The Forest as Seen by its Trees

Terry R. McConnell
Syracuse University

October 24, 2012

Abstract

We consider estimating the density of points in a point process by a count of the number of points in a unit volume, and the bias that results when the volume is centered at a point of the point process rather than at a randomly chosen location.

1 Introduction

A common problem in spatial statistics is to estimate the density of points in a point process. If the process is translation invariant, then its intensity measure - the expected number of points in a unit volume set - is a multiple of Lebesgue measure. In this case one seeks a good point estimator of the multiplier μ.

Assuming the point process to be 2 dimensional for the sake of discussion, a reasonable approach would be to count the number of points in a randomly chosen unit area. The authors of [1] used this method to estimate the density of trees in a forest using tree stem counts from a collection of circular plots that had been chosen to study the environment around sampled trees. Thus each plot was centered on a tree rather than on a randomly selected point. Since the resulting density estimate would likely be biased in the direction of overestimating the true density of trees, particularly if there were a significant tendency for trees to grow in clumps, it was necessary to estimate and correct for this bias. It is easy to imagine other scenarios that would lead to the same issue. For example, one might attempt to estimate the density of stars from old photographic plates, each of which would naturally have been centered upon some object of interest.

If we assume trees are distributed according to a spatially homogenous Poisson point process, then it is not difficult to show [1] that the bias is exactly 1 tree. On the other hand, the Poisson model does not seem to fit the actual distribution of trees very well. There is a tendency for tree stems to be more dispersed than they would be in the Poisson model due to competition among trees.

In section 2 we formulate and prove a result on the bias that generalizes the Poisson case in a way that may be more realistic for applications. In section 3, we obtain upper and lower bounds of a similar form that hold assuming only translation invariance.

2 A Bias Identity

Let F (the *forest*) be a large sphere in d-dimensions that contains points (the *trees*) distributed according to a random point process. Let $N(A)$ denote the number of trees in a given set A. Let U be a unit volume sphere centered at a point chosen at random, uniformly from F, and V a unit volume sphere centered at a point chosen at random, uniformly from the set of trees in F. (We will assume the underlying point process is translation invariant with an intensity measure that is a finite positive multiple μ of Lebesgue measure. Thus, in particular, the number of trees in the forest is finite with probability one.)

Let $\nu = EN(V)$. We can also calculate ν as the expected value of $N(U)$ conditioned on the event that there is a tree at the center of U, under any set of assumptions where this interpretation makes sense, and this is the way it will be calculated below. Also, since the point process is translation invariant, we may and do assume that U is the unit volume sphere centered at the origin.

Given that $N(U) = n$, let $X_1, X_2, \ldots, X_n$ denote the positions of the trees in U. Marginally, the X_i are uniformly distributed on U, but not generally independent unless the point process is Poisson. (See, for example, [2] for background on point processes, and chapter 2 of that reference for an introduction to Poisson point processes.) As mentioned above, actual tree locations tend not to be independent. We shall assume that the underlying point process is such that the joint distribution of the X_i has a density with a continuous version. In particular, this allows for the conditioning interpretation of ν. Let B_ϵ denote a small sphere of volume ϵ centered at the origin. We shall further assume that there is $\epsilon_0 > 0$ such that for any n and distinct i and j in the range $1 \ldots n$ we have

$$P_n(X_i \in B_\epsilon, X_j \in B_\epsilon) \leq P_n(X_i \in B_\epsilon) P_n(X_j \in B_\epsilon), \epsilon < \epsilon_0. \tag{1}$$

(Here P_n denotes the conditional probability given $\{N(U) = n\}$.) Thus, a pair of trees is less likely to be found in the same small volume than it would be assuming the tree locations were independent. We also assume that the distribution of individual tree locations under P_n is uniform. This prohibits examples such as a random shift of the integer lattice points.

The final assumption we need is that the probability function f of $N(U)$ is such that $N(U)$ has a finite third moment. Under this, and the other assumptions above, we shall now prove the following identity relating means μ and ν, and the variance σ^2 of $N(U)$:

$$\nu = \mu + \frac{\sigma^2}{\mu}. \tag{2}$$

To see this, first note that for any $0 < \epsilon < \epsilon_0$ and $n \geq 1$ we have

$$n\epsilon - n^2\epsilon^2 < P_n\left(\bigcup_{i=1}^{n}\{X_i \in B_\epsilon\}\right) \leq n\epsilon. \tag{3}$$

The right hand inequality is obvious. The left-hand inequality follows from (1) and the inclusion-exclusion formula.

