Journal of Mathematical Imaging and Vision, 4, 273–290 (1994).

Algebraic Analysis of the Generating Functional for Discrete Random Sets and Statistical Inference for Intensity in the Discrete Boolean Random-Set Model*

N.D. SIDIROPOULOS AND J.S. BARAS
Systems Research Institute for and Department of Electrical Engineering, University of Maryland, College Park, MD 20742

C.A. BERENSTEIN
Systems Research Institute for and Department of Mathematics, University of Maryland, College Park, MD 20742

Abstract. We consider binary digital images as realizations of a uniformly bounded discrete random set, a mathematical object that can be defined directly on a finite lattice. In this setting we show that it is possible to move between two equivalent probabilistic model specifications. We formulate a restricted version of the discrete-case analog of a Boolean random-set model, obtain its probability mass function, and use some methods of morphological image analysis to derive tools for its statistical inference.

Key words. image modeling and analysis, Boolean model and random sets, mathematical morphology, morphological skeleton, decision theory

1 Introduction

The *Boolean random set* is arguably the most important random-set model to date. It has received considerable attention in the literature (for example, see [2], [5], [6], [10], [11], [25], [26], [32] and the references therein), in terms of both theory and practice. Typical applications include, but are not limited to, random clumping of dust or powder particles; modeling of geological structures, bomb fields, patterns in photographic emulsion, colloids in gel form, and structural inhomogeneities in amorphous matter [32, p. 68]; tumor growth [5]; and the spatial pattern of heather [6]. Other potential applications include particle counting and size analysis in images of cell cultures and modeling of clutter in infrared imaging.

Informally, a Boolean random set is constructed by centering a simple random shape (set), such as a disc of random size, at each point of a Poisson field of points in the plane and then taking the union of the resulting sets. Random shapes centered at different points of the Poisson field are assumed to be independent and statistically equivalent. The points of the Poisson field are known as the *germs*, and the Poisson field itself is sometimes called the *germ process*. The random shapes are known as the *primary grains*.

When it comes to modeling, random-set recipes are hard to come by. To quote Cressie and Laslett [5], "The choice of mathematical models available to the data analyst is often governed by their tractability, rather than their applicability. When the data are sets, this leaning is even more profound." The Boolean model is important because it is one of the precious few random-set models that are both tractable and applicable. In many applications there exists no physical interpretation of the germ-grain construction; in others there is strong evidence that the germs and the grains correspond to actual physical entities (e.g., modeling of bomb fields, where the germs correspond to the points

*This research was partially supported by National Science Foundation grant NSFD CDR 8803012 through the Engineering Research Centers program.

of impact and the primary grains model random dispersion). In the former case, the Boolean model is simply used as a device to generate and analyze randomness; in the latter, it also provides some insight into the mechanism of image-data generation.

The transition from continuous-domain random sets to discrete-domain random sets is troublesome [26]. In practice, one usually deals with finitely many samples of a portion of a binary image that is contained within a fixed window (i.e., a realization of a uniformly bounded discrete random set). Most automated image-analysis systems operate on finite spaces. Given the current rate of advances in digital computers, it seems fair to say that this trend will only continue to grow in the foreseeable future.

There are essentially three ways to introduce discrete random sets. Before we present our choice, let us discuss the alternatives. The first is through sampling (discretization) of continuous-domain random sets [20]. This poses several technical problems. For example, sampling with a regular lattice introduces lattice-dependent artifacts (e.g., the sampling does not preserve the Euler–Poincare characteristic [25]). In general, it is impossible to say anything about several important features (e.g., convexity) of the underlying continuous-domain structure, solely on the basis of its sampled realizations. The second approach is based on the theory of random point fields [22]. A special class of random point fields, namely, Markov random fields (MRFs), has been successfully used to model random texture. However, these models cannot easily describe complicated geometrical structure; i.e., they generally fail to capture the morphological aspects of image data.[1] Random-set theory, on the other hand, is closely related to mathematical morphology, a nonlinear image algebra that specifically addresses the problem of quantitative shape description. As a result of this connection, random-set theory provides a unified framework that allows the modeling of both morphological (syntactical) *and* statistical characteristics of images. Thus the need to develop a theory for discrete random sets along the lines of the corresponding continuous-domain theory becomes apparent.

We have chosen to define uniformly bounded discrete random sets directly on a finite lattice and to base subsequent developments on this definition. This axiomatic approach has many advantages. It avoids technicalities and enables us to focus on problems that are important in practice. In particular, it allows us to talk about probability mass and inference. Certain important results of random-set theory can be significantly strengthened in the case of uniformly bounded discrete random sets. Of course, these benefits come at a certain price. In this case we ignore the details of the underlying continuous physical structure that fall beneath our resolution. However, at any rate, this loss of detail is forced on us by the limitations of the digital imaging system; we might as well accept it and live with it. Our findings suggest that, from the point of view of applications, such an axiomatic approach is significantly more flexible.[2]

The cornerstone of statistical image analysis is the development of models that summarize the most important characteristics of images by a few parameters. Once we have a bagful of discrete random-set models at our disposal, the obvious next step is to fit the models to the data. Parametric models whose parameters cannot be fitted to the data are useless for image analysis (although they can be of some use in image synthesis and computer graphics). Thus enters the need for statistical inference. Statistical-inference techniques are very scarce in random-set theory. Most of the existing literature is concerned with parameter estimation for the Boolean model, based on some variant of the method of moments [5], [6], [11], [23], [25], [26].[3] This can be partially attributed to the lack of a total set ordering, which makes for an extremely complicated "distribution function." The method of moments, although well known and widely practiced, is largely considered to be a last resort, when maximum-likelihood (ML) or *maximum-a-posteriori* (MAP) inference is not possible. We will show that within a uniformly bounded discrete-random-set framework we can make some progress towards the ML inference of the Boolean model. We develop an axiomatic formulation of uniformly bounded, discrete random sets, strengthen a general characterization

theorem of Choquet, Kendall, and Matheron, and consider a restricted version of the discrete case analog of the Boolean random set and develop some tools for its statistical inference.

2 Fundamentals of Discrete Random Sets

DEFINITION 1. Let B be a bounded subset of $\mathbf{Z}^2$. Assume that B contains the origin. Let $\Sigma(\Omega)$ denote the σ-algebra on Ω. Let $\Sigma(B)$ denote the power set (i.e., the set of all subsets) of B, and let $\Sigma(\Sigma(B))$ denote the power set of $\Sigma(B)$. A uniformly bounded discrete random set, or, for brevity, *discrete random set* (DRS), X on B is a measurable mapping of a probability space $(\Omega, \Sigma(\Omega), P)$ into the measurable space $(\Sigma(B), \Sigma(\Sigma(B)))$. A DRS X on B induces a unique probability measure P_X on $\Sigma(\Sigma(B))$.

DEFINITION 2. The functional

$$T_X(K) = P_X(X \cap K \neq \emptyset), \qquad K \in \Sigma(B),$$

is called the *capacity functional* of the DRS X.

DEFINITION 3. The functional

$$Q_X(K) = P_X(X \cap K = \emptyset) = 1 - T_X(K), \qquad K \in \Sigma(B),$$

is called the *generating functional* of the DRS X.

The generating functional is to the DRS what the cumulative distribution function (cdf) is to scalar discrete random variables. The following lemma will be useful.

LEMMA 1 (variant of Möbius inversion for Boolean algebras; see [1] for basic Möbius inversion). Let v be a function on $\Sigma(B)$. Then v can be represented as

$$v(A) = \sum_{S \subseteq A^c} u(S); \qquad \text{"external decomposition."}$$

The function u is uniquely determined by v; namely,

$$u(S) = \sum_{C \subseteq S} (-1)^{|C|} v(S^c \cup C); \qquad \text{"inversion."}$$

where the superscript c denotes complement with respect to B.

Proof. Uniqueness: Assume that the external decomposition formula holds. The right-hand side of the inversion formula may be written

$$\begin{aligned}
&\sum_{C \subseteq S} (-1)^{|C|} v(S^c \cup C) \\
&\quad = \sum_{C \subseteq S} (-1)^{|C|} \sum_{D \subseteq S \cap C^c} u(D) \\
&\quad = \sum_{C \subseteq S} (-1)^{|C|} \sum_{D \subseteq S \setminus C} u(D) \\
&\quad = \sum_{C \subseteq S} \sum_{D \subseteq S \setminus C} (-1)^{|C|} u(D) \\
&\quad = \sum_{D \subseteq S} \sum_{C \subseteq S \setminus D} (-1)^{|C|} u(D) \\
&\quad = \sum_{D \subseteq S} u(D) \sum_{C \subseteq S \setminus D} (-1)^{|C|} = u(S)
\end{aligned}$$

since

$$\sum_{C \subseteq S} (-1)^{|C|} = \begin{cases} 0, & S \neq \emptyset, \\ 1, & S = \emptyset. \end{cases}$$

Existence: Assume that the inversion formula holds. The right-hand side of the external decomposition formula may be written

$$\begin{aligned}
\sum_{S \subseteq A^c} u(S) &= \sum_{S \subseteq A^c} \sum_{C \subseteq S} (-1)^{|C|} v(S^c \cup C) \\
&= \sum_{S \subseteq A^c} \sum_{C \subseteq S} (-1)^{|C|} v((S \setminus C)^c) \\
&= \sum_{D \subseteq A^c} \sum_{C \subseteq A^c \setminus D} (-1)^{|C|} v(D^c) \\
&= \sum_{D \subseteq A^c} v(D^c) \sum_{C \subseteq A^c \setminus D} (-1)^{|C|} \\
&= v((A^c)^c) = v(A),
\end{aligned}$$

as for the uniqueness part.

