RECONSTRUCTING PATTERNS FROM SAMPLE DATA[1]

By Paul Switzer

Stanford University

1. Introduction. A Euclidean k-dimensional region A having unit volume is partitioned into m Lebesgue-measurable subregions $A_1, A_2, \cdots, A_m$ each having positive volume. The results of this partition we will call a pattern. It is convenient to assign an identifying color to each of these subregions so that a pattern may be visualized as a k-dimensional mosaic. This pattern is unobservable except at a fixed set of n points $s_1, s_2, \cdots, s_n$ in A, i.e., we will know the color of each of these n points which we call sample-points.

This data, the color observations at the sample-points, might be used to estimate the volumes of some or all of the subregions $\{A_i\}$. In particular, if n_i is the number of sample-points observed to fall in A_i, then a naive estimator of $\mu(A_i)$, the volume of A_i, is n_i/n. Several authors have derived the variance of this estimator for a fixed set of sample-points using a model wherein the partition of A is a realization of a random process with the following stationarity and isotropy properties:

(1) (i) for all points $s \varepsilon A$, $\Pr(s \varepsilon A_i) = p_i$ for $i = 1, \cdots, m$; and
(ii) for all pairs of points $s, s' \varepsilon A$ with distance $|s - s'| = d$ between them, $\Pr(s' \varepsilon A_i \mid s \varepsilon A_j) = P_{ij}(d)$ for $i, j = 1, \cdots, m$.

The variance criterion is then used to compare alternative methods for locating, sample-points in A. A good reference is Matérn [2].

However, in some geological and perhaps other applications (e.g., mineral deposits, soils, vegetation, land-use), the same data are sometimes used to produce an estimated reconstruction of the pattern, i.e., the region A is partitioned into m subregions $\hat{A}_1, \hat{A}_2, \cdots, \hat{A}_m$. Often this will be a "freehand" reconstruction, but where an explicit rule is used it is often the "nearest-point" rule which says: a point $s \varepsilon A$ is assigned to $\hat{A}_i$ if the sample-point nearest to s is observed to fall in A_i.

It is then natural to ask how accurately does the estimated pattern $\{\hat{A}_i\}$ represent the unobserved pattern $\{A_i\}$? If it is reasonable to measure accuracy by the extent to which the two patterns fail to coincide then as a loss function we might take

$$L = 1 - \sum_{j=1}^{m} \mu(A_j \cap \hat{A}_j), \tag{2}$$

which is the proportion of the volume of A that is incorrectly assigned. Using a

Received 10 January 1966.

[1] Part of this work was supported by a grant from the National Science Foundation GS-341 while the author was at Harvard University and by ONR contract Nonr 225(52) at Stanford University.

randomness model for pattern generation with the properties (1) we can, in principle, compute the expected loss EL; EL can then be used as a criterion to compare alternative methods for locating sample-points in A, to gauge the effect of increasing or decreasing the sampling density, and to explore modifications of the "nearest-point" reconstruction rule. These are the major tasks of this paper.

Typically in the principal applications that come to mind the region A is planar and the number of subregions is two ($k = m = 2$), e.g., presence-absence, high-low, light-dark patterns. Therefore, we will make a point of specializing most of our results to this case.

2. The simple nearest-point rule. Let $s_1, \cdots, s_n$ be the fixed locations of n sample-points in the region A. For an arbitrary point $s \varepsilon A$ let $N(s)$ denote the unique nearest sample-point, i.e., $N(s) = s_i$ where $|s - s_i| < |s - s_j|$ for all $j \neq i$. In general, there will be a non-empty set A_0 having zero volume and consisting of points in A having no unique nearest sample-point. If the estimated partition $\hat{A}_1 + \hat{A}_2 + \cdots + \hat{A}_m = A$ is contructed according to the rule $\hat{A}_i = \{s \varepsilon A : N(s) \varepsilon A_i\}$ it will be called the simple nearest-point rule. Strictly speaking $\{\hat{A}_i\}$ is a partition not of A but of $A - A_0$ but it will be clear that for our purposes this is an unimportant distinction.

An alternative equivalent formulation of this rule will prove to be more useful. Let S_i be the set of points closer to the sample-point s_i than to any other sample-point, i.e., $S_i = \{s \varepsilon A : N(s) = s_i\}$. We will call S_i *the region of influence* of the sample-point s_i. It is clear that $\{S_i\}$ are a partition of $A - A_0$, that each S_i is a convex polyhedron and that the simple nearest-point rule has the alternative definition

$$\hat{A}_i = \sum_{s_j \varepsilon A_i} S_j . \tag{3}$$

Here and elsewhere in context $\sum$ denotes the union of disjoint sets.

THEOREM 1. *If the simple nearest-point rule is used to reconstruct the pattern, if the loss function is given by* (2), *and if the randomness model has properties* (1), *then for any fixed set of sample-points* $s_1, \cdots, s_n$ *the risk* EL *is given by*

$$1 - \sum_{j=1}^{m} p_j \sum_{i=1}^{n} \int_{S_i} P_{jj}(|s - s_i|)\, d\mu(s). \tag{4}$$

PROOF. We use a theorem by Robbins [3] which says: if $\mu(B)$ is the volume of a random Lebesgue-measurable subregion B of A, then

$$E\mu(B) = \int_A \Pr(s \varepsilon B)\, d\mu(s) \tag{5}$$

provided $\Pr(s \varepsilon B)$ is defined for almost all points $s \varepsilon A$. For the random set B we take $A_j \cap \hat{A}_j$. For any point $s \varepsilon A - A_0$ we have $\Pr(s \varepsilon A_j \cap \hat{A}_j) = \Pr\{s \varepsilon A_j, N(s) \varepsilon A_j) = \Pr(s \varepsilon A_j) \Pr\{N(s) \varepsilon A_j \mid s \varepsilon A_j\} = p_j P_{jj}(|s - N(s)|)$ by (1). So by Robbin's theorem

$$\begin{aligned} E\mu(A_j \cap \hat{A}_j) &= \int_A \Pr(s \varepsilon A_j \cap \hat{A}_j)\, d\mu(s) = \int_{A-A_0} \Pr(s \varepsilon A_j \cap \hat{A}_j)\, d\mu(s) \\ &= p_j \int_{A-A_0} P_{jj}(|s - N(s)|)\, d\mu(s) = p_j \sum_{i=1}^{n} \int_{S_i} P_{jj}(|s - s_i|)\, d\mu(s) \end{aligned}$$

since $\{S_i\}$ is a partition of $A - A_0$ and since $N(s) = s_i$ for all points $s \varepsilon S_i$. The theorem is now proved because $EL = 1 - \sum E\mu(A_j \cap \hat{A}_j)$ by (2).

