ON TWO METHODS OF DISTANCE SAMPLING

by

R. Johnson and N. A. Langberg

FSU Statistics Report M450

January, 1978
The Florida State University
Department of Statistics
Tallahassee, Florida 32306

Key Words: strong consistency, asymptotic normality.
ON TWO METHODS OF DISTANCE SAMPLING

by

R. Johnson and N. A. Langberg

ABSTRACT

We consider the estimation of the number of individuals in a unit area for phenomena where individuals are spread in a region.

Two methods of distance sampling are studied, estimators and their optimal properties are presented and a comparison between the two methods is conducted.
1. **Introduction and Summary.**

Trees in a forest, stars in a galaxy, and bacteria on a petri dish, are a few examples of a natural spread of individuals in a region, (say R). In the pursuit of a better understanding of these phenomena, attempts were made to use probabilistic and statistical tools.

We start with a short representation of a probabilistic model, useful in analyzing some aspects of the phenomena.

Let \( \{X(A)\}_{A \in \Sigma} \) be a stochastic process, where \( R \) is a Euclidian space, \( \Sigma \) the collection of Borel sets of finite Lebesgue measure (say \( \lambda \)) and \( X(A) \) the number of individuals in the set \( A \), \( (A \in \Sigma) \).

On the stochastic process we impose the following two conditions:

The distribution of \( X(A) \) varies only with \( \lambda(A) \).  

\[
(1.1)
\]

For every \( m \) disjoint sets in \( \Sigma \), \( A_1, A_2, \ldots, A_m \), the associated random variables \( X(A_1), \ldots, X(A_m) \) are independent.

\[
(1.2)
\]

Conditions (1.1) and (1.2) insure (1.3).

The processes \( \{Y_p(t)\}_{t>0} \) and \( \{\sum_{i=1}^{N(t)} Z_i\}_{t>0} \) are equivalent.

\[
(1.3)
\]

[Where \( Y_p(t) \) is the number of individuals in a sphere of Lebesgue measure \( t \) and arbitrary fixed center \( P \), \( \{N(t)\}_{t>0} \) is a Poisson process with parameter \( \mu \), \( \{Z_i\}_{i=1, 2, \ldots, n \ldots} \) is a i.i.d. sequence of random variables, independent of \( \{N(t)\}_{t>0} \), with values in \( \{1, 2, 3, \ldots, n \ldots\} \).]

Two distributions frequently used for \( X(A) \), \( (A \in \Sigma) \), are the Poisson, where \( Z_1 \equiv 1 \) [Eberhardt (1967), Pollard (1971)]. and the Negative Binomial, where

\[
P(Z_1 = k) = \frac{-\theta^k}{k \cdot \ln(1 - \theta)} \quad k = 1, 2, \ldots, n \ldots, \quad 0 < \theta < 1. \quad \text{[Patil and Stiteler}
(1974).] We conclude by noting that (1.1) and (1.2) suffice to determine uniquely the process \( \{X(A), A \in \mathcal{L}\} \), in terms of \( \{Y_p(t)\}_{t \geq 0} \).

For inference purposes statisticians use two sampling methods: (a) Plot sampling, where disjoint regions are randomly selected and the number of individuals in each region is counted. [Pielow (1969). Patil and Stiteler (1974).] (b) Distance sampling, where distances between individuals or between points and individuals are measured. [Eberhardt (1967), Pielow (1969), Pollard (1971).]

Our objective is to estimate the average number of individuals in a region of Lebesgue measure one. Formally, by Wold identity and (1.3), our parameter of interest is \( \mu \in \mathbb{E} \mathcal{Z}_1 \), (say \( \theta \)). Naturally we assume that
\[
0 < \mathbb{E} X(A) < \infty \text{ for every } A \in \mathcal{L}.
\] (1.4)

We propose two different schemes, both in the category of distance sampling.

In scheme A we measure distances from a fixed point \( P_A \) to the \( n \) nearest clusters and count the number of individuals in each of the \( n \) clusters. In scheme B we measure distances from a fixed point \( P_B \), to enough successive clusters, stopping when for the first time we reach a total of at least \( L \) individuals. We again count the number of individuals in each cluster.

In Sections 2 and 3, we suggest estimators for \( \theta \), based on the two schemes respectively and present some of their optimal properties.

