

Approximate Likelihood Procedures for the Boolean Model Using Linear Transects

by

John C. Handley
Xerox Corporation
800 Phillips Road, MS 128-27E
Webster, NY 14580
jhandley@crt.xerox.com

Abstract

The Boolean model is a random closed set process consisting of a Poisson point process (producing germs) coupled with an independent random shape process (grains). Origins of grains are translated to germs to produce an arrangement of overlapping (or interpenetrating) shapes. An accurate discrete approximation to the continuous linear Boolean model offers computationally efficient likelihood procedures including maximum likelihood estimation and likelihood ratio tests. The discrete approximation allows covering probabilities to be calculated using recursive formulas, which approach continuous densities as the sampling rate increases. Inference for higher dimensional Boolean models is handled by linear transects. Two two-dimensional estimation examples demonstrate the efficacy of this method.

1 Introduction

The Boolean model (BM) is a fundamental random closed set process applied to materials science, cellular communications networks, food science, queueing theory, and many other fields [6], [16], [17], [18]. It is comprised of a homogeneous Poisson point (*germ*) process $\Phi = \{\xi_1, \xi_2, \dots\}$ in R^n , $n \geq 1$, with intensity λ and an independent compact random set (*grain*) process S_0 with outcomes $\{S_1, S_2, \dots\}$, compact subsets of R^n . (S_0 is called the *primary grain*.) The Boolean model is

$$\mathcal{B} = \bigcup_{n=1}^{\infty} (S_n + \xi_n). \quad (1)$$

The BM is also called the *Poisson germ-grain model*. What one observes is a pattern of overlapping (or interpenetrating) random bodies.

Estimation is guided by formulas linking measurements of the germ-grain observations to probability formulas for germ and grain processes. Statements about the grain process involve Minkowski functionals of the shapes, often in terms of mean area and perimeter. These quantities are difficult to obtain since the notion of a continuous shape is quite general even without introducing randomness. Most applications assume sufficiently regular

shapes so that random shapes are parameterized by random variables [1]. Reference [12] is an entire book devoted to estimation for the BM.

Our approach is to approximate the continuous 1D model with a discrete one and analyze it directly. This results in recursive formulas for coverage probabilities that can be used for approximate likelihood statements. We then present a 2D estimation scheme comprised of (nearly) independent 1D samples of a 2D process.

Attempts to formulate a discrete version of the BM (more appropriately called the *binomial germ-grain model*) have encountered formidable combinatorial difficulties because the continuous formalism was translated directly to the discrete setting [15]. In a series of papers, Dougherty and Handley developed and exploited recursive expressions for coverage probabilities in the one-dimensional (1D) binomial case [4], [11]. Using these basic results, García and Petrou applied the 1D binomial model to texture analysis in image processing [7], [8].

The 1D binomial germ-grain approximates a 1D BM in a straight-forward way and it was demonstrated that the binomial model offers extremely fast, compact, and stable algorithms for problematic lineal path probability densities [10], [14]. We use this approximation here to develop likelihood procedures for the continuous BM. Our approach to inference for higher dimensional BMs is to analyze 1D slices of realizations because a basic consequence of the projection theorem for Poisson processes is that under mild conditions a 1D slice of a two or higher-dimensional BM is again a 1D BM [17].

We introduce the 1D BM, the 1D binomial germ-grain model, and then show how it approximates the 1D BM. Next, we demonstrate likelihood procedures with two example 2D Boolean models.

2 The One-Dimensional Boolean Model

The 1D BM is thoroughly discussed in Chap. 2 of [9]. We review several important facts to set up the development. Consider a Poisson process on the line with intensity λ and outcomes $\{\xi_1, \xi_2, \dots\}$. For a grain process, consider a positive random variable X with distribution C . Strictly speaking, we associate the random discrete set (interval) $[0, x]$ with the outcome $\{X = x\}$. The 1D or *linear* BM is

$$\Xi = \bigcup_{i=1}^{\infty} [\xi_i, \xi_i + x_i]. \quad (2)$$

An observation of Ξ consists of an alternating sequence of spacings and clumps. Spacings are independent and exponentially distributed with mean λ^{-1} . Clump-lengths are also independent and identically distributed according to a distribution we denote by G .

