Forest Science, Vol 41, No 3, pp 470–476

A Sampling Strategy to Estimate the Area and Perimeter of Irregularly Shaped Planar Regions

TIMOTHY G. GREGOIRE
HARRY T. VALENTINE

ABSTRACT. The length of a randomly oriented ray emanating from an interior point of a planar region can be used to unbiasedly estimate the region's area and perimeter. Estimators and corresponding variance estimators under various selection strategies are presented. FOR. SCI. 41(3):470–476.

ADDITIONAL KEY WORDS. Random radius, unbiased estimation.

CONSIDER THE REGION PICTURED IN FIGURE 1 to be the region of heartrot fungus in a horizontal section of a tree stem; or the cross-section of the tree stem; or the projection onto the ground of a tree's crown; or a contiguous area within a stand inhabited by a rare species of biota. For whatever reason the shaped region may be of interest, we presume a research or survey need to estimate either or both the region's area and perimeter. The methods presented here are simple, yet they provide estimators of area and perimeter that are either unbiased or biased but consistent. Moreover, their mean square error decreases with increasing sampling intensity, so that the accuracy of estimation can be linked to its expense or worth. The following situations are examples of where these procedures might be used advantageously.

Horizontal importance sampling of stem volume as expounded by Valentine et al. (1992) is unbiased if the stem cross-sectional area is measured or unbiasedly estimated at each point of measurement on the stem. For irregularly shaped boles, which would be difficult to measure, the method expounded here can be used to provide an unbiased estimate. As a further example, the projected area of a tree crown is often computed as a circular area using an average crown "width" measurement. The method of this paper dispenses with the need for an assumption of circularity, which likely never is upheld. As a final motivating example, the length of edge in patches of vegetation is of interest in ecological surveys (Schreuder and Gregoire 1995), yet its exact field measurement may be prohibitive. Sampling can provide an affordable alternative in some cases.

AREA ESTIMATION

Let $\mathscr{X}$ denote a specific, fixed point within the region $\mathbb{R}$ of Figure 1. Let $r(\tau)$ denote the ray emanating from $\mathscr{X}$ at an angle τ with respect to a fixed baseline.

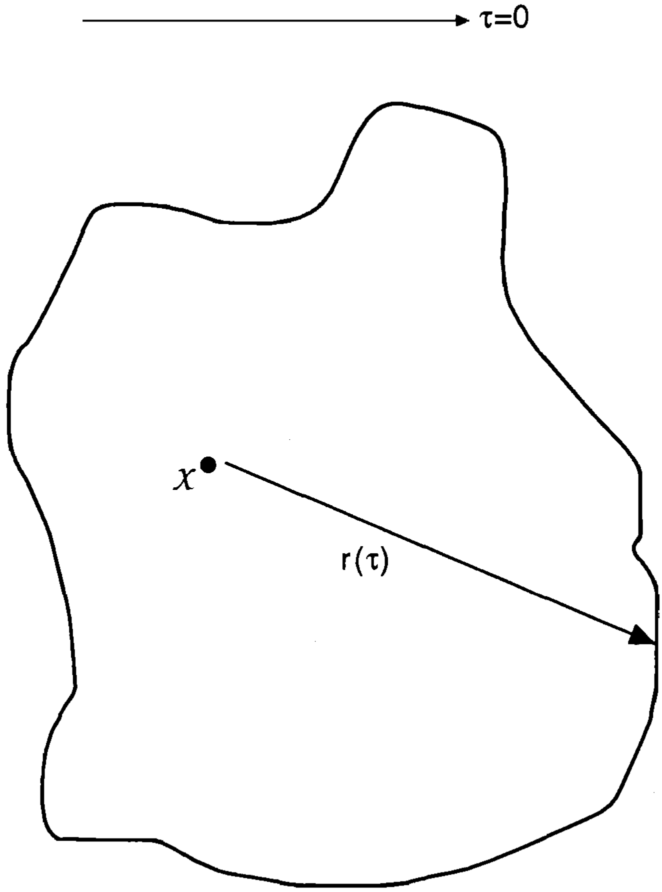

FIGURE 1. Planar region ℝ for which a ray emanating from a fixed interior point 𝔛 can intersect the boundary only once.

