Parametric Estimation of a Boolean Random Field

S. Sundhar Ram and D. Manjunath

***Abstract*— We develop generalized method of moments estimators to estimate the parameters of a two-dimensional Boolean random field from measurements made on the coverage process induced on a straight line in the field. This is distinct from earlier studies e.g. [1], [2], where the parameters of a two-dimensional Boolean field are obtained from the coverage properties of a two-dimensional set. This problem has applications in radio-active field monitoring.**

***Index Terms*— Radioactive Field Estimation, Boolean Field, Parameter Estimation, Health of a Wireless Sensor Network**

I. Introduction and Preliminaries

A two-dimensional Boolean random field is a collection of random sets $\{X_i + C_i\}$, where $X_i \in \Re^2$ are the points of a Poisson process, C_i are i.i.d. random subsets in $\Re^2$ and $X_i + C_i =: \{X_i + x : x \in C_i\}$ [1]. In this paper, our interest is in estimating the parameters of a two-dimensional Boolean random field from measurements made on the coverage process induced on a straight line (one dimensional set) in the field. This is distinct from earlier studies e.g., [1], [2], where the parameters of a two-dimensional (n-dimensional) Boolean field are obtained from the coverage properties of a two-dimensional (respectively, n-dimensional) set. This is also distinct from the non-parametric field estimation and detection problems studied in the context of sensor networks, e.g., [3], [4].

The motivations for the results obtained in this paper are many but we focus on two problems. The first problem is that of sensing radioactive deposits in a field [5]. Consider a sensor traveling along a straight line path in the field, where it is activated by the radioactive deposits. The pattern of activation of the sensors is to be used to estimate the spatial density and the strength of the deposits. By suitably modeling the radioactive field, the problem can be reduced to the Boolean parameter estimation problem. Let the radioactive deposits be distributed according to a spatial Poisson process $\{X_i\}$ where the i is just any systematic indexing of the particles. The *radiation-exposure region*, the region within which a sensor can be activated, of the i^{th} deposit can be modeled as a random set C_i. The mapping to the Boolean field parameter estimation problem is obvious with the Poisson density of $\{X_i\}$ measuring the quantity of the deposit and the size of $\{C_i\}$ measuring the strength of the deposit. Another motivation is in monitoring the health of a randomly deployed sensor network by estimating the density of the active sensors and their effective range. We discuss this monitoring problem in more detail in Section III.

The first author is a first year graduate student at the Electrical and Computer Engineering Deptt., University of Illinois at Urbana-Champaign. He can be contacted at ssriniv5@uiuc.edu. The second author is with the Electrical Engineering Deptt., Indian Institute of Technology, Bombay, India. He can be contacted at dmanju@ee.iitb.ac.in.

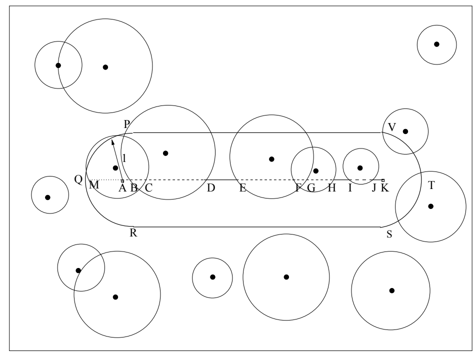

Fig. 1. L is the line-segment AK. Segments AC, BD, EG, FH and IJ are the covered components and the dashed segments AD, EH and IJ are the clumps. The region between the clumps, i.e., DE and HI, are the holes. The induced coverage of L is the union of the clumps i.e., $AD \cup EH \cup IJ$. The clump AD extends beyond L. Only points in $PQRSTV$ may affect the coverage of L.

A. Preliminaries

Consider a Boolean field in $\Re^2$. A point x in the two-dimensional field is said to be *covered* by the Boolean random field if $x \in \cup_i (X_i + C_i)$. For a set $A \subset \Re^2$, we will refer to $\bigcup_i A \cap (X_i + C_i)$ as the *coverage* of A and each $A \cap (X_i + C_i)$ as the *covered component* i (see Fig. 1). Our interest is in the case when A is a straight line-segment. In this case, $\bigcup_i A \cap (X_i + C_i)$, will be a union of disjoint line segments. We call each of these disjoint line-segments as *clumps* in A and the total length of $\bigcup_i A \cap (X_i + C_i)$ as the *covered length* of A (see Figure 1). Thus the clumps are unions of overlapping covered components. The segments between the clumps are called *holes*.