Formula (4) is just the probability that a point chosen with uniform probability over A will have a color different from its nearest sample-point. A simplification of this formula occurs if the sample-points are *systematically* arranged defined as follows: Let the origin of k-dimensional Euclidean space be located at one of the sample-points, say s_0, and let S_0 denote its region of influence; if for every other sample-point s_i there is a translation and rotation which carries s_i into s_0 and S_i into S_0 then the arrangement of sample-points will be called systematic. For such arrangements it follows that the integral of formula (4) has the same value for all i, hence we can write

$$EL = 1 - n\sum_{j=1}^{m} p_j \int_{S_0} P_{jj}(|s|)\, d\mu(s). \tag{6}$$

Consider now the special case of two-color patterns, that is $m = 2$. Then $p_2 = 1 - p_1$ and it is easily shown that $P_{22}(d) = 1 - p_1(1 - p_1)^{-1}[1 - P_{11}(d)]$; for let $s, s' \varepsilon A$ be any two points with $|s - s'| = d$ then using properties (1) and elementary arguments we get $P_{22}(d) = \Pr(s \varepsilon A_2 \mid s' \varepsilon A_2) = (1 - p_1)^{-1} \cdot \Pr(s \varepsilon A_2, s' \varepsilon A_2) = (1 - p_1)^{-1}[\Pr(s \varepsilon A_2) - \Pr(s \varepsilon A_2, s' \varepsilon A_1)] = (1 - p_1)^{-1}[(1 - p_1) - p_1 \Pr(s \varepsilon A_2 \mid s' \varepsilon A_1)] = 1 - p_1(1 - p_1)^{-1}[1 - \Pr(s \varepsilon A_1 \mid s' \varepsilon A_1)] = 1 - p_1(1 - p_1)^{-1}[1 - P_{11}(d)]$. Using this result the risk formula (4) specialized to two-color patterns can be expressed as

$$EL/2p_1 = 1 - \sum_{i=1}^{n} \int_{S_i} P_{11}(|s - s_i|) d\mu(s), \tag{7}$$

and the risk formula (6) for systematic sample-point arrangements specializes to

$$EL/2p_1 = 1 - n \int_{S_0} P_{11}(|s|)\, d\mu(s). \tag{8}$$

Speaking of two-color patterns, it is interesting but not surprising to note that the loss function (2) we have been using can be regarded as a mean-square-error function. Specifically, let $f(s)$ be the indicator function for the subregion A_1, i.e., $f(s) = 1$ if $s \varepsilon A_1$ and zero otherwise; and let $\hat{f}(s)$ be the indicator function for $\hat{A}_1$. Then, regardless of the rule used to construct the estimated pattern $\{\hat{A}_1, \hat{A}_2\}$,

$$L = \int_A [\hat{f}(s) - f(s)]^2\, d\mu(s). \tag{9}$$

This is easily shown by noting that the integrand is the indicator function for the region $\hat{A}_2 \cap A_1 + \hat{A}_1 \cap A_2 = A - A_1 \cap \hat{A}_1 - A_2 \cap \hat{A}_2 - A_0$. Hence, the right-hand side of (9) equals $1 - \mu(A_1 \cap \hat{A}_1) - \mu(A_2 \cap \hat{A}_2)$ which is the expression for L given in (2) for the case $m = 2$.

A somewhat more general loss function than the one heretofore considered results by associating different rates of loss with different kinds of errors. That is, let $A_{ij} = A_i \cap \hat{A}_j$, i and $j = 1, \cdots, m$, and let L_{ij} be the loss per unit volume of A_{ij} with the restriction that $L_{ij} \geqq 0$ for $i \neq j$ and $L_{ij} \leqq 0$ for $i = j$. It is clear

that the m^2 regions $\{A_{ij}\}$ are a partition of $A - A_0$, and our generalized loss function can be written as

$$L' = \sum_{i,j} L_{ij}\mu(A_{ij}). \tag{10}$$

It is easily seen that the special loss function L in (2) is obtained by taking $L_{ij} = 1$ for $i \neq j$ and $L_{ij} = 0$ for $i = j$. The following theorem demonstrates a simple relation between the general expected loss EL' and the special expected loss EL.

THEOREM 2. *If the simple nearest-point rule is used to reconstruct a two-color pattern generated by a randomness model with properties* (1), *then*

$$EL' = \tfrac{1}{2}EL(L_{12} + L_{21} - L_{11} - L_{22}) + p_1L_{11} + (1 - p_1)L_{22}. \tag{11}$$

PROOF. Since $A_1 + A_2 = \hat{A}_1 + \hat{A}_2 + A_0 = A$, it follows that for any point $s \varepsilon A$, $\Pr(s \varepsilon A_{12}) = \Pr(s \varepsilon A_{21}) + [p_1 - \Pr(s \varepsilon \hat{A}_1)]$, $\Pr(s \varepsilon A_{11}) = p_1 - \Pr(s \varepsilon A_{12})$, $\Pr(s \varepsilon A_{22}) = (1 - p_1) - \Pr(s \varepsilon A_{21})$. But using the simple nearest-point rule gives $\Pr(s \varepsilon \hat{A}_1) = \Pr\{N(s) \varepsilon A_1\} = p_1$, so for this rule $\Pr(s \varepsilon A_{21}) = \Pr(s \varepsilon A_{12}) = p_1 - \Pr(s \varepsilon A_{11}) = (1 - p_1) - \Pr(s \varepsilon A_{22})$ for all $s \varepsilon A$. Now by Robbin's theorem (5),

$$E\mu(A_{ij}) = \int_A \Pr(s \varepsilon A_{ij})\, d\mu(s),$$

hence $E\mu(A_{12}) = E\mu(A_{21}) = p_1 - E\mu(A_{11}) = (1 - p_1) - E\mu(A_{22})$. From the definition (2) we now get $EL = 1 - [E\mu(A_{11}) + E\mu(A_{22})] = 2[p_1 - E\mu(A_{11})]$, and from (10) we get $EL' = L_{11}E\mu(A_{11}) + L_{12}E\mu(A_{12}) + L_{21}E\mu(A_{21}) + L_{22}E\mu(A_{22})$. The right-hand side of this last expression for EL' can now be expressed in terms of $E\mu(A_{11})$ which in turn can be expressed in terms of EL. Making the appropriate substitutions yields statement (11) of the theorem.

3. Applications to sample arrangement and sample size problems. In the preceding section we derived expressions for the risk EL when the simple nearest-point rule was used to construct an estimated pattern $\{\hat{A}_1, \cdots, \hat{A}_m\}$ from observations on a set of sample-points $s_1, \cdots, s_n$ in a region $A = \sum \hat{A}_i + A_0$ (where A_0 has zero volume). This is a commonly used rule and it is simple to handle mathematically. In particular, we have seen that for this rule EL depends on the underlying randomness model only through the m numbers $\{p_j\}$ and the m functions $\{P_{jj}(d)\}$ for values of d in some neighborhood of zero. For the special case of two-color patterns EL was seen to depend only on p_1 and $P_{11}(d)$, the same being true for the general risk EL'.

In principal we could now use EL as a criterion for the comparison of alternative arrangements of a given number n of sample-points. Or, for a given type of arrangement, we may determine the number of sample-points needed to reach a certain level of the EL criterion. For the remainder of this section we will be looking at the special case of two-color patterns on the planar region A.

The result of any comparison we make will, of course, depend on the functional form of $P_{11}(d)$. Recall that the quantity $p_1P_{11}(d)$ is the probability that a pair of

points in A, separated by a distance d, are both in the subregion A_1. In many circumstances it is natural, therefore, to require that $P_{11}(d)$ decay monotonically with increasing d to the value p_1, the unconditional probability of a single point falling in A_1. (By virtue of Robbin's theorem (5) p_1 is also the expected area of A_1.) Of necessity $P_{11}(0) = 1$; when the boundary of A_1 has finite expected length, Matérn [2] has shown that $P_{11}(d)$ must be convex in the vicinity of $d = 0$ and for random processes in two or more dimensions $P_{11}(d)$ must be everywhere continuous. The model

$$P_{11}(d) = p_1 + (1 - p_1)e^{-cd}, \qquad c > 0, \tag{12}$$

is consistent with these requirements and we will use it for illustrative comparisons below; a pattern-generating procedure with precisely this $P_{11}(d)$ function is described briefly in Section 5 and in greater detail in Switzer [4].

