

CONFIDENCE INTERVALS FOR STEREOLOGICAL ESTIMATORS WITH INFINITE VARIANCE*

JOEL N. FRANKLIN†

Abstract. A statistical estimator is discussed for using two-dimensional electron-microscope data to estimate N_V , the number of organelles per unit volume. Under general assumptions, the estimator is shown to be the unique unbiased estimator of N_V . Though the estimator has infinite variance, large samples are shown to yield an approximately normally distributed statistic from which confidence intervals for N_V can be obtained.

1. Introduction. In cell biology, an important problem of morphometric cytology is to estimate the number of cell nuclei, or other organelles, per unit volume. This number is usually called N_V , but we shall call it simply n . The number n is to be estimated from observations with an electron microscope. The microscope produces pictures of random plane sections; the biologist observes disks where the random plane has cut some of the nuclei. From repeated independent two-dimensional observations the biologist compiles statistics from which he wishes to estimate the number n .

Recent papers [1], [2], [3] have discussed and applied estimators with infinite variance. These papers show how to construct an unbiased estimator for n , but they do not show how to construct confidence intervals. Since the estimators have infinite variance, the need for confidence intervals is critical.

We will show that if no assumption is made on the sizes of the nuclei, then the unbiased estimator is unique, with infinite variance. Nevertheless, we will show that large independent samples may be used as if they had been taken from a Gaussian population with unknown but finite variance. Then confidence intervals can be obtained from standard statistical tables.

2. Definition of the estimator. Assume that the unit cube contains n balls. Denote the cube by

$$(2.1) \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1, \quad 0 \leq z \leq 1.$$

The problem is to estimate n by observing a random plane slice $x = X$, where X is a uniformly distributed random variable, $0 < X < 1$.

Let the balls have radii $r_i > 0$ and centers \mathbf{c}_i ($i = 1, \dots, n$). The slice $x = X$ intersects the i th ball in a disk with positive area A_i if

$$(2.2) \quad -r_i < X - \xi_i < r_i$$

where $\mathbf{c}_i = (\xi_i, \eta_i, \zeta_i)$. In this event

$$(2.3) \quad A_i = \pi[r_i^2 - (X - \xi_i)^2] > 0.$$

We define the estimator $u(X)$ as follows. If the slice intersects at least one ball in a disk with area $A_i > 0$, we define u as the sum

$$(2.4) \quad u(X) = \sum_{A_i > 0} (\pi A_i)^{-1/2}.$$

* Received by the editors March 18, 1980, and in final form May 5, 1980. This research was partially supported by the National Science Foundation under grant MCS78-02198. Reproduction in whole or in part is permitted for any purpose of the United States Government.

† Department of Applied Mathematics, California Institute of Technology, Pasadena, California 91125.

If the slice intersects none of the balls, we define $u(X) = 0$. This estimator was introduced by R. L. Fullman [4].

In [1], it is proved that $u(X)$ is an unbiased estimator of the number of balls:

$$(2.5) \quad E u(X) = n.$$

Now the strong law of large numbers implies that, for independent samples of large size m , with probability one,

$$(2.6) \quad \lim_{m \rightarrow \infty} \frac{1}{m} [u(X_1) + \dots + u(X_m)] = n.$$

The estimator $u(X)$ is applied to cell biology in [3]. Unfortunately, as it is observed in [1], the estimator has infinite variance. In [1], [2], [3] no attempt was made to obtain confidence intervals. If the variance of $u(X)$ were finite, one could apply the central limit theorem in the usual way to obtain confidence intervals. But the following argument implies that if no assumption is made about the sizes of the n balls, then an unbiased estimator of n must have infinite variance.

3. Uniqueness of the estimator. Suppose $v(X)$ is any unbiased estimator of n . Then, in particular, it must work for $n = 0$ and for $n = 1$.

Let us assume that the estimator is a function of the positive areas A_i cut out of the n balls by the random plane slice $x = X$. If $n = 0$, then of course no areas will be cut out, and so we require $v(X) = 0$. In general, n is unknown, but if a random slice cuts out no areas $A_i > 0$, we should set the estimator $v(X) = 0$ to cover the contingency $n = 0$.