Section 4 is devoted to a comparison between the two schemes.
2. The analysis of scheme A.

The information collected in sampling scheme A contains two parts: n continuous random variables, \( W_1 < W_2 \ldots < W_n \), where \( W_i \) is the Lebesgue measure of the sphere determined by the center \( P_A \) and distance to the ith nearest cluster, \( i = 1, 2, \ldots, n \) and \( n \) independent discrete random variables, \( K_1, K_2, \ldots, K_n \), where \( K_i \) is the number of individuals in cluster \( i \), \( i = 1, 2, \ldots, n \).

From (1.3) we derive the following three conclusions:

The random vectors \((W_1, W_2, \ldots, W_n)\) and \((K_1, K_2, \ldots, K_n)\) are independent. \( (2.1) \)

\( W_1, W_2 - W_1, \ldots, W_n - W_{n-1} \) are i.i.d. exponential random variables with mean \( 1/\mu \). \( (2.2) \)

The random vectors \((K_1, K_2, \ldots, K_n)\) and \((Z_1, Z_2, \ldots, Z_n)\) are identically distributed. \( (2.3) \)

We start by estimating \( \theta \). We do so by estimating \( \mu \) by \( \hat{\mu}_n \) and \( \mu Z \) (say \( m \)) by \( \hat{m}_n \) and then use their product (say \( \hat{\theta}_n \)) to estimate \( \theta \).

Let

\[
\hat{\mu}_n = n/W_n, \quad \hat{m}_n = \frac{1}{n} \sum_{i=1}^{n} K_i/n \quad \text{and} \quad \hat{\theta}_n = \hat{\mu}_n \cdot \hat{m}_n. \]

\( (2.4) \)

The Kolmogorov strong law of large numbers, (2.2) and (2.3), imply the strong consistency of \( \hat{\mu}_n, \hat{m}_n \) and \( \hat{\theta}_n \). In addition \( \hat{m}_n \) is unbiased and \( \hat{\mu}_n \) is asymptotically (when \( n \to \infty \)) unbiased. For the sample \((W_1, W_2, \ldots, W_n)\), \( \hat{\mu}_n \) is sufficient and a maximum likelihood estimator of \( \mu \). Since no additional information on the distribution of \( Z_1 \) is available, \( \hat{m}_n \) is the "best" we can do. Our next objective is to prove the asymptotic normality of \( \hat{\theta}_n \).
To obtain the asymptotic normality of $\hat{\theta}_n$, we apply classical central limit theory and (2.2) to $\sqrt{n}(\hat{\theta}_n - \theta)$ and then use the transformation $1/x$ to get that $\sqrt{n}(\hat{\mu}_n - \mu)$ converges in law when $n \to \infty$ to a centered normal random variable with variance $\mu^2$.

To show the asymptotic normality of $\hat{\theta}_n$, we add the obvious assumption that $Z_1$ is a nondegenerate random variable. Since for every natural number $k$

$$\lim_{h \to 0} \frac{1}{h} \mathbb{P}(Y \Phi_h = k) = \mu \mathbb{P}(Z_1 = k),$$

we add condition (2.5).

For at least one natural $k$

$$0 < \lim_{h \to 0} \frac{1}{h} \mathbb{P}(Y \Phi_h = k) < 1. \tag{2.5}$$

Since $(\hat{\theta}_n - \theta)\sqrt{n} = m(\hat{\mu}_n - \mu) + \mu(\hat{\theta}_n - m) + A_n$ [where $A_n = \sqrt{n}(\hat{\mu}_n - \mu)(\hat{m}_n - m)$] and $A_n \xrightarrow{P} 0$, when $n \to \infty$, it suffices to consider the asymptotic behavior of $m(\hat{\mu}_n - \mu) + \mu(\hat{m}_n - m)$. Statement (2.1) insures the validity of (2.6).

If (2.5) is satisfied, then $\sqrt{n}(\hat{\theta}_n - \theta)$ converges in law to a centered normal random variable with variance $m^2\mu^2 + \mu^2 \sigma^2$. [where $\sigma^2 = \text{Var}(Z_1)$.]
3. The analysis of scheme B.