The first comparison to be made is among alternative *systematic* arrangements of the sample-points as previously defined. It is clear that the shape of A determines which, if any, systematic arrangements are possible. In order to bypass this complication we will allow its shape to very in order to accommodate any particular arrangement, but keeping its area always equal to unity.

A frequently used design places the n sample-points at the vertices of a square grid. The nearest-point regions of influence, the $\{S_i\}$, are therefore n squares centered at the n sample-points each having area $1/n$. Using the model (12) in formula (8), the average incorrectly assigned area EL is given by

$$EL/2p_1(1 - p_1) = 1 - 8 \int_0^{\frac{1}{2}} \int_0^y \exp\{-c((x^2 + y^2)/n)^{\frac{1}{2}}\}\, dx\, dy. \tag{13}$$

The right-hand side of this expression has been computed for selected values of $cn^{-\frac{1}{2}}$; the results are given in the first column of Table I. So, for example, if for a given sample-size the average incorrectly assigned area is $.30p_1(1 - p_1)$ then in order to halve this risk we will need about $(.40/.20)^2 = 4$ times as many sample-points; whereas to reduce a risk of $1.50p_1(1 - p_1)$ by half we will need about $(4.00/1.20)^2 = 11$ times as many sample-points.

An alternative systematic arrangement places the n sample-points at the vertices of an equilateral triangular grid. The nearest-point regions of influence are n regular hexagons centered at the sample points, each having area $1/n$. A formula similar to (13) can be written down for this triangular design; it turns out that, for those values of $cn^{-\frac{1}{2}}$ used in Table I, the improvement this design provides over the square-grid design does not exceed .004 in units of $EL/2p_1(1 - p_1)$.

If we are required to travel from sample-point to sample-point in order to take our observations, then the cost of travel should influence the choice of sample arrangement. Suppose the n sample-points are placed at the vertices of a rectangular grid so that their regions of influence are all rectangles with dimensions $1/r \times r/n$ where $1/r \geqq r/n$, say. Then it is not hard to see that, for a fixed sample size, less travel is involved as the grid becomes more elongated, i.e., as r gets smaller. In particular, let the region A be a unit square and let r be an integer,

TABLE I

Values of $EL/2p_1(1 - p_1)$ *for* n *sample-points on a* $r^{-1} \times rn^{-1}$ *rectangular grid, using model (12) and simple nearest-point rule*

$cn^{-\frac{1}{2}}$	nr^{-2} = length:width				
	1.0	2.0	5.0	10.0	50.0
.20	.073	.080	.108	.145	.283
.30	.108	.117	.157	.206	.384
.40	.141	.152	.202	.262	.465
.50	.172	.186	.244	.313	.531
.60	.202	.218	.283	.359	.586
.80	.259	.278	.353	.439	.668
1.00	.311	.332	.415	.505	.726
1.20	.359	.381	.469	.560	.769
1.40	.403	.426	.516	.607	.801
1.60	.443	.467	.558	.646	.825
1.80	.481	.505	.594	.680	.844
2.00	.515	.539	.627	.708	.860
2.50	.590	.613	.693	.764	.889
3.00	.651	.672	.744	.805	.908
4.00	.744	.760	.814	.858	.931

then the travel path would consist essentially of r parallel traverses spaced $1/r$ units apart and the number of sample-points on each traverse would be n/r. The distance traveled is then essentially equal to r with a maximum of $r = n^{\frac{1}{2}}$ corresponding to a square-grid arrangement.

But while travel distance decreases as the sampling grid becomes more elongated, it is reasonable to expect that the risk EL increases when this happens. Using the model (12) with formula (8) yields an expression for $EL/2p_1(1 - p_1)$ which is a function of r and $cn^{-\frac{1}{2}}$. The values of the expression for selected values of r and $cn^{-\frac{1}{2}}$ are exhibited in Table I. So, for example, if a square-grid yields an average incorrectly assigned area (EL) of $.40p_1(1 - p_1)$ then the same number of sample-points arranged on a 2:1 rectangular grid would yield an EL about 10 per cent higher with a saving of about 30 per cent in the distance traveled. If α is the cost of making an observation and β is the cost of travel per unit distance, then for a fixed total budget B it will be possible to increase the sample size the more we elongate the sampling grid; in fact, for a given α, β, B there will be an optimum value for r yielding the minimum possible EL, and in general, this optimum sample design will not be a square grid.

In some applications the cost of making an observation is negligible compared to travel costs, so that all that really matters cost-wise is how many traverses are made. In such a situation it is obvious that the more densely we sample along a traverse the smaller will be the risk EL. Let σ be the spacing between parallel traverses and let $f \leqq \sigma$ be the spacing between sample-points along any traverse.

Keeping σ fixed and letting $f \to 0$ we get the limiting case of a continuous record along each traverse. This will be called *line sampling* with spacing σ.

THEOREM 3. *If A is an $h \times 1/h$ rectangular region ($h \geqq 1$), if a two-color pattern is generated on A by randomness model with properties (1), if A is line-sampled by k traverses with spacing $1/hk = \sigma$, and if the simple nearest-point rule is used to reconstruct the pattern, then the risk is given by*

$$EL/2p_1 = 1 - (2/\sigma) \int_0^{\sigma/2} P_{11}(u)\, du. \tag{14}$$

PROOF. While proving Theorem 2 we showed that for any kind of sampling $EL = 2[p_1 - E\mu(A_{11})]$, where by Robbin's theorem (5)

$$E\mu(A_{11}) = \int_A \Pr\,(s \,\varepsilon\, A_1 \cap \hat{A}_1)\, d\mu(s).$$

But $\Pr\,(s \,\varepsilon\, A_1 \cap \hat{A}_1) = \Pr\,\{s \,\varepsilon\, A_1\,, N(s) \,\varepsilon\, A_1\} = p_1 P_{11}(|s - N(s)|)$ by virtue of property (1). So we can write

$$EL = 2p_1\{1 - \int_A P_{11}(|s - N(s)|)\, d\mu(s)\}.$$

Now for definiteness let the Cartesian (x, y) co-ordinates of the vertices of the rectangle A be $(\frac{1}{2}\sigma, 0)$, $(k\sigma + \frac{1}{2}\sigma, 0)$, $(\frac{1}{2}\sigma, h)$, $(k\sigma + \frac{1}{2}\sigma, h)$. Then the sampling traverses are segments of the lines $x = i\sigma$, $i = 1, \cdots, k$. Let A be partitioned into k rectangles $R_1\,, R_2\,, \cdots, R_k$ with the vertices of R_i given by $(i\sigma - \frac{1}{2}\sigma, 0)$, $(i\sigma + \frac{1}{2}\sigma, 0)$, $(i\sigma - \frac{1}{2}\sigma, h)$, $(i\sigma + \frac{1}{2}\sigma, h)$. The sample-point $N(s)$ nearest to an arbitrary point $s = (x, y) \,\varepsilon\, R_i$ will be the projection of s on the traverse line $x = i\sigma$, i.e., $N(s) = (i\sigma, y)$ and $|s - N(s)| = |x - i\sigma|$. Hence we can write

$$\begin{aligned}\int_A P_{11}(|s - N(s)|)\, d\mu(s) &= \sum_{i=1}^{k} \int_0^h \int_{i\sigma-\frac{1}{2}\sigma}^{i\sigma+\frac{1}{2}\sigma} P_{11}(|x - i\sigma|)\, dx\, dy \\ &= kh \int_{-\sigma/2}^{\sigma/2} P_{11}(|u|)\, du = (2/\sigma) \int_0^{\sigma/2} P(u)\, du,\end{aligned}$$

since by definition $h = 1/k\sigma$. This completes the proof of the theorem.