Let $s(\tau)$ be the length of $r(\tau)$ within ℝ. At present we consider only those regions for which $r(\tau)$ can intersect the boundary of ℝ at most once.

The area (A) of ℝ is deducible from the calculus as

$$A = 0.5 \int_0^{2\pi} s^2(\tau) d\tau$$

Let the random angle θ be selected uniformly, $\theta \sim U[0,2\pi]$, for example by drawing a $U[0,1]$ random number and multiplying by 2π. An unbiased estimator of A is

$$\hat{A} = \frac{1}{m} \sum_{i=1}^{m} a_i$$

where $a_i = \pi s^2(\theta_i)$ and where θ_i, $i = 1, \ldots, m$, are independent realizations of θ. For $m > 1$, the variance of $\hat{A}$ is unbiasedly estimated by

$$\widehat{\text{var}}(\hat{A}) = \frac{1}{m(m-1)} \sum_{i=1}^{m} (a_i - \hat{A})^2$$

If the angles θ_i are not selected independently, $\hat{A}$ remains unbiased as long as θ_i has a uniform marginal distribution, but the variance estimator does not. (Proofs of unbiasedness are available in a separate addendum from the authors.)

This estimator of A was derived by Matérn (1956) and mentioned by Husch et al. (1982) without elaboration, but it seems to be unknown or unused by most foresters and ecologists. Valentine et al. (1986) proposed its use in connection with two-stage sampling to estimate bole increment. The variance estimator of $\hat{A}$ is straightforward to derive but has not been presented previously.

The location of $\mathscr{X}$ is arbitrary; however, the variance of $\hat{A}$ is directly related to the variation among the distances $s(\tau)$ Therefore, the more central the location of $\mathscr{X}$, the more precisely A will be estimated. For some shapes, the central location may not even be within the interior of $\mathbb{R}$—consider, for example, a region which is shaped like a torus.

If $\mathscr{X}$ is located on the boundary of $\mathbb{R}$ and θ is selected from $U[0,2\pi]$, the resulting ray may not pass through $\mathbb{R}$, in which case $s(\theta) = 0$. If the location of $\mathscr{X}$ on the boundary is unavoidable, then it would be prudent to measure the angular range, λ, which encompasses $\mathbb{R}$, and then to select θ as a $U[0,\lambda]$ random variable. All selected rays then will intersect $\mathbb{R}$. In this situation it would be necessary to substitute $\lambda/2$ for π in a_i. For θ near 0 or λ, $s(\theta)$ may be subject to greater measurement error than would be the case for rays that cut more clearly across $\mathbb{R}$.

If the shape of $\mathbb{R}$ is oblong then rays that are perpendicular to each other will tend to have an average length that is closer to the mean length than either one individually. In this case it may be worthwhile to generate θ in antithetic pairs. For example, θ_i and θ_i^*, where θ_i is as defined above and where $\theta_i^* = \theta_i + \pi/2$, would be an antithetic pair of measurement angles. For m such pairs,

$$\hat{A}' = \frac{1}{m} \sum_{i=1}^{m} \bar{a}_i, \quad \text{and} \quad \widehat{\text{var}}(\hat{A}') = \frac{1}{m(m-1)} \sum_{i=1}^{m} (\bar{a}_i - \hat{A}')^2$$

where $\bar{a}_i = \pi(s^2(\theta_i) + s^2(\theta_i^*))/2$.