Let $R(L)$ be the stopping rule associated with scheme B. From (1.3) it follows that $R(L)$ equals

$$
\min \left\{ k \mid \sum_{i=1}^{k} Z_i \geq L \right\}.
$$

(3.1)

The information collected under scheme B consists of $R(L)$ continuous random variables, $W_1 < W_2 \ldots < W_{R(L)}$, [where $W_i$ is the Lebesgue measure of a sphere centered at $P_B$ and determined by the distance to the $i$th nearest cluster, $i = 1, 2, \ldots, R(L)$] and $R(L)$ discrete random variables, $K_1, K_2, \ldots, K_{R(L)}$, [where $K_i$ is the number of individuals in the $i$th cluster, $i = 1, 2, \ldots, R(L)$.]

From (1.3) we derive the following three conclusions:

(3.2)

$$
(W_1, W_2, \ldots, W_{R(L)}) \mid R(L) \text{ and } (K_1, K_2, \ldots, K_{R(L)}) \mid R(L) \text{ are independent.}
$$

(3.3)

$W_1, W_2 - W_1, \ldots, W_{R(L)} - W_{R(L)-1}$, are i.i.d. exponential random variables with mean $1/\mu$.

$$
(K_1, K_2, \ldots, K_{R(L)}, R(L)) \text{ and } (Z_1, Z_2, \ldots, Z_{R(L)}, R(L)) \text{ are identically distributed.}
$$

(3.4)

Let

$$
\hat{\mu}_L = R(L)/W_{R(L)}, \quad \hat{m}_L = \frac{R(L)}{\sum_{i=1}^{R(L)} K_i} \quad \text{and} \quad \hat{\theta}_L = \hat{\mu}_L \cdot \hat{m}_L
$$

(3.5)

be the respective estimators of $\mu$, $m$ and $\theta$ ($\theta = EZ_1$).

We claim that if, for some positive number $\delta$, $EZ_1^{2+\delta} < \infty$, then the estimators defined in (3.5) are strongly consistent.

We show first that (3.6), (3.7), and (3.8) hold.

$$
\lim_{L \to \infty} \frac{K_{R(L)}}{L} = 0 \text{ (a.s.)}.
$$

(3.6)
\[
\lim_{L \to \infty} \frac{R(L)}{L} = 1 \quad \text{(a.s.)}.
\]

(3.7)

\[
\lim_{L \to \infty} \frac{R(L)}{L} = 1/m \quad \text{(a.s.)}.
\]

(3.8)

Since \( \sum_{K=1}^{\infty} \frac{1}{L^{1+\delta}} \) is finite, (3.6) is a corollary of (3.9) (with \( \beta = 1 \)).

\[
P(K_{R(L)} \geq L^{\beta} \epsilon) \leq \frac{1}{L} \sum_{k=1}^{L} P(Z_k \geq L^{\beta} \epsilon) \leq \frac{EZ^{2+\delta}/(\epsilon^{2+\delta}L^{2+\delta} \beta - 1)}{L^{2+\delta} \beta - 1}, \quad (\epsilon > 0).
\]

(3.9)

Statement (3.7) is a consequence of (3.6) and (3.10).

\[
R(L) \leq \sum_{i=1}^{L} K_i \leq L + K_{R(L)} - 1.
\]

(3.10)

Let \( \frac{R(L_n)}{L_n} \) converge to \( a \) for some sample point \( \omega \) and for \( \{L_n\} \) a subsequence of the natural numbers. Clearly \( 0 \leq a \leq 1 \), since \( 0 \leq \frac{R(L)}{L} \leq 1 \). For \( n \) sufficiently large, \( \frac{1}{n} \sum_{i=1}^{n} Z_i(w) \leq \frac{1}{n} \sum_{i=1}^{n} Z_i(w) \leq \frac{1}{n} \sum_{i=1}^{n} Z_i(w) \). [Where \( [u] \) is the greatest integer less than or equal to \( u \) and \( \epsilon > 0 \).] By the strong law of large numbers and (3.7) we conclude that \( a = 1/m \). Hence (3.8) holds.

The consistency of \( \hat{n}_L \) follows from (3.7) and (3.8).