Using the exponential model (12), the result of Theorem 3 for line sampling specializes to

$$EL/2p_1(1 - p_1) = 1 - 2(1 - e^{-\frac{1}{2}c\sigma})/c\sigma. \tag{15}$$

For example, if for a given number of equally spaced traverses we would get a value for the risk EL of $.60p_1(1 - p_1)$, then by doubling the number of traverses we could reduce the risk by about $\frac{1}{3}$.

For a fixed spacing σ between traverses it is interesting to note how quickly EL approaches its limiting minimum value as f, the spacing between sample-points along a traverse, goes to zero. Using the model (12) for $P_{11}(d)$, Table II indicates how little is to be gained by using f spacings less than $\sigma/4$.

4. The optimum nearest-point rule. We now return to the general case of m-color patterns $\{A_1\,, \cdots, A_m\}$ on a k-dimensional Euclidean region $A = \sum A_i$ of unit volume. Heretofore, the simple nearest-point rule has been used in this paper to construct the estimated pattern $\{\hat{A}_1\,, \cdots, \hat{A}_m\}$ based on observations on n sample-points in A. This rule assigns a point $s \,\varepsilon\, A$ to the subregion $\hat{A}_j$

TABLE II

Values of $EL/2p_1(1 - p_1)$ *for* $1/\sigma f$ *sample-points on a* $\sigma \times f$ *rectangular grid, using model (12) and simple nearest-point rule*

$c\sigma$	$f = \sigma$	$f = \sigma/4$	$f = 0$
.40	.141	.099	.094
1.00	.311	.226	.213
2.00	.515	.390	.368
4.00	.744	.602	.568

if and only if the sample-point nearest to s, $N(s)$, is observed to fall in A_j. As we have seen, this rule leads in a straightforward way to quite simple expressions for the risk EL when the pattern generating model is assumed to have properties (1). In fact, EL depends on the model only through the set of probabilities $\{p_i\}$ and the $P_{ij}(d)$ functions in a vicinity of $d = 0$.

The same simplicity property is characteristic of any nearest-point rule defined in the following manner: For an arbitrary point $s \varepsilon A$ let the unique sample-point nearest to it, $N(s)$, be observed to fall in A_j; then s is assigned to $\hat{A}_{j'}$ where j' may depend only on s and j. (Once again there is a zero-volume set A_0 of points in A not having a unique nearest sample-point on which any nearest-point rule is not defined or may be arbitrarily defined.) For any rule of this class we get from the definition (2) and Robbin's theorem (5)

$$(16)\quad EL = 1 - \sum_{j=1}^{m} \int_A \Pr(s \varepsilon A_j \dot{\cap} \hat{A}_j)\, d\mu(s)$$
$$= 1 - \sum_{j=1}^{m} \int_A \sum_{k=1}^{m} \Pr\{s \varepsilon A_j \cap \hat{A}_j \mid N(s) \varepsilon A_k\} \Pr\{N(s) \varepsilon A_k\}\, d\mu(s).$$

The event $\{s \varepsilon \hat{A}_j \mid N(s) \varepsilon A_k\}$ cannot occur unless $j = k'$ in which case it must occur by our definition of nearest-point rules. So by exchanging the orders of summation and integration and using the model properties (1) we get

$$EL = 1 - \sum_{k=1}^{m} \int_A \Pr\{s \varepsilon A_{k'} \mid N(s) \varepsilon A_k\} \Pr\{N(s) \varepsilon A_k\}\, d\mu(s)$$
$$= 1 - \sum_{k=1}^{m} p_k \int_A P_{k'k}\{|s - N(s)|\}\, d\mu(s).$$

It is clear that if for each s and k we choose that k' which makes $P_{k'k}\{|s - N(s)|\}$ largest, then EL will be minimized in the class of nearest-point rules. This method of choosing k' therefore defines the *optimum* nearest-point rule. In other words, the optimum rule assigns to each point its most probable color given the observed color of the sample-point nearest to it—an intuitively reasonable conclusion. We remark that at some points there may be "ties" for most probable color and that the set of such points may have positive volume; any choice of color among these tied colors yields the same minimum value of the risk so that any choice corresponds to an optimum rule.

Such a rule usually implies that $E\mu(A_i) \neq E\mu(\hat{A}_i)$, thus bias is introduced in the *areal* representation of the colors. In fact, the areal biases can easily be large, especially when $E\mu(A_i) = p_i$ is small and $P_{ii}(d)$ decays rapidly.

A disadvantage of the optimum rule is that we are unable even to construct the estimated pattern unless there is some knowledge of the $P_{ij}(d)$ functions. On the other hand, the simple rule of the previous sections was defined by taking $j' = j$ for all points s hence the construction of the estimate could proceed independently of the model. We will show that, for a further restricted class of models, the simple rule *is* the optimum rule if the number of sample-points is sufficiently high. Furthermore, a sure method will be given of improving the simple rule when it is not optimum.

The restricted class of models we will consider is defined by imposing the following conditions on the $P_{ij}(d)$ functions:

(17)
(i) Uniform preference: $p_h > p_i$ implies $P_{hj}(d) \geqq P_{ij}(d)$ for all $d \varepsilon [0, D]$ and all i, $h \neq j$, where D is the diameter of the region A.
(ii) Monotone effect of distance: $P_{hj}(d)$ is a monotone function on $[0, D]$ for all h and j. Necessarily, $P_{jj}(0) = 1$ so that $P_{jj}(d)$ is non-increasing, and for $h \neq j$ $P_{hj}(0) = 0$ so that $P_{hj}(d)$ is non-decreasing.

In the next section specific pattern-generating models having these properties will be demonstrated. In the meantime we observe that for any such restricted model there is an optimum nearest-point rule with a quite simple structure.

THEOREM 4. *Let $p_M = \max_i p_i$, assumed unique,* i.e., *A_M is the random subregion of A having largest expected volume. If the randomness model is of the isotropic stationary type* (1) *with the uniformity and monotonicity properties* (17), *then we can find m numbers $\{r_j\}$ on $[0, D]$ which define an optimum nearest-point rule as follows: "If $N(s)$, the sample-point nearest to a point $s \varepsilon A - A_0$, is observed to fall in A_j then place $s \varepsilon \hat{A}_j$ if $|s - N(s)| < r_j$ and otherwise place $s \varepsilon \hat{A}_M$."*

PROOF. As we showed before, an optimum rule places a point s into $\hat{A}_{j'}$ whenever $N(s)$ is in A_j where j' is chosen so that $P_{j'j}\{|s - N(s)|\} \geqq P_{hj}\{|s - N(s)|\}$ for $h \neq j'$. By the uniform preference property (i) it follows that $P_{Mj}(d) \geqq P_{hj}(d)$ for M, $h \neq j$ and all $d \varepsilon [0, D]$. Hence, we may restrict ourselves to the two possibilities $j' = j$ and $j' = M$.

