Journal of Statistical Physics, Vol. 18, No. 1, 1978

Number Fluctuation Analysis of Random Locomotion. Statistics of a Smoluchowski Process

Stephen L. Brenner,[1] Ralph J. Nossal,[1] and George H. Weiss[1]

Received June 27, 1977

We analyze a scheme, originally suggested by Smoluchowski, by which a diffusion coefficient D can be estimated by measuring the number of particles occupying a fixed region of a surface at various times. An expression is derived relating the variance of the estimated value $\hat{D}$ to several experimental parameters. This expression is evaluated numerically to determine how statistical uncertainty depends on adjustable variables. Particular attention is given to experiments involving locomotion of migrating leukocytes.

KEY WORDS: Number fluctuation spectroscopy; leukocyte mobility; Smoluchowski process; diffusion.

1. INTRODUCTION

During the last sixty years there have been several applications of a counting scheme now known as number fluctuation spectroscopy (NFS). In this technique the fluctuating number of particles in a fixed region of space is determined as a function of time, and kinetic parameters then are derived from the statistical properties of the resultant stochastic signal. An early application of this method was the measurement of diffusion coefficients of colloidal particles,[1,2] based on a theoretical analysis due to Smoluchowski,[3] which has been summarized in reviews by Chandrasekhar[4] and Kac.[5] Related schemes recently have been used to measure diffusion coefficients and reaction rates of macromolecules by fluorescence correlation spectroscopy[6,7] and to obtain swimming speeds of flagellated bacteria by analysis of fluctuating scattered light.[8,9] The technique has also been used to estimate mobility parameters of migrating white blood cells.[10]

Because of the wide applicability of NFS, it is of interest to design such

[1] National Institutes of Health, Physical Sciences Laboratory, Division of Computer Research and Technology, Bethesda, Maryland.

0022-4715/78/0100-0001$05.00/0

experiments so as to minimize the sampling error, i.e., the error due to having only statistical estimates rather than exact knowledge of the relevant statistical functions. Similar analyses have been given by Jakeman *et al.*,[11] DeGiorgio and Lastovka,[12] Saleh and Cardoso,[13] and Koppel,[14] among others, for a variety of experiments involving autocorrelations in photon counting experiments and fluorescence correlation spectroscopy. In the present paper we consider the design of such autocorrelation experiments in two dimensions, with the specific object of designing experiments to measure leukocyte motility in terms of correlation functions. There are a number of experiments showing that leukocyte migration on a surface, in the absence of chemoattractants, can be characterized as a diffusion-like process.[15] Experiments by Nossal and Chang[10] on leukocyte locomotion used time-lapse cinemicrography to record the number of cells in a fixed area at different times. An analysis following that of Smoluchowski[13] was used to estimate an average diffusion constant (or mobility coefficient) $\hat{D}$ from calculated autocorrelation functions.

Several parameters need to be set in such an experiment. The following are of interest: (1) size of the viewing area; (2) cell population density or, equivalently, expected number of cells in the viewing area; (3) number of observations; (4) time interval between successive observations; and (5) number of points used in curve fitting the autocorrelation function. Additionally, one must take into account that biological parameters change with time, effectively limiting the maximum duration of the experiment. This contrasts with correlation experiments in physics that have no such limitations.

Let $\hat{D}$ be the estimate of D, and let $\mathrm{var}(\hat{D})$ be the variance of that estimate. We will study the behavior of

$$\epsilon^2 = \mathrm{var}(\hat{D})/D^2 \tag{1}$$

as a function of the parameters enumerated above. In Section 2 we derive an expression for $\mathrm{var}(\hat{D})$ on the assumption that the estimate $\hat{D}$ is made by a linearized least squares fit to the theoretical autocorrelation function. Section 3 contains numerical computations based on the theory in Section 2. Several details of the analysis are presented in appendices.

2. THEORY

We assume that noninteracting particles move as random walkers in two dimensions in an isotropic, infinite medium, i.e., one for which edge effects can be neglected. The basic idea of NFS is to choose a region Ω and to count the number of particles in Ω at times $\tau, 2\tau, 3\tau, \ldots$. These numbers, denoted by $n_i = n(i\tau)$, are assumed to be measured exactly, and form a

stochastic process. The scheme suggested by Smoluchowski starts from the observation that the theoretical autocorrelation function $R(t)$ satisfies

$$R(t) \equiv \langle n(t)n(0)\rangle - \langle n\rangle^2 = \langle n\rangle Q(t) \tag{2}$$

where $Q(t)$ is the conditional probability that a particle in Ω at $t = 0$ is in Ω at time t, and $\langle n\rangle$ is the expected number of particles in Ω at any time. In the present case we choose Ω to be a square of side X. Under the assumptions of diffusive motion, the transition probability density can be written

$$P(\mathbf{r}_2, t|\mathbf{r}_1, 0) = (4\pi Dt)^{-1} \exp[(1/4Dt)|\mathbf{r}_2 - \mathbf{r}_1|^2] \tag{3}$$

and $Q(t)$ can be expressed in terms of the dimensionless variable $z = X/(4Dt)^{1/2}$ as

$$\begin{aligned} Q(t) &= (1/X^2) \int_\Omega \int P(\mathbf{r}_2, t|\mathbf{r}_1, 0)\, d^2\mathbf{r}_1\, d^2\mathbf{r}_2 \\ &= \{\operatorname{erf} z + (1/z\sqrt{\pi})[\exp(-z^2) - 1]\}^2 \end{aligned} \tag{4}$$

where $\operatorname{erf} z \equiv (2/\pi^{1/2}) \int_0^z \exp(-u^2)\, du$. A graph of Q plotted as a function of z^{-1} is shown in Fig. 1.