THEOREM 1. Given $Q_X(K), \forall K \in \Sigma(B)$, $P_X(A), \forall A \in \Sigma(\Sigma(B))$, is uniquely determined and, in fact, can be recovered by using the measure-reconstruction formulas

$$P_X(A) = \sum_{K \in A} P_X(X = K),$$

with

$$P_X(X = K) = \sum_{K' \subseteq K} (-1)^{|K'|} Q_X(K^c \cup K').$$

Proof. The functional Q_X can be expressed in terms of P_X as

$$Q_X(K) = \sum_{K' \subseteq K^c} P_X(X = K').$$

This observation, along with Lemma 1, establishes the validity of the theorem.

The *uniqueness* part of this theorem, originally due to Choquet [4], has been independently introduced in the context of continuous-domain random-set theory by Kendall [16] and Matheron [19], [20]. Related results were also obtained by Ripley [22]. However, the measure-reconstruction formulas are essentially applicable only within a uniformly bounded DRS formulation. In the case of (uncountably or countably) infinite observation sites, the uniqueness result relies heavily on Kolmogorov's extension theorem, which is nonconstructive.

3 Mathematical Morphology

Let H be a small, primitive subset of B, which contains the origin, $\{\bar{0}\}$. In the terminology of mathematical morphology[4] H is called a *structuring element*. For our purposes H will be assumed to be convex.[5] Let X_h denote the translate of X by the vector h, and let

$$H^s = \{-h \mid h \in H\}.$$

DEFINITION 4. The *erosion* $X \ominus H^s$ of a set $X \subset \mathbf{Z}^2$ by a structuring element H is defined as

$$X \ominus H^s = \bigcap_{h \in H} X_{-h} = \{z \in \mathbf{Z}^2 \mid H_z \subseteq X\}.$$

DEFINITION 5. The *dilation* $X \oplus H^s$ of a set $X \subset \mathbf{Z}^2$ by a structuring element H is defined as

$$X \oplus H^s = \bigcup_{h \in H} X_{-h} = \{z \in \mathbf{Z}^2 \mid H_z \cap X \neq \emptyset\}.$$

DEFINITION 6. The *opening* $X \circ H$ of a set $X \subset \mathbf{Z}^2$ by a structuring element H is defined as

$$X \circ H = (X \ominus H^s) \oplus H = \bigcup_{z \in \mathbf{Z}^2 \mid H_z \subseteq X} H_z.$$

The opening is an idempotent (stable) operation in the sense that $(X \circ H) \circ H = X \circ H$. A set $X \subset \mathbf{Z}^2$ is said to be H-open if and only if $X \circ H = X$. The set X is H-open if and only if it can be written as $X = L \oplus H$ for some $L \subset \mathbf{Z}^2$ [20].

4 DRS Analog of the Boolean Model

Let us now define the Boolean DRS. In doing so, we proceed by analogy with the continuous case. We first define the class of germ-grain DRSs. Then we refine this class to derive the Boolean DRS. Our work focuses on a particular class of Boolean DRS models and is closely related to the morphological analysis of shape and the associated theory of shape–size distributions [19], [20].

PROPOSITION 1. Let $\{X_i\}_{i=0}^{\infty}$ be a sequence of (not necessarily independent) DRSs on B. Then for any finite N, $\bigcup_{i=0}^{N} X_i$ is a DRS on B and, furthermore, the limit $\bigcup_{i=0}^{\infty} X_i$ always exists and is a DRS on B.

Proof. Observe that the space of realizations $\Sigma(B)$ is itself a σ-algebra. The result follows easily.

DEFINITION 7. Let Ψ be a DRS on B, and let $\{G_1, G_2, \ldots\}$ be a set of nonempty independent and identically distributed DRSs on B, characterized by the generating functional Q_G. Define

$$X = \bigcup_{i=1,2,\ldots} G_i \oplus \{y_i\},$$

where $\Psi = \{y_1, y_2, \ldots\}$. Then X will be called a *germ-grain DRS*. The points $\{y_1, y_2, \ldots\}$ will be called the germs, and the DRSs $\{G_1, G_2, \ldots\}$ will be called the primary grains of the germ-grain DRS X.

Remark. For brevity, we assume from this point on that the result of a $\oplus$ operation is automatically restricted to B. Therefore since $\{y_i\}$ is a singleton, $G_i \oplus \{y_i\}$ is simply the translate of G_i by the vector y_i, restricted to B. Also, by convention, the superscript c denotes complement with respect to B.

In general, we cannot compute the generating functional Q_X of the germ-grain DRS X in terms of the generating functional Q_G, which characterizes the primary grains. This is a significant drawback. Nevertheless, this computation is possible for a restricted class of germ-grain DRS models.

DEFINITION 8. Let Ψ be a generalized Bernoulli lattice process (or Bernoulli DRS or binary Bernoulli random field) on B, constructively defined in the following manner: each point $z \in B$ is contained in Ψ with probability $\lambda_s(z)$, independently of all others. Let $\{G_1, G_2, \ldots\}$ be a set of nonempty independent and identically distributed DRSs on B, characterized by the generating functional Q_G. Define

$$X = \bigcup_{i=1,2,\ldots} G_i \oplus \{y_i\},$$

where $\Psi = \{y_1, y_2, \ldots\}$. Then X will be called a *discrete Boolean random set* (DBRS) and will be denoted by (λ_s, Q_G)-DBRS. The function λ_s will be called the intensity function (or simply the *intensity*) of both the DBRS and the underlying Bernoulli lattice process.

PROPOSITION 2. Let $\{X_1, X_2, \ldots, X_N\}$ be an independent sequence of nonempty DRSs, characterized by the generating functional $Q_{X_1}, \ldots, Q_{X_N}$, respectively. Define

$$Y = \bigcup_{i=1}^{N} X_i.$$

Then

$$Q_Y(K) = \prod_{i=1}^{N} Q_{X_i}(K) \qquad \forall K \in \mathcal{E}(B).$$

Proof.

$$\begin{aligned} Q_Y(K) &= \Pr(Y \cap K = \emptyset) \\ &= \Pr(X_1 \cap K = \emptyset, \ldots, X_N \cap K = \emptyset) \\ &= \Pr(X_1 \cap K = \emptyset) \cdots \Pr(X_N \cap K = \emptyset) \\ &= \prod_{i=1}^{N} Q_{X_i}(K), \end{aligned}$$

and the proof is complete.

PROPOSITION 3. The generating functional of a (λ_s, Q_G)-DBRS X is given by

$$\begin{aligned} Q_X(K) &= \prod_{z \in B} [1 - \lambda_s(z) + \lambda_s(z) Q_{G \oplus \{z\}}(K)] \\ &= \prod_{z \in B} [1 - \lambda_s(z) + \lambda_s(z) Q_G(K \oplus \{-z\})]. \end{aligned}$$

Proof. The proof follows from independence and from Proposition 2.

In the discrete case the notion of size of a convex structuring element can be normalized through the operation of set dilation:

$$rH = \begin{cases} \{\bar{0}\} \oplus H \oplus H \oplus \cdots \oplus H & \\ (r \text{ times}) & r = 1, 2, \ldots, \\ \{\bar{0}\}, & r = 0. \end{cases}$$

DEFINITION 9. Let Ψ be a generalized Bernoulli lattice process on B, of intensity $\lambda_s(z)$. Let $\{G_1, G_2, \ldots\}$ be a set of nonempty, convex, independent and identically distributed DRSs on B, each given by $G_i = R_i H$, where $\{R_1, R_2, \ldots\}$ form an sequence of $\mathcal{Z}_+$-valued random variables that is independent of Ψ and each R_i is distributed according to a probability mass function (pmf) $f_R(r)$ that is compactly supported on $\{0, 1, \ldots, \bar{R}\}$. Define

$$X = \bigcup_{i=1,2,\ldots} G_i \oplus \{y_i\},$$

where $\Psi = \{y_1, y_2, \ldots\}$. Then X will be called a *discrete radial Boolean random set* (*DRBRS*) and will be denoted by (λ_s, H, f_R)-DRBRS.