Suppose $n = 1$. Let the unit cube contain one ball, with radius r and center (ξ, η, ζ) . The random slice $x = X$ misses the ball with probability $1 - 2r$, and in this event we have decided to set $v(X) = 0$. But if X lies in the interval $\xi - r < X < \xi + r$, then the slice cuts out the positive area A given by (2.3). If the estimator $v(X)$ is a function of A , set $v(X) = h(A)$; then the estimator has the expected value

$$(3.1) \quad E v(X) = \int_{\xi-r}^{\xi+r} h(A(x)) dx.$$

We require $E v(X) = 1$ if $n = 1$. Then the last equation becomes

$$(3.2) \quad 1 = \int_{\xi-r}^{\xi+r} h(\pi[r^2 - (x - \xi)^2]) dx.$$

This is an integral equation for an unknown function $h(A)$. We will show that it is an Abel equation, and we will solve it. (Similar uses of Abel equations in stereology appear in [7] and [9].)

If we introduce A as a variable of integration in (3.2), we see that A traverses the interval $0 < A < \pi r^2$ twice. Then (3.2) becomes

$$(3.3) \quad 1 = \pi^{-1/2} \int_0^a h(A)(a - A)^{-1/2} dA,$$

where we define $a = \pi r^2$. The integral is Abel's fractional integral of order $\frac{1}{2}$. We deduce

$$(3.4) \quad \pi^{-1/2} \int_0^a 1 \cdot (a - A)^{-1/2} dA = \int_0^a h(A) dA,$$

which implies, at points of continuity,

$$(3.5) \quad h(A) = (\pi A)^{-1/2}.$$

This is the estimator $u(X)$ defined by (2.4) if the random slice cuts out one positive area $A > 0$. Thus, the choice of the unbiased estimator is forced: $v(X) = u(X)$. For $n > 1$, the estimator (2.4) is obtained by establishing the additivity of an unbiased estimator as follows:

Let the symbol B_i represent the i th ball ($i = 1, \dots, n$). An estimator of n must depend on the random slice $x = X$ and on the n balls B_1, \dots, B_n , and so we may denote the estimator by $v(X; B_1, \dots, B_n)$ if $n \geq 1$ or by $v(X; \phi)$ if $n = 0$. Let $u(X; \dots)$ be the estimator defined by (2.4); we have just shown that an unbiased estimator $v(X; \dots)$ must equal $u(X; \dots)$ if $n = 0$ or 1. If $n > 1$, we write the two identities

$$E v(X; B_1, \dots, B_n) = n, \quad E v(X; B_2, \dots, B_n) = n - 1.$$

Subtraction gives

$$(3.6) \quad E [v(X; B_1, \dots, B_n) - v(X; B_2, \dots, B_n)] = 1.$$

If we regard the balls B_2, \dots, B_n as fixed, the quantity $[\dots]$ depends only on X and on the variable first ball, B_1 ; it suffices to fix the center of B_1 and to let the radius r vary in a small range $0 < r_1 < \epsilon$. Then the uniqueness of the estimator for $n = 1$ implies

$$(3.7) \quad [v(X; B_1, \dots, B_n) - v(X; B_2, \dots, B_n)] = u(X; B_1).$$

Similarly, if $n > 2$ we deduce

$$[v(X; B_2, \dots, B_n) - v(X; B_3, \dots, B_n)] = u(X; B_2).$$

and so on. Summarizing, we find, for all $n \geq 1$,

$$(3.8) \quad v(X; B_1, \dots, B_n) = u(X; B_1) + \dots + u(X; B_n),$$

which is our additive estimator (2.4).

The variance is infinite because all terms in (3.8) are nonnegative, and so

$$E v^2(X; B_1, \dots, B_n) \geq E u^2(X; B_1),$$

and now (3.5) implies

$$(3.9) \quad \begin{aligned} E u^2(X; B_1) &= \int_{\xi-r}^{\xi+r} h^2(A(x)) dx \\ &= \pi^{-1/2} \int_0^a (\pi A)^{-1} (a-A)^{-1/2} dA = +\infty. \end{aligned}$$

Nevertheless, we will derive for the estimator (2.4) a usable central limit theorem.

4. Asymptotic distribution. The estimator $u(X)$ defined by (2.4) is a random variable; it has a probability distribution

$$(4.1) \quad F(u) = \Pr \{u(X) \leq u\}.$$

To obtain a central limit theorem for the singular estimator $u(X)$, we will first show this: For large u there is a probability density $f(u) = F'(u)$ which has the form

$$(4.2) \quad f(u) = 2bu^{-3} + O(u^{-4}),$$

where b is a positive constant if the cube contains at least one ball ($n \geq 1$).