We consider a specific case of the above problem. Let B denote a two-dimensional Boolean field, where $\{X_i\}$ has density λ and the sets C_i are circles with random radii δR_i. Here δ, $0 \leq \delta \leq 1$, is a scaling constant and R_i are i.i.d random variables distributed as R with density function $f_R(r)$. We assume that the sensing radius has a finite support. Equivalently, we take the support of R, to be in $[0, 1]$. For the radiation deposit example, δ can be viewed as a measure of the deposit strength and R as a reference condition, e.g., initial strength. It is reasonable to assume that the statistics of R are known.

Our primary interest is to estimate λ and δ from the number of clumps and the clump lengths on a straight line segment L of length l_0. Note that this is equivalent to estimating the Boolean parameters from the hole lengths and the no. of holes. In the radioactive deposit measurement example, this corresponds to the case when the monitoring sensor travels along L and at any location it cannot distinguish the number deposits that are activating it. The observation by the sensors

1-4244-0728-1/07/$20.00 ©2007 IEEE

here is similar to a *paralyzable (or Type II)* counter [6]. Other sensor activation models are discussed in Section III.

Let L be a straight line path of length l_0. Let $\mathcal{L}$ be the extension of L to infinity on both sides. To obtain the estimators we first need to understand the coverage process on $\mathcal{L}$. The clumps on $\mathcal{L}$ correspond to the coverage of $\mathcal{L}$ induced by the two-dimensional Boolean process B described earlier. It was proved in [7] that the random process on $\mathcal{L}$ is an $M/G/\infty$ queue (equivalently an one-dimensional Boolean process). Specifically,

1) The coverage process induced on $\mathcal{L}$ has the same statistics as the one-dimensional Boolean process, $\{\bar{X}_i + \bar{C}_i\}$ where $\bar{X}_i$ and $\bar{C}_i$ are as follows. $\{\bar{X}_i\}$ is a Poisson arrival process on $\mathcal{L}$ of density $\bar{\lambda} = 2\lambda\beta\delta$, $\bar{C}_i$'s are the random i.i.d intervals $(0, 2\delta\bar{R}_i]$ and the $\bar{R}_i$'s are i.i.d random variables distributed as $\bar{R}$ with density
$$f_{\bar{R}}(\bar{r}) = \begin{cases} \frac{\bar{r}}{\beta}\int_{\bar{r}}^{1} \frac{f_R(r)}{\sqrt{r^2-\bar{r}^2}}\, dr & \text{for } 0 \le \bar{r} \le 1 \\ 0 & \text{otherwise} \end{cases}.$$
Observe that the $\bar{C}_i$s are essentially the covered components and $\mathsf{E}(\bar{R}) = \frac{\pi\mathsf{E}(R^2)}{4\mathsf{E}(R)}$.
2) The clump lengths, Z_i, are i.i.d. random variables distributed as Z, with mean
$$\mathsf{E}(Z) = \frac{e^{\pi\mathsf{E}(R^2)\delta^2\lambda} - 1}{2\lambda\mathsf{E}(R)\,\delta}. \tag{1}$$
3) The total covered length, C_L, has mean
$$\mathsf{E}(C_L) = l_0\left(1 - e^{-\pi\delta^2\lambda\mathsf{E}(R^2)}\right). \tag{2}$$

II. A Generalized Method of Moments Estimator

A closed form expression for the distribution and the variance of the clump length is not available. Hence, developing a maximum likelihood estimator or a method of moments estimator seems impractical. We instead devise a *generalized method of moments* (GMM) estimator. GMM estimators have many desirable properties including consistency, asymptotic unbiasedness and asymptotic normality [8].