We now proceed to compute the generating functional of a (λ_s, H, f_R)-DRBRS. Define

$$d^H(z, K) = \min_{k \in K} \|z - k\|_H,$$

where

$$\|z - k\|_H = \min\{n \geq 0 \,|\, (\{z\} \oplus nH) \cap \{k\} \neq \emptyset\}.$$

Observe that for $z \in K$, $d^H(z, K) = 0$, since H contains the origin. By independence of the germs and the grains it follows that a DRBRS X is a union of independent DRSs. Thus by Proposition 3 it follows that $Q_X(K)$ is of product form, the product index ranging over all primary grains that can hit K, i.e., grains that are centered at $z \in K \oplus \overline{R}H^s$ (observe that $\{z\} \oplus rH \cap K \neq \emptyset$ if and only if $z \in K \oplus rH^s$ and that $\overline{R}$ is the maximum possible grain radius). For each such z the corresponding product term is the probability that the primary grain centered at z does not hit K. This probability is equal to the sum of two terms: the probability that the particular grain does not appear at all $(1 - \lambda_s(z))$ plus the probability of it appearing but not hitting K, i.e., appearing with a smaller radius than is needed to hit $K(\lambda_s(z)F_R(d^H(z, K) - 1))$. It follows that,

$$Q_X(K) = \prod_{z \in K \oplus \check{R}H^s} [(1 - \lambda_s(z)) + \lambda_s(z)F_R(d^H(z, K) - 1)],$$

where

$$F_R(m) = \sum_{l=0}^{m} f_R(l)$$

and $F_R(-1) = 0$ by convention.

We can now use Theorem 1 to compute $P_X(X = K)$ in terms of the model parameters.

$$P_X(X = K) = \sum_{K' \subseteq K} (-1)^{|K'|} \prod_{z \in (K^c \cup K') \oplus \check{R}H^s} [(1 - \lambda_s(z)) + \lambda_s(z)F_R(d^H(z, K^c \cup K') - 1)].$$

5 Statistical Inference of the DRBRS Model

Even though we have been able to write down an expression for the likelihood function, we are still faced will a complicated formula that is difficult to work with. In particular, largely because of the highly oscillatory Möbius kernel $(-1)^{|K'|}$, it is not directly amenable to optimization, which immediately rules out direct ML parameter estimation. Furthermore, the computational complexity associated with a brute-force calculation of the likelihood is exponential in $|K|$. One would therefore be interested in obtaining tight bounds on $P_X(X = K)$. To be useful these bounds must be reasonably well behaved and relatively easy to compute. For the simple case of a DRBRS model of constant intensity $\lambda_s(z) = p = 1 - q \ \forall z \in B$, and of primary grains of fixed size (one, by convention), the generating functional is simply given by

$$Q_X(K) = q^{|K \oplus H^s|}$$

We have the following result for this model.

PROPOSITION 4. For all $q \in [0, 1]$ and all *realizable*[6] $K \in \Sigma(B)$, $K \neq \emptyset, B$,

$$L_q(K) \leq P_X(X = K) \leq U_q(K),$$

with

$$L_q(K) = q^{|K^c \oplus H^s|}(1 - q)^{|(K^c \oplus H^s)^c|}$$

and

$$U_q(K) = \frac{1}{2}q^{|K^c|}\left[(1 + q)^{|K|} + (1 - q)^{|K|}\right] - 2^{|K|-1}q^{|K^c \oplus H^s| + |K \oplus H^s|}.$$

Both bounds are *polynomials* in q, they are equal to zero at the endpoints $q = 0, 1$, they are strictly positive for all $q \in (0, 1)$, and they are *unimodal* in $(0, 1)$. The mode of the lower bound is located at

$$\hat{\underline{q}}(K) = \frac{|K^c \oplus H^s|}{|B|}.$$

Proof. See appendix A.

The measurements $|K|$, $|K \oplus H^s|$, and $|K^c \oplus H^s|$ can be interpreted as crude statistical summaries of the observation. Bounds of this type can be used to obtain a *feasible region* of the ML estimate of q, given the observation K. Since

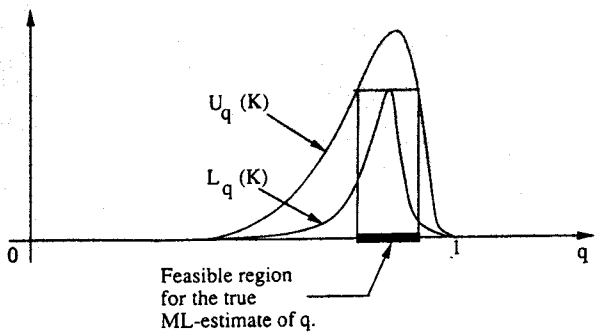


Fig. 1. The concept of the feasible region.

these bounds are typically very-high-degree polynomials; their (unique) modes are very sharp, leading to very accurate localization of the true ML estimate of q. Specifically, since the bounds are *unimodal polynomials*, the ML estimate of q must be within the closed interval of q-values delimited by the two q-values at which the upper bound is equal to the peak of the lower bound. This situation is illustrated in figure 1. The same technique can also be used for binary hypothesis testing between two values of q. If these two values are sufficiently far apart, the bounds will indicate that the corresponding likelihoods are in disjoint regions, in which case a decision procedure based on these bounds is as good as one based on the exact likelihood but is much faster. We will return to hypothesis testing later.

The mode $\hat{\underline{q}}(K)$ of the lower bound $L_q(K)$ *underestimates* q; i.e., $\hat{\underline{q}}(X)$ is a *biased* estimator of q on the basis of the observation X. This can be seen as follows.

$$E\hat{\underline{q}}(X) = \frac{E|X^c \oplus H^s|}{|B|}.$$

Let L denote the germ-point process. Then $X^c \oplus H^s \subseteq L^c$. Thus

$$|X^c \oplus H^s| \leq |L^c|$$

and

$$E|X^c \oplus H^s| \leq E|L^c|.$$

So

$$E\hat{\underline{q}}(X) \leq \frac{E|L^c|}{|B|} = \frac{q|B|}{|B|} = q.$$

The computation of the exact bias of the mode of the lower bound is difficult because of enumeration problems. However, this bias can be learned in an unbiased and consistent manner from an independent and identically distributed sequence of training samples of the DRBRS under consideration. In any case, and exactly because of this nonnegative bias, the mode of the lower bound per se is not a good estimator of q. It underestimates q in a manner analogous to that in which the so-called connected-component estimator (CCE) overestimates q (i.e., the former estimator overfits grains to the data, just as the later estimator underfits grains to the data). The resulting bias in $\hat{\underline{q}}$ is small in absolute terms, but it can be significant in relative terms. This is why we prefer the fuzzy answer provided by the feasible *region*, rather than choosing a particular representative *value* (e.g., $\hat{\underline{q}}$) for q. The CCE of q on the basis of X is explicitly given by

$$\hat{\bar{q}}(X) = \frac{|B| - CC(X)}{|B|},$$

where $CC(X)$ is the number of connected components of X. Obviously, this estimator overestimates q, because the number of germs (points) of any particular realization of L is always greater than or equal to the number of connected components of the corresponding realization of $X = L \oplus H$. Similarly, the mode of the lower bound provides an estimate of q based on the maximum number of germs that can be fit into the data (this can be easily seen from the explicit formula for the mode of the lower bound). Thus all reasonable estimates of q must remain essentially within the region delimited by the mode of the lower bound $\hat{\underline{q}}$ and the estimate produced by the CCE, i.e., $\hat{\bar{q}}$. This knowledge may at times permit better q-localization by reducing the size of the feasible region. Note that the feasible region provides q-localization based on an entirely different principle, i.e., containing the ML estimate of q on the basis of X. The ML estimator of q on the basis of X is not guaranteed to be unbiased, and thus the two criteria cannot be easily related. In particular, even though $\hat{\underline{q}}$ is known to be within the feasible region, the location of $\hat{\bar{q}}$ relative to the right endpoint of the feasible region can

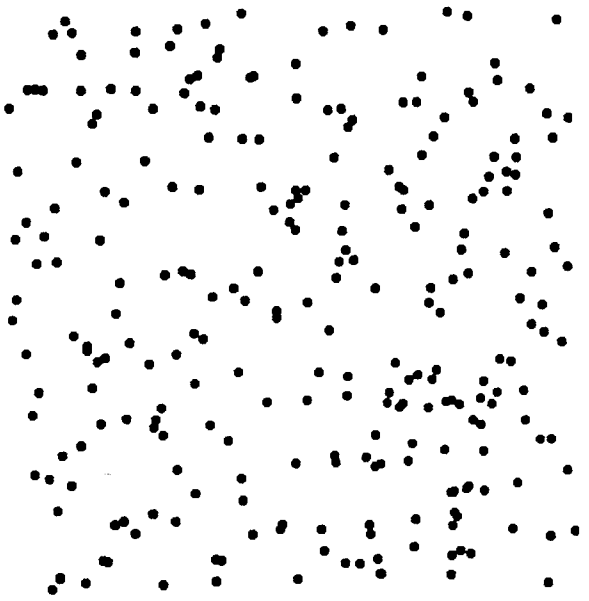


Fig. 2. Realization of a Boolean model of constant intensity and fixed primary grain.

not be *a priori* known (see figure 1). However, the two criteria can be heuristically combined to improve localization.