Observations of the Smoluchowski process cannot furnish values of Q, but only estimates of that function. Nossal and Chang[10] fit their data to

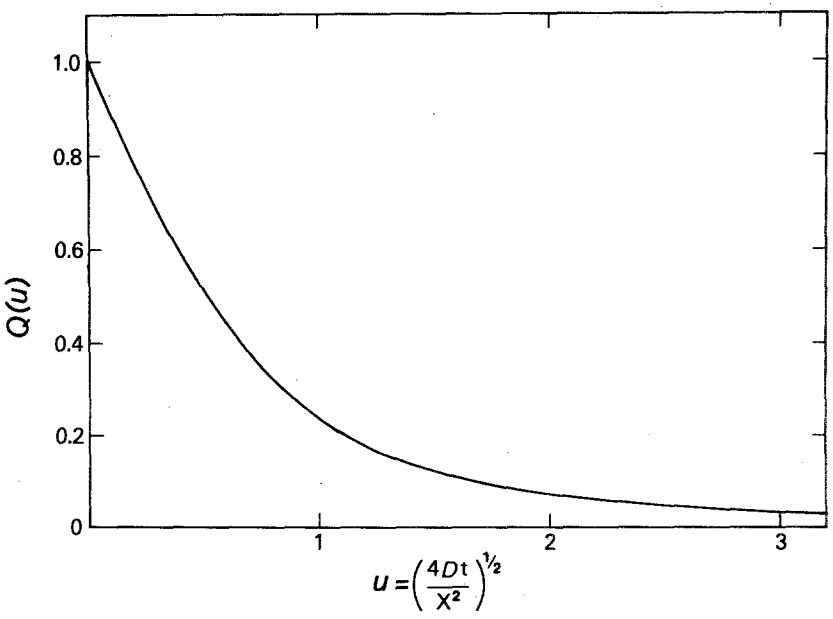

Fig. 1. The probability $Q(t)$ as defined in Eq. (4), plotted as a function of the dimensionless variable $(4Dt/X^2)^{1/2}$.

the linear part of the estimated Q, that is, to the early part of the curve in Fig. 1. Alternatively, one can use a nonlinear least squares technique to fit a larger range in z, presumably leading to a more accurate estimate of D. However, the problem of optimizing design parameters has not yet been considered.

Our method for calculating var($\hat{D}$) starts from a linearized least squares function whose minimization yields an expression for $\hat{D}$. We point out that it might be appropriate to start by minimizing a quadratic form derived from the maximum likelihood formalism,[13] but the resulting computations become extremely complicated with no assurance that the inherent Gaussian approximation is applicable. We start with the simpler problem of determining $\hat{D}$ by minimizing the quadratic form

$$f(\tilde{D}) = \sum_{i=1}^{M} [\hat{Q}_i - Q(i, \tilde{D})]^2 \tag{5}$$

in which $\hat{Q}_i$ is the experimental estimate of $Q(i\tau)$, $Q_i \equiv Q(i, \tilde{D})$ is the theoretical form of Eq. (4) with $t = i\tau$, and M is the number of points to be used for the curve fit. The parameter $\hat{D}$ is therefore the solution to

$$\sum_{i=1}^{M} [Q(i, \hat{D}) - \hat{Q}_i] \frac{\partial Q_i}{\partial \tilde{D}}\bigg|_{\tilde{D}=\hat{D}} = 0 \tag{6}$$

Let us denote the residuals by δD and δQ_i, i.e.,

$$\delta D = D - \hat{D}, \qquad \delta Q_i = Q(i, \hat{D}) - \hat{Q}_i \tag{7}$$

and make the further assumption that the experiment is sufficiently accurate that only first-order terms in the δQ_i need be retained. This leads to the approximation

$$\delta D = -\sum_{i=1}^{M} \frac{\partial Q_i}{\partial D} \delta Q_i \Big/ \sum_{i=1}^{M} \left(\frac{\partial Q_i}{\partial D}\right)^2 + \text{higher order terms} \tag{8}$$

and allows us to write ϵ^2 [Eq. (1)] as

$$\epsilon^2 = \frac{\sum_{i=1}^{M} \sum_{i=1}^{M} C_i C_j Q_i^{1/2} Q_j^{1/2} \langle \delta Q_i \, \delta Q_j \rangle}{(\sum_{i=1}^{M} C_i^2 Q_i)^2} \tag{9}$$

where C_j is defined by

$$C_j = (1/y)(j/\pi)^{1/2}[\exp(-y^2/j) - 1] \tag{10}$$

Notice that the cross-terms $\langle \delta Q_i \, \delta Q_j \rangle$ must be retained since we are not dealing with measurement error, which usually is assumed to be uncorrelated.

In order to use the result in Eq. (9), we must specify the form of the $\hat{Q}_i$ and calculate the covariances $\langle \delta Q_i \, \delta Q_j \rangle$. If we define the estimators

$$\hat{n} = \frac{1}{N} \sum_{i=1}^{N} n_i, \qquad \hat{G}_j = \frac{1}{N-j} \sum_{i=1}^{N-j} n_i n_{i+j} \tag{11}$$

then an obvious estimator for Q_i is

$$\hat{Q}_i = (\hat{G}_i - \hat{n}^2)/\hat{n} \tag{12}$$

The analysis of ϵ^2 that results from using this nonlinear estimator is extremely complicated. To simplify the analysis, we will use a linear approximation adopted by Koppel.[14] The approximation is valid, as we will see, in the limit of large N. A derivation of the linearized estimator, which we denote by $\hat{q}_i$, is as follows: Write $\hat{n} = \langle n \rangle + \delta n$ and substitute this into Eq. (12), keeping terms that are first order in δn. This leads to

$$\begin{aligned} Q_i &\approx \frac{1}{\langle n \rangle} \left\{ [\hat{G}_i - \langle n \rangle^2] \left[1 - \frac{\delta n}{\langle n \rangle} \right] - 2\langle n \rangle \, \delta n \right\} \\ &= \frac{1}{\langle n \rangle} \left\{ (\hat{G}_i - \langle n \rangle^2) \left[2 - \frac{\hat{n}}{\langle n \rangle} \right] - 2\langle n \rangle \, \delta n + 2\langle n \rangle^2 \right\} \end{aligned} \tag{13}$$

We justify the neglect of second-order terms by noticing that $\langle n \rangle = O(1)$, $\langle \delta n \rangle = 0$, and $\langle \delta n^2 \rangle = O(\ln N/N)$. The order of $\langle \delta n^2 \rangle$ has been estimated by noting

$$\begin{aligned} \langle \delta n^2 \rangle &= \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} (n_i - \langle n \rangle)(n_j - \langle n \rangle) \\ &= \frac{1}{N^2} \left(N\sigma^2 + \langle n \rangle \sum_{j=1}^{N-1} (N-j) Q_j \right) \end{aligned} \tag{14}$$

where $\sigma^2 = \langle n \rangle^2 - \langle n \rangle^2$ and Q_j is given by Eq. (4). The logarithmic dependence arises from the fact that Q_j varies as j^{-1} for large values of j. A second step in the derivation of $\hat{q}_i$ involves replacing the term $2 - \hat{n}/\langle n \rangle$ by 1. If one writes

$$2 - \hat{n}/\langle n \rangle = 1 + \beta \tag{15}$$

it then follows that $\langle \beta \rangle = 0$, $\langle \beta^2 \rangle = \langle \delta n^2 \rangle / \langle n \rangle$, so that setting $\beta = 0$ in Eq. (15) involves an error that is comparable to that used in deriving Eq. (13). When such an approximation is made we find

$$\hat{q}_i = \langle n \rangle^{-1} (\hat{G}_i - 2\hat{n}\langle n \rangle + \langle n \rangle^2) \tag{16}$$

from which it follows that $\langle q_i \rangle = Q_i = Q(i\tau)$. Note that although Eq. (15) is used in the error analysis, the Q_i defined by Eq. (12) should be used for the actual data reduction.