First, suppose $n = 1$. Let the ball have the center (ξ, η, ζ) and the radius r . If $|X - \xi| < r$, the random plane $x = X$ cuts out the positive area

$$(4.3) \quad A(X) = \pi[r^2 - (X - \xi)^2] > 0,$$

and then

$$(4.4) \quad (\pi r)^{-1} \leq u(X) = \pi^{-1}[r^2 - (X - \xi)^2]^{-1/2} < \infty.$$

Therefore, $u(X)$ has the probability distribution function

$$(4.5) \quad F(u) = \begin{cases} 0 & \text{if } u < 0, \\ 1 - 2r & \text{if } 0 \leq u < (\pi r)^{-1}, \\ 1 - 2r + 2[r^2 - (\pi u)^{-2}]^{1/2} & \text{if } (\pi r)^{-1} \leq u < \infty. \end{cases}$$

For large u the derivative has the asserted form (4.2) with $b = \pi^{-2}r^{-1}$.

If $n = 1$ the estimator $u(X)$ is large only if X lies in two narrow bands:

$$(4.6) \quad \xi - r < X < \xi - r + \varepsilon \quad \text{and} \quad \xi + r - \varepsilon < X < \xi + r.$$

As (4.4) shows, $u(X) \rightarrow \infty$ as $X \rightarrow \xi - r + 0$ or as $X \rightarrow \xi + r - 0$. For this reason we will call $\xi \pm r$ *singularities* of the uniformly distributed sampling variable X . The ball lies in the slab between the two singularities.

For $n > 1$ the i th ball lies in the slab $\xi_i - r_i \leq x \leq \xi_i + r_i$, and so the i th ball causes the estimator $u(X)$ to be large in two narrow bands:

$$(4.7) \quad \xi_i - r_i < X < \xi_i - r_i + \varepsilon \quad \text{and} \quad \xi_i + r_i - \varepsilon < X < \xi_i + r_i.$$

Call $\xi_i - r_i$ a *left* singularity; call $\xi_i + r_i$ a *right* singularity. Two of the balls share a singularity if they have a common tangent plane $x = \text{constant}$. Let the union of the n left singularities be the distinct values $a_1 < \dots < a_p$; let the union of the n right singularities be the values $b_1 < \dots < b_q$. Then the estimator $u(X)$ is large only for X in $p + q$ narrow bands:

$$(4.8) \quad a_i < X < a_i + \varepsilon \quad \text{and} \quad b_j - \varepsilon < X < b_j.$$

These bands are disjoint if $0 < \varepsilon \ll 1$.

Let us observe $u(X)$ in the left band $a_i < X < a_i + \varepsilon$. Here we see

$$(4.9) \quad u(X) = \sum \pi^{-1}[r_k^2 - (X - \xi_k)^2]^{-1/2} + w(X),$$

where the sum \sum is taken over those balls for which a_i is a left singularity, so that $a_i = \xi_k - r_k$. The remainder $w(X)$ comes from those balls cut by $X = a_i$, so that $\xi_k - r_k < a_i < \xi_k + r_k$; the function $w(X)$ is analytic in the neighborhood of $X = a_i$. The sum \sum has the form

$$(4.10) \quad \begin{aligned} \sum &= \sum \pi^{-1}[r_k^2 - (X - a_i - r_k)^2]^{-1/2} \\ &= \sum \pi^{-1}[2r_k(X - a_i) - (X - a_i)^2]^{-1/2} \\ &= (X - a_i)^{-1/2}v(X), \end{aligned}$$

where $v(X)$ is analytic near a_i , with $v(a_i) > 0$.

Now $u(X)$ takes the form

$$(4.11) \quad u(X) = (X - a_i)^{-1/2}v(X) + w(X) \quad \text{for } (a_i < X < a_i + \varepsilon).$$

As $X \rightarrow a_i + 0$, $u(X)$ goes to infinity like a positive constant times $(X - a_i)^{-1/2}$. It is now clear that the event $u(X) > t$ has probability of the order t^{-2} as $t \rightarrow \infty$:

$$(4.12) \quad \Pr \{u(x) > t, a_i < X < a_i + \varepsilon\} = c_2 t^{-2} + c_3 t^{-3} + \dots,$$

with $c_2 > 0$.

This is the probability that $u(X) > t$ and that X lies in one of the disjoint bands (4.8). Since these events are disjoint, the probability of their union is the sum of their

probabilities, and so we find an expression

$$(4.13) \quad \Pr \{u(X) > t\} = b_2 t^{-2} + b_3 t^{-3} + \dots$$

as $t \rightarrow \infty$, with $b_2 > 0$. Taking minus the derivative, we obtain the asymptotic form of the probability density:

$$(4.14) \quad f(t) = 2b_2 t^{-3} + 3b_3 t^{-4} + \dots,$$

where the series converges for large t . This implies the asserted form (4.2) if we set $b = b_2$.