In practice, for typical observations all these estimates are close to each other. As an example, figure 2 depicts a realization of a DRBRS of constant intensity and fixed primary grain. For this example, $q = 0.999$, and the computed estimates are $\underline{\hat{q}} = \text{`}998$, $\hat{\bar{q}} = 0.99918$, whereas the feasible region is [0.9978, 0.9992].

5.1 *Noisy Observations*

Quite often we do not have the luxury of observing a noise-free realization of X. In practice, images are typically corrupted by sensor noise, sampling errors, and transmission errors. Therefore it is of interest to investigate the robustness of estimation tools in a noisy environment. To do so we need to assume a reasonable degradation mechanism and to get a grasp on the sensitivity of the relevant statistics as a function of a suitable parameterization of the noise.

One particular degradation model is the *independent union noise model.* It assumes that the observable DRS Y is the union of the signal DRS X with a noise DRS N, which is independent of X. The statistics of N can be arbitrary. If N can be modeled as another DRBRS of constant intensity and fixed primary grain (different from that of X), then we can work out similar bounds on the probability of the observation. However, in this case the upper bound is not unimodal. Alternatively, we can try to break up the solution into a signal-estimation step and a parameter-estimation step. The signal-estimation step provides an estimate, $\hat{X}(Y)$ of the signal X on the basis of the observation Y, whereas the parameter-estimation step computes the necessary statistics on the estimate $\hat{X}(Y)$. This approach is clearly suboptimal. However, if the estimate of X remains reasonable close to X, then we expect the overall procedure to be nearly optimal. Since our approach is essentially ML-based, it makes sense to use an ML estimator to perform the signal-estimation step. Towards this end, we have the following lemma.[7]

LEMMA 2. Let $O_H(B)$ denote the collection of all H-open subsets of B. Assume that the signal DRS X on B induces a probability mass function on $\Sigma(B)$ that has the following property:

$$P_X(X = K) \neq 0 \quad \text{iff } K \in O_H(B).$$

This assumption means that the signal is almost surely (a.s.) H-open; i.e., with probability one, the signal is a union of translates of H. Furthermore, assume that the observable DRS is $Y = X \cup N$, where N is a homogeneous Bernoulli lattice process of intensity $r \in [0, 1)$ (i.e., each point $z \in B$ is included in N with probability r independently of all other points), which is independent of X. Then $Y \circ H$ is the unique ML estimate of X on the basis of Y, regardless of the specific value of r.

Proof. Let $\hat{X}_{\mathrm{ML}}(Y)$ denote the ML estimate of X on the basis of Y. Then by definition

$$\begin{aligned}\hat{X}_{\mathrm{ML}}(Y) &= \arg \max_{K \in O_H(B)} \{\Pr(Y \mid X = K)\} \\ &= \arg \max_{K \in O_H(B),\, K \subseteq Y} \{\Pr(Y \mid X = K)\} \\ &= \arg \max_{K \in O_H(B),\, K \subseteq Y} \{r^{|Y|-|K|}(1-r)^{|B|-|Y|}\}\end{aligned}$$

$$= \arg \max_{K \in O_H(B),\, K \subseteq Y} \{r^{-|K|}\}$$
$$= \arg \max_{K \in O_H(B),\, K \subseteq Y} \{|K|\}.$$

So $\hat{X}_{\mathrm{ML}}(Y)$ is the largest H-open subset of Y, which is by definition the opening of Y by H, i.e.,

$$\hat{X}_{\mathrm{ML}}(Y) = Y \circ H,$$

and the proof is complete.

Observe that the proof crucially depends on $|B|$ being finite. Now it can be readily seen that (modulo some unavoidable edge effects) a DRBRS X of constant intensity and fixed primary grain H satisfies the condition of lemma 2. Thus if the observable DRS Y can be modeled as the union of the signal DRS X with an independent realization of a homogeneous Bernoulli lattice process (an assumption that is often reasonable), then the ML estimate of X on the basis of Y is simply $Y \circ H$. It is worth noting that this ML signal-estimation step does not assume knowledge of r; i.e., it is independent of the noise level. Although this is hardly surprising to the seasoned field veteran, it is pleasant and rather useful: it means that to estimate the signal we need not bother to measure the noise level; thus we avoid a potentially difficult task, given that we do not assume availability of a separate training sequence of the noise process. Note that this opening (ML) estimator is biased, in that $X \subseteq (X \cup N) \circ H, \forall X \in O_H(B)$, and a theoretical analysis of the behavior of the size of this bias as a function of r (the intensity of the noise process) seems difficult (it is currently under investigation). However, in practice the opening (ML) estimate remains very close to X, even when the intensity r of the noise process is high (e.g., $r = 0.5$). Moreover, simulation experiments have demonstrated that the statistics $|\hat{X}_{\mathrm{ML}}(Y)|$, $|\hat{X}_{\mathrm{ML}}(Y) \oplus H^s|$, and $|(\hat{X}_{\mathrm{ML}}(Y))^c \oplus H^s|$ are very robust; i.e., they remain very close to $|X|$, $|X \oplus H^s|$, and $|X^c \oplus H^s|$, respectively, for up to 80% noise, i.e., for $r = 0.8$. The modes of the bounds themselves are rather insensitive to small perturbations of the statistics. For example, the mode of the lower bound

$$\underline{\hat{q}}(X) = \frac{|X^c \oplus H^s|}{|B|}$$

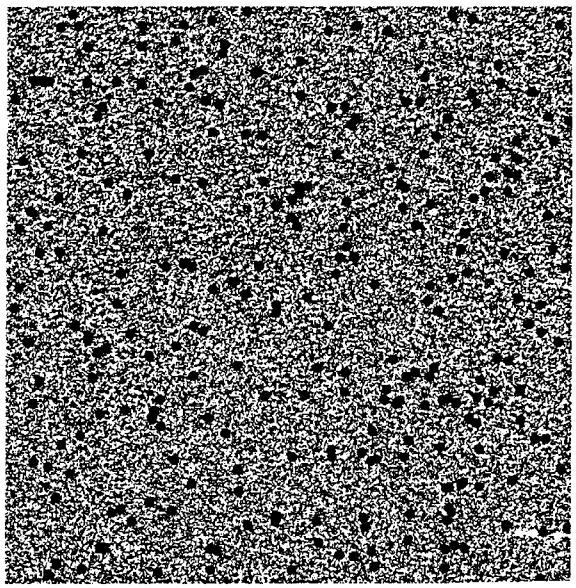


Fig. 3. The realization of figure 2, corrupted by i.i.d. union noise of intensity 0.5.

is robust under small perturbations of $|X^c \oplus H^s|$. Let us illustrate this approach. Consider figure 3, which depicts a realization of the observable DRS Y obtained by taking the union of the signal DRS X, depicted in figure 2, with an independent realization of a homogeneous Bernoulli lattice process of intensity $r = 0.5$. The ML estimate of X on the basis of the realization of figure 3 is depicted in figure 4. As can be seen, there is hardly any discernible difference between the realizations of figure 2 and 4. The statistics $|\hat{X}_{\mathrm{ML}}(Y)|$, $|\hat{X}_{\mathrm{ML}}(Y) \oplus H^s|$, and $|(\hat{X}_{\mathrm{ML}}(Y))^c \oplus H^s|$ all differ by less than 0.4% from their nominal values, i.e., $|X|$, $|X \oplus H^s|$, and $|X^c \oplus H^s|$, respectively.

5.2 *Nonconstant Radii*

These bounds can be extended to the case of a DRBRS model of constant intensity, $\lambda_s(z) = p = 1 - q, \forall z \in B$, and primary grains of random size by using the following approximation of the corresponding generating functional.