These results allow us to approximate $\langle \delta Q_i \, \delta Q_j \rangle$ by

$$\begin{aligned} \langle \delta Q_i \, \delta Q_j \rangle &\approx \langle \hat{q}_i \hat{q}_j \rangle - Q_i Q_j \\ &= (1/\langle n \rangle^2) \langle \hat{G}_i \hat{G}_j \rangle - (2/\langle n \rangle) \langle \hat{n}(\hat{G}_i + \hat{G}_j) \rangle \\ &\quad + 4\langle \hat{n}^2 \rangle - \langle n \rangle^2 + \langle n \rangle (Q_i + Q_j) - Q_i Q_j \end{aligned} \tag{17}$$

The first term in Eq. (17) requires evaluation of four-time correlation functions since

$$\langle \hat{G}_i \hat{G}_j \rangle = \frac{1}{(N-i)(N-j)} \sum_{r=1}^{N-i} \sum_{s=1}^{N-j} \langle n_r n_{r+i} n_s n_{s+j} \rangle \tag{18}$$

and the second term requires evaluation of three-time correlation functions. Expressions for such quantities are derived in Appendix A. After performing the requisite algebra we then find that $\langle \delta Q_m \, \delta Q_l \rangle$ has the form

$$\langle \delta Q_m \, \delta Q_l \rangle = N^{-1} \sum_{i=1}^{3} f_i(m, l) \langle n \rangle^{i-2}, \qquad l \geqslant m \tag{19}$$

where explicit expressions for the coefficients $f_i(m, l)$ are given in Appendix B [Eqs. (B6)–(B9)]. As seen next, this simple dependence upon $\langle n \rangle$ enables us to obtain an analytic expression for the optimal number of particle to be used for a given set of experimental conditions. Consequently, it greatly facilitates numerical analysis of how the uncertainty in $\hat{D}$ depends upon experimental parameters.

3. RESULTS AND DISCUSSION

We now use Eqs. (9) and (19) to explore how uncertainty in the estimated D is influenced by various experimental parameters. The uncertainty ϵ^2 is a function of five variables that can be set by the experimenter. These are: (1) X, the size of the square; (2) τ, the interval between successive measurements of $n(t)$; (3) M, the number of points used in the least squares evaluation of $\hat{D}$; (4) $\langle n \rangle$, the expected number of particles in the viewing area, or equivalently, the particle density; and (5) N, the total number of measurements.

In the present calculation we assume known an initial estimate of D. This type of assumption is customary in the statistical literature on design of experiments,[16] but other approaches taking into account an initial uncertainty are also possible.[17] Our procedure is reasonable as long as ϵ^2 is a slowly varying function of D.

3.1. General Comments

We have seen that X, τ, and D always appear together in the dimensionless form $y = (X^2/4D\tau)^{1/2}$. The root-mean-square distance moved by a

particle in an interval of time τ is $(4D\tau)^{1/2}$, so that y is the ratio of the size of the viewing field to a length representative of the average distance a particle migrates in a fundamental time unit. Intuitively, we might expect y to be on the order of unity for the number fluctuation experiment to be effective. In a recent experimental study of leukocyte locomotion, Nossal and Chang[10] found a mobility of 7.5×10^{-8} cm^2/sec, having empirically chosen a square viewing field of side 0.3 mm and having counted cells 180 successive times at intervals of 150 sec. This corresponds to a value of $y \approx 4.5$.

The optimum choice for $\langle n \rangle$, i.e., that which minimizes the variance in $\hat{D}$, can be determined analytically. From Eqs. (9) and (19) we obtain

$$\epsilon^2 = A_1\langle n \rangle^{-1} + A_2 + A_3\langle n \rangle \tag{20}$$

where the A_i are independent of $\langle n \rangle$ and are explicitly defined in Appendix B.

Minimizing ϵ^2 with respect to $\langle n \rangle$ yields for the optimal value $\langle n \rangle^* = (A_1/A_3)^{1/2}$. This value was used to reduce the dimensionality of the multivariate parameter space. Also, unless otherwise stated, M was adjusted to the value M^* that minimizes ϵ^2.

3.2. Effect of the Number of Observations N

The uncertainty in $\hat{D}$ can be reduced by increasing N, the total number of observations. Two cases are of interest, the first being that of unconstrained

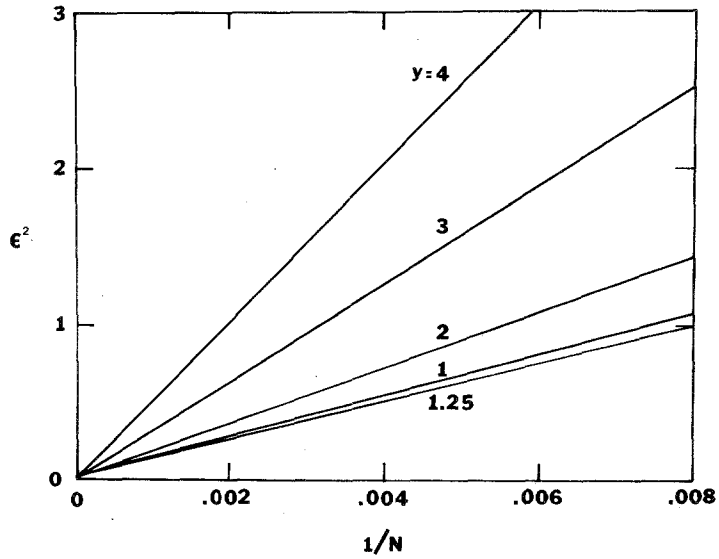

Fig. 2. Plot of ϵ^2, the relative variance in $\hat{D}$, as a function of the inverse of the number of observation N. Curves are drawn for several values of y.