For applications in cell biology, it is important that we have made no assumption about the distribution of the n balls within the cube. The n balls are considered fixed and nonrandom, though their sizes, positions, and number are unknown.

5. A central limit theorem. Let $u(X)$ be the stereological estimator (2.4). Let the unit cube be cut by m parallel random planes $x = X_1, \dots, X_m$, where the X_i are uniformly distributed and statistically independent. Form the empirical mean

$$(5.1) \quad \bar{U} = \frac{1}{m}(U_1 + \dots + U_m).$$

Then \bar{U} has infinite variance, but has expected value $\mu = n =$ the number of balls in the cube.

Form the random variable

$$(5.2) \quad Y = \left(\frac{m}{\log m}\right)^{1/2} (\bar{U} - \mu).$$

We will now prove that as $m \rightarrow \infty$, Y becomes Gaussian, with mean zero and with the finite variance b , where $2b$ is the leading coefficient in the asymptotic probability density (4.2).

This differs from the usual central limit theorem because of the logarithmic term in Y . The logarithmic term counteracts the infinite variance of the finite sample means \bar{U} so that the limiting distribution has finite variance.

The proof uses Lévy's theorem for characteristic functions. Define the expected value

$$(5.3) \quad \phi(t) = E \exp(itu(X)).$$

This is the characteristic function of our estimator $u(X)$. We will now show that for small $t > 0$,

$$(5.4) \quad \phi(t) = 1 + i\mu t + bt^2 \log t + O(t^2),$$

where $\mu = E u(X) = n$, and where b appears in (4.2).

If $u(X)$ has the distribution function $F(u)$, then

$$\phi(t) = \int_0^\infty e^{itu} dF(u).$$

Let M be a large constant such that the density $f(u)$ obeys (4.2) for $u \geq M$. Then,

$$\begin{aligned} \phi(t) &= \int_0^\infty (1 + itu) dF(u) + \int_0^\infty (e^{itu} - 1 - itu) dF(u) \\ &= 1 + i\mu t + \int_0^\infty (e^{itu} - 1 - itu) dF(u), \end{aligned}$$

where $\mu = E u(X)$.

For the last integral, we have

$$(5.5) \quad \int_0^\infty (e^{itu} - 1 - itu) dF(u) = \int_0^M + \int_M^\infty.$$

The first part is $O(t^2)$ as $t \rightarrow +0$. The second part is

$$(5.6) \quad \int_M^\infty (e^{itu} - 1 - itu)(2bu^{-3} + O(u^{-4})) du.$$

Here

$$(5.7) \quad \int_M^\infty (e^{itu} - 1 - itu) O(u^{-4}) du = \int_M^\infty O(t^2 u^2) O(u^{-4}) du = O(t^2).$$

The last three formulas imply

$$(5.8) \quad \int_0^\infty (e^{itu} - 1 - itu) dF(u) = O(t^2) + 2b \int_M^\infty (e^{itu} - 1 - itu) u^{-3} du.$$

From [6, p. 228], we obtain the formula

$$(5.9) \quad \int_1^\infty e^{iut} u^{-3} du = \frac{1}{2} + it + \frac{t^2}{2} \log t + O(t^2)$$

as $t \rightarrow +0$. From the same source, we obtain the formula

$$(5.10) \quad \int_1^\infty e^{iu^2 t} u^{-3} du = \frac{1}{2} - \frac{it}{2} \log t + O(t),$$

which we shall require later.

Since M is constant, we have

$$(5.11) \quad \begin{aligned} \int_M^\infty (e^{itu} - 1 - itu) u^{-3} du &= O(t^2) + \int_1^\infty (e^{itu} - 1 - itu) u^{-3} du \\ &= O(t^2) + \int_1^\infty e^{itu} u^{-3} du - \frac{1}{2} - it \\ &= O(t^2) + \frac{t^2}{2} \log t, \end{aligned}$$

by (5.9). Now (5.8) implies the asserted formula (5.4).