The generalized moments that we use are the expected total covered length (stated in Eqn. 2) and the expected clump length (stated in Eqn. 1). If c_L and $\bar{z}$ are the sample values of the covered length and the sample mean clump length, then the GMM estimators, $\hat{\delta}_1$ and $\hat{\lambda}_1$, are obtained as the minimizers of $\epsilon_1 = \left(\bar{z} - \frac{e^{\pi\mathsf{E}(R^2)\delta^2\lambda}-1}{2\lambda\beta\delta}\right)^2 + \left(c_L - l_0\left(1 - e^{-\pi\delta^2\lambda\mathsf{E}(R^2)}\right)\right)^2$. We get,
$$\hat{\delta}_1 = \frac{2\mathsf{E}(R)\,(l_0 - c_L)\;\bar{z}}{\pi\mathsf{E}(R^2)\,c_L}\ln\left(\frac{l_0}{l_0 - c_L}\right) \tag{3}$$
$$\hat{\lambda}_1 = \frac{\ln\left(\frac{l_0}{l_0 - c_L}\right)}{\pi\mathsf{E}(R^2)\,\hat{\delta}_1^2} = \frac{c_L}{2\mathsf{E}(R)\,(l_0 - c_L)\;\bar{z}\;\hat{\delta}_1}. \tag{4}$$

Let n_z be the number of clumps with the sample clump lengths being $z_1, z_2, \cdots, z_{n_z}$. The sample of the covered length, c_L, is $c_L = \sum_{i=1}^{n_z} z_i$. $\bar{z}$, is obtained as the sample mean of only clumps that are completely within L. Thus we will not include clumps that include the beginning and end of path L in obtaining $\bar{z}$. For example, in Fig. 1 the clumps that are observed are AD, EH and IJ. However the clump AD extends beyond L and is a part of the clump MD. Therefore, we ignore the length of AD in calculating the sample mean.

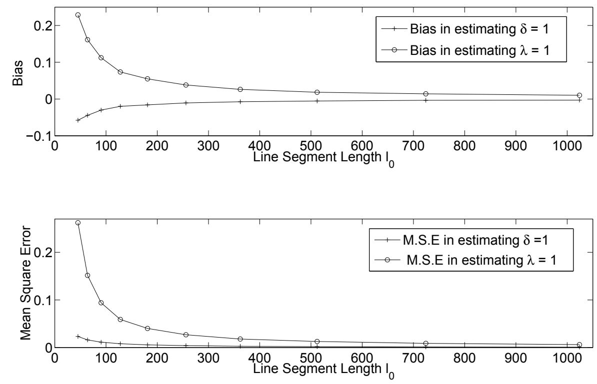

Fig. 2. The bias and mean square error in $\hat{\delta}_1$ and $\hat{\lambda}_1$ as a function of l_0.

We now study the estimator in Eqns. 3 and 4 through the following simulation model. Sensor nodes are deployed according to a Poisson process in the rectangle with diagonals at $(-1,-1)$ and $(l_0+1, 1)$. $f_R(r)$ is uniform in $[0,1]$, $\delta = 1$ and $\lambda = 1$. $\bar{z}$ and c_L are obtained from the coverage of $(0, l_0)$ of the x-axis. Fig. 2 plots the bias and the mean square error in the estimators as a function of l_0. These are averaged over $10,000$ samples. Note that bias in estimating a parameter θ using an estimator $\hat{\theta}$ is $\mathsf{E}\left(\hat{\theta}\right) - \theta$. We observe that the bias in $\hat{\delta}_1$ and $\hat{\lambda}_1$ when $l_0 = 50$ is approximately 2% and 10% respectively. As expected, the estimator bias and mean square error approach 0 with increasing l_0. Also observe that the bias in $\hat{\delta}_1$ is negative while $\hat{\lambda}_1$ has a positive bias. In the following we prove that this is indeed true. Solving for δ from Eqns. 1, 2 we get,
$$\delta = \frac{2\mathsf{E}(R)\,(l_0 - \mathsf{E}(C_L))\;\mathsf{E}(Z)}{\pi\mathsf{E}(R^2)\,\mathsf{E}(C_L)}\ln\left(\frac{l_0}{l_0 - \mathsf{E}(C_l)}\right)$$
Since the Z_is are i.i.d. random variables, $\mathsf{E}(Z) = \mathsf{E}(\bar{z})$. Also, $\mathsf{E}(C_L) = \mathsf{E}(c_L)$. Therefore,
$$\mathsf{E}\left(\hat{\delta}_1\right) - \delta = \frac{2\mathsf{E}(R)}{\pi\mathsf{E}(R^2)}\left(\mathsf{E}\left(\frac{(l_0 - c_L)\bar{z}}{c_L}\ln\left(\frac{l_0}{l_0 - c_L}\right)\right) - \frac{(l_0 - \mathsf{E}(c_L))\mathsf{E}(\bar{z})}{\mathsf{E}(c_L)}\ln\left(\frac{l_0}{l_0 - \mathsf{E}(c_L)}\right)\right)$$
Observe that the above expression is of the form $\mathsf{E}(f(c_L)\bar{z}) - f(\mathsf{E}(c_L))\mathsf{E}(\bar{z})$. $\bar{z}$ and c_L are positively correlated random variable. Also $f(c_L)$ is an decreasing function of c_L and hence $f(c_L)$ and $\bar{z}$ are negatively correlated. Therefore $\mathsf{E}(f(c_L)\bar{z}) \le \mathsf{E}(f(c_L))\,\mathsf{E}(\bar{z})$. Hence
$$\mathsf{E}\left(\hat{\delta}_1\right) - \delta \le \frac{2\mathsf{E}(R)\,\mathsf{E}(\bar{z})}{\pi\mathsf{E}(R^2)}\left(\mathsf{E}(f(c_L)) - f(\mathsf{E}(c_L))\right).$$
It can be proved that $f(\cdot)$ is a concave function. From Jensen's inequality it follows that $\mathsf{E}(f(c_L)) - f(\mathsf{E}(c_L)) \le 0$ and hence