LEMMA 3. For a DRBRS X of constant intensity, $p = 1 - q$, and q sufficiently close to 1, the following approximation is valid:

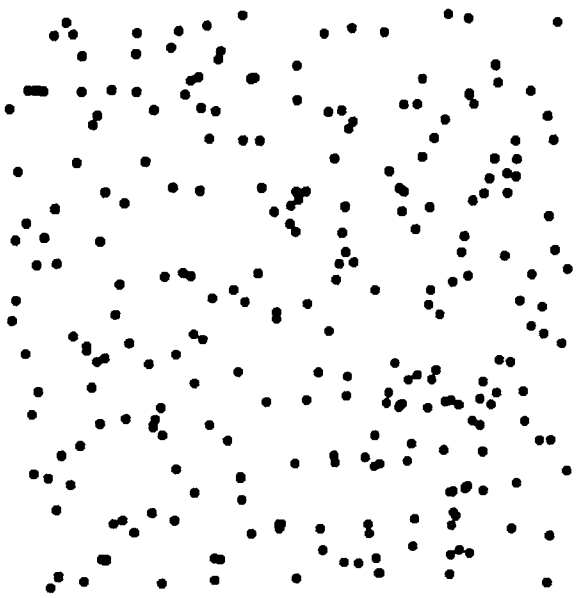


Fig. 4. The ML estimate of the signal of figure 2, on the basis of the observation depicted in figure 3.

$$Q_X(K) \cong q^{E|K \oplus RH^*|},$$

where the expectation is taken with respect to the probability mass function f_R of the radii.

Proof. See appendix A.

Using this approximation, which is asymptotically good as q goes to 1, and Theorem 1, we can obtain the same upper bound, but this time on the *approximate* (instead of the *actual*) probability. In this case, H in the expression for the upper bound is replaced by $\overline{R}H$, where $\overline{R}$ is the maximum possible radius. A unimodal lower bound for this case can be obtained by using the *morphological skeleton transform*.

5.3 *Morphological Skeletonization as a Method of Obtaining a Consistent Realization of the Embedded Marked-Point Process*

Another approach to the problem of estimating the probability of a given observation is suggested by looking at it from the viewpoint of shape analysis. The idea is that we can use certain morphological shape-description schemes to obtain one realization of the underlying marked-point process (i.e., the germ points marked by their corresponding radii), which can give rise to the observed realization K of the DRBRS X. Then we can obtain a lower bound on $P_X(X = K)$ simply by computing the probability of this realization of the marked-point process. If the grains of K are disconnected[8] and are contained in B, then there exists *a unique* realization of the underlying marked-point process that can give rise to K. In this case the unique realization can be recovered and the exact probability $P_X(X = K)$ can be computed. This approach can lead to good estimation and hypothesis-testing procedures if the data are sufficiently sparse. As an example, let us consider the simple-versus-simple hypothesis-testing problem

$$H_0 : X \sim (\lambda_s^{(0)}, H, f_R^{(0)})\text{-DRBRS}$$

versus

$$H_1 : X \sim (\lambda_s^{(1)}, H, f_R^{(1)})\text{-DRBRS}.$$

In principle, given any observation $K \in \Sigma(B)$, the probability of this observation under each one of the two hypotheses can be computed by using Theorem 1, and the Bayesian rule of choice can be implemented. In practice, the computational cost associated with this brute-force method limits its applicability. We therefore pursue an alternative approach. The key idea is the following. Suppose that instead of the DRBRS realization K, we were given the realization of the germ points $\{y_1, y_2, \ldots\}$ and the associated radii $\{R_1, R_2, \ldots\}$ that produced K. Let these data be represented by an ordered list of collections of sites $\{L_0, \ldots, L_{\hat{R}}\}$, corresponding to radii $\{0, \ldots, \overline{R}\}$, respectively. Note that, for one or more $n \in \{0, \ldots, \overline{R}\}$, L_n may be empty. The log-likelihood-ratio test for these data is simply given by

$$\begin{aligned}\log \frac{\Pr_1\{L_0, \ldots, L_{\hat{R}}\}}{\Pr_0\{L_0, \ldots, L_{\hat{R}}\}} &= \sum_{z \in B \mid z \notin \cup_{n=0}^{\hat{R}} L_n} \log\left(\frac{1-\lambda_s^{(1)}(z)}{1-\lambda_s^{(0)}(z)}\right) \\ &\quad + \sum_{n=0}^{\hat{R}} |L_n| \log\left(\frac{f_R^{(1)}(n)}{f_R^{(0)}(n)}\right) \\ &\quad + \sum_{n=0}^{\hat{R}} \sum_{z \in L_n} \log \frac{\lambda_s^{(1)}(z)}{\lambda_s^{(0)}(z)} \underset{H_0}{\overset{H_1}{\gtrless}} t, \end{aligned} \tag{1}$$

where the optimal threshold t is a function of the prior probabilities of the two hypotheses and the losses incurred when different kinds of decision errors are made. Therefore we can easily classify the observation, according to Bayesian decision theory. However, the recovery of marked-point-process data from the observation K is an ill-posed problem.

Simply put, the *morphological skeleton* [13], [18] of a binary shape K with respect to a structuring element H is the locus of the centers of all *maximal inscribable replicas* of H in K.[9] A *replica* of H is a scaled and shifted version of H. A replica of H is *maximal* in K if and only if it cannot be properly contained in any other replica of H that can be inscribed in K. The *morphological-skeleton function* (MSF) of K with respect to H is the function whose support is the morphological skeleton of K with respect to H, and its value at each skeleton point is equal to the radius of the corresponding maximal inscribable replica of H. The morphological skeleton is explicitly given by

$$SK(K) = \bigcup_{n=0}^{N} S_n(K)$$
$$= \bigcup_{n=0}^{N} [(K \ominus nH^s) - (K \ominus nH^s) \circ H],$$

where

$$N = \max\{n \mid K \ominus nH^s \neq \emptyset\}.$$

The set $S_n(K)$ is the locus of centers of maximal inscribable replicas of size n, and it is called the nth skeleton subset of K. Given all the skeleton subsets, K can be reconstructed by means of

$$K = \bigcup_{n=0}^{N} S_n(K) \oplus nH. \tag{2}$$

Given all the skeleton subsets, the MSF is uniquely determined. Conversely, given the MSF, all the skeleton subsets are uniquely determined.

From (2) it is clear that the MSF provides *one* realization of the germ points (the support set of the MSF), along with their associated radii (the values of the MSF), which can give rise to K. We propose the application of the log-likelihood-ratio test (1) to these data (i.e., $L_n = S_n$, $n = 0, \ldots, N$, and $L_n = \emptyset$, $N < n \leq \overline{R}$) as a decision rule for the simple hypothesis-testing problem under consideration.[10] If the grains of K are disconnected (a situation that arises with high probability if the intensity of the germ process is uniformly low and $\overline{R}$ is small) and contained in B, then the true (unique) realization of the underlying marked-point process is actually recovered and the proposed decision rule is exact maximum likelihood. The reason is that the nth skeleton subset of a union of disconnected sets with respect to a convex structuring element is the union of the nth skeleton subsets of the disconnected sets and the MSF of $nH \oplus \{z\}$ with respect to H is equal to n at z and is zero elsewhere. The overall procedure can be efficiently implemented (in polynomial time) thanks to the existence of fast morphological skeletonization algorithms [18]. Figure 5 depicts a realization of a DRBRS and its skeleton. Simulation results have been very encouraging, even when the primary grains overlap substantially. These simulations suggested that for hypothesis testing between two DRBRS models of different intensities, the size of the skeleton is an important statistic in the sense of possessing high discriminatory power. This prompted us to investigate whether it is possible to make ML decisions between DRBRS models of different intensities (but otherwise identical) based solely on the size of the skeleton. As it turns out, this is a move in the right direction. In fact, the important statistic is the size of a superset of the skeleton. This is the subject of the following theorem.

THEOREM 2. Consider the simple-versus-simple hypothesis-testing problem

$$H_0 : X \sim (p_0, H, f_R)\text{-DRBRS}$$

versus

$$H_1 : X \sim (p_1, H, f_R)\text{-DRBRS},$$

where p_0, p_1 are constants in $(0, 1)$, $p_1 > p_0$, and $f_R(r)$ (the common size distribution) is zero

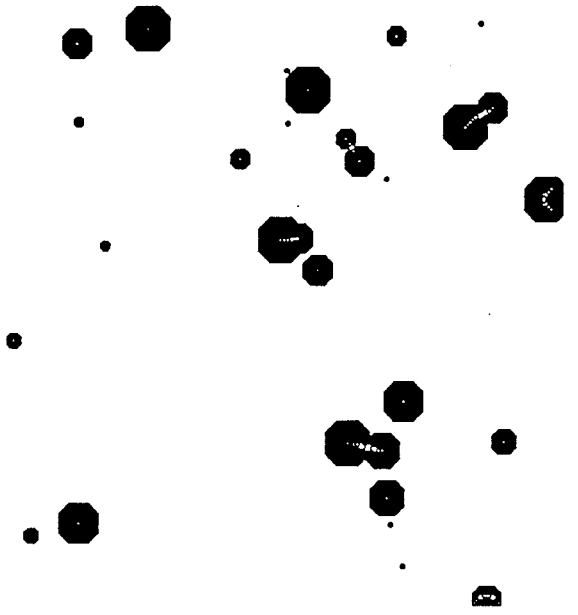


Fig. 5. Realization of a DRBRS and its skeleton. The skeleton points are the white points highlighted within the primary grains.

outside $\{\underline{R}, \ldots, \overline{R}\}$, where $\underline{R} \geq 0$. Define

$$\gamma(X) \triangleq \frac{|(X^c \oplus \underline{R}H^s)^c|}{|B|}.$$

Let K be the observation, and let $P_0(X = K)$ and $P_1(X = K)$ denote the probability of the observation under the null and alternative hypotheses, respectively. If

$$\gamma(K) < l(p_0, p_1) \triangleq \frac{\log(1-p_1) - \log(1-p_0)}{\log(p_0(1-p_1)) - \log(p_1(1-p_0))},$$

then $P_0(X = K) > P_1(X = K)$.