If the shape of $\mathbb{R}$ is irregular or if $\mathscr{X}$ is located far from the region's center, a systematic approach may be yield more precise estimates. To sample systematically and still be able to estimate variance, we propose to generate m sets of angles. The first angle, $\theta_{i,1}$, of the ith set is generated as a $U[0,2\pi/n]$ random variable. The other angles of the set are defined thus: $\theta_{i,j} = \theta_{i,j-1} + 2\pi/n$, for $j = 2, \ldots, n$. Given m sets of systematically selected samples,

$$\hat{A}'' = \frac{1}{m} \sum_{i=1}^{m} \check{a}_i \quad \text{and} \quad \widehat{\text{var}}(\hat{A}'') = \frac{1}{m(m-1)} \sum_{i=1}^{m} (\check{a}_i - \hat{A}'')^2$$

where $\check{a}_i = (\pi/n) \sum_{j=1}^{n} s^2(\theta_{i,j})$. As a further variation on this theme, the number of angles in each systematic set can vary among sets.

The indented region of Figure 2 may be problematic, depending on the

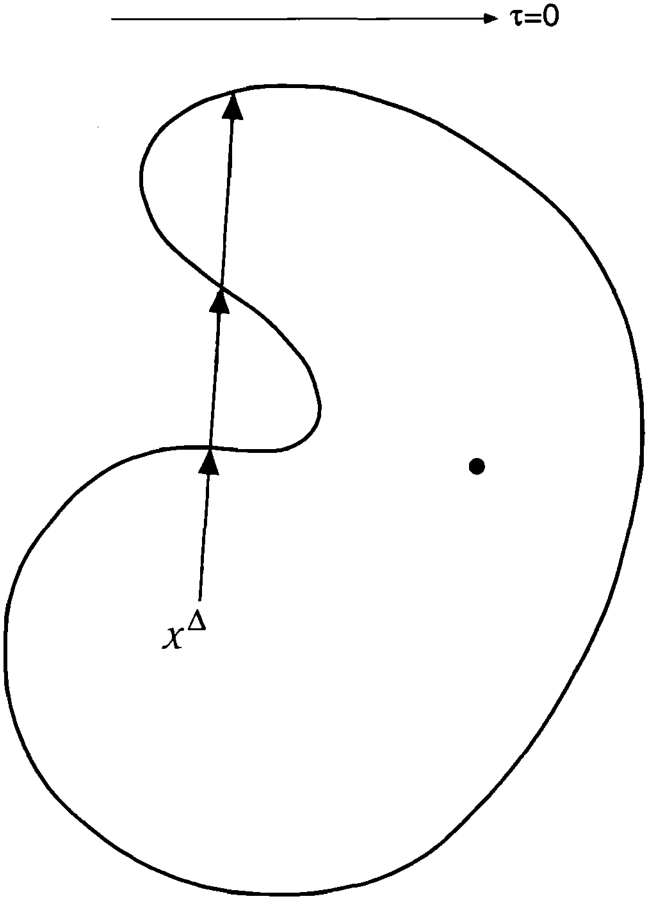

FIGURE 2. Planar region $\mathbb{R}$ for which a ray emanating from a fixed interior point $\mathscr{X}$ can intersect the boundary more than once.

location of $\mathscr{X}$. If it is located at the solid bullet, then the sampling method and estimators presented above may be used. If it is located at the triangle, however, modifications to both are needed whenever $r(\tau)$ intersects the boundary of $\mathbb{R}$ multiple times. For the ray shown in Figure 2, let $s_1(\tau)$, $s_2(\tau)$, and $s_3(\tau)$, respectively, be the distances from $\mathscr{X}$ to the first, second, and third points of intersection of $r(\tau)$ with the boundary. Therefore, $s_1(\tau)$ is the length of $r(\tau)$ in the lower lobe of $\mathbb{R}$, and $s_3(\tau) - s_2(\tau)$ is its length in the upper lobe. An unbiased estimator of A in this case is

$$\tilde{A} = \frac{1}{m} \sum_{i=1}^{m} \tilde{a}_i$$

where

$$\tilde{a}_i = \pi(s_1^2(\theta_i) - s_2^2(\theta_i) + s_3^2(\theta_i))$$

We speculate that $\tilde{A}$ remains unbiased for even more convoluted regions which give rise to more than three intersections of $r(\tau)$ with the boundary. However, our advice when dealing with such regions is to locate $\mathcal{B}$ judiciously to avoid the possibility that $r(\tau)$ has multiple intersections with the boundary. Alternatively, one can partition $\mathbb{R}$ and then estimate the area of each part separately using $\hat{A}$.