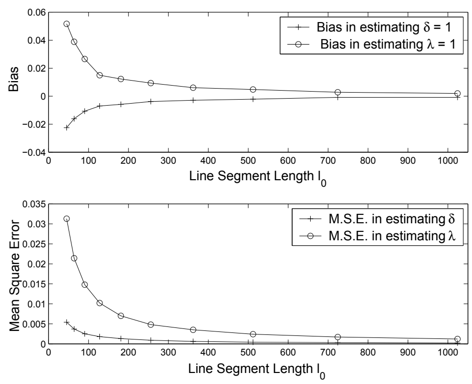

Fig. 3. The bias and mean square error in $\hat{\delta}_2$ and $\hat{\lambda}_2$ as a function of l_0.

$\mathsf{E}\left(\hat{\delta_1}\right) - \delta \leq 0$. Similarly, we can show $\hat{\lambda}_1$ overestimates.

III. Other Measurement Models

The next measurement model we consider has an application in network monitoring. Typically, sensors are driven by a fixed battery supply of finite energy that cannot be replaced. In any sensor network, with time and use there is a gradual decrease in the available energy and when it reduces below a threshold, the sensor becomes non-functional. There are also unpredictable sensor losses due to sensor faults and enemy action. Sensor network monitoring will involve identifying such sensor losses and monitoring other critical network properties.

We consider a system where external mobile agents, called scanners, travel into the network and make local observations about the communication capabilities of the sensor nodes in the network. The scanner communicates with a sensor when it is in the communication region of that sensor. The scanner travels along L and measures the number of sensors it communicates with, N_s, and the distances on L for which it could communicate with the sensors, $T_1, \cdots, T_{N_s}$. These are then used to obtain the sensor density and the parameters of the communication region. Such holistic approaches to network monitoring, as against node level fault detection, were first discussed in [9], where periodic updates of the aggregated 'health' of the network were obtained. Also, the idea of using external mobile agents to perform critical network activities is not new and has been discussed before as *mobile agent based distributed sensor network* (MADSN), e.g., in [10].