From (5.4) we find, for $t \rightarrow +0$,

$$(5.12) \quad E \exp it[u(X) - \mu] = e^{-itu} \phi(t) = 1 + bt^2 \log t + O(t^2) \equiv \psi(t).$$

Then $\bar{U} - \mu$ has the c.f.

$$(5.13) \quad E \exp it(\bar{U} - \mu) = \left[\psi\left(\frac{t}{m}\right) \right]^m.$$

Now the definition (5.2) implies

$$(5.14) \quad E e^{itY} = E \exp \left[it \left(\frac{m}{\log m} \right)^{1/2} (\bar{U} - \mu) \right] = \left[\psi \left(\frac{t}{m \log m} \right)^{1/2} \right]^m.$$

If we keep $t \geq 0$ fixed and let m become large, we find, from (5.12) and (5.14),

$$\begin{aligned}
 (5.15) \quad E e^{itY} &= \left[1 + b \left(\frac{t^2}{m \log m} \right) (\log t - \frac{1}{2} \log m - \frac{1}{2} \log \log m) + O \left(\frac{1}{m \log m} \right) \right]^m \\
 &= \left[1 - \frac{1}{2} \frac{bt^2}{m} + o \left(\frac{1}{m} \right) \right]^m.
 \end{aligned}$$

If we take the limit as the sample size $m \rightarrow \infty$, we get

$$(5.16) \quad \lim_{m \rightarrow \infty} e^{itY} = \exp \left(-\frac{1}{2} bt^2 \right).$$

This is the characteristic function of a Gaussian random variable with mean zero and variance b , and now Lévy's theorem implies

$$(5.17) \quad \lim_{m \rightarrow \infty} \Pr \{ Y \leq y \} = (2\pi b)^{-1/2} \int_{-\infty}^y \exp \left(-\frac{1}{2} \frac{v^2}{b} \right) dv.$$

In order to use this result to compute confidence intervals, we shall need to know the limit variance b . Section 6 will show how to estimate b from large samples U_1, \dots, U_m .

G. S. Watson [8] proved a limit theorem for the sums $\sum A_i^{-1/2}$ (where our $A_i > 0$ equals his $\pi y_i^2/4$). He assumed a large number of *independent, identically distributed random variables* $A_i^{-1/2}$. We have not made this assumption.

To take an extreme case, we may let the unit cube (2.1) contain just the two balls with these radii and centers:

$$r_1 = .2, \quad \mathbf{c}_1 = (.5, .5, .3); \quad r_2 = .1, \quad \mathbf{c}_2 = (.5, .5, .7).$$

As the random plane $x = X$ sweeps through the cube, we shall observe

$$\begin{aligned}
 A_1 > 0 &\quad \text{if } .3 < X < .7, \\
 A_2 > 0 &\quad \text{if } .4 < X < .6.
 \end{aligned}$$

Here the random variables A_1 and A_2 are distributed neither identically nor independently.

For us the independent random variables are not the areas $A_i > 0$ but the samples X_1, \dots, X_m of the uniformly distributed random variable X . Then the formula (2.4) produces independent samples U_1, \dots, U_m of the random variable $u(X)$.

6. Estimation of the limit variance. The constant b appears in the asymptotic form (4.2) of the probability density $f(u)$. We will now show that for large samples the following limit occurs with probability one:

$$(6.1) \quad \lim_{m \rightarrow \infty} \frac{U_1^2 + \dots + U_m^2}{m \log m} = b.$$

This formula can be used to estimate b . But we shall prefer the following formula:

$$(6.2) \quad \lim_{m \rightarrow \infty} (\log m)^{-1} (m-1)^{-1} \sum_{j=1}^m (U_j - \bar{U})^2 = b.$$

The two limit formulas are equivalent because $\bar{U} \rightarrow \mu$ with probability one. We shall later observe a practical reason for preferring the second formula in an approximation with finite m .

To prove the first formula we will use the characteristic function of the random variable U^2 :

$$(6.3) \quad \phi_2(t) = E \exp(itu^2(X)),$$

where $u(X) = U$ is the stereological estimator (2.4).