3 The One-Dimensional Binomial Germ-Grain Model

The one-dimensional binomial germ-grain model consists of a Bernoulli marking process with marking probability p and a natural-number-valued random variable X with distribution C . The marking process marks integer values on the line. As in the BM, outcome $\{X = x\}$ is identified with set (interval) $[0, x - 1]$. It is shown in [11] that spacings are iid geometric, clump-lengths are independent of spacings and each other, and the runlength (or clump-length) density is

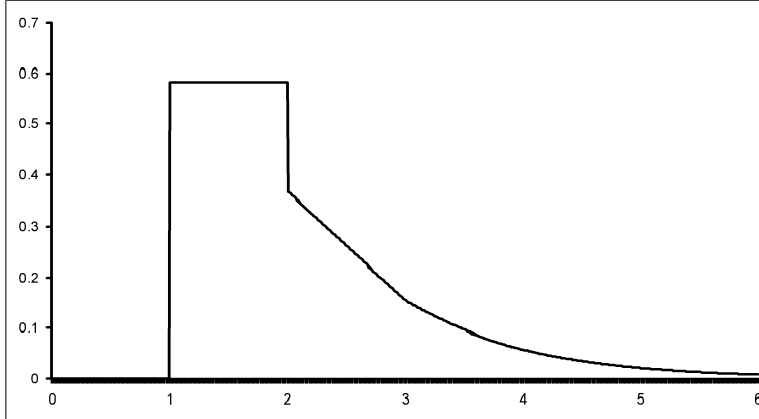


Figure 1: Approximate clump-length density for constant segment length 1 and intensity 1.

$$P(K = m) = \sum_{j=1}^m [F(m) - F(j-1)] \times \prod_{i=1}^{j-1} F(i-1)P(K = m-j), \quad (3)$$

where $F(i) \equiv 1 - p + pC(i)$ and $P(K = 0) \equiv (1 - p)/p$. The essence of Eq. 3 is a recursive “accounting” of all possible covering events leading to a length m run of covered integers, given that a clump-length occurs. This direct approach to decomposing the event space deviates from the usual transform approach to stochastic processes.

If we sample the real line into n samples per unit distance and approximate the Poisson intensity $\lambda = pn$, then the discrete clump-length density is

$$P(K = x) = \sum_{j=1}^{\lfloor nx \rfloor} [F(x) - F(\frac{j-1}{n})] \times \prod_{i=1}^{j-1} F(\frac{i-1}{n})P(K = x - j/n), \quad (4)$$

where $F(x) = 1 - \frac{\lambda}{n} + \frac{\lambda}{n}C(x)$. Equation 4 is an approximation of the continuous clump-length density. Although Eq. 4 is recursive, if intermediate results of $P(K = x)$ and $C(x)$ are stored in arrays, the computation time of $P(K = x)$ is all but negligible.

In [9] the clump-length density for a BM with constant segment length a and intensity λ is computed, perhaps the only case where a closed form can be obtained

$$f(x) = \lambda(e^{a\lambda} - 1)^{-1} \times \left[1 + \sum_{j=1}^{\lfloor (x/a)-1 \rfloor} \frac{(-1)^j}{j!} \{\lambda(x - (j+1)a)\}^{j-1} e^{-ja\lambda} \{\lambda(x - (j+1)a) + j\} \right],$$

for $x > a$ and $f(x) = 0$ on $(0, a]$. Figure 1 shows a clump-length density approximation with $n = 1000$ of the BM with constant segment length $a = 1$ and intensity $\lambda = 1$, computed in a few seconds with a desktop PC; compare with Figure 2.3 of [9].