Proof. Let L be a realization of the germ points that can give rise to the observation K. The probability of this realization under p_0 is $\Pr_0(L) = p_0^{|L|}(1-p_0)^{|B|-|L|}$, whereas under p_1 it is $\Pr_1(L) = p_1^{|L|}(1-p_1)^{|B|-|L|}$. It is easy to see that

$$\Pr_0(L) > \Pr_1(L) \iff \frac{|L|}{|B|} < l(p_0, p_1).$$

But any L that can give rise to K necessarily satisfies

$$L \subseteq (K^c \oplus \underline{R}H^s)^c;$$

thus

$$\frac{|L|}{|B|} \leq \gamma(K) < l(p_0, p_1) \quad \text{by assumption.}$$

Therefore $\Pr_0(L) > \Pr_1(L)$ uniformly over all L that can give rise to K. Hence since the two models have the same primary grain and size distribution, we conclude that $P_0(X = K) > P_1(X = K)$.

Some remarks are in order. This theorem states that if $\gamma(K) < l(p_0, p_1)$, then we can safely decide in favor of the null hypothesis H_0 in the sense that our decision, coincides with the ML decision. That is, if $\gamma(K) < l(p_0, p_1)$, then we make a computationally cheap decision, which also happens to be statistically sound. This technique may have potential for application in the automated screening of cell samples, where the alternative hypothesis corresponds to an abnormally high average number of cells per unit area. Then most of the observed samples can be classified with minimal effort whereas the few samples that do not meet the criterion of Theorem 2 can be examined in greater detail by either a machine or a human expert.

By symmetry, if the size of the smallest L that can give rise to the observation K satisfies $|L|/|B| > l(p_0, p_1)$, then we can conclude that $P_1(X = K) > P_0(X = K)$ and we can safely decide in favor of the alternative hypothesis H_1. However, there exists no known efficient (polynomial-time) algorithm that can determine the size of the smallest L that can give rise to K. The only way we know how to do this is by exhaustive search, whose complexity is exponential in $|K|$. In this case we might as well compute the exact likelihoods of the observation under the two hypotheses and compare them.

The behavior of l as a function of p_0, p_1 is of considerable interest because it determines the rate at which Theorem 2 can be used to simplify ML decisions. It can be shown that l is *roughly* halfway between p_0 and p_1; i.e.,

$$l(p_0, p_1) \cong \frac{p_0 + p_1}{2}.$$

Let $E_0\gamma(X)$ denote the expectation of $\gamma(X)$ under p_0. It can be proved, along the lines of proof of Proposition 1, that $E_0\gamma(X) \geq p_0$. In practice, we can estimate both $E_0\gamma(X)$ and the standard deviation of $\gamma(X)$ under p_0 from a set of training data. Denote these by $\hat{\gamma}_0$, and $\hat{\sigma}_0$, respectively. Then, as a rule of thumb, if $p_1 > 2(\hat{\gamma}_0 + \hat{\sigma}_0) - p_0$, we will be able to classify most of the normal samples optimally and with minimal effort.

6 Conclusions

In this paper we have taken the approach of modeling binary digital image data as realizations of a uniformly bounded discrete random set, a mathematical object that can be axiomatically defined directly on a finite sample space. We have argued the merits of such an approach, most notably the ability to *recover* the associated probability measure by means of a Möbius-type transformation and knowledge of the generating functional. On the basis of this result and some tools of morphological shape analysis, we have developed a restricted version of the discrete-case analog of the Boolean random set, obtained its complete probabilistic specification, and provided various tools for its statistical inference. Although, in reality, binary digital image data are sampled versions of an underlying physical process, which lives in a continuum, the data per se can assume only a finite number of realizations. This is the case in many applications, in which physical barriers limit the available resolution. Although a discrete random-set approach may ignore the fine letter of the underlying physical structure, it provides a useful, and, most importantly, *tractable* idealization, which, as demonstrated, can lead to practical inference procedures.

Appendix A

A.1 Proof of Proposition 4

Upper bound:

$$P_X(X = K) = \sum_{K' \subseteq K} (-1)^{|K'|} q^{|(K^c \cup K') \oplus H^s|}$$

$$= \sum_{K' \subseteq K, |K'| = \text{even}} q^{|(K^c \cup K') \oplus H^s|} - \sum_{K' \subseteq K, |K'| = \text{odd}} q^{|K^c \cup K') \oplus H^s|}.$$

Observe that, by distributivity of dilation over union and by using the union bound,

$$|(K^c \cup K') \oplus H^s| = |(K^c \oplus H^s) \cup (K' \oplus H^s)| \leq |K^c \oplus H^s| + |K' \oplus H^s|.$$

Furthermore, since H is assumed to contain the origin,

$$|(K^c \cup K') \oplus H^s| \geq |K^c \cup K'| = |K^c| + |K'|.$$

Therefore, since q is a probability,

$$\begin{aligned} P_X(X = K) &\leq \sum_{K' \subseteq K, |K'| = \text{even}} q^{|K^c| + |K'|} - \sum_{K' \subseteq K, |K'| = \text{odd}} q^{|K^c \oplus H^s| + |K' \oplus H^s|} \\ &= q^{|K^c|} \sum_{K' \subseteq K, |K'| = \text{even}} q^{|K'|} - q^{|K^c \oplus H^s|} \times \sum_{K' \subseteq K, |K'| = \text{odd}} q^{|K' \oplus H^s|} \\ &\leq q^{|K^c|} \sum_{K' \subseteq K, |K'| = \text{even}} q^{|K'|} - q^{|K^c \oplus H^s|} \times \sum_{K' \subseteq K, |K'| = \text{odd}} q^{|K \oplus H^s|} \\ &= q^{|K^c|} \sum_{K' \subseteq K, |K'| = \text{even}} q^{|K'|} - q^{|K^c \oplus H^s| + |K \oplus H^s|} \times \sum_{K' \subseteq K, |K'| = \text{odd}} 1. \end{aligned}$$

Thus

$$P_X(X = K) \leq q^{|K^c|} \left(\sum_{i = \text{even}} \binom{|K|}{i} q^i \right) - q^{|K^c \oplus H^s| + |K \oplus H^s|} \times \left(\sum_{i = \text{odd}} \binom{|K|}{i} \right).$$

Using the fundamental identity

$$\sum_{i=0}^{|K|} \binom{|K|}{i} z^i = (1 + z)^{|K|} \quad \forall z \in C$$

and successively setting $z = -1, 1$, we obtain

$$\sum_{i=\text{odd}} \binom{|K|}{i} = 2^{|K|-1}.$$

Similarly, replacing z by qz and then setting $z = -1, 1$, we obtain

$$\sum_{i=\text{even}} \binom{|K|}{i} q^i = \frac{1}{2}\left[(1+q)^{|K|} + (1-q)^{|K|}\right],$$

from which we finally obtain the expression for the upper bound.

We will need the following lemma.

LEMMA 4 (Descartes' rule of signs [21, pp. 36–43]). Let $p(x)$ be a polynomial of a real variable with real coefficients

$$p(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_n x^n.$$

Let C denote the number of changes of sign of the sequence of its coefficients (for each $m \geq 1$, if $\alpha_{m-1}\alpha_m < 0$, then (α_{m-1}, α_m) constitute a change of sign). Let Z be the number of *positive* real zeros of $p(x)$ (a zero of multiplicity k is counted as k zeros). Then

$$C - Z \geq 0$$

and $C - Z$ is an even number.

By using once more the identity

$$\sum_{i=0}^{|K|} \binom{|K|}{i} z^i = (1+z)^{|K|} \quad \forall z \in C,$$

it can be seen that $U_q(K)$ can be written as

$$U_q(K) = \sum_{i=\text{even}}^{|K|} \binom{|K|}{i} q^{|K^c|+i} - 2^{|K|-1} q^{|K^c \oplus H^s| + |K \oplus H^s|}.$$

Since all the coefficients of this polynomial are strictly positive except for the coefficient of the highest degree, which is strictly negative, Descartes' rule of signs leads us to conclude that $U_q(K)$ has at most one zero in $(0, \infty)$. But $U_1(K) = 0$, and, therefore, this is the unique zero in $(0, \infty)$. Hence $U_q(K) > 0$, $\forall q \in (0, 1)$.