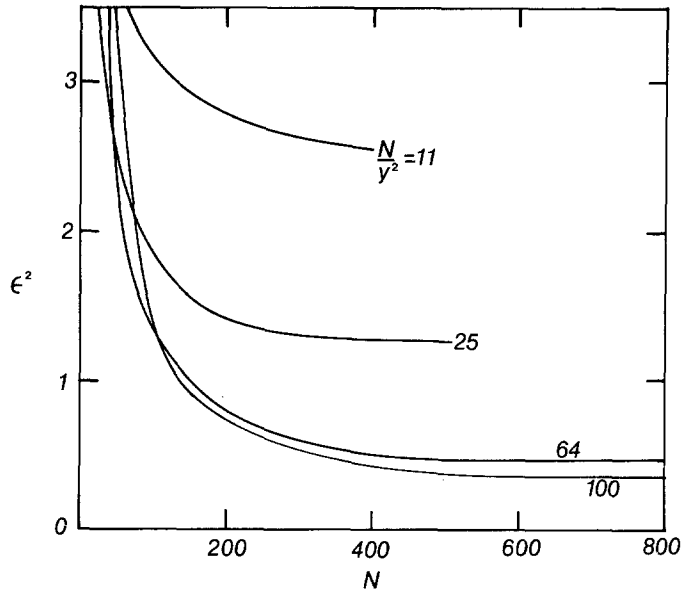

Fig. 3. The relative variance in $\hat{D}$ as a function of the number of observations N for several fixed-time experiments characterized by different values of N/y^2.

N and τ, and the second being the time-constrained experiment in which $N\tau$ is held fixed. Since ϵ^2 is a sampling variance, we might expect it to be roughly proportional to N^{-1}. Although this dependence is not obvious from the detailed expression for ϵ^2, it is, in fact, obeyed quite closely in the unconstrained experiment. Such dependence is shown in Fig. 2. Therefore, by doubling the number of observations, one can halve the uncertainty in $\hat{D}$. In the time-constrained experiment we fix $N\tau = T$, which is equivalent to fixing N/y^2. In Fig. 3 we plot ϵ^2 as a function of N for several values of N/y^2. In contrast to the situation for unconstrained experiments, ϵ^2 here tends toward a constant value characteristic of continuous sampling.

3.3. Effect of the Field Dimension X and Sampling Interval τ

The parameters X and τ always appear in the combination $X/\tau^{1/2}$, i.e., through the variable y. Hence a plot of ϵ^2 as a function of y shows the effect either of increasing the size of the field or decreasing the interval between observations. Figure 4 gives such a plot for the value $N = 100$. To convert this to other values of N, one need only multiply the plotted curve by $(100/N)$ as suggested by Fig. 2. The position of the minimum is therefore invariant with respect to changes in N.

As y increases, the box becomes so big or the experiment so short that

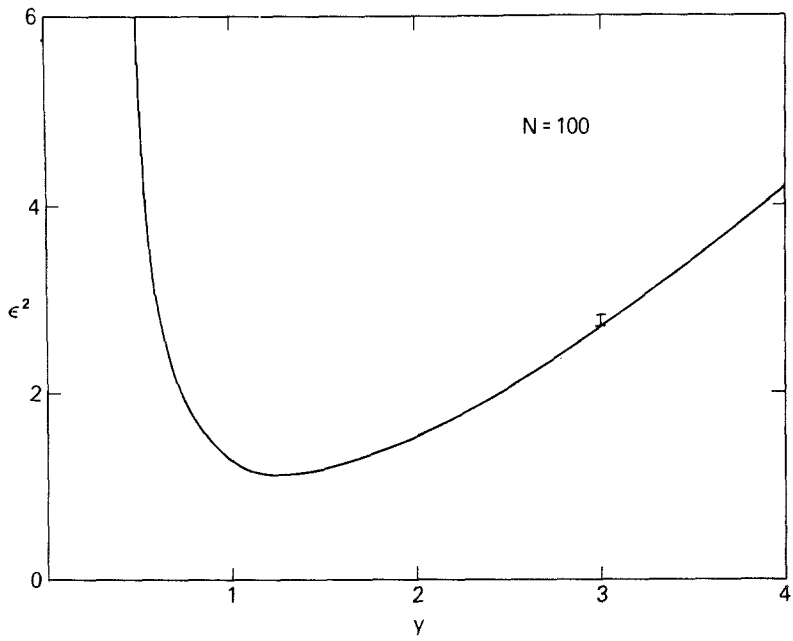

Fig. 4. The relative variance of $\hat{D}$ as a function of the dimensionless variable $y = (X^2/4D\tau)^{1/2}$. The bar "I" at $y = 3$ denotes the range of ϵ^2 that results from choosing M between 8 and 14, while the optimal choice for this value of y is $M = 11$.

fluctuations of $n(t)$ tend to be rare. Conversely, as y gets small, either Ω becomes so small or the time between observations so long that the number of particles present in Ω at any two times may be unrelated. Thus, for very large or very small y, one expects ϵ^2 to be large. The position of the minimum is quite close to $y = 1.25$. Since the value of D is initially unknown, the exact value of the optimum y is not interesting by itself. However, if one avoids experiments for which $y < 0.5$, the variation of ϵ^2 with y is not great. If the actual value of y is anywhere between 0.6 and 3.8, ϵ will be less than twice its minimum value. The asymmetry about the minimum suggests that it is least damaging to choose y too large, that is, it is better to have few correlated fluctuations rather than many uncorrelated ones.

The vertical line on the curve in Fig. 4 is another indication of the insensitivity of the standard error ϵ to the parameters. The optimal number of points to use in the least squares fit is 11 for $y = 3$. The line shows the effect of choosing values of M from 8 to 14.

For the fixed-time experiment we have set $N\tau = T$. This is equivalent to fixing $N/y^2 = \alpha = \text{const}$. In Fig. 3 we have plotted ϵ^2 vs. N for several time constants. For N sufficiently large the curves of ϵ^2 reach a plateau

value which is relatively insensitive to α. The plateau is reached when $N \gtrsim 8\alpha$.

If N is chosen too small, the value of ϵ^2 is considerably larger than its plateau value. The sharp drop in ϵ^2 in the curves of Fig. 3 can be understood in terms of the curve for the unconstrained case (Fig. 4). Since we have fixed $N/y^2 = \alpha$, an increase in N implies an increase in y. Hence, when $y < y^* \approx 1.25$, an increase in N decreases ϵ^2 on two accounts, the increase in N and that in y. On the other hand, when $y > y^*$, an increase in N decreases ϵ^2 because of the larger sample number, but there is a compensating increase in ϵ^2 due to y.