Since our application is binary image processing, spaces and clumps are white and black runlengths, respectively: $\{W_i\}$, $\{B_i\}$. We assume that black and white runlengths

form pairs (ignoring incomplete runlengths): $\mathcal{S} = \{(B_i, W_i) : i = 1, \dots, N\}$. Assuming C has (vector) parameter $\boldsymbol{\theta}$, the approximate likelihood function for observation \mathcal{S} is

$$L(\boldsymbol{\theta}; \mathcal{S}) = \prod_{i=1}^N (\lambda/n)(1 - \lambda/n)^{W_i-1} P(K = B_i; \boldsymbol{\theta}). \quad (5)$$

Compare Eq. 5 with the approximate continuous likelihood developed in [9], pp. 102–107 (see also [3], pp. 528–533)

$$\tilde{L} \propto \lambda^N \exp\left(-\lambda \sum_{i=1}^N W_i\right) \prod_{i=1}^N g(B_i; \boldsymbol{\theta}), \quad (6)$$

where g is the continuous clump-length density. The difficulty with Eq. 6 is that we must obtain g through the Laplace-Stieljes transform ([3], pp. 531–532, [9], p. 91)

$$\gamma(s) = 1 + (s/\lambda) - \left(\lambda \int_0^\infty \exp\left[-st - \lambda \int_0^t \{1 - C(x)\} dx\right] dt\right)^{-1}. \quad (7)$$

Equation 5 offers a stable, computationally efficient approximate likelihood for general segment-length distributions C . We show two examples next.

4 Examples

We illustrate an approximate maximum likelihood estimation with two disparate examples. The first is a “classic” BM of random-radius disks. This model has received the most attention in the literature. One can imagine this model being useful for “soft” grains. The second uses crystal-like square grains.

4.1 Disks With Lognormally Distributed Radii

Consider a 2D BM with Poisson intensity ρ and grains that are disks with random radii distributed according to density ϕ . The i^{th} noncentral moment is

$$M_i = \int_0^\infty r^i \phi(r) dr.$$

For the disk process $M_i = R_0^i \exp\{i^2 \beta^2 / 2\}$. The vacancy (theoretical proportion not covered) is $v = \exp\{-\pi \rho M_2\}$. The induced 1D BM has Poisson intensity $\lambda = 2\rho M_1$ and segment length distribution

$$C(x) = 1 - \frac{1}{M_1} \int_{x/2}^\infty \sqrt{r^2 - x^2/4} \phi(r) dr. \quad (8)$$

In this example, the radii are lognormally distributed

$$\phi(r) = \frac{1}{r\beta\sqrt{2\pi}} \exp\left\{-\frac{1}{2} [\log(r/R_0)/\beta]^2\right\}.$$

The parameters of interest are (ρ, R_0, β) . Figure 2 shows two such BM realizations.

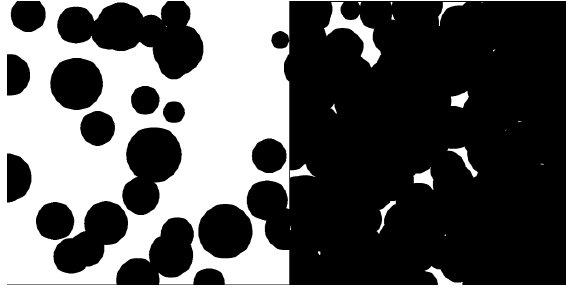


Figure 2: Log-normally distributed radii BMs with $\beta = 0.25, R_0 = 1$: $\rho = 0.1$ (left), $\rho = 1.0$ (right).