Form the function $Q_j(d) = P_{jj}(d) - P_{Mj}(d)$. Then an optimum rule puts $j' = j$ if $Q_j(|s - N(s)|) \geqq 0$ and $j' = M$ otherwise. By the monotonicity condition (ii) it follows that $Q_j(d)$ is non-increasing on $[0, D]$. Since $Q_j(0)$ is necessarily non-negative, the set of values of $d \varepsilon [0, D]$ for which $Q_j(d)$ is non-negative must be an interval $[0, r_j)$ or $[0, r_j]$ for some unique number $r_j \varepsilon [0, D]$. Hence, an optimum rule is to take $j' = j$ or $j' = M$ according as $|s - N(s)| < r_j$ or $|s - N(s)| > r_j$. Since the set $\{s\}$ for which $|s - N(s)| = r_j$ has zero volume we may arbitrarily put such points in $\hat{A}_M$, and the proof of the theorem is complete. Note that if the $P_{ij}(d)$ functions are continuous on $[0, D]$ then the $Q_j(d)$ functions are continuous and r_j is the smallest root of $Q_j(d) = 0$; if this equation has no root in $[0, D]$ then $r_j = D$.

A definition analogous to (3) for the simple nearest-point rule can be written down for the optimum rule in the case where Theorem 4 applies. As in Section 2, let $S_1, \cdots, S_n$ be the regions of influence of the sample-points $s_1, \cdots, s_n$. Let

$T_{i,j}$ $(i = 1, \cdots, n; j = 1, \cdots, m)$ be the spherical region of radius r_j centered at the sample-point s_i. Then Theorem 4 says that an optimum nearest-point rule for reconstructing a pattern can be defined by

$$\hat{A}_j = \sum_{s_i \varepsilon A_j} S_i \cap T_{i,j}, \qquad j \neq M; \tag{18}$$

$$\hat{A}_M = A - A_0 - \sum_{j \neq M} \hat{A}_j .$$

This formulation is convenient for the statement of the next theorem.

THEOREM 5. *Let* $V_{i,j} = S_i \cap T_{i,j}$. *If the pattern is generated by a model with properties* (1) *and* (17), *then the risk EL for the optimum nearest-point rule is given by*

$$1 - \sum_{i=1}^{n} \sum_{j=1}^{m} p_j \{ \int_{V_{i,j}} P_{jj}(|s - s_i|)\, d\mu(s) + \int_{s_i - V_{i,j}} P_{Mj}(|s - s_i|)\, d\mu(s) \}. \tag{19}$$

PROOF. Since the $\{S_i\}$ are a partition of $A - A_0$, since $\{V_{i,j}, S - V_{i,j}\}$ is a partition of S_i for each j, and since $N(s) = s_i$ for all points $s \varepsilon S_i$, we can put formula (16) into the form

$$EL = 1 - \sum_{j=1}^{m} p_j \sum_{i=1}^{n} [\int_{V_{i,j}} P_{j'j}(|s - s_i|)\, d\mu(s) + \int_{s - V_{i,j}} P_{j'j}(|s - s_i|)\, d\mu(s)].$$

This last formula is valid for any nearest-point rule. But in Theorem 4 we showed that the optimum rule puts $j' = j$ for all points $s \varepsilon V_{i,j}$ and puts $j' = M$ for all points $s \varepsilon S_i - V_{i,j}$. Substituting these results into the last formula completes the proof of the theorem except for interchanging the order of the two summations.

If the sample-points are arranged so that for every i and every point $s \varepsilon S_i$ we have $|s - N(s)| \leqq \min_{j \neq M} r_j$, then $T_{i,j} \supset S_i$ for all i and all $j \neq M$. In this case the definition (18) of the optimum rule reduces to (3) which defines the simple nearest-point rule of Section 2. So we have a case where the simple and optimum rules coincide, and necessarily the expression (19) for the risk will reduce to (4). For a sufficiently large number of sample-points there will always be a sample-point arrangement with the above-stated property provided none of the r_j's are zero. (None of the r_j's will be zero provided the boundary of every A_j has finite expected length since in that case $P_{jj}(d)$ is continuous at $d = 0$; see Matérn [2].) More will be said later about the coincidence of these rules in the special situation of two-color patterns in the plane.

If the sample-points have a systematic arrangement as defined in Section 2, then formula (19) reduces to

$$EL = 1 - n \sum_{j=1}^{m} p_j \{ \int_{V_{0,j}} P_{jj}(|s|)\, d\mu(s) + \int_{s_0 - V_{0,j}} P_{Mj}(|s|)\, d\mu(s) \}, \tag{20}$$

where S_0 is the nearest-point region of influence of the sample-point located at the origin, and $V_{0,j} = S_0 \cap T_{0,j}$ where $T_{0,j}$ is a sphere of radius r_j centered at the origin.

Now we consider the specialization of the optimum nearest-point rule to the case $m = 2$, i.e., two-color patterns generated by models with properties (1) and

(17). It is easily checked that the properties (17) obtain provided only that $P_{11}(d)$ is non-increasing. Take $p_1 < \frac{1}{2}$ for definiteness, then $p_2 = 1 - p_1 > p_1$ and in our earlier notation $A_M = A_2$. The radius r_1 of the spherical regions $T_{i,1}$ is the supremum of the set of values for d such that $P_{11}(d) \geqq P_{21}(d) = 1 - P_{11}(d)$, i.e., the "largest" value for d such that $P_{11}(d) \geqq \frac{1}{2}$. The optimum rule for two-color patterns can therefore be stated as follows: A point $s \in A - A_0$ is placed in $\hat{A}_1$ if $N(s) \in A_1$ *and* $P_{11}\{|s - N(s)|\} \geqq \frac{1}{2}$; otherwise, it is placed in $\hat{A}_2$. Alternatively, $\hat{A}_1 = \sum_{s_i \in A_1} S_i \cap T_{i,1}$, $\hat{A}_2 = A - A_0 - \hat{A}_1$.

The risk EL for two-color patterns using the optimum rule can be put in the form

$$(21) \quad EL/p_1 = 1 - 2\sum_{i=1}^{n} \int_{V_{i,1}} [P_{11}(|s - s_i|) - \tfrac{1}{2}]\, d\mu(s); \qquad p_1 < \tfrac{1}{2}.$$

This expression can be derived by making the substitutions $p_2 = 1 - p_1$, $P_{22}(d) = 1 - p_1(1 - p_1)^{-1}[1 - P_{11}(d)]$, and $P_{21}(d) = 1 - P_{11}(d)$ in formula (19). The same substitutions made in formula (20) yield the corresponding risk for systematic sample-point arrangements, viz.,

$$(22) \qquad EL/p_1 = 1 - 2n \int_{V_{0,1}} [P_{11}(|s|) - \tfrac{1}{2}]\, d\mu(s); \qquad p_1 < \tfrac{1}{2}.$$

We indicated earlier that in the general m-color case the simple and optimum rules coincided under certain conditions. For the two-color case with $p_1 < \frac{1}{2}$ the condition is that no point in A be further than the distance r_1 from a sample point. When the two rules do not coincide then by definition the optimum rule will yield a smaller EL than the simple rule. However, the application of the optimum rule requires knowledge of the $P_{11}(d)$ function sufficient to determine the radius r_1. The next theorem shows what we might do if we are vague about r_1.