Next, consider the derivative of the upper bound with respect to q. After some algebraic manipulation it can be written as

$$\frac{d}{dq}U_q(K) = q^{|K^c|-1} \times \left[|K^c| + \sum_{i=2, i=\text{even}}^{|K|} \left[\binom{|K|}{i}(|K^c|+i)\right] q^i - 2^{|K|-1}[|K^c \oplus H^s| + |K \oplus H^s|] \times q^{|K^c \oplus H^s| + |K \oplus H^s| - |K^c|}\right].$$

Again, since all the coefficients of this polynomial are strictly positive except for the coefficient of the highest degree, which is strictly negative, by using Descartes' rule of signs we conclude that $(d/dq)U_q(K)$ has at most one zero in $(0, \infty)$. But

$$\left.\frac{d}{dq}U_q(K)\right|_{q=0} = 0,$$

$$\left.\frac{d}{dq}U_q(K)\right|_{q=0^+} = 0,$$

and

$$\left.\frac{d}{dq}U_q(K)\right|_{q=1} = |K^c|2^{|K|-1} + |K|2^{|K|-2} - 2^{|K|-1}[|K^c \oplus H^s| + |K \oplus H^s|] < 0 \quad \forall K \neq \emptyset.$$

Therefore by continuity, we conclude that $(d/dq)U_q(K)$ has at least one zero in $(0, 1)$, which must also be unique. Hence since its derivative has only one zero crossing in $(0, 1)$, $U_q(K)$ must be *unimodal* in $(0, 1)$.

Lower bound: It can be easily seen that *one* possible germ configuration that can give rise to the observation K is given by the set of points $(K^c \oplus H^s)^c$. In particular, let L denote the germ-point process (which is itself a DRS). Then X can be written as $X = L \oplus H$. By simple geometric arguments $(X^c \oplus H^s)^c \oplus H = X$ and $L \subseteq (X^c \oplus H^s)^c$; i.e., $X^c \oplus H^s \subseteq L^c$. Hence

$$P_X(X = K) \geq q^{|K^c \oplus H^s|}(1-q)^{|K^c \oplus H^s)^c|}.$$

The lower bound is strictly positive for all $q \in (0, 1)$ (by inspection). One can show that it is unimodal by looking at its derivative and using Descartes' rule of signs. After some manipulation

$$\frac{d}{dq}L_q(K) = \sum_{i=0}^{|K^c \oplus H^s)^c|} \binom{|K^c \oplus H^s)^c|}{i} (-1)^i (|K^c \oplus H^s| + i) q^{|K^c \oplus H^s|+i-1}.$$

Since $K \neq \emptyset$ and K is realizable (meaning that it can be written as $K = L \oplus H$ for some $L \in \mathcal{F}(B)$), it follows that $|(K^c \oplus H^s)^c| \geq 1$ and $|K^c \oplus H^s| \geq 1$, in which case the corresponding sequence of coefficients of the polynomial $(d/dq)L_q(K)$ has exactly $|(K^c \oplus H^s)^c|$ sign changes and, therefore, $C = |(K^c \oplus H^s)^c|$. Hence the number of positive zeros is $Z \leq C = |(K^c \oplus H^s)^c|$. However, $(d/dq)L_q(K)$ can also be written as

$$\frac{d}{dq}L_q(K) = (1-q)^{|(K^c \oplus H^s)^c|-1} \times \left[|K^c \oplus H^s| q^{|K^c \oplus H^s|-1}(1-q) - q^{K^c \oplus H^s|}|(K^c \oplus H^s)^c|\right],$$

from which it is obvious that it has a zero of multiplicity $|(K^c \oplus H^s)^c| - 1$ at $q = 1$. Therefore since $C - Z$ must be an even number, there remains one more positive zero to be accounted for. By inspection of the last formula, $(d/dq)L_q(K) \neq 0$ for $q > 1$. Hence this last zero is in $(0, 1)$. Since the derivative has one and only one zero crossing in $(0, 1)$, we have established unimodality of $L_q(K)$ in $(0, 1)$. The mode location is obtained by simple differentiation of the logarithm of the lower bound.

A.2 Proof of Lemma 3

For convenience, let $S \triangleq RH$ denote the generic primary grain, centered at the origin. Then

$$Q_X(K) = \prod_{z \in K \oplus \check{R}H^s} (1 - pT_S(K \oplus \{-z\})),$$

with

$$T_S(K) = 1 - F_R(d^H(\{(0, 0)\}, K) - 1).$$

By definition of the distance metric d^H, we can extend the product domain to the entire base frame B:

$$Q_X(K) = \prod_{z \in B}(1 - pT_S(K \oplus \{-z\})).$$

Thus

$$\log Q_X(K) = \sum_{z \in B} \log(1 - pT_S(K \oplus \{-z\})).$$

Taking the derivative with respect to p and evaluating at $p = 0$, we obtain

$$\begin{aligned}\frac{d}{dp}\log Q_X(K)\Big|_{p=0} &= -\sum_{z \in B} T_S(K \oplus \{-z\}) \\ &= -\sum_{z \in B} \Pr(S \cap (K \oplus \{-z\}) \neq \emptyset) \\ &= -\sum_{z \in B} E\,1(S \cap (K \oplus \{-z\}) \neq \emptyset) \\ &= -E\sum_{z \in B} 1(S \cap (K \oplus \{-z\} \neq \emptyset) \\ &= -E|K \oplus S^s|.\end{aligned}$$

Since

$$p = 0 \mapsto Q_X(K) = 1 \mapsto \log Q_X(K) = 0,$$

we have the following first-order Taylor-series approximation *of the logarithm of* $Q_X(K)$.

$$\log Q_X(K) \cong \left(\frac{d}{dp}\log Q_X(K)\Big|_{p=0}\right) p;$$

i.e.,

$$\begin{aligned}\log Q_X(K) &\cong -pE|K \oplus S^s| \\ Q_X(K) &\cong e^{-pE|K \oplus S^s|} = \left(e^{-p}\right)^{E|K \oplus S^s|} \\ &= q^{E|K \oplus S^s|}\end{aligned}$$

since, for p close to 0, $e^{-p} \cong (1 - p) = q$.

Acknowledgments

The authors thank the referees and the editor for their insightful suggestions, which helped improve the quality of this paper. The first author

also thanks John Goutsias for his encouragement and feedback.

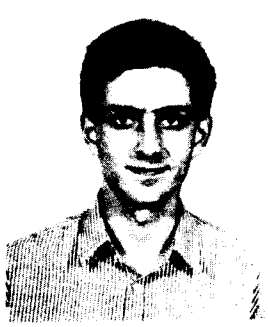
Notes

1. Some preliminary work on modeling simple geometrical structure by using MRFs has been recently reported in [3]. The basic idea is to incorporate geometrical constraints into the MRF clique structure in a suitable approximate sense. This idea may prove promising.
2. A similar idea has been concurrently and independently developed in [19] and [15]. There is a fundamental difference between the two formulations: we consider uniformly bounded discrete random sets, whereas Goutsias et al. consider discrete random sets on the infinite lattice $\mathbf{Z}^2$. In a recent paper [12] Goutsias has reconciled the two approaches by looking at *finite observations* of discrete random sets defined on $\mathbf{Z}^2$ and using certain key elements of our analysis in [28] to develop some tools for the morphological analysis of discrete random shapes.
3. The exception is an approximate ML approach proposed in [10]. However, there are problems with this approach, as pointed out in [2].
4. Refer to [26] for a thorough introduction to the principles of mathematical morphology. Here we merely reproduce some basic definitions.
5. In digital topology [14], [17], [26] the *convex hull* of a bounded set, $H \subset \mathbf{Z}^2$, is defined as the intersection of the convex hull of H in the topology of $\mathcal{R}^2$ with $\mathbf{Z}^2$. A bounded set, $H \subset \mathbf{Z}^2$, is convex if it is identical to its convex hull.
6. Being realizable means that K can be written as $K = L \oplus H$ for some $L \in \Sigma(B)$. If K cannot be written this way, then it is not a realization of the DRBRS model under consideration and, therefore, its probability is zero.
7. This is a statistical signal-filtering result for a particular class of DRS models. See [27] for a recent survey of morphological filtering, and see [9] for a tutorial overview of the field. Statistical approaches have also been investigated; for a sample see [7], [8], [24], [28], [29], [30], [31].
8. Here, "disconnected" refers to the chessboard-block, or nearest-neighbor, sense.
9. Many other related notions of skeleton exist. However, the definition given is sufficient for our purposes.
10. This idea has been concurrently and independently developed in [15] as a means of performing shape-size analysis and synthesis of a different class of DRS models.

