencies, a	0	1	2	3	4
Relative frequency, estimate of $P(A_c=a)$	.391	.296	.214	.084	.015
Estimate of $\lambda(a)$	.222	.344	.404	.511	1.252

Table III gives the estimates of $\lambda(a_1,a_2)$.

TABLE III

Estimates of Conditional Expected Number of Occurrences for Border Cells, $\lambda(a_1,a_2)$

a_1	a_2				
	0	1	2	3	4
0	0.222	0.277	0.305	0.323	0.337
1		0.345	0.369	0.389	0.411
2			0.402	0.433	0.473
3				0.512	0.626
4					1.267

3. COMPUTATION OF PROBABILITIES OF VARIOUS CONFIGURATIONS

We are now in a position to use the model to estimate the probabilities of various configurations. We illustrate with the configuration of Figure 1. Table IV illustrates the computation.

It may be helpful to elucidate the steps in the computation, since other researchers may wish to employ a similar two-dimensional Poisson-multinomial Markov model in problems they are working on. The steps are as follows.

(1) Count the number of adjacencies for each cell.

(2) Look up the corresponding expected number of occurrences in Table III.

(3) Compute the Poisson probability of this number of occurrences:

(4) Note the types of characteristics in the cell, if any.

(5) Compute the multinomial probability of this configuration of types in a cell.

(6) Multiply the results of steps (3) and (5) to obtain the probability for the cell.

(7) Multiply the results for all cells to obtain the probability for the whole configuration.

The result, which will be a very small probability, is best expressed on a negative log scale. [This means that in step (7) we add the logs of the numbers rather than multiply the numbers themselves.]

This deals with the assignment of a probability P to the occurrence of a given configuration in a given set of cells. For inferential purposes it is necessary to estimate the probability that a person has this configuration somewhere on his fingers. This problem is discussed in Section 3.2 of Osterburg, Parthasarathy, Raghavan and Sclove (1977). Further problems of inference concerning the identity of the person who left a print are considered in Section 4 of the same paper.

Now we carry out the above steps for the configuration of Figure 1 (Table IV).

TABLE IV

Computation of Estimate of Probability of Configuration of Figure 1

a_1, a_2	λ	Number in cell	Poisson prob. of this number in cell	Types in cell	Multinomial prob. of these types	Cell probability	Number of such cells
0,0	0.222	0	$e^{-0.222}=.801$	0	1	.801	6
0,0	0.222	1	$0.222e^{-0.222}=.178$	E	.497	.178(.497)=.188	3
0,0	0.222	1	$0.222e^{-0.222}=.178$	F	.159	.178(.159)=.028	1
0,1	0.277	0	$e^{-0.277}=.758$	0	1	.758	2
0,2	0.305	0	$e^{-0.305}=.737$	0	1	.737	4
0,2	0.305	0	$0.305e^{-0.305}=.225$	E	.497	.225(.497)=.112	1
0,3	0.323	0	$e^{-0.323}=.724$	0	1	.724	5
0,4	0.337	0	$e^{-0.337}=.714$	0	1	.714	1
1,1	0.345	0	$e^{-0.345}=.708$	0	1	.708	13
1,1	0.345	1	$0.345e^{-0.345}=.244$	F	.159	.244(.159)=.039	1
1,2	0.369	0	$e^{-0.369}=.691$	0	1	.691	4
1,3	0.389	0	$e^{-0.389}=.678$	0	1	.678	1
2,2	0.402	0	$e^{-0.402}=.669$	0	1	.669	1
							43

$$-\log_{10}\hat{P} = -6\log_{10}.801 - 3\log_{10}.088 - \ldots - \log_{10}.669 = 12.9$$

The estimated negative log probability is 12.9. Compare this with the figure of 11.4 given [Osterburg, Parthasarathy, Raghavan and Sclove (1977)] by the approximation based on independence; the ratio of probabilities is about 32:1. This difference is unimportant since we are interested only in order of magnitude.

As a second example, consider the configuration of Figure 2, in which the occurrences are still all singletons but are tightly clustered. The negative log probability, estimated by the method of this paper, is 11.2, compared to 11.4 for the method based on independence. The ratio of these two probability estimates is 1.6:1.

	1	2	3	4	5	6
a	0	0	0	0	0	0
b	0	0	F	E	E	0
c	0	0	E	F	0	0
d	0	0	0	E	0	0
e	0	0	0	0	0	0
f	0	0	0	0	0	0
g	0	0	0	0	0	0
h			0			

FIG. 2

Configuration of 43 cells with 4 ending ridges and 2 forks, as in Fig. 1, but here the six occurrences are clustered.