To demonstrate maximum likelihood estimation, we fix $R_0 = 1$ and generate a BM with $\rho = 0.4$ and $\beta = 0.5$ (vacancy is $v = 0.126$). Figure 3 shows a particular realization. The image was generated using a pseudorandom number generator to output 671 positions with uniformly distributed x and y coordinates on a 4096×4096 square. This assumes a sampling rate of 100 pixels per unit distance horizontally and vertically. Pseudorandom variates were generated for radii at each position and the resulting coordinates and radii were written as rendering commands into a PostScript file. The PostScript file was rendered as a black and white 4096×4096 raster image. Runlengths were sampled from this image every 200 pixels horizontally and vertically, resulting in a total of 154 complete white and black runlength pairs. Sample spacing was chosen large enough to reasonably assure independent binomial germ-grain processes, but small enough to produce many runlengths. Clearly, the optimal spacing depends on the process in a complicated way; the one chosen here was for convenience and produces a reasonable result. Also, samples were taken orthogonally to double the sample size at the expense of introducing a small amount of dependence; in practice, this appears to be a fruitful trade-off. Optimization was done using the downhill simplex method because of its simplicity and robustness. The elementary trapezoidal rule was used for numerical integration, once again, for its simplicity [13]. Using numerical optimization, the maximum likelihood estimate is $(\hat{\rho}, \hat{\beta}) = (0.36, 0.51)$. Figure 4 shows a two-parameter approximate 95% likelihood ratio-based confidence region $\{(\rho, \beta) : -2(l(\rho, \beta) - l(\hat{\rho}, \hat{\beta})) \leq 5.99\}$.

4.2 Randomly-Oriented Squares

In this example, we posit a BM composed of two independent BMs, Ξ_1 and Ξ_2 , each having randomly rotated (uniformly) squares as grains. The processes differ in fixed square sizes, D_1 and D_2 , and intensities, ρ_1 and ρ_2 . Let $\Xi = \Xi_1 \cup \Xi_2$ be the BM from which we sample. For a randomly-oriented-square BM, the induced linear process has intensity $\lambda_i = 4\rho_i D_i / \pi$ and segment-length distribution [14]

$$C_i(x) = \begin{cases} x/2D_i, & x \leq D_i \\ 1 - \frac{x^2 - 2D_i\sqrt{x^2 - D_i^2}}{2xD_i}, & D_i \leq x \leq D_i\sqrt{2} \\ 1, & \text{otherwise.} \end{cases}$$

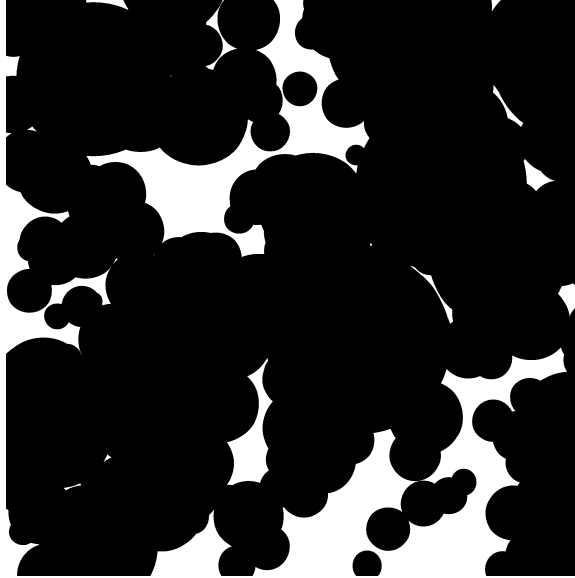


Figure 3: BM realization with lognormally distributed radii: $(\rho, R_0, \beta) = (0.4, 1.0, 0.5)$.