References

1. M. Aigner, *Combinatorial Theory*, Springer-Verlag: New York, 1979.
2. G. Ayala, J. Ferrandiz, and F. Montes, "Boolean models: ML estimation from circular clumps," *Biomed. J.*, vol. 32, pp. 73–78, 1990.
3. F. Chen and P.A. Kelly, "Algorithms for generating and segmenting morphologically smooth binary images," in *Proc. 26th Annual Conference on Information Sciences and Systems*, Princeton University, Princeton, NJ, vol. II, pp. 902–906, 1992.
4. G. Choquet, "Theory of capacities," *Ann. Inst. Fourier*, vol. 5, pp. 131–295, 1953.
5. N. Cressie and G.M. Laslett, "Random set theory and problems of modeling," *SIAM Rev.*, vol. 29, pp. 557–574, 1987.
6. P.J. Diggle, "Binary mosaics and the spatial pattern of heather," *Biometrics*, vol. 37, pp. 531–539, 1981.
7. E. Dougherty, "Optimal mean square N-observation digital morphological filters, I: optimal binary filters," *Comput. Vis., Graph. Image Process.*, vol. 55, pp. 36–54, 1992.
8. E. Dougherty, "Optimal mean square N-observation digital morphological filters, II: optimal gray-scale filters," *Comput. Vis., Graph. Image Process.*, vol. 55, pp. 55–72, 1992.
9. E. Dougherty and C. Giardina, *Morphological Methods in Image and Signal processing*, Prentice-Hall: Englewood Cliffs, NJ, 1988.
10. V. Dupac, "Parameter estimation in the Poisson field of discs," *Biometrika*, vol. 67, pp. 187–190, 1980.
11. J. Serra (ed.), *Theoretical Advances*, Image Analysis and Mathematical Morphology, vol. 2, Academic Press: San Diego, CA, 1988.
12. J. Goutsias, "Morphological analysis of discrete random shapes," *J. Math. Imag. Vis.*, vol. 2, pp. 193–215, 1992.
13. J. Goutsias and D. Schonfeld, "Morphological representation of discrete and binary images," *IEEE Trans. Signal Process.*, vol. SP-39, pp. 1369–1379, 1991.
14. J. Goutsias and C. Wen, "Modeling discrete random shapes: a random set theory approach, Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD, Tech. Rep. JHU/ECE 90–13, 1990.
15. J. Goutsias and C. Wen, "Discrete random set models for shape synthesis and analysis," *Proc. Soc. Photo-Opt. Instrom. Eng.*, vol. 1606, pp. 174–185, 1991.
16. D.G. Kendall, "Foundations of a theory of random sets," in *Stochastic Geometry*, E.F. Harding and D.G. Kendall, ed., Wiley: London, 1974, pp. 322–376.
17. T.Y. Kong and A. Rosenfeld, "Digital topology: introduction and survey," *Comput. Vis., Graph., Image Process.*, vol. 48, pp. 357–393, 1989.
18. P. Maragos and R.W. Schafer, "Morphological skeleton representation and coding of binary images," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, pp. 1228–1244, 1986.
19. G. Matheron, *Elements pour une theorie des Milieux Poreux*, Masson: Paris, 1967.
20. G. Matheron, *Random Sets and Integral Geometry*, Wiley: New York, 1975.
21. G. Polya and G. Szego, *Problems and Theorems in Analysis, vol. II*, Springer-Verlag: New York, 1976.

22. B.D. Ripley, "Locally finite random sets: foundations for point process theory," *Ann. Probab.*, vol. 4, pp. 983–994, 1976.
23. M. Schmitt, "Estimation of the density in a stationary Boolean model," *J. Appl. Probab.*, vol. 28, pp. 702–708, 1991.
24. D. Schonfeld and J. Goutsias, "Optimal morphological pattern restoration from noisy binary images," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. PAMI-13, pp. 14–29, 1991.
25. J. Serra, "The Boolean model and random sets," *Comput. Graph. Image Process.*, vol. 12, pp. 99–126, 1980.
26. J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press: New York, 1982.
27. J. Serra and L. Vincent, "An overview of morphological filtering, *Circuits, Syst., Signal Process.*, vol. 11, pp. 47–108, 1992.
28. N.D. Sidiropoulos, "Statistical inference, filtering, and modeling of discrete random sets," Ph.D. thesis, University of Maryland, College Park, MD, 1992.
29. N.D. Sidiropoulos, J.S. Baras, and C.A. Berenstein, "Discrete random sets: an inverse problem, plus tools for the statistical inference of the discrete Boolean model," *Proc. Soc. Photo-Opt. Instrum. Eng.*, vol. 1769, pp. 32–43, 1992.
30. N.D. Sidiropoulos, J.S. Baras, and C.A. Berenstein, "Optimal mask filtering of discrete random sets under a union/intersection noise model," in *Proc. 26th Annual Conference on Information Sciences and Systems*, Princeton University, Princeton, NJ, vol. I, pp. 313–318, 1992.
31. N.D. Sidiropoulos, J.S. Baras, and C.A. Berenstein, "Optimal morphological filters for discrete random sets under a union or intersection noise model," *Proc. Soc. Photo-Opt. Instrum. Eng.*, vol. 1818, pp. 402–413, 1992.
32. D. Stoyan, W.S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*, Wiley: Berlin, 1987.



Nicholas D. Sidiropoulos received the Diploma in electrical engineering from the Aristotelian University of Thessaloniki, Greece, in 1988 and the M.S. and Ph.D. degrees in electrical engineering, both from the University of Maryland at College Park, in 1990 and 1992, respectively. From 1988 to 1992 he was a Fulbright fellow and a research assistant at the Center for Systems Research of the University of Maryland. He is currently on military duty, serving as scientific assistant and lecturer at the Hellenic Air Force Academy. His current research interests are random set theory, image and signal processing, and statistical communications. He is a member of IEEE and the Technical Chamber of Greece.

John S. Baras received the B.S. degree in electrical engineering with highest distinction from the National Technical University of Athens, Greece, in 1970. He received the M.S. and Ph.D. degrees in applied mathematics from Harvard University in 1971 and 1973, respectively. Since 1973 he has been with the Department of Electrical Engineering, University of Maryland at College Park, where he is currently professor and member of the applied mathematics faculty. From 1985 to 1991 he was the founding director of the Systems Research Center, now the Institute for Systems Research. In February 1990 he was appointed to the Martin Marietta Chair in Systems Engineering. Since 1991 he has been the codirector of the Center for Satellite and Hybrid Communication Networks, a NASA Center for the Commercial Development of Space, which he cofounded. Dr. Baras has held visiting research scholar positions at Stanford University, MIT, Harvard University, the Institut National de Recherche en Informatique et en Automatique, and the University of California at Berkeley. He has numerous publications in control and communication systems and is the coeditor of *Recent Progress in Stochastic Calculus*, published by Springer-Verlag in 1990. His current research interests include stochastic systems, signal processing, and understanding, with emphasis on speech and image signals, real-time architectures, symbolic computation, intelligent control systems, robust nonlinear control, distributed-parameter systems, hybrid communication network simulation, and management.

Among his awards are a 1978 U.S. Naval Research Laboratory Research Publication Award, the 1980 Outstanding Paper Award of the IEEE Control Systems Society, and a 1983 Alan Berman Research Publication Award from the Naval Research Laboratory. He has served in the IEEE Engineering Research and Development Committee, the Aerospace Industries Association Advisory committee on advanced sensors, the IEEE fellow evaluation committee, and the IEEE Control Systems Society Board of Governors (1991–1993). He is currently serving on the editorial

boards of *Mathematics of Control, Signals, and Systems, of Systems and Control: Foundations and Applications,* of *IMA Journal of Mathematical Control and Information*, and of *Systems Automation – Research and Applications*, and he is the managing editor of the series *Progress in Automation and Information Systems* from Springer-Verlag. He is a fellow of the IEEE and a member of Sigma Xi, the American Mathematical Society, and the Society for Industrial and Applied Mathematics.

Carlos A. Berenstein received the Licenciado en Matematicas degree from the University of Buenos Aires in 1966 and the Ph.D. degree in mathematics from the Courant Institute, New York University, in 1970. Since 1970 he has taught at Harvard University and Brandeis University, and he is currently a professor of mathematics and a member of the Institute for Systems Research at the University of Maryland at College Park. He has also been a visiting professor at several foreign universities, including the University of Paris (Bordeaux) and the Scuola Normale Superiore (Pisa). He has written several monographs and research articles on Fourier analysis, function theory in one and several complex variables, convolution equations, difference–differential equations, boundary-value problems integral, geometry, exponential polynomials and functional analysis, and digital geometry. His current interests include applications of the preceding mathematical theories to tomography, filtering theory, image processing, and complexity theory.