We assume that the sensor nodes are distributed according to a spatial Poisson process of density λ and the communication range of sensor node i is a circle of random radius δR_i with the R_is being i.i.d. We can then model the network communication map of the sensor network as a Boolean field similar to B. Observe that $T_1, \cdots, T_{N_s}$ will be the lengths of the covered components. Recall that the covered component i is the length of the coverage of sensor i. From Section II we have $\mathsf{E}(N_s) = 2\lambda\delta\mathsf{E}(R)$ and the T_i's are i.i.d. random variables distributed as T with mean $\mathsf{E}(T) = \frac{\delta\pi\mathsf{E}\left(R^2\right)}{2\mathsf{E}(R)}$. Therefore if n_s is the sample of N_s and $\bar{t}$ the sample mean covered components lengths, the GMM estimators, $\hat{\delta}_2$ and $\hat{\lambda}_2$, are the minimizers of $\epsilon_2 = (n_s - 2\lambda\delta\mathsf{E}(R))^2 + \left(\bar{t} - \frac{\delta\pi\mathsf{E}\left(R^2\right)}{2\mathsf{E}(R)}\right)^2$. As in the previous case, we use the lengths of only those covered components that are completely within L to obtain $\bar{t}$. We get

$$\begin{aligned} \hat{\delta}_2 &= \frac{2\mathsf{E}(R)\,\bar{t}}{\pi\mathsf{E}(R^2)} \\ \hat{\lambda}_2 &= \frac{n_s}{2\mathsf{E}(R)\,\hat{\delta}_2}. \end{aligned}$$

Fig. 3 plots the bias and mean square error in the estimator as a function of l_0, averaged over 10000 iterations. f_R is uniform in $[0,1]$, $\delta = 1$ and $\lambda = 1$. From Figs. 2 and 3 we observe that the errors in δ_2 and λ_2 are almost an order smaller than the errors in δ_1 and λ_1. We present an intuitive argument for the same. In obtaining the GMM estimator we are approximating the expected value of the observations by their sample values. Since a clump is the union of covered components, $\mathsf{VAR}(n_z) \geq \mathsf{VAR}(n_s)$ and $\mathsf{VAR}(Z) \geq \mathsf{VAR}(T)$. Therefore $\bar{T}$ and n_s will be closer to their expected values than will $\bar{Z}$ and n_z. This explains the smaller errors in δ_2 and λ_2.

For the sake of completeness, we consider another observation model that may be only of theoretical interest and is motivated by the *type-I* or *non-paralyzable counter* [6]. In this model, if the starting points of the covered components are considered as arrival times and the end points as departure times, the random process 'seen' by the mobile agent is an $M/G/1/1$ queue, i.e., there is only one server and any new arrival is dropped if the server is busy. We estimate the Boolean field parameters from the total number, N_b, and the length, $B_1, \cdots, B_{N_b}$, of these busy periods in L. Unlike in the first observation model where the random process seen by the mobile agent is an $M/G/\infty$ queue and the busy period lengths are statistically equivalent to the clump lengths, in this model the busy lengths are equivalent to the covered component lengths. Since the arrival process is a Poisson point process, the time from the instant an arrival was completely serviced to the next arrival is exponentially distributed with arrival rate $\bar{\lambda}$. Therefore, the idle periods will be exponentially distributed with parameter $\bar{\lambda}$. By considering a busy period followed by an idle period as a renewal period, a strong law result can be obtained for the number of busy periods N_b in a time interval of duration l_0. From Theorem 3.5 of [6], with probability 1

$$\lim_{l_0 \to \infty} \frac{N_b}{l_0} = \frac{1}{1/\bar{\lambda} + \mathsf{E}\left(\bar{B}\right)} = \frac{2\lambda\delta\mathsf{E}(R)^2}{\mathsf{E}(R) + \pi\lambda\delta^2\mathsf{E}(R)\,\mathsf{E}(R^2)} \quad (5)$$

Let n_b and b_i be the samples of N_b and B_i respectively. Let $\bar{b}$ be the sample mean of the busy periods obtained by averaging the lengths of only those busy periods that are completely within L. Then, from Eqn. 5 and the fact that the busy periods (covered components) have mean $\frac{\delta\pi\mathsf{E}\left(R^2\right)}{2\mathsf{E}\left(R^2\right)}$, the following GMM estimator can be obtained.

$$\begin{aligned} \hat{\delta}_3 &= \frac{2\mathsf{E}(R)\,\bar{b}}{\pi\mathsf{E}(R^2)} \\ \hat{\lambda}_3 &= \frac{n_b}{2\mathsf{E}(R)\,l_0\hat{\delta}_3 - \pi n_b\mathsf{E}(R^2)\,\hat{\delta}_3^2} \end{aligned}$$