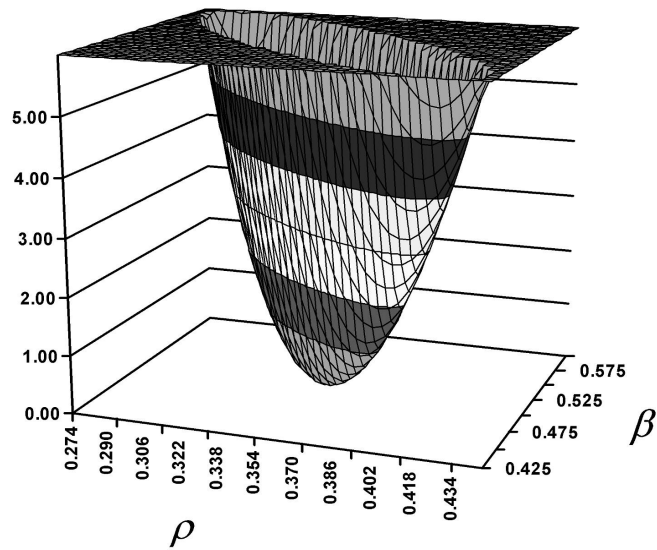


Figure 4: Approximate 95% confidence region for $(\hat{\rho}, \hat{\beta}) = (0.36, 0.51)$ where $(\rho, \beta) = (0.4, 0.5)$.

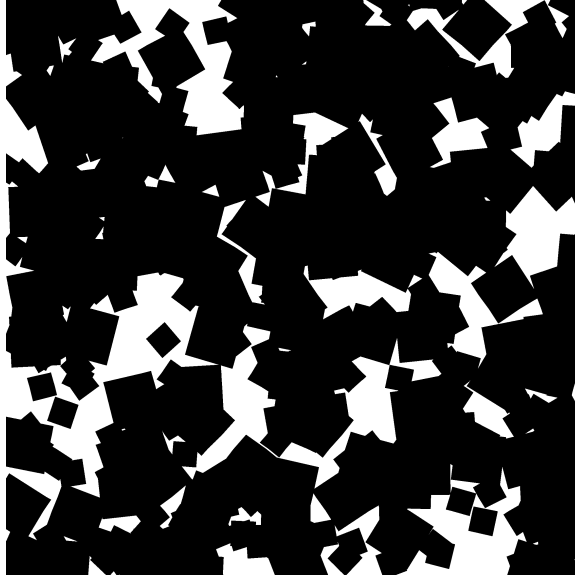


Figure 5: Mixed-square BM realization with $D_1 = 1$, $D_2 = 2$, $\rho_1 = 0.4$, and $\rho_2 = 0.3$.

We assume that the square sizes are known (they are easily measured from a realization) and that intensities ρ_1 and ρ_2 are to be estimated. Figure 5 shows a realization of the mixed-square BM where $D_1 = 1$, $D_2 = 2$, $\rho_1 = 0.4$, and $\rho_2 = 0.3$ (generated similarly to the random-disk example). As before we take ten equally-spaced slices horizontally and vertically from the realization for a total of 394 complete black/white runlength pairs. To approximate the likelihood of the mixture, in Eq. 4 let $F = F_1 F_2$ where $F_i(x) = 1 - \frac{\Delta_i}{n} + \frac{\Delta_i}{n} C_i(x)$. Optimizing the likelihood function, we obtain $(\hat{\rho}_1, \hat{\rho}_2) = (0.38, 0.29)$. Figure 6 shows an approximate 95% likelihood-ratio-based two-parameter confidence region, $\{(\rho_1, \rho_2) : -2(l(\rho_1, \rho_2) - l(\hat{\rho}_1, \hat{\rho}_2)) \leq 5.99\}$.

5 Summary

By approximating the 1D BM with a discrete counterpart, recursive formulas for runlength probabilities provide computationally efficient likelihood procedures. These procedures depart fundamentally from the usual and more difficult continuous methods. Another difference is that the discrete approach relies on direct decomposition of 1D covering events rather than transform methods. Estimation for higher dimensional models is done using linear samples as demonstrated by two examples of 2D estimation.

References

- [1] Cressie, N. and G. M. Laslett, “Random set theory and problems of modeling,” *SIAM Review*, Vol. 29, 1987, pp. 557–574.