We observe that by (3.8), for almost every sample point,

\[
\left[ \frac{L}{m} (1 - \epsilon) \right] \sum_{j=1}^{L} (W_j - W_{j-1}) \leq W_{R(L)} \leq \left[ \frac{L}{m} (1 - \epsilon) \right] \sum_{j=1}^{L} (W_j - W_{j-1}), \quad (\epsilon > 0, W_0 = 0).
\]

Consequently, the consistency of \( \hat{n}_L \) follows from (3.3), (3.8) and the strong law of large numbers.

We summarize the results in Theorem 3.1.
Theorem 3.1. If, for some positive number $\delta$, $E Z_1^{2+\delta} < \infty$, then $\hat{\mu}_L$, $\hat{m}_L$ and $\hat{\theta}_L$ are strongly consistent estimators of $\mu$, $m$ and $\theta$, respectively.

We note that for the sample $(W_1, W_2, \ldots, W_{R(L)})|_{R(L)}$, $\hat{u}_L$ is sufficient and a maximum likelihood estimator of $\mu$.

Our next objective is to establish the asymptotic normality of $\hat{\theta}_L$.

From (3.10) we get that $0 \leq \sqrt{L}(\hat{m}_L - m) - \sqrt{L}(\frac{L}{R(L)} - m) \leq \frac{K}{\sqrt{L}} \frac{L}{R(L)} - \frac{1}{\sqrt{R(L)}} \frac{L}{R(L)}$.

Consequently, by (3.8) and (3.9) (with $\beta = \frac{1}{2}$) we conclude that

$$\sqrt{L}(\hat{m}_L - m) - \sqrt{L}(\frac{L}{R(L)} - m) \xrightarrow{D} 0, \text{ when } L \to \infty. \quad (3.11)$$

Let $u$ be a real number and $A(L) = [L/(\frac{\mu}{\sqrt{L}} + m)]$. Then $P(\sqrt{L}(\frac{L}{R(L)} - m) < u) = P(R(L) > A(L)) = P(\sum_{i=1}^{L} Z_i < L)$. Since $\lim_{L \to \infty} \frac{L - mA(L)}{\sqrt{A(L)}} = \frac{u}{\sqrt{m}}$, Lemma 3.2 follows from the central limit theorem.

Lemma 3.2. If, for some positive number $\delta$, $E Z_1^{2+\delta} < \infty$ and (2.5) is satisfied, then $\sqrt{L}(\hat{m}_L - m)$ and $\sqrt{L}(\frac{L}{R(L)} - m)$ converge in law when $L \to \infty$ to a centered normal random variable with variance $\sigma^2$, $(\sigma^2 = \text{Var}(Z_1))$.

From the Berry and Esseen bound, [Loève (1963), pp. 288.] it follows that for every natural number $k$, that $\sup_{u} |P((-\frac{k}{k} - \mu)^{-1} \mu \sqrt{R} \leq u) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-t^2/2} dt| \leq \frac{C}{\sqrt{k}}$.

[Where $S_k$ is the sum of $k$ i.i.d. exponential random variables with mean $1/\mu$ and $C$ a positive constant.] By (3.2) and (3.3) $P(\alpha \sqrt{R(L)}(\hat{m}_L - \mu) + \beta \sqrt{R(L)}(\hat{m}_L - m) \leq u) = \sum_{j=1}^{\infty} P(R(L) = j) \int_{-\infty}^{\infty} P(\alpha \sqrt{\frac{S_j}{j}} - \mu \leq u - \beta \sigma \sqrt{R(L)}(\hat{m}_L - m) \leq w) \bigg|_{R(L) = j}$. [Where $S_j$ is defined above.] Since $\lim_{L \to \infty} \frac{1}{\sqrt{R(L)}} = 0$, [By the dominated convergence theorem]
Theorem and (3.8).] it follows from the Berry and Esseen bound that

\[ \lim_{L \to \infty} \text{Pr}(\hat{\boldsymbol{\theta}}_L - \mu - 1) + \beta \sqrt{\text{Var}(\hat{\mu}_L)} \leq u) = \lim_{L \to \infty} \text{Pr}(\alpha \hat{Z}_L + \beta \sqrt{\text{Var}(\hat{\mu}_L)} \leq u). \]

[Where \( Z \) is a standard normal random variable independent of \( \hat{\mu}_L \).]