PERIMETER ESTIMATION

The perimeter (P) of $\mathbb{R}$ is deducible from the calculus as

$$P = \int_0^{2\pi} \sqrt{s^2(\tau) + (ds(\tau)/d\tau)^2}\, d\tau.$$

An unbiased estimator of P is

$$\hat{P} = \frac{1}{m}\sum_{i=1}^{m} p_i$$

where $p_i = 2\pi \sqrt{s^2(\theta_i) + (ds(\theta_i)/d\theta_i)^2}$ and where θ_i, $i = 1, \ldots, m$, again are independent realizations of $\theta \sim U[0,2\pi]$. For $m > 1$, the variance of $\hat{P}$ is unbiasedly estimated by

$$\widehat{\mathrm{var}}(\hat{P}) = \frac{1}{m(m-1)}\sum_{i=1}^{m}(p_i - \hat{P})^2.$$

This estimator of P is not likely to be feasible in many situations because $ds(\theta_i)/d\theta_i$ cannot be measured, in general. If $s(\tau)$ is measured at $\tau = \theta_i$ and $\tau = \theta_i + \Delta\theta$ for some small, measurable increment $\Delta\theta$, then the tangent $ds(\theta_i)/d\theta_i$ can be approximated by the slope $\Delta s(\theta_i)/\Delta\theta$, where $\Delta s(\theta_i) = s(\theta_i) - s(\theta_i + \Delta\theta)$. The resulting estimator

$$\mathring{P} = \frac{1}{m}\sum_{i=1}^{m} \mathring{p}_i$$

where

$$\mathring{p}_i = 2\pi\sqrt{s^2(\theta_i) + (\Delta s(\theta_i)/\Delta\theta)^2}$$

is an approximately unbiased and consistent estimator of P and its variance is unbiasedly estimated by

$$\widehat{\mathrm{var}}(\mathring{P}) = \frac{1}{m(m-1)}\sum_{i=1}^{m}(\mathring{p}_i - \mathring{P})^2$$

whenever $m > 1$. This estimator and the following one require the measurement of not only $s(\theta_i)$ but also $s(\theta_i + \Delta\theta)$ and $\Delta\theta$. To avoid bias due to subjective selection, it is preferable to select $\Delta\theta$ *a priori*.

If angles are generated in antithetic pairs,

$$\overset{\circ}{P}' = \frac{1}{m} \sum_{i=1}^{m} \bar{p}_i$$

where

$$\bar{p}_i = \pi(\sqrt{s^2(\theta_i) + (\Delta s(\theta_i)/\Delta\theta)^2} + \sqrt{s^2(\theta_i^*) + (\Delta s(\theta_i^*)/\Delta\theta)^2})$$

is approximately unbiased and its variance is unbiasedly estimated by

$$\widehat{\mathrm{var}}(\overset{\circ}{P}') = \frac{1}{m(m-1)} \sum_{i=1}^{m} (\bar{p}_i - \overset{\circ}{P}')^2.$$

For regions where antithetic selections may improve the estimation of A, estimation of P may not be improved to the same degree, depending on the influence of the terms involving $\Delta s(\theta_i)$ and $\Delta s(\theta_i^*)$.

For the systematic scheme described above in regard to estimation of A, the estimators $\overset{\circ}{P}''$ and vâr $(\overset{\circ}{P}'')$ follow analogously.