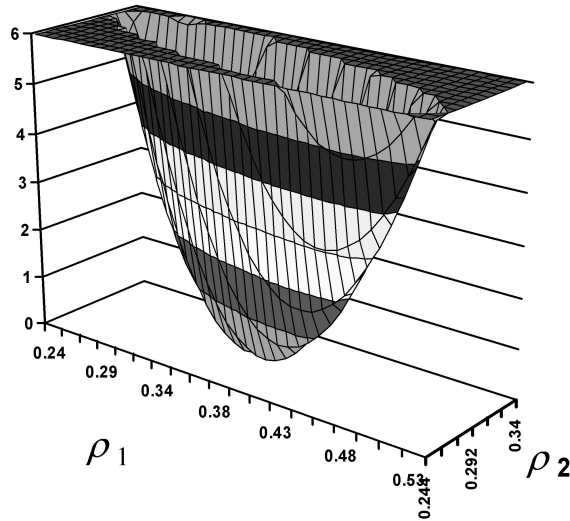


Figure 6: Approximate 95% confidence region for the maximum likelihood estimate $(\hat{\rho}_1, \hat{\rho}_2) = (0.38, 0.29)$; population value is $(0.3, 0.4)$.

- [2] Diggle, P. J., “Binary mosaics and the spatial pattern of heather,” *Biometrics*, Vol. 37, 1981, pp. 531–539.
- [3] Dougherty, E. R., *Random Processes for Image and Signal Processing*, SPIE/IEEE Press, Bellingham/NY.
- [4] Dougherty, E. R. and J. C. Handley, “Recursive maximum-likelihood estimation in the one-dimensional discrete Boolean random set model,” *Signal Processing*, Vol. 43, No. 1, 1995, pp. 1–15.
- [5] Dupač, V., “Parameter estimation in the Poisson field of disks,” *Biometrika*, Vol. 67, 1980, pp. 187–190.
- [6] Frey, A. and V. Schmidt, “Marked point processes in the plane I – A survey with applications to spatial modeling of communications networks,” *Advances in Performance Analysis*, Vol. 1, No. 1, 1998, pp. 65–110.
- [7] García, P., *Texture Analysis Using the One-Dimensional Boolean Model*, Ph.D. Dissertation, Universitat Jaume I, Departament d’Informàtica, Castelló, Spain, October 1999.
- [8] García, P. and M. Petrou, “Classification of binary textures using the one-dimensional Boolean model,” *IEEE Transactions on Image Processing*, Vol. 8, No. 10, 1999, pp. 1457–1462.
- [9] Hall, P., *Introduction to the Theory of Coverage Processes*, John Wiley, NY, 1988.

- [10] Handley, J. C., “Discrete approximation of the linear Boolean model of heterogeneous materials,” *Physical Review E*, Vol. 60, No. 5, 1999, pp. 6150–6152.
- [11] Handley, J. C. and E. R. Dougherty, “Optimal nonlinear filter for signal-union-noise and runlength analysis in the directional one-dimensional discrete Boolean random set model,” *Signal Processing*, Vol. 51, No. 3, 1996, pp. 147-166.
- [12] Molchanov, I. S., *Statistics of the Boolean Models for Practitioners and Mathematicians*, John Wiley, Chichester, 1996.
- [13] Press, W., B. Flannery, S. Teukolusky and W. Vetterling, *Numerical Recipes in C*, Cambridge University Press, NY, 1988.
- [14] Quintanilla, J. and S. Torquato, “Lineal measures of clustering in overlapping particle systems,” *Physical Review E*, Vol. 54, No. 4, 1996, pp. 4027–4036.
- [15] Sidiropoulos, N. D., J. S. Baras and C. A. Berestein, “Algebraic analysis of the generating functional for discrete random sets and statistical inference for intensity in the discrete Boolean random-set model,” *Journal of Mathematical Imaging and Vision*, Vol. 4, 1994, pp. 273–290.
- [16] Stadge, W., “The busy period of the queueing system $M/G/\infty$,” *Journal of Applied Probability*, Vol. 22, 1985, pp. 697–704.
- [17] Stoyan, D., W.S. Kendall and J. Mecke, *Stochastic Geometry and Its Applications*, John Wiley, NY, 1987.
- [18] Torquato, S., *Random Heterogeneous Materials*, Springer-Verlag, NY, 2001.