3.4. Effect of Choosing M

The estimation of $\hat{D}$ involves a least squares fit to M values of the $\hat{Q}_j$ [Eq. (4)]. If M is too small, not enough values of $Q(t)$ are sampled. If M is too large, values of $Q(t)$ with large sampling errors will be used. A curve of M^*, the optimal value of M for a given y, is plotted as a function of y in Fig. 5. Although M^* is an integer, we have plotted a smooth curve, with the understanding that the integer closest to the curve is to be chosen. Additional

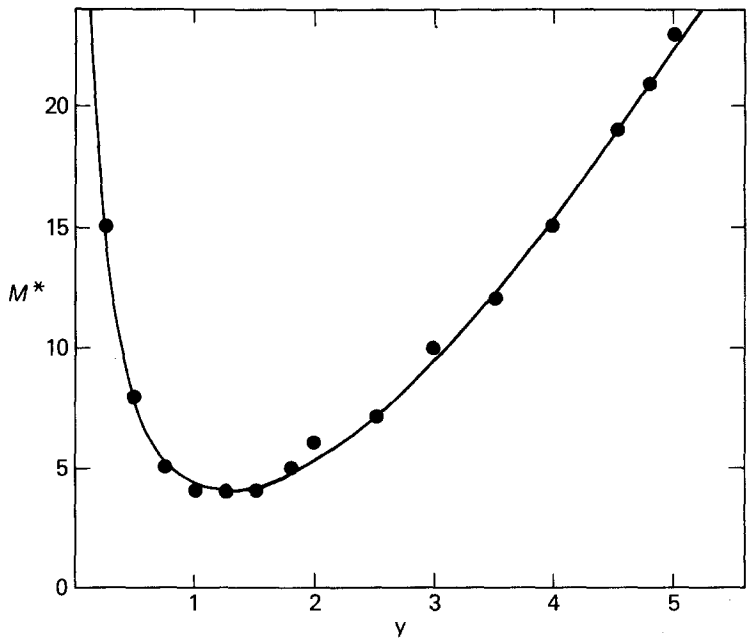

Fig. 5. Plot of M^*, the optimal number of points to use in the least squares fit to $Q(t)$, as a function of y. The points are found to be independent of the choice of $\langle n \rangle$ or N.

calculations show that ϵ^2 is insensitive to variations in M. For example, when $y = 2$ the optimizing value of M is $M^* = 6$, but ϵ^2 is at most 10% greater than its minimum value for $3 \leqslant M \leqslant 12$.

3.5. Dependence of $\langle n \rangle^*$ on N and y

At the beginning of this section we derived an equation for $\langle n \rangle^*$, the optimal number of particles to be found in the viewing area. This quantity is important when designing an experiment and we now indicate its dependence on other variables.

We have observed from numerical calculations that $\langle n \rangle^*$ increases with N according to

$$\langle n \rangle^* = A(y)N^{\beta(y)} \tag{21}$$

where $\beta(y)$ is only weakly dependent upon y, decreasing from a value of 0.6 at $y = 5$ to a value 0.5 at $y = 1$. The coefficient $A(y)$ is shown in Fig. 6.

Figure 7 illustrates the dependence of $\langle n \rangle^*$ on y for a fixed number of

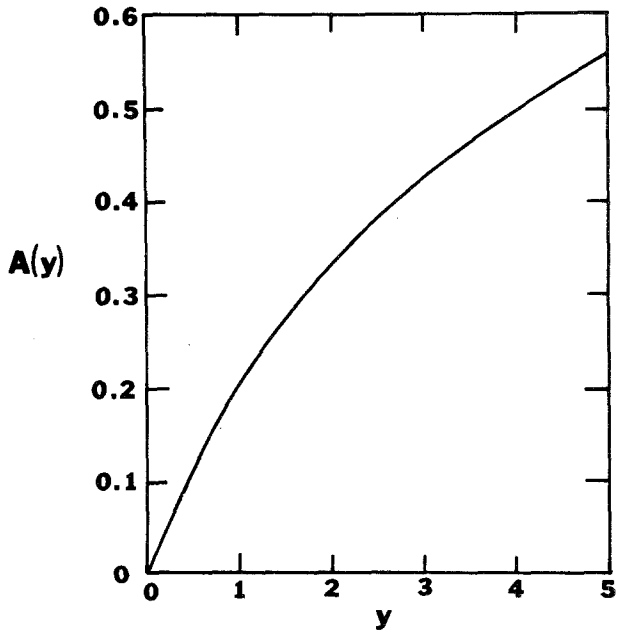

Fig. 6. The coefficient $A(y)$ as a function of y [see Eq. (21)]. The number of points in the least squares fit is consistently taken as $M = 4$.

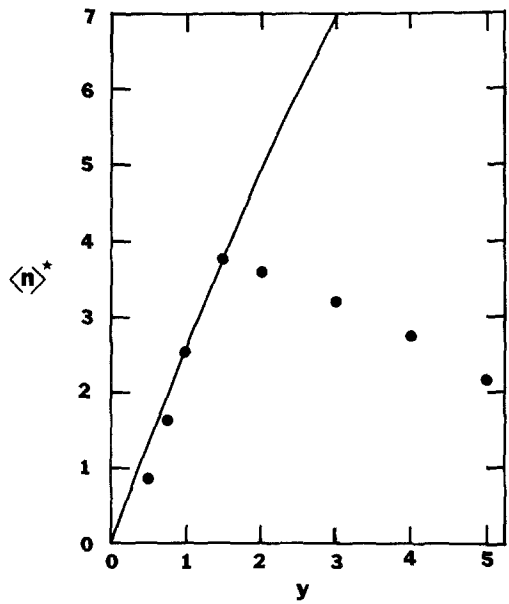

Fig. 7. The optimal number of particles in the viewing region $\langle n \rangle^*$ shown as a function of y. The solid line is calculated with $M = 4$ for all y. The circles indicate values of $\langle n \rangle^*$ that follow upon setting $M = M^*$. The number of observations here is taken to be $N = 100$ [see Eq. (21)].

observations. The solid line pertains to the case $M = 4$. We observe that the optimal number of particles increases with X. This is reasonable since particles near the boundaries of Ω are the ones likely to contribute to the fluctuations in $n(t)$. If M is always adjusted to its optimal value for a given y, the behavior indicated by the circles in Fig. 7 results. Although $\langle n \rangle^*$ varies with M, separate calculations show that the normalized variance ϵ^2 is quite insensitive to the expected number of particles in the viewing region. For a typical case, when $\langle n \rangle^* = 2$, variations in $\langle n \rangle$ between 0.5 and 8 lead to an increase in ϵ of no more than 20%.

4. CONCLUDING REMARKS

Suppose we were to consider a number fluctuation experiment where the total number of observations and the size of the viewing area were fixed for reasons related to instrumental design. In this case the mobility coefficient should be estimated, and the interval between observations chosen so that $y \simeq 1.25$–2. In contrast, for an experiment utilizing leukocytes—in which the total observation time T is fixed by the limited viability of the cells—one

might proceed as follows: (1) estimate D and select a field size X small enough that $\alpha \equiv 4DT/X^2 \geqslant 10$, yet large enough that X will be 10–20 times the average persistence length of the individual segments of the random walk executed by the cells in order to be reasonably certain that the cell motion may be viewed as a diffusion process; (2) determine the number of observations N from the relationship $N \geqslant 8\alpha$ and, thus, the time between observations τ from the allowed total time $T = N\tau$; (3) estimate the cell density $\langle n \rangle$ from Eq. (21), with $A(y)$ given by the curve in Fig. 6. Ideally, $\langle n \rangle$ should be large enough so that one might vary the size of the viewing area about the optimal value and perform redundant analyses of the data in order to increase the reliability of the extracted mobility coefficient (see, e.g., Fig. 3 of Ref. 10).