In terms of the distribution $F(u)$, we have

$$(6.4) \quad \phi_2(t) = \int_0^\infty \exp(itu^2) dF(u) = 1 + \int_0^\infty [\exp(itu^2) - 1] dF(u).$$

Let M be a positive constant for which, as in (4.2),

$$(6.5) \quad F'(u) = f(u) = 2bu^{-3} + O(u^{-4}) \quad \text{for } u \geq M.$$

On the finite range $0 \leq u \leq M$, we have

$$(6.6) \quad \int_0^M [\exp(itu^2) - 1] dF(u) = O(t) \quad \text{as } t \rightarrow 0.$$

And for $u > M$, we have

$$(6.7) \quad \begin{aligned} \int_M^\infty [\exp(itu^2) - 1] dF(u) &= \int_M^\infty [\exp(itu^2) - 1](2bu^{-3} + O(u^{-4})) du \\ &= 2b \int_1^\infty [\exp(itu^2) - 1]u^{-3} du + O(t) \\ &= -ibt \log t + O(t), \end{aligned}$$

by (5.10). Combining formulas (6.1)–(6.5), we find

$$(6.8) \quad \phi_2(t) = E \exp(itU^2) = 1 - ibt \log t + O(t) \quad \text{as } t \rightarrow +0.$$

Now let U_1, \dots, U_m be an independent sample of size m . Then (6.8) implies, for fixed $t \geq 0$ as $m \rightarrow \infty$,

$$(6.9) \quad \begin{aligned} E \exp\left(\frac{it(U_1^2 + \dots + U_m^2)}{m \log m}\right) &= \left[\phi_2\left(\frac{t}{m \log m}\right)\right]^m \\ &= \left[1 + \frac{ibt}{m} + o\left(\frac{1}{m}\right)\right]^m. \end{aligned}$$

In the limit as $m \rightarrow \infty$ this expression becomes $\exp(ibt)$, which is the c.f. of a variate that equals b with probability one. This completes the proof of the asserted limit formula (6.1). The equivalent formula (6.2) now follows.

7. Construction of confidence intervals. We wish to use the stereological estimator (2.4) to estimate the expected value $\mu = n$ = the unknown number of balls in the cube. Suppose we observe an independent sample of large size m :

$$(7.1) \quad U_i = u(X_i), \quad i = 1, \dots, m.$$

We will construct a confidence interval for μ .

First we compute the sample mean $\bar{U} = \sum U_i/m$. According to (5.2), the random variable

$$(7.2) \quad Y = \left(\frac{m}{\log m}\right)^{1/2} (\bar{U} - \mu)$$

is approximately Gaussian, with mean zero and with the unknown finite limit variance b . If m is large we may use formula (6.2) to approximate b by the random variable

$$(7.3) \quad B = (\log m)^{-1}(m - 1)^{-1} \sum_{j=1}^m (U_j - \bar{U})^2.$$

If we divide Y by the approximate standard deviation \sqrt{B} obtained from (7.3), we obtain the random variable

$$(7.4) \quad Z = [m(m - 1)]^{1/2}(\bar{U} - \mu) \left[\sum_{j=1}^m (U_j - \bar{U})^2 \right]^{-1/2}.$$

Then the distribution of Z is approximately Gaussian with mean 0 and variance 1.

We have computed $Z = Y/\sqrt{B}$. In this computation the logarithmic terms cancel, and the resulting quotient (7.4) looks exactly like Student's t -score with $m - 1$ degrees of freedom, which becomes standard Gaussian as $m \rightarrow \infty$. This is a useful coincidence. In morphometric cytology, there is a practical lower limit $A_0 > 0$ for experimentally observable cross-section areas A_i . Thus, the sums $\sum A_i^{-1/2}$ may have a practical upper bound in the stereological estimator (2.4). This will have the effect of producing a slightly biased estimator with a finite variance. But then the expression Z in (7.4) is appropriate, and as usual we can say that Z is approximately standard Gaussian.

It is for this reason that we have preferred the approximation (6.2) to the theoretically equivalent expression (6.1). The expression (6.1) would replace Z by

$$(7.5) \quad Z' = m(\bar{U} - \mu) \left[\sum_{i=1}^m U_i^2 \right]^{-1/2}.$$

In our theoretical case, in which $E U_i^2 = \infty$, the variate Z' , like Z in (7.4), becomes standard normal as $m \rightarrow \infty$. But in a practical application, with $\text{var}(U) = \sigma^2 < \infty$, as $m \rightarrow \infty$ the variate Z' becomes normal with variance

$$(7.6) \quad \frac{\sigma^2}{[\sigma^2 + (E U_i)^2]}.$$

Since the quantity in (7.6) is unknown, we cannot construct confidence intervals using Z' . But we can easily use the preferred estimator Z , as the following example shows.

Example. If X is uniformly distributed for $0 < X < 1$, then $U = X^{-1/2}$ has the probability density

$$(7.7) \quad f(u) = 2u^{-3}, \quad (1 < u < \infty),$$

with $\mu = E(U) = 2$ and variance infinity. Thus, U has a density of the form (4.2), with $b = 1$.