THEOREM 6. *For the two-color patterns generated by a random model with properties* (1) *and* (17) *with* $p_1 < \frac{1}{2}$ *and* $P_{11}(d)$ *non-increasing, let* r_1^* *be any overestimate of* r_1, *then the sub-optimum pattern reconstruction rule which is the same as the optimum rule with* r_1 *replaced by* r_1^* *never gives larger values of the risk* EL *than the simple rule.*

PROOF. Let $T_{i,1}^*$ be the spherical region of radius r_1^* centered at the sample-point s_i, and let $V_{i,1}^* = S_i \cap T_{i,1}^*$ where S_i is the region of influence of the point s_i as earlier defined. Then, by analogy with (21), the formula for the risk using the sub-optimum rule is

$$EL/p_1 = 1 - 2\sum_{i=1}^{n} \int_{V_{i,1}^*} H_i(s)\, d\mu(s),$$

where $H_i(s) = [P_{11}(|s - s_i|) - \frac{1}{2}]$. We can write

$$\int_{V_{i,1}^*} H_i(s)\, d\mu(s) = \int_{S_i} H_i(s)\, d\mu(s) - \int_{S_i - V_{i,1}^*} H_i(s)\, d\mu(s).$$

But since $P_{11}(d)$ is a non-increasing function and $r_1^* > r_1$ by assumption, $P_{11}(r_1^*) \leqq P_{11}(r_1) \leqq \frac{1}{2}$. Furthermore, all points $s \in S_i - V_{i,1}^*$ will be not less than distance r_1^* from s_i by the definition of $V_{i,1}^*$, hence, for all such points

$P_{11}(|s - s_i|) \leqq P_{11}(r_1^*) \leqq \frac{1}{2}$, i.e., $H_i(s) \leqq 0$ for $s \varepsilon S_i - V_{i,1}^*$. Therefore,

$$\int_{V_{i,1}^*} H_i(s)\, d\mu(s) \geqq \int_{S_i} H_i(s)\, d\mu(s)$$

and

$$EL/p_1 \leqq 1 - 2\sum_{i=1}^n \int_{S_i} H_i(s)\, d\mu(s) = 2 - 2\sum_{i=1}^n \int_{S_i} P_{11}(|s - s_i|)\, d\mu(s)$$

since $\sum \mu(S_i) = \mu(A - A_0) = \mu(A) = 1$. But the right-hand side is just the expression for EL/p_1 when the simple nearest-point rule is used and the theorem is thus proved. Note that if $P_{11}(r_1^*)$ is strictly less than $P_{11}(r_1)$ then the sub-optimum rule of the theorem has strictly smaller risk than the simple rule—provided these two rules do not coincide.

As earlier indicated the optimum and simple rules are likely to coincide when the number of sample-points n gets large, provided they are fairly evenly distributed over A. This occurs because, whereas r_1 is a fixed number not depending on n, the furthest distance to the nearest sample-point will generally shrink as n gets larger—and once this furthest distance falls below r_1 the two rules become equivalent.

On the other hand, for small n the minimum distance between any two sample-points may exceed $2r_1$ if the sample-points are fairly evenly distributed over A. It is easily seen that this is a sufficient condition for the spherical region $T_{i,1}$ (with radius r_1 centered at the sample-point s_i) to be wholly contained within S_i (the region of influence of s_i) for each $i = 1, \cdots, n$. For consider any point $s \varepsilon T_{i,1}$ then $|s - s_i| \leqq r_1$; let s_j be the sample-point nearest to s, then necessarily $|s - s_j| < |s - s_i| \leqq r_1$ whence $|s - s_j| + |s - s_i| \leqq 2r_1$; but if $i \neq j$ then by the triangle inequality $|s - s_j| + |s - s_i| \geqq |s_i - s_j| > 2r_1$ by assumption; therefore, we have a contradiction unless $i = j$ and we conclude that $s \varepsilon T_{i,1}$ implies $s \varepsilon S_i$, i.e., $S_i \supset T_{i,1}$ for all i.

In this case of large spacing between sample-points we have seen that $S_i \cap T_{i,1} = T_{i,1}$ for all i. So we see from (18) that the optimum rule with $p_1 < \frac{1}{2}$ is defined by taking $\hat{A}_1 = \sum_{s_i \varepsilon A_1} T_i$, i.e., $\hat{A}_1$ is a union of disjoint k-dimensional spheres of radius r_1 centered at those sample-points which are observed to fall in A_1. Formula (21) for the risk EL becomes

$$(23) \qquad EL/p_1 = 1 - 2\sum_{i=1}^n \int_{T_{i,1}} [P_{11}(|s - s_i|) - \tfrac{1}{2}]\, d\mu(s)$$
$$= 1 - [4n\pi^{k/2}/\Gamma(k/2)] \int_0^{r_1} [P_{11}(u) - \tfrac{1}{2}] u^{k-1}\, du.$$

It is interesting to observe that the risk is linear in the sample-size for this case. Furthermore, this last expression in no way depends on the arrangement of the sample-points, so we can conclude that if the spacing between sample-points is sufficiently large and if $P_{11}(d)$ is non-increasing, then the risk EL associated with the optimum rule ceases to depend on the sample-point arrangement in any other way.

We now return to briefly consider the more general loss function L' as defined in (10). For the case of two-color patterns it was shown in Theorem 2 how the

risk EL' could be directly obtained from EL when the simple nearest-point rule is used to construct the estimated pattern. A similar connection between EL' and EL exists also when the optimum rule is used with two color-patterns, viz.,

$$(24)\quad 2EL' = (L_{12} + L_{21} - L_{11} - L_{22})EL + (L_{12} - L_{21} + L_{11} - L_{22})p_1 + (-L_{12} + L_{21} + L_{11} - L_{22})np_1\mu(V_{0,1}) + 2L_{22},$$

a formula valid for systematic sample-point arrangements and pattern-generating models of the type (1) with $p_1 < \frac{1}{2}$; the region $V_{0,1}$ was defined in the explanation of formula (20). To prove (24) recall that $L' = L_{12}\mu(A_{12}) + L_{21}\mu(A_{21}) + L_{11}\mu(A_{11}) + L_{22}\mu(A_{22})$ where $A_{ij} = A_i \cap \hat{A}_j$; as in the proof of Theorem 2, it is easily shown that when the optimum rule is used with a systematic arrangement

$$(25)\quad \begin{aligned} E\mu(A_{21}) &= E\mu(A_{12}) - p_1[1 - n\mu(V_{0,1})], \\ E\mu(A_{11}) &= p_1 - E\mu(A_{12}), \\ E\mu(A_{22}) &= (1 - p_1) - E\mu(A_{21}); \end{aligned}$$

hence, EL' is linear in $E\mu(A_{12})$, say; but EL is the special case of EL' with $L_{11} = L_{22} = 0$ and $L_{12} = L_{21} = 1$, so EL is linear in $E\mu(A_{12})$ whence EL' is linear in EL; the particular linear relation (24) can be found by repeated use of (25).