As a third example, let us still consider a configuration of 43 cells with 4 ending ridges and 2 forks, as in Figures 1 and 2, but now suppose some of these occur not as singletons but rather as multiple occurrences in cells (Figure 3).

	1	2	3	4	5	6
a	0	0	0	0	0	0
b	0	0	F	EF	0	0
c	0	0	0	EEE	0	0
d	0	0	0	0	0	0
e	0	0	0	0	0	0
f	0	0	0	0	0	0
g	0	0	0	0	0	0
h			0			

FIG. 3

Configuration of 43 cells with 4 ending ridges and 2 forks, as in Figures 1 and 2, but with the six occurrences as a singleton, a doubleton, and a triplet.

The new feature of this example is the multiple occurrences in cells (b,4) and (c,4). Finding the probability of the doubleton in cell (b,4) involves the following computation.

Number of adjacencies = 1

Expected number of occurrences = 0.345

Poisson probability of 2 occurrences $= e^{-0.345}\,\frac{0.345^2}{2!} = 4.21 \times 10^{-2}$

Multinomial probability that the 2 occurrences are E and F $= \binom{2}{1,1,0,\ldots,0} .497^2 .159^1 .0135^0 \cdots .0558^0$

$= 7.85 \times 10^{-2}$

Prob. for cell (b,4) $= (7.85 \times 10^{-2})(4.21 \times 10^{-2}) = 3.31 \times 10^{-3}$

The computation of the probability of the triplet EEE in cell (c,4) is as follows.

Number of adjacencies = 2

Expected number of occurrences = 0.402

Poisson probability of 3 occurrences $= e^{-0.402}\,\frac{0.402^3}{3!} = 7.24 \times 10^{-3}$

Multinomial probability that the 3 occurrences are E,E,E $= \binom{3}{3,0,\ldots,0} .497^3 .159^0 \cdots .0558^0$

$= .123$

Prob. for cell (c,4) $= .123(7.24 \times 10^{-3}) = 8.89 \times 10^{-4}$

Computing the probabilities for all the cells and multiplying gives $-\log_{10}\hat{P} = 11.7$ for this configuration. The method given in Osterburg, Parthasarathy, Raghavan and Sclove (1977) lumps all multiple occurrences together, assigning a cell with amy such multiple occurrence a probability of .0355. A fork has a probability of .0382, an empty cell, a probability of .766; hence $\hat{P} = .766^{40} .0355^2 .0382^1$, $-\log_{10}\hat{P} = 9.0$. The difference is negative log probabilities provided by the two methods is $11.7 - 9.0 = 2.7$; $10^{2.7} = 500$. The difference begins to be sizable but is not particularly important in terms of order of magnitude, not when one is more interested simply in discriminating between probabilities on the order of one in a thousand and those on the order of one in a million millions.

As a fourth example, let us consider again a total of 43 cells with 6 occurrences, two forks and four ridge endings, but <u>all in one cell</u>. This is to be considered as a mathematical extreme and not an empirical example: We never observed more than 5 occurrences in a single cell and these were combinations of ridge endings and dots. The conditional expected number of occurrences in a single cell, with no adjacent cells occupied, is 0.222; the corresponding Poisson probability of 6 occurrences is $e^{-0.222}0.222^6/6! = 1.33 \times 10^{-7}$. Obtaining the relative frequencies from Table I, we compute the multinomial probability of 2 forks and 4 ridge endings, given 6 occurrences, as

$$\binom{6}{4,2,0,\ldots,0}.497^4.159^2.0135^0\cdots.0558^0 = 15(.497)^4(.159)^2$$
$$= 2.31 \times 10^{-2}.$$

The probability for the cell is thus $(1.33 \times 10^{-7})(2.31 \times 10^{-2}) = 3.08 \times 10^{-9}$. Computing the probabilities for the other cells by following the steps outlined and combining the results for all cells gives $-\log_{10}\hat{P} = 13.4$ for this configuration. The method given in Osterburg, Parthasarathy, Raghavan and Sclove (1977) gives $-\log_{10}\hat{P} = 6.32$ for this configuration. That method gives such a high probability compared to the method of the present paper because that method lumps all multiple occurrences together, as mentioned above, assigning a cell with any such multiple occurrence a probability of .0355, or a weight of $1.45 = -\log_{10}.0355$. Using that method, one would assign that probability to the cell containing the multiple occurrence EEEEFF. The method of the present paper gives a probability of 3.08×10^{-9}, i.e., a weight of $8.51 = -\log_{10}\hat{P}$ to this multiple occurrence. This underscores the advance made by the present method, compared to the method in Osterburg, Parthasarathy, Raghavan and Sclove (1977). A criticism of the present method is that it is too dependent upon the placement of the grid of cells, which is arbitrary, in that multiple occurrences in a single cell could have been placed in two cells if the grid had been placed differently. However, the effect of this appears not to be large. For example, Figures 4(a) and (b) show portions of two configurations, both with 4 occurrences. In Fig. 4(a) the 4 occurrences are all in one cell;