Using a standard congruential method to generate pseudorandom numbers with a computer, I generated a sample X_1, \dots, X_{20} of size $m = 20$. Setting $U_i = X_i^{-1/2}$, I computed

$$(7.8) \quad \begin{aligned} \bar{U} &= \frac{1}{m}(U_1 + \dots + U_m) = 1.947, \\ \text{SD} &= \left[\frac{1}{m-1} \sum_{i=1}^m (U_i - \bar{U})^2 \right]^{1/2} = 1.419. \end{aligned}$$

According to (7.4) and (7.8),

$$(7.9) \quad Z = \frac{\sqrt{m}(\bar{U} - \mu)}{\text{SD}}.$$

If Z were normal with mean 0 and variance 1, then Z would lie between ± 1.645 with probability 90%. Then (7.8) and (7.9) would yield this 90% confidence interval for μ :

$$(7.10) \quad 1.425 < \mu < 2.469.$$

As we should expect, the true value $\mu = 2$ does lie in this interval.

This example illustrates that for large samples U_1, \dots, U_m , confidence intervals for the expected value μ may be constructed as if the U_i were Gaussian, with finite variance.

In general, about how large will a confidence interval be? Suppose there are n balls in the unit cube. If there were only one ball, (4.5) shows the limit variance $b = \pi^{-2}r^{-1}$. For n balls, if the balls do not have common tangent planes $x = X$, the limit variance b may be roughly approximated by a sum:

$$(7.11) \quad b \sim \pi^{-2}(r_1^{-1} + \dots + r_n^{-1}).$$

For definiteness, let us look at the 90% confidence interval. If Z were standard normal, then with 90% probability $|Z| < 1.645$. Then (7.4) gives this confidence interval for the unknown $\mu = n$:

$$(7.12) \quad |\mu - \bar{U}| < 1.645[m(m-1)]^{-1/2} \left[\sum_{j=1}^m (U_j - \bar{U})^2 \right]^{1/2}.$$

From (7.3), this becomes

$$(7.13) \quad |\mu - \bar{U}| < 1.645(m^{-1} \log m)^{1/2} B^{1/2},$$

where B is an unbiased estimator of b . Replacing B in (7.13) by the approximation (7.11), we get the interval

$$(7.14) \quad |\mu - \bar{U}| < 1.645(m^{-1} \log m)^{1/2} \pi^{-1} \left(\sum_{i=1}^n r_i^{-1} \right)^{1/2}.$$

Let r be the harmonic mean of the n radii r_i . Then (7.14) becomes

$$(7.15) \quad |\mu - \bar{U}| < 0.52r^{-1/2}(n \cdot m^{-1} \log m)^{1/2}.$$

For example, suppose the unit cube contains $\mu = n = 100$ balls with harmonic mean radius $r = 0.1$. Let \bar{U} be the arithmetic mean of an independent sample U_1, \dots, U_m of size $m = 20$. Then the approximate 90% confidence interval (7.15) is $|\mu - \bar{U}| < 6.36$. Thus, with probability 90% the unknown $\mu = 100$ will be estimated with an error less than 6.36%.

8. Numerical check of the limit theorem. The limiting Gaussian behavior of $Y = (\bar{U} - \mu)\sqrt{(m/\log m)}$ depends on the formula (5.15), in which there is a term $o(1/m)$. Similarly, our formula (6.1) for the limit variance b depends on the formula (6.9), in which there is a term $o(1/m)$. In both cases, as the derivations show, the terms $o(1/m)$ are of the order

$$(8.1) \quad o\left(\frac{1}{m}\right) = O\left(\frac{\log \log m}{m \log m}\right) \quad \text{as } m \rightarrow \infty.$$

Indeed, these terms go to zero faster than $1/m$, but just barely. The factor $(\log \log m)/\log m$ goes to zero very slowly (for $m = 10^4$ it equals 0.24; for $m = 10^{10}$ it equals 0.14).

This suggests that the limits may be attained very slowly, perhaps too slowly to be useful in practice. To dispel this doubt, it is necessary to do some numerical experiments.