This section is concluded with an example based on two-color patterns on a plane region A. Let the n sample-points be systematically arranged on the vertices of a square $n^{-\frac{1}{2}} \times n^{-\frac{1}{2}}$ grid; the nearest-point regions of influence will therefore all be $n^{-\frac{1}{2}} \times n^{-\frac{1}{2}}$ squares with a sample-point at the center of each. We will let the shape of A be altered so that this will always be possible, but A will always have unit area. For illustrative purposes, we will once again use the model (12), viz., $P_{11}(d) = p_1 + (1 - p_1)e^{-cd}$, $c > 0$. Since this $P_{11}(d)$ is non-increasing the main results of this section are applicable. Assume for definiteness that $p_1 < \frac{1}{2}$.

Solving the equation $P_{11}(r_1) = \frac{1}{2}$, we get

$$(26)\quad r_1 = (1/c)\log_e((1 - p_1)/(\tfrac{1}{2} - p_1)).$$

The furthest distance from an arbitrary point in A to a sample-point in A is easily seen to be $(2n)^{-\frac{1}{2}}$. Therefore, the simple rule becomes the optimum nearest-point rule provided the sample size is greater than $(2r_1^2)^{-1}$ where r_1 is given by (26).

For example, suppose we want a sample size n large enough so that the risk EL will be less than $p_1(1 - p_1)$. Then it may be surmised from Table I for the simple rule that this risk level can be achieved only if n is large enough to make the simple rule an optimum rule—for any p_1 greater than about $\frac{1}{3}$ and less than $\frac{1}{2}$. In fact, if we want to achieve a level for EL of $.6p_1(1 - p_1)$ or less, then for any p_1 the sample-size will always need to be so large as to make the simple and optimum nearest-point rules equivalent. To see this we first note from Table I that to achieve such values of EL the quantity $cn^{-\frac{1}{2}}$ must be less than .98 approximately if the simple rule is used, i.e., n must be greater than $1.04c^2$; now the maximum

value of $(2r_1^2)^{-1}$ is also approximately $1.04c^2$, so $n > 1.04c^2$ implies $n > (2r_1^2)^{-1}$ always; but as noted earlier $n > (2r_1^2)^{-1}$ makes the simple and optimum rules equivalent for this example.

Formula (23) for widely spaced sample-points specialized to this example becomes

$$EL/p_1 = 1 - (n\pi/c^2)[2 - r_1(1 - 2p_1)(2c + c^2r_1)] \tag{27}$$

where r_1 is given in (26). In fact, as pointed out earlier in a more general context, the expression (27) is valid for *any* arrangement of the n sample-points provided the distance between any pair of sample-points is always greater than $2r_1$. The superiority of the optimum rule compared to the simple rule is most pronounced in this case of widely spaced sample-points, the difference being more noticeable as p_1 gets smaller. Table III exhibits a comparison of the two rules for this example.

TABLE III

Values of EL(optimum rule)/EL(simple rule) for n points on a $n^{-\frac{1}{2}} \times n^{-\frac{1}{2}}$ square grid, using model (12)

$cn^{-\frac{1}{2}}$	$p_1 \to 0$	$p_1 = .10$	$p_1 = .20$
2.00	.87	.93	.98
3.00	.73	.80	.88
4.00	.65	.72	.80

5. Some models of random patterns. Models for patterns of more than two colors seem to have received little or no previous attention. We will therefore begin with some examples of models for two-color patterns.

First, we have the so-called "bombing" models, an example of which can be specified as follows: Points (bombs) are dropped by a Poisson process with intensity δ into k-dimensional Euclidean space E_k and each such point generates (destroys) a spherical region of fixed radius r with itself at the center. By this procedure, a fixed region $A \subset E_k$ is randomly partitioned into $\{A_1, A - A_1\}$ where A_1 consists of that part of A which is contained in the union of spherical regions generated by the points of the Poisson process, i.e., A_1 is the bombed-out portion of A. For this model $\Pr(s \in A_1) = p_1 = 1 - \exp\{-\delta r^k C_k\}$ identically for points $s \in A_1$, and

$$\Pr(s' \in A_1, s \in A_1) = p_1 P_{11}(d) = (2p_1 - 1) + (1 - p_1)^{H(k,d/r)} \tag{28}$$

identically for all pairs of points $s, s' \in A$ with $|s - s'| = d$, where $H(k, u) = 2 - C^*(k, u)/C_k$, C_k is the volume of a unit radius k-sphere, and $C^*(k, u)$ is the volume common to two-unit k-spheres when their centers are distance u apart. A version of formula (28) has been given essentially by Matérn [2]. For any fixed r and δ, $P_{11}(d)$ decreases continuously for $d \leqq 2r$, and for $d \geqq 2r$, $P_{11}(d) = p_1$. In one dimension $C_1 = 2$, $C^*(1, u) = 2 - u$ for $u \leqq 2$, and $C^*(1, u) = 0$ for

$u \geqq 2$; in two dimensions $C_2 = \pi, C^*(2, u) = \pi - u(1 - u^2/4)^{\frac{1}{2}} - 2 \arcsin(u/2)$ for $u \leqq 2$, and $C^*(2, u) = 0$ for $u \geqq 2$.

Since this bombing model does not treat A_1 and A_2 symmetrically, it is not clear how one should generalize it to m-color patterns. Another two-color model discussed by Matérn [2] does treat A_1 and A_2 symmetrically and can be easily generalized to m colors. This model, which we will call an "occupancy" model, can be specified as follows: Two Poisson processes drop points into E_k ; one process with intensity δ_1 drops black points and the other process, with intensity δ_2, drops white points. Each of these Poisson points spreads out to occupy a cell consisting of all points in E_k which are nearer to it than to any other one of the Poisson points. This procedure randomly partitions the fixed region $A \subset E_k$ into $\{A_1, A - A_1\}$ where A_1 is the subregion of A occupied by black cells and $A_2 = A - A_1$ is the subregion occupied by white cells. (Points on cell boundaries can be assigned to A_1 to complete the specification of the model.) Equivalently, a Poisson process with intensity $\delta = \delta_1 + \delta_2$ drops points in E_k ; the points induce a partition of E_k into cells in the manner described and the cells are then independently colored black or white with fixed probabilities δ_1/δ and δ_2/δ, respectively; then A_1 is the intersection of A with the union of black cells. For every point $s \varepsilon A$, $\Pr(s \varepsilon A_1) = \delta_1/\delta = p_1$. For every pair of points $s, s' \varepsilon A$, $\Pr(s \varepsilon A_1, s' \varepsilon A_1) = p_1 \Pr(s, s' \text{ in same cell}) + p_1^2 \Pr(s, s' \text{ not in same cell})$. It is easily seen that $\Pr(s, s' \text{ in same cell})$ depends on s, s' only through $|s - s'| = d$, say, so we can write

$$\Pr(s \varepsilon A_1 \mid s' \varepsilon A_1) = P_{11}(d) = p_1 + (1 - p_1)W(d) \tag{29}$$

where $W(d)$ is the probability that any two points which are d apart are both in the same cell. An expression for $W(d)$ is given by Matérn, but it has an elementary form only in the one-dimensional case, viz., $W(d) = (1 + \delta d) \exp\{-2\delta d\}$.