We proved the following:

**Lemma 3.3.** If, for some positive \( \delta \), \( EZ_1^{2+\delta} < \infty \) and (2.5) is satisfied, then

\[ \sqrt{L}(\hat{\mu}_L - \mu) \Rightarrow \text{normal random vector with variances } \frac{m}{\mu^2} \text{ and } \frac{m \sigma^2}{\mu^2}, \]

respectively, and independent components, \( \sigma^2 = \text{Var}(Z_1) \).

In order to establish the asymptotic normality of \( \hat{\theta}_L \), it suffices to consider the asymptotic behavior of \( m \sqrt{L}(\hat{\mu}_L - \mu) + \mu \sqrt{L}(\hat{m}_L - m) \). [see Section 2.]

By applying the transformation \( m \sqrt{x + \mu y} \) and making use of Lemma 3.3, the asymptotic normality of \( m \sqrt{L}(\hat{\mu}_L - \mu) + \mu \sqrt{L}(\hat{m}_L - m) \) follows. We summarize the results in

**Theorem 3.4.** If, for some positive \( \delta \), \( EZ_1^{2+\delta} < \infty \) and (2.5) is satisfied, then

\[ \sqrt{L}(\hat{\mu}_L - \mu), \sqrt{L}(\hat{m}_L - m) \text{ and } \sqrt{L}(\hat{\theta}_L - \theta) \text{ converge in law when } L \to \infty \text{ to centered normal random variables, with respective variances } \frac{m}{\mu^2}, \frac{m \sigma^2}{\mu^2} \text{ and } \frac{m^2 \sigma^2 (m^2 + \sigma^2)}, \]

\[ \sigma^2 = \text{Var}(Z_1). \]
4. **Comparisons between the two schemes.**

In this section we compare the two methods based on the size of their asymptotic variances and the cost to perform them.

Let us assume that scheme A was used and that the total number of sampled individuals was $L_n$.

$$L_n = \sum_{i=1}^{n} Z_i.$$  \hspace{1cm} (4.1)

Since $L_n/\sqrt{n}$ converges with probability one to $m$, $(m = EZ_1)$, when $n \to \infty$, the asymptotic results presented in Section 2 can be restated as follows.

$$\sqrt{L_n} (\hat{\mu}_n - \mu), \sqrt{L_n} (\hat{\sigma}_n - \sigma)$$

and $\sqrt{L_n} (\hat{\theta}_n - \theta)$ converge in law to centered normal random variables with respective variances $\mu^2$, $\sigma^2$ and $\mu^2(\sigma^2 + m^2)$. \hspace{1cm} (4.2)

Hence, the asymptotic variances of (4.2) are equal, respectively, to those presented in Theorem (3.4) for scheme B with $L = L_n$.

The same argument applies when scheme B is used. The results of Theorem 3.4 can be restated in the following form.

$$\sqrt{R(L)} (\hat{\mu}_L - \mu), \sqrt{R(L)} (\hat{\sigma}_L - \sigma)$$

and $\sqrt{R(L)} (\hat{\theta}_L - \theta)$ converge in law to centered normal random variables with respective variances $\mu^2$, $\sigma^2$ and $\mu^2(\sigma^2 + m^2)$. \hspace{1cm} (4.3)

[see (3.8).] The asymptotic variances of (4.3) are respectively equal to those of scheme A with $n = R(L)$.

**Conclusion 4.1.** Based on asymptotic variance comparisons the two methods are equivalent.

Let us assume that $c$ and $d$ are the respective costs of sampling a cluster and an individual and that the average number of clusters in scheme B is equal to $n$.

$$ER(L) = n.$$  \hspace{1cm} (4.4)
The average cost of performing scheme A is

\[(c + dm)n.\]  \hspace{1cm} (4.5)

For scheme B the average cost is

\[cER(L) + dmER(L).\] \hspace{1cm} (4.6)

Hence, under (4.4) the costs of the two schemes are equal.

**Conclusion 4.2.** Based on cost comparison the two schemes are equivalent.

We note that in the Poisson model, where \(Z_1 \equiv 1\), the two schemes coincide.

**Conclusion 4.1.** remains valid if we equate \(L\) of scheme B to the average number of individuals in scheme A (\(nm\)).

The choice between scheme A and scheme B is a matter of practical convenience. If one's major desire is to control the number of clusters to be counted, then scheme A will be preferred. If the major desire is to control the total number of individuals to be counted, then scheme B will be used.
REFERENCES