Let A denote the event that at least one tree lies in B_ϵ. Then by Bayes' formula we have, for any $n \geq 1$,

$$P(N(U) = n|A) = \frac{P_n(A)f(n)}{\sum_{k=1}^{\infty} P_k(A)f(k)}. \tag{4}$$

Use the left-hand inequality in (3) in the numerator and the right-hand inequality in each term of the denominator to obtain the inequality $P(N(U) = n|A) \geq \frac{(n-n^2\epsilon)f(n)}{\mu}$. Multiply both sides of this inequality by n and sum from $n = 1$ to infinity to obtain $E(N(U)|A) \geq \frac{EN(U)^2}{\mu} - \frac{\epsilon}{\mu}EN(U)^3$. Letting ϵ tend to zero, we have $\nu \geq \mu + \frac{\sigma^2}{\mu}$. The opposite inequality is proved similarly, using the left side of (3) in the denominator of (4) and the right side of (3) in the numerator of (4). This completes the proof of (2).

3 General Upper and Lower Bounds

In this section we retain the notation introduced above, but here it is more convenient to assume that U and V are unit volume cubes rather than spheres. Also assume that the forest, F, is a very large cube. Denote by αU the cube having the same center as U, but α times the side length. Denote by $|A|$ the Lebesgue measure of a set A.

We assume as before that trees are distributed according to a translation invariant point process with intensity measure that is a positive finite multiple of d-dimensional Lebesgue measure, but impose no further assumptions than these. We shall prove in this setting that

$$c_d\left(\mu + \frac{\sigma^2(N(\frac{1}{2}U))}{\mu}\right) \quad \leq \quad \nu \quad \leq \quad C_d\left(\mu + \frac{\sigma^2(N(2U))}{\mu}\right) \tag{5}$$

holds in the limit as the size of the forest becomes infinite. (Here c_d and C_d are constants depending only on dimension.)

To prove (5), let B be the event that the center of U lies in $\frac{1}{2}V$, or equivalently, that the center of V lies in $\frac{1}{2}U$. Note that we have

$$P(B) = \frac{1}{2^d|F|}, \tag{6}$$

which can be seen by first conditioning on the location of V. Also, note that B is independent of $N(V)$. (On the other hand, B is *not* independent of $N(U)$, since the occurrence of B entails having at least one tree in U.)

Now we have

$$\nu \;=\; \frac{E(N(V);B)}{P(B)} \;\leq\; \frac{E(N(2U);B)}{P(B)}, \tag{7}$$

since on the event B we have $V \subseteq 2U$.

But,

$$P(B|N(2U)) \;\leq\; P(V \text{ is centered in } 2U|N(2U)) \;=\; \frac{N(2U)}{N}, \tag{8}$$

where N is the number of trees in the forest. Using this in (7) above, we have

$$\nu \;\leq\; \frac{EN^2(2U)}{NP(B)}, \tag{9}$$

and the right inequality in (5) with $C_d = 2^{3d}$ follows, since by (6) we have $NP(B) = 2^{-d}\mu$. The left inequality in (5), with $c_d = 2^{-d}$, can be proved by a similar argument: One has $\frac{1}{2}U \subseteq V$ on the event B, hence

$$\nu \;\geq\; \frac{E(N(\frac{1}{2}U);B)}{P(B)} = \frac{E(P(B|N(\frac{1}{2}U))N(\frac{1}{2}U))}{P(B)} \;=\; \frac{EN^2(\frac{1}{2}U)}{NP(B)}, \tag{10}$$

and the rest follows as before.

References

[1] D.G. Ray, R.D. Yanai, R.D. Nyland, and T.R. McConnell, Growing-Space Relationships in Young Even-Aged Northern Hardwood Stands Based on Individual-Tree and Plot-Level Measurements, *Northern Journal of Applied Forestry*, submitted.

[2] J. Møller and R.P. Waagepetersen, *Statistical Inference and Simulation for Spatial Point Processes*, Chapman Hill, Boca Raton, 2004