Finally, it is to be noted that the preceding work concerns minimization of the uncertainty ϵ^2 when the kinetic parameter is extracted from a *single* number fluctuation experiment. If r independent experiments are performed, the uncertainty in the mobility ϵ^2 is decreased to $1/r$th its value for a single experiment. In the case of the leukocyte migration studies, the viewing area Ω is obtained by masking off a small portion of a time-lapse motion picture and several experiments effectively can be performed by marking off different areas of the same film record. These will not be truly independent experiments, but if the viewing regions are far apart, we expect the decrease in ϵ^2 to be roughly proportional to the inverse of the number of viewing areas studied.

APPENDIX A. EXPRESSIONS FOR MULTIPLE-TIME CORRELATION FUNCTIONS

Multiple-time correlation functions $\langle n(t_1)n(t_2) \cdots n(t_r)\rangle$ are defined by

$$\langle n(t_1)n(t_2) \cdots n(t_r)\rangle = \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \cdots \sum_{n_r=0}^{\infty} n_1 n_2 \cdots n_r W_r(n_1, t_1; n_2, t_2; \ldots; n_r, t_r) \tag{A1}$$

where W_r is the joint probability for finding n_1 particles in Ω at time t_1, n_2 at t_2, and so on. Kac[5] has derived a generating function from which it is possible to find the multiple-time correlation functions by differentiation.

If $n_1 \equiv n(t_1)$ and $Q(t_2 - t_1)$ is designated $Q_{2\text{-}1}$, then the first four correlation functions are

$$\begin{aligned}
\langle n_1 n_2 \rangle &= \langle n \rangle^2 + \langle n \rangle Q_{2\text{-}1} \\
\langle n_1 n_2 n_3 \rangle &= \langle n \rangle^3 + \langle n \rangle^2 (Q_{2\text{-}1} + Q_{3\text{-}1} + Q_{3\text{-}2}) + \langle n \rangle g_{123} \\
\langle n_1 n_2 n_3 n_4 \rangle &= \langle n \rangle^4 + \langle n \rangle^3 (Q_{2\text{-}1} + Q_{3\text{-}1} + Q_{4\text{-}1} + Q_{3\text{-}2} + Q_{4\text{-}2} \\
&\quad + Q_{4\text{-}3}) + \langle n \rangle^2 (g_{123} + g_{124} + g_{134} + g_{234} + Q_{2\text{-}1}Q_{4\text{-}3} \\
&\quad + Q_{3\text{-}1}Q_{4\text{-}2} + Q_{4\text{-}1}Q_{3\text{-}2}) + \langle n \rangle g_{1234}
\end{aligned} \tag{A2}$$

where we have used the ordering convention $t_1 < t_2 < t_3 < t_4$. The quantities g_{ijk} and g_{ijkl} are defined as

$$g_{ijk} = (1/|\Omega|) \iiint P(\mathbf{r}_2, t_j; \mathbf{r}_3, t_k|\mathbf{r}_1, t_i)\, d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3$$
$$g_{ijkl} = (1/|\Omega|) \iiiint P(\mathbf{r}_2, t_j; \mathbf{r}_3, t_k; \mathbf{r}_4, t_l|\mathbf{r}_1, t_i)\, d\mathbf{r}_1\, d\mathbf{r}_2\, d\mathbf{r}_3\, d\mathbf{r}_4 \qquad \text{(A3)}$$

where the integrations are performed over Ω. Approximations for evaluating these functions are discussed in Appendix C.

APPENDIX B. DERIVATION OF EQS. (19) AND (20)

Substitution of Eq. (11) for $\hat{G}$ and $\hat{n}$ into Eq. (16) yields

$$\langle \delta Q_m\, \delta Q_l \rangle \approx \frac{1}{(N\langle n\rangle)^2} \sum_{i=1}^{N} \sum_{j=1}^{N} [\langle n_i n_{i+m} n_j n_{j+l}\rangle - \langle n\rangle^4 + 4\langle n_i n_j\rangle\langle n\rangle^2 - 2\langle n\rangle(\langle n_i n_{i+m} n_j\rangle + \langle n_j n_{j+l} n_i\rangle)] \qquad \text{(B1)}$$

Equation (B1) can be simplified by considering separately the terms in the double sum where $i = j$, $i < j$, and $i > j$. We first note that there are N identical terms which arise where $i = j$. For $i < j$ define $k = j - i$ and for $i > j$ define $k = i - j$ and note that there are $N - k$ identical terms for each possible k. Thus, Eq. (B1) reduces to

$$\langle \delta Q_m\, \delta Q_l \rangle = \frac{1}{N\langle n\rangle^2} \left[D + \sum_{k=1}^{N} \left(1 - \frac{k}{N}\right) T_k \right] \qquad \text{(B2)}$$

where

$$D = \langle n_0^2 n_m n_l\rangle - \langle n\rangle^4 + 4\langle n\rangle^2\langle n^2\rangle - 2\langle n\rangle(\langle n_0^2 n_m\rangle + \langle n_0^2 n_l\rangle) \qquad \text{(B3)}$$

and

$$T_k = \langle n_0 n_m n_k n_{k+l}\rangle + \langle n_0 n_l n_k n_{k+m}\rangle - 2\langle n\rangle(\langle n_0 n_m n_k\rangle + \langle n_0 n_k n_{k+m}\rangle) - 2\langle n\rangle(\langle n_0 n_l n_k\rangle + \langle n_0 n_k n_{k+l}\rangle) + 8\langle n\rangle^2\langle n_0 n_k\rangle - 2\langle n\rangle^4 \qquad \text{(B4)}$$