An alternative sampling strategy and estimator of P involves measurement of arc length, say $t(\theta_i)$, along the perimeter of $\mathbb{R}$ from θ_i to $\theta_i + \Delta\theta$. Providing that 2π is an integer multiple of $\Delta\theta$ (say, $M = 2\pi/\Delta\theta$), the estimator

$$\overset{\triangle}{P} = \frac{M}{m} \sum_{i=1}^{m} t(\theta_i)$$

is an unbiased estimator of P. This estimator may be particularly suitable for convoluted regions, such as that shown in Figure 2, when a $r(\tau)$ can intersect the boundary of $\mathbb{R}$ multiple times. In this situation, it is not apparent to us how to construct an unbiased estimator of P analogous to $\bar{A}$. However, $\overset{\triangle}{P}$ is a feasible alternative. Whenever $m > 1$, the variance of $\overset{\triangle}{P}$ is unbiasedly estimated by

$$\widehat{\mathrm{var}}(\overset{\triangle}{P}) = \frac{1}{m(m-1)} \sum_{i=1}^{m} (t(\theta_i) - \overset{\triangle}{P})^2 .$$

DISCUSSION

As an alternative to sampling, numerical methods may be employed. Schreuder and Gregoire (1995) discuss the estimation of the length of boundary between two regions of a fixed-area circular plot and estimation of the area of each region. One of the proposed methods involves the partitioning of a subregion into successive triangles; area and length are estimated by numerical integration. Numerical integration could also be used after partitioning $\mathbb{R}$ into n segments emanating from $\mathscr{X}$, each of angular width $2\pi/n$. Compared to the estimators of area and perimeter from the systematic sampling approach outlined earlier, the estimators based on numerical integration are biased, in general, but have no sampling variance. The mean squared error of estimation diminishes as n increases with both approaches. With the sampling approach, the sampling error can be estimated.

If $\mathbb{R}$ is systematically sampled, then it is possible to approximate $ds(\tau)/d\tau$ in $\hat{P}$ by fitting a cubic spline to the measurements $s(\theta_{i,j})$, differentiating the fitted spline, and evaluating it at each point $\tau = \theta_{i,j}$. The measurements, $s(\theta_{i,j})$, would need to be sorted by θ. In order to provide derivatives at the end points, we suggest augmenting the sorted data vector by prefixing those measurements of $s(\tau)$ for which $-\pi/2 \leq \tau \leq 0$ and suffixing those measurements for which $0 \leq \tau \leq \pi/2$. The statistical properties of the resulting estimator of P are inscrutable, however.

LITERATURE CITED

HUSCH, B., C.I. MILLER, and T.W. BEERS. 1982. Forest mensuration. Ed. 3. Wiley, New York.

MATÉRN, B. 1956. On the geometry of the cross-section of a stem. Meddelanden Från Statens Skogsforskningsinstitut., Stockholm, 46.

SCHREUDER, H.T., and T.G. GREGOIRE. 1995. Estimating condition areas and boundary lengths in mapped designs. Can. J. For. Res. (In press).

VALENTINE, H.T., T.G. GREGOIRE, G.M. FURNIVAL, and D.S. SOLOMON. 1986. Unbiased estimation of bole increment. P. 67–77 *in* Environmental influences on tree and stand increment, Solomon, D.S., and T.B. Brann, (eds.). Maine Agric. Exp. Stn. Misc. Publ. 601.

VALENTINE, H.T., C. BEALLE, and T.G. GREGOIRE. 1992. Comparing vertical and horizontal modes of importance and control-variate sampling for bole volume. For. Sci. 38:160–172.

Manuscript received February 8, 1994

AUTHORS AND ACKNOWLEDGMENTS

Timothy G. Gregoire is Associate Professor of Forest Biometrics, Department of Forestry, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0324; and Harry T. Valentine is Research Forester, USDA Forest Service, Louis C. Wyman Forest Sciences Laboratory, P.O. Box 640, Durham, NH 03824-0640.