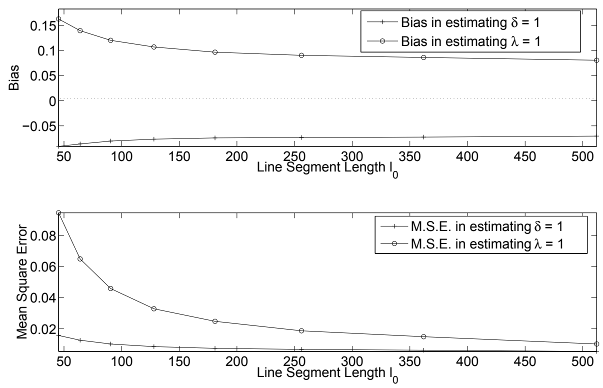

Fig. 4. The bias and mean square error in $\hat{\delta}_3$ and $\hat{\lambda}_3$ as a function of l_0.

Fig. 4 plots the bias and mean square error in the estimator as a function of l_0, averaged over 10000 iterations. f_R is uniform in $[0, 1]$, $\delta = 1$ and $\lambda = 1$. As expected the bias and the mean square decrease to 0 as l_0 increases. Further, from Figs. 3 and 4 we observe that the errors are greater for δ_3 and λ_3. This is because the number of complete covered components observed in this model will be smaller as we drop arrivals when the server is busy.

IV. Discussions

In this paper we have developed simple estimators to estimate the parameters of a Boolean field from its one-dimensional coverage properties. A comparison of the estimators developed in this paper with other estimation procedures, discussed in literature, is not instructive as these procedures use different measurements to estimate the parameters. The analyses of homogeneous Boolean fields that we have presented can also be extended to non homogeneous Boolean fields. Since $\delta \leq 1$ and the support of R is in $[0, 1]$, only Poisson points within a distance of 1 affect the coverage properties of a line segment. For example, in Fig. 1 only points in $PQRSTV$ may affect the coverage of the L. Thus, if the density of the non-homogeneous field changes slowly we can make the assumption that the density is uniform over the area that can potentially cover L. Such estimates from sensors monitoring different parts of the field can then be stitched together to estimate the complete non-homogeneous field. Also, along similiar lines it is possible to estimate the parameters of a $k-$dimensional Boolean field from the coverage properties of a $l-$dimensional set in the field.

References

[1] P. Hall, *Intoduction to the Theory of Coverage Process*, John Wiley and Sons, 1988.

[2] I. Molchanov and D. Stoyan, “Asymptotic Properties for Paramters of the Boolean Model,” *Advances in Applied Probability*, 2004.

[3] L. Heinrich and M. Werner, “Kernel Estimation of the Diameter Distribution in Boolean models with Spherical Grains,” *Journal of Nonparametric Statistics*, 2000.

[4] C. Chiasserini, A. Nordio, and E. Viterbo, “On Data Acquisition and Field Reconstruction in Wireless Sensor Networks,” in *Tyrrhenian International Workshop on Digital Communications*, July 2005.

[5] A. Jeremić and A. Nehorai, “Landmine Detection and Localization Using Chemical Sensor Array Processing,” *IEEE Transactions on Signal Processing*, vol. 48, 2000.

[6] S. Ross, *Applied Probability Models with Optimization Applications*, Dover Publications, 1970.

[7] S. Sundhar Ram, D. Manjunath, S. K. Iyer, and D.Yogeshwaran, “On the Path Tracking Properties of Random Sensor Networks,” 2006, To appear in IEEE Transactions on Mobile Computing. Available at `http://www.ifp.uiuc.edu/~ssriniv5` .

[8] W. K. Newey and D. McFadden, “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics*, vol. 4, chapter 36, pp. 2111–2245. Elsevier, 1994.

[9] Y. J. Zhao, R. Govindan, and D. Estrin, “Residual Energy Scan for Monitoring Sensor Networks,” in *IEEE Wireless Communication and Networking Conference*, March 2002, vol. 1, pp. 356–362.

[10] H. Qi, S. S. Iyengar, and K. Chakrabarty, “Multi-Resolution Data Integration Using Mobile Agents in Distributed Sensor Networks,” *IEEE Transactions on Systems Man and Cybernetics PART C*, vol. 31, no. 3, 2001.