Inserting the expression given by Eqs. (A4)–(A5) into the latter, we find after some algebraic manipulations

$$N\langle \delta Q_m\, \delta Q_l \rangle = \sum_{i=1}^{3} f_i(m, l)\langle n\rangle^{i-2}, \qquad l \geqslant m \qquad \text{(B5)}$$

where

$$f_i(m, l) = d_i + \sum_{k=1}^{m} a_k \gamma_i(k; 0, l, m) + \sum_{k=m+1}^{N-1} a_k \gamma_i(k; 0, l, 2k - m) + \sum_{k=1}^{l-m} a_k \gamma_i(k; 1, m, l) + \sum_{k=l-m+1}^{l} a_k \gamma_i(k; 0, m, l) + \sum_{k=l+1}^{N-1} a_k \gamma_i(k; 0, m, 2k - l) \qquad \text{(B6)}$$

and where $a_k = (1 - k/N)$. Here

$$d_1 = g_{0ml}, \qquad d_2 = 2g_{0ml} + Q_{l-m} + Q_m Q_l - Q_m - Q_l, \qquad d_3 = 1 - Q_m - Q_l + Q_{l-m} \tag{B7}$$

$$\gamma_1(k; x, y, z) = g_{0,\xi,\nu,\rho}$$
$$\gamma_2(k; x, y, z) = g_{0,\nu+\xi-k,\rho} + g_{\xi,\nu,\rho} - g_{0,\xi,\zeta} - g_{0,k,y+k} + Q_k Q_\pi + Q_{y+k} Q_{z-k}$$
$$\gamma_3(k; x, y, z) = Q_\pi - Q_{y+k} - Q_{z-k} + Q_k \tag{B8}$$

and

$$\pi = (k - l)(1 - 2x) - m(1 + 2x) + 2y$$
$$\xi = k + \tfrac{1}{2}(l + m - z - y), \qquad \zeta = \tfrac{1}{2}(l + m + z - y) \tag{B9}$$
$$\nu = \zeta + x(k + m - l), \qquad \rho = k + y - x(k + m - l)$$

In Appendix C we show that the probabilities g_{ijk} and g_{ijkl} [see Eq. (A6)] are approximated remarkably well by

$$g_{ijk} \approx Q_{j-i} Q_{k-j}, \qquad g_{ijkl} \approx Q_{j-i} Q_{k-j} Q_{l-k} \tag{B10}$$

Introducing these approximations into Eq. (B6) results in negligible loss of accuracy while providing large savings in computation time.

Finally, when Eqs. (B5) and (B6) are used in conjunction with Eq. (9), we find that the relative variance of $\hat{D}$ can be expressed as $\epsilon^2 = A_1 \langle n \rangle^{-1} + A_2 + A_3 \langle n \rangle$, where the A_i are given as

$$A_i = \frac{1}{N} \frac{\sum_{m=1}^{M} \sum_{k=1}^{M} C_m C_l Q_m^{1/2} Q_l^{1/2} f_i(m, l)}{(\sum_{l=1}^{M} C_l^2 Q_l)^2} \tag{B11}$$

APPENDIX C. EXACT AND APPROXIMATE FORMS FOR g_{ijk} AND g_{ijkl}

The integrals g_{ijk} and g_{ijkl} given in Eq. (A3) are to be evaluated when calculating ϵ^2. It will be shown below that both multiple integrals can be reduced to single integrals requiring numerical evaluation. Since many of these integrals are needed for a single calculation of ϵ^2, we also produce an extremely simple approximation that has been used in our calculations, and show that errors in ϵ^2 due to inaccuracy of the approximation are generally less than 1%.

We first consider the reduction of g_{ijkl} to a simpler form. The function g_{ijkl} is given by

$$g_{ijkl} = \frac{abc}{\pi^3 X^2} \int \cdots \int_{\Omega} \exp[-a|\mathbf{r}_1 - \mathbf{r}_2|^2 - b|\mathbf{r}_2 - \mathbf{r}_3|^2 - c|\mathbf{r}_3 - \mathbf{r}_4|^2] \, d\mathbf{r}_1 \, d\mathbf{r}_2 \, d\mathbf{r}_3 \, d\mathbf{r}_4 \tag{C1}$$

where $a = (4Dt_{ji})^{-1}$, $b = (4Dt_{kj})^{-1}$, $c = (4Dt_{lk})^{-1}$, and the integration is over Ω for each r_i. The difficulty presented by Eq. (C1) is that the limits of integration are finite. To overcome this difficulty, we extend the limits of integration to infinity while multiplying the integrand by $H(\mathbf{r}_1)H(\mathbf{r}_2)H(\mathbf{r}_3)H(\mathbf{r}_4)$ where $H(\mathbf{r}) = 1$ for $\mathbf{r}$ in Ω and $H(\mathbf{r}) = 0$ for $\mathbf{r}$ outside Ω. Then, an application of Parseval's theorem leads to the form

$$g_{ijkl} = \frac{1}{8\pi^2 X^2} \int_{-\infty}^{\infty} \cdots \int \Gamma(\boldsymbol{\omega}_1)\Gamma(\boldsymbol{\omega}_2)\Gamma(\boldsymbol{\omega}_3)\Gamma(\boldsymbol{\omega}_4)\, \delta\left(\sum_{i=1}^{4} \boldsymbol{\omega}_i\right) \times \left[\exp\left(-\frac{\omega_1^2}{4a} - \frac{\omega_4^2}{4c}\right)\right]\left[\exp\left(-\frac{1}{4b}|\boldsymbol{\omega}_1 + \boldsymbol{\omega}_2|^2\right) + \exp\left(-\frac{1}{4b}|\boldsymbol{\omega}_3 + \boldsymbol{\omega}_4|^2\right)\right] d\boldsymbol{\omega}_1\, d\boldsymbol{\omega}_2\, d\boldsymbol{\omega}_3\, d\boldsymbol{\omega}_4 \tag{C2}$$

where

$$\Gamma(\boldsymbol{\omega}) = \frac{2}{\pi} \frac{\sin \omega_x R}{\omega_x} \frac{\sin \omega_y R}{\omega_y} \tag{C3}$$

in which $R = X/2$. A change of variables to $\boldsymbol{\eta}_1 = \boldsymbol{\omega}_1 + \boldsymbol{\omega}_2$ and $\boldsymbol{\eta}_2 = \boldsymbol{\omega}_3 + \boldsymbol{\omega}_4$ leads to the representation

$$g_{ijkl} = (1/4\pi^2 X^2) \int_{-\infty}^{\infty} \int I(\eta_x, a) I(\eta_y, a) I(\eta_x, c) I(\eta_y, c) \times \exp[-\eta^2/(4b)]\, d\boldsymbol{\eta} \tag{C4}$$

where

$$I(\eta, a) = \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\sin \omega R}{\omega} \frac{\sin(\eta - \omega)R}{\eta - \omega} \exp\left(-\frac{\omega^2}{4a}\right) d\omega \tag{C5}$$