A generalization of the two-color occupancy model to m colors is immediately obvious. A partition of E_k into cells is induced by a Poisson point process in the manner previously described. The cells are then each independently given one of m colors according to a fixed set of probabilities $p_1, p_2, \cdots, p_m$. The random subregion A_j of the fixed region $A \subset E_k$ is defined to be the intersection of A with the union of j-colored cells. For this model, $p_j = \Pr(s \varepsilon A_j)$ for every point $s \varepsilon A$, and for every pair of points $s, s' \varepsilon A$ with $|s - s'| = d$,

$$\Pr(s' \varepsilon A_j \mid s \varepsilon A_j) = P_{jj}(d) = p_j + (1 - p_j)W(d)$$

and

$$\Pr(s' \varepsilon A_i \mid s \varepsilon A_j) = P_{ij}(d) = p_i[1 - W(d)], \qquad i \neq j. \tag{30}$$

Note that $P_{ij}(d)$ does not depend on j.

Actually, the occupancy model can be conceived as one of a large class of models for m-color patterns which we will call "cell-structure" models. All such models begin with a random countable partition of E_k into cells. The partitioning

process must have the property that for any two points in the fixed region $A \subset E_k$, the probability $W(d)$ that they are both in the same cell of the partition may depend only on the distance d between them. Each cell is then independently given one of m colors according to a fixed set of probabilities $p_1, \cdots, p_m$, $\sum p_i = 1$. The random subregion A_j is defined to be the intersection of A with the union of j-colored cells. It is clear that formulas (30) hold for *any* model of the cell-structure type.

THEOREM 7. *If an optimum nearest-point rule is used to reconstruct an m-color pattern on a region A generated by a cell-structure model in which all the cells are convex, then the risk EL can be expressed as*

$$(31)\quad (1 - p_M) - \sum_{j \neq M} p_j \sum_{i=1}^{n} \{\int_{V_{i,j}} W(|s - s_i|)\, d\mu(s) - (p_M - p_j) \cdot \int_{V_{i,j}} [1 - W(|s - s_i|)]\, d\mu(s)\},$$

where to review some notation, p_M is the assumed unique maximum of the $\{p_i\}$, the $\{s_i\}$ are the sample-point locations in A, and $V_{i,j}$ is the region $S_i \cap T_{i,j}$—where S_i is the set of points in A nearer to s_i than to any other sample-point, $T_{i,j}$ is the spherical region of radius r_j centered at s_i, and r_j is the largest value of d for which $P_{jj}(d) \geqq P_{Mj}(d)$.

PROOF. The expression (31) is obtained upon simplification by direct substitution of (30) into the expression (19) of Theorem 5. This substitution is valid provided that the $P_{ij}(d)$ functions of (30) satisfy the conditions (i) and (ii) of (17).

Since for any cell-structure model $P_{ij}(d) = p_i[1 - W(d)]$ for all $i \neq j$ and all d, it is clear that the uniform preference condition (i) is satisfied whether or not the cells are convex. Convexity is introduced here as a simple sufficient condition for the monotonicity property (ii). It will be enough to show that convexity of the cells implies that $W(d)$ is non-increasing since $P_{ij}(d)$ is linear in $W(d)$ for all i, j and for all cell-structure models.

Let a, b, c be three distinct collinear points in A, let $|a - b| = d_1$, $|b - c| = d_2$, and $|a - c| = d_1 + d_2 = d_0$, and let xy denote the event that the points x and y are both in the same cell of the random partition of A. By convexity of the cells $P(ac) = P(ab, bc)$. Then provided $W(d_1) > 0$, $W(d_1 + d_2) = P(ac) = P(ab, bc) = P(bc|ab)W(d_1) \leqq W(d_1)$. If $W(d_1) = 0$ then $P(ab) = 0$ so $W(d_1 + d_2) = P(ab, bc) = 0$. Hence, $W(d_1 + d_2) \leqq W(d_1)$ for every $d_1, d_2 \geqq 0$, i.e., $W(d)$ is non-increasing and the proof of the theorem is complete.

In addition to the occupancy model discussed earlier, we now give two further examples of cell-structure models both for patterns in a plane region. In the first example the plane is partitioned checkerboard style into $h \times h$ squares; the origin of the plane is chosen with uniform probability over any $h \times h$ square and the checkerboard is then rotated through an angle chosen with uniform probability on $(0, 2\pi)$. The cells are the squares of the checkerboard and the cells are now independently colored with fixed probabilities. For this cell-structure model of

the checkerboard type

$$
\begin{aligned}
W(d) &= 1 - \pi^{-1}(4\,d/h - d^2/h^2), && d \leqq h, \\
&= 1 - \pi^{-1}\{2 + 4\cos^{-1}(h/d) + d^2/h^2 - 4(d^2/h^2 - 1)^{\frac{1}{2}}\}, && h \leqq d \leqq h2^{\frac{1}{2}}, \\
&= 0, && d \geqq h2^{\frac{1}{2}}.
\end{aligned} \tag{32}
$$

In the second example the plane is partitioned into cells by "Poisson" lines defined thus: A Poisson point process with intensity δ chooses points (θ, p) in the infinite rectangular strip $0 \leqq \theta < \pi$, $-\infty < p < \infty$; each point so chosen corresponds to a "Poisson" line $x \cos\theta + y \sin\theta - p = 0$ given in terms of an (x, y) Cartesian co-ordinate system. This method of partitioning the plane is explored by Kendall and Moran [1]. The resulting cells are then independently colored with fixed probabilities. For this cell-structure model of the Poisson type it can be shown that

$$W(d) = e^{-2\delta d}. \tag{33}$$

This gives $P_{11}(d) = p_1 + (1 - p_1)e^{-2\delta d}$, which is precisely the model used for the illustrative comparisons of Sections 2 and 3 involving two-color patterns. So we see that patterns with this $P_{11}(d)$ function can actually be generated.

6. Conclusion. We have been principally concerned with the risk structure for the problem of reconstructing a pattern from point observations. Some sample design comparisons were made. A somewhat obvious method for improving the common nearest-point rule of reconstructing a pattern was proposed and shown to be optimum in a suitably restricted context. Some pattern-generating models were briefly explored.

We did not at all discuss the statistical problem of estimating the risk in the case where the generating model is not completely specified. For example, if we use the exponential model (12) for two-color patterns then we may wish to leave p_1 and c as free parameters to be estimated from the data. Some tentative computer sampling experiments have yielded estimates of the risk, based on naive data estimates of p and c, which were surprisingly close to the actual loss even for moderate sample sizes. But no attempt has yet been made on the general estimation problem.

Another neglected aspect with important potential applications concerns the use of designs in which the sample-point locations are chosen in some sequential manner, e.g., a two-stage design.

REFERENCES

[1] KENDALL, M. G. and MORAN, P. A. P. (1963). *Geometrical Probability*. Hafner, New York.
[2] MATÉRN, B. (1960). Spatial variation. *Medd Statens Skogsforskningsinstit.* **49** 1–144.
[3] ROBBINS, H. E. (1944). On the measure of a random set. I. *Ann. Math. Statist.* **15** 70–74.
[4] SWITZER, P. (1965). A random set process with a Markovian property. *Ann. Math. Statist.* **36** 1859–1863.