As in the example (7.7), I used the random variable $U = 1/\sqrt{X}$, where X is uniformly distributed over $0 < X < 1$. I computed 20 independent runs, each run producing an independent sample of size $m = 20$. For each run I computed four numbers:

$$\begin{aligned}
 \bar{U} &= \frac{1}{20}(U_1 + \cdots + U_{20}), \\
 \text{SD} &= \left[\frac{1}{19} \sum_{i=1}^{20} (U_i - \bar{U})^2 \right]^{1/2}, \\
 Z &= \sqrt{20}(\bar{U} - 2)/\text{SD}, \\
 F(Z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^Z \exp(-\frac{1}{2}z^2) dz.
 \end{aligned}
 \tag{8.2}$$

For example, the first run produced the numbers \bar{U} and Z that appeared in (7.8); it also produced the numbers

$$Z = -0.1677, \quad F(Z) = 0.4334. \tag{8.3}$$

The 20 runs produced 20 numbers $W = F(Z)$. If the 20 numbers Z were an independent sample from the normal distribution, then the 20 numbers W would be an independent sample from the uniform distribution.

To check the uniformity of the 20 numbers W , I computed the Kolmogorov-Smirnov statistic

$$D_{20} = \sup_{0 \leq x \leq 1} |F_{20}(x) - x|, \tag{8.4}$$

where $f_{20}(x)$ is the empirical distribution function of the W 's:

$$F_{20}(x) = \frac{1}{20} \sum_{w \leq x} 1, \quad 0 \leq x \leq 1. \tag{8.5}$$

I computed the value $D_{20} = 0.2137$. To check the significance of this value, I used the table in [5, p. 324], which says

$$\Pr \{D_{20} > 0.231\} = 0.20. \tag{8.6}$$

Therefore, if the 20 numbers Z had come from the normal distribution, the Kolmogorov-Smirnov statistic D_{20} would have exceeded the experimental value 0.2137 with probability greater than 0.20. Thus, if a statistician were testing the hypothesis that these 20 numbers Z came from the standard normal distribution, he would accept the hypothesis at the 20% significance level.

9. Generalization. For definiteness, we have discussed only the case of spheres. The paper [1] shows how to define unbiased estimators $U = u(X)$ for the number n of general convex bodies in the unit cube. No assumption is made on the size distribution; one assumes only the knowledge of the mean shape constant J defined in [1], formula (4.2).

If the bodies are smooth, the shape constant J is finite, and our stereological estimator (2.4) becomes

$$u(X) = J^{-1} \sum_{A_i > 0} A_i^{-1/2}. \tag{9.1}$$

(For spheres, the shape constant J equals $\sqrt{\pi}$.) The variance of the estimator is infinite, but its expected value equals the number of bodies in the cube.

In every case it is easy to see, as in § 4 of this paper, that the estimator $u(X)$ has a probability density of order u^{-3} for large u , as in formula (4.2). But that is all we needed to establish our limit theorem. And so we have the result for general smooth convex bodies: *for large independent samples U_1, \dots, U_m the quantity Z in (7.4) is approximately standard Gaussian, with mean 0 and variance 1.* In applications, this limit theorem enables us to construct intervals for the statistical estimation of the number of bodies in the cube.

REFERENCES

- [1] J. N. FRANKLIN, *Some stereological principles in morphometric cytology*, this Journal, 33 (1977), pp. 267–278.
- [2] ———, *Estimation of three-dimensional cytological parameters from two-dimensional measurements made on cell sections*, J. Cell Biology, 74 (1977), pp. 485–489.
- [3] J. W. POSAKONY, J. M. ENGLAND AND G. ATTARDI, *Mitochondrial growth and division during the cell cycle in HeLa cells*, J. Cell Biology, 74 (1977), pp. 468–491.
- [4] R. L. FULLMAN, *Measurement of particle sizes in opaque bodies*, J. Metals, 5 (1953), pp. 447–452.
- [5] B. W. LINDGREN, G. W. MCEL RATH AND D. A. BERRY, *Probability and Statistics*, 4th Ed., Macmillan, New York, 1978.
- [6] M. ABRAMOWITZ AND I. A. STEGUN, eds., *Handbook of Mathematical Functions*, National Bureau of Standards Applied Mathematics Series, vol. 55, Washington, DC, 1964.
- [7] M. G. KENDALL AND P. A. P. MORAN, *Geometric Probability*, Griffin, London, 1963.
- [8] G. S. WATSON, *Estimating functionals of particle size distributions*, Biometrika, 58 (1971), pp. 483–490.
- [9] L. SANTALÓ, *Integral Geometry and Geometric Probability*, Encyclopedia of Mathematics and Its Applications, vol. 1, Gian-Carlo Rota, ed., Addison-Wesley, Reading, MA, 1976.
- [10] L.-M. CRUZ ORIVE, *Particle size-shape distributions: the general spheroid problem*, J. Microscopy, 107 (1976), pp. 235–253.