in Fig. 4(b) they are in two cells. The Poisson probability for the 12 cells of Fig. 4(a) is $(e^{-0.222})^7(.222e^{-0.222})^1(e^{-0.345})^4$ giving $-\log_{10}\hat{P} = 2.02$. The Poisson probability for the 12 cells of Fig. 4(b) is $(e^{-0.222})^5(.222e^{-0.222})^1(e^{-0.345})^3(0.345e^{-0.345})^1(e^{-0.402})^2$, $-\log_{10}\hat{P} = 2.64$. The difference is 2.64 - 2.02 = 0.62; $10^{0.62} = 4.2$, not much difference at all.

0	0	0	0
0	xxxx	0	0
0	0	0	0

FIG. 4(a)

0	0	0	0
0	xx	xx	0
0	0	0	0

FIG. 4(b)

Portions of two configurations. (a): 4 occurrences in a single cell. (b): 4 occurrences, 2 in one cell and 2 in the next. (The symbol x indicates an occurrence of some type.)

APPENDIX: MARGINAL DISTRIBUTION OF THE CHARACTERISTICS

The distribution of the number of occurrences per cell is given in Table V.

TABLE V

Distribution of Number of Occurrences

Number of occurrences	0	1	2	3	4	5	Total
Number of cells	6584	1594	320	72	19	2	8591
Proportion of cells	.766	.185	.0372	.00838	.0022	.00023	1.00

This distribution is well fit by a negative binomial distribution. A Poisson or geometric distribution is not adequate. An interpretation of this finding is that what is involved is a gamma-type mixture of Poisson distributions [see, e.g., Parzen(1962), p. 57], resulting in a negative binomial distribution, as in the classical accident studies of Greenwood and Yule (1920). We remark that we proceeded with parameter estimation and the fitting of distributions as if we had strict between-cell independence though we have shown that this does not hold strictly. We do note, however, that our parameter estimates are based upon (at most) the first two sample moments and that the sample moment is unbiased for the corresponding population moment, even without independence. The chi-square goodness-of-fit tests, on the other hand, are based on an assumption of independent trials.

Empirical support for the assumption of a local Poisson process, i.e., a Poisson process where the intensity parameter varies (a non-homogeneous Poisson process) is given, then, by the fact that a negative binomial distribution fits. Theoretical support for the assumption is demonstrated by the plausibility of the following assumptions. [These are the usual aximoms for a Poisson process (see, e.g., Parzen (1962), p. 119), generalized to two dimensions.] Given any set in the (x,y)-plane, Let N(S) be the number of occurrences in S and let a(S) be the area of S. Given any point (x,y), let $\{S_n\}$ be a sequence of sets tending to (x,y): $\lim_{n\to\infty} S_n = \{(x,y)\}$. Then the following assumptions are plausible. There is a positive number $\lambda(x,y)$ such that

$$\lim_{n\to\infty} \frac{1 - P[N(S_n)=0]}{a(S_n)} = \lambda(x,y),$$

$$\lim_{n\to\infty} \frac{P[N(S_n) = 1]}{a(S_n)} = \lambda(x,y),$$

and

$$\lim_{n\to\infty} \frac{P[N(S_n) \geq 2]}{a(S_n)} = 0 .$$

ACKNOWLEDGEMENTS

The initial investigation, reported in Osterburg, Parthasarathy, Raghavan and Sclove (1977), was supported under a contract with the Center for Research in Criminal Justice, University of Illinois at Chicago Circle. Work on the current report was supported under Grant AFOSR 77-3454 from the Air Force Office of Scientific Research. The author gratefully acknowledges these sources of support.

BIBLIOGRAPHY

Galton, F. (1892). Finger Prints. London: Macmillan; republication (1965), New York: DaCapo Press.

Greenwood, M. & Yule, G.U. (1920). An inquiry into the nature of frequency-distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. Jour. Royal Statist. Soc. 83, 255.

Osterburg, J.W., Parthasarathy, T., Raghavan, T.E.S & Sclove, S.L. (1977). Development of a mathematical formula for the calculation of fingerprint probabilities based on individual characteristics. Jour. Amer. Statist. Assoc. 72, 772-8.

Parzen, E. (1962). Stochastic Processes. San Francisco: Holden-Day.

Sclove, S.L. (1978). The occurrence of fingerprint characteristics as a two-dimensional process. Technical Report, AFOSR Grant 77-3454, February 15, 1978.