This integral can also be put into a more convenient form by applying Parseval's theorem. One obtains

$$I(\eta, a) = \left(\frac{4a}{\pi}\right)^{1/2} \frac{1}{\eta} \int_0^X [\exp(-au^2)][\sin \eta R + \sin \eta(R - u)]\, du \tag{C6}$$

which can be substituted into the expression for g_{ijkl} in Eq. (C4). When this is done, the resulting integrals can be evaluated in a straightforward way and the final result can be expressed as

$$g_{ijkl} = (g_1 + g_2 + g_3 + g_4)^2 \tag{C7}$$

To define the g's we require the dimensionless lengths

$$\alpha = X\sqrt{a}, \qquad \beta = X\sqrt{b}, \qquad \gamma = X\sqrt{c}, \qquad \delta = (\alpha^2 + \gamma^2)^{1/2} \tag{C8}$$

The g's can be written in terms of these quantities as

$$g_1(\alpha, \beta, \gamma) = \tfrac{1}{4} \operatorname{erf} \alpha \operatorname{erf} \gamma \; Q^{1/2}(\beta)$$

$$g_2(\alpha, \beta, \gamma) = (\gamma/2\sqrt{\pi}) \operatorname{erf} \alpha \int_0^1 u Q^{1/2}(\beta u)\{\exp[-\gamma^2(1-u)^2] - \exp(-\gamma^2 u^2)\}\, du \tag{C9}$$

$$g_3(\alpha, \beta, \gamma) = g_2(\gamma, \beta, \alpha)$$

$$g_4(\alpha, \beta, \gamma) = (\alpha\gamma/2\delta\sqrt{\pi}) \int_0^1 u Q^{1/2}(\beta u) I(\alpha, \gamma, u)\, du$$

with

$$\begin{aligned} I(\alpha, \gamma, u) = {} & \left\{\exp\left[-\left(\frac{\alpha\gamma u}{\gamma}\right)^2\right]\right\}\left(\operatorname{erf}\frac{\alpha^2 u}{\delta} + \operatorname{erf}\frac{\gamma^2 u}{\delta}\right. \\ & - \operatorname{erf}\left(\delta - \frac{\gamma^2 u}{\delta}\right) - \operatorname{erf}\left(\delta - \frac{\alpha^2 u}{\delta}\right) \\ & + \left\{\exp -\left[\frac{\alpha\gamma}{\delta}(1+u)\right]^2\right\}\left(\operatorname{erf}\frac{\alpha^2 - \gamma^2 u}{\delta} + \operatorname{erf}\frac{\gamma^2 - \alpha^2 u}{\delta}\right)\Bigg) \\ & + \left\{\exp -\left[\frac{\alpha\gamma}{\delta}(1-u)\right]^2\right\}\left(\operatorname{erf}\frac{\alpha^2(1-u)}{\delta} + \operatorname{erf}\frac{\gamma^2(1-u)}{\delta}\right) \end{aligned} \tag{C10}$$

A similar derivation for g_{ijk} yields

$$g_{ijk} = \tfrac{1}{4}\left\{\int_0^1 (\operatorname{erf} \alpha u)\, [\operatorname{erf} \beta u + \operatorname{erf} \beta(1-u)]\, du\right\}^2 \tag{C11}$$

Since hundreds of these integrals may be needed in an evaluation of ϵ^2, it is clearly advantageous to have simple expressions for g_{ijk} and g_{ijkl}. Suitable approximations are suggested by examining the short-time behavior of the conditional probability density defined in Eq. (3);

$$\lim_{t \to 0} P(\mathbf{r}_2, t|\mathbf{r}_1, 0) = \delta(\mathbf{r}_2 - \mathbf{r}_1) \tag{C12}$$

The corresponding behavior of g_{ijk} is found to be

$$\lim_{t_{ji} \to 0} g_{ijk} = Q_{k-j}, \qquad \lim_{t_{kj} \to 0} g_{ijk} = Q_{j-i} \tag{C13}$$

The approximation

$$g_{ijk} \approx Q_{j-i} Q_{k-j} \tag{C14}$$

is exact in the limit of short times, and can be shown to be exact in the limit of long times. Similarly, one has in both these limits

$$g_{ijk} \approx Q_{j-i} Q_{k-j} Q_{l-k} \tag{C15}$$

Table I. Effect of Using Uncoupling Approximations for g_{ijk} and g_{ijkl}

N	y	M^*	$\langle n\rangle^*_{\text{approx}}$	$\langle n\rangle^*_{\text{exact}}$	ϵ^2(approx)	ϵ^2(exact)
25	4	15	1.18	1.18	15.9302	15.9287
25	1	4	1.19	1.19	6.1177	6.1192
50	2	6	2.38	2.39	3.6108	3.6134
100	1.25	4	3.16	3.16	1.2835	1.2839
100	2	6	3.57	3.58	1.7901	1.7910

We have made several test calculations of the approximations of Eqs. (C14) and (C15) for parameters typical of the leukocyte mobility experiments.[10] In this case α, β, and γ are found to be in the range 0.1–5.0. The approximations of Eqs. (C14) and (C15) always underestimate values of g_{ijk} and g_{ijkl}. The worst agreement in g_{ijk} is for $\alpha = \beta = 3.3$, where the error is 5.8%; for the g_{ijkl} it is 12.9%, for $\alpha = \beta = \gamma = 3.5$. The approximation continues to improve as these parameters increase.

The three- and four-time probabilities are not in themselves of great interest, but their combination in the calculation of ϵ^2 is. It appears that errors tend to cancel, leading to results typified by those in Table I. In no case are the errors in the approximate values of ϵ^2 or of $\langle n\rangle^*$ as large as 1%.

ACKNOWLEDGMENT

The authors would like to thank Nancy A. Crawford for her cheerful assistance in preparing the manuscript.

REFERENCES

1. T. Svedberg, *Z. Phys. Chem.* **77**:147 (1911).
2. A. Westgren, *Ark. Mat. Ast. Fys.* **11**:8, 14 (1916).
3. M. v. Smoluchowski, *Wien. Ber.* **123**:2381 (1914).
4. S. Chandrasekhar, *Rev. Mod. Phys.* **15**:1 (1943).
5. M. Kac in *Probability and Related Topics in Physical Sciences* (Interscience, New York), 1959, p. 132.
6. D. Magde, E. L. Elson, and W. W. Webb, *Biopolymers* **13**:29 (1974).
7. E. L. Elson and W. W. Webb, *Ann. Rev. Biophys. Bioeng.* **4**:311 (1975).
8. D. W. Schaefer, *Science* **180**:1293 (1973).
9. D. W. Schaefer and B. J. Berne, *Biophys. J.* **15**:785 (1975).
10. R. Nossal and Y. T. Chang, *J. Mechanochem. Cell Motil.* **3**:247 (1976).
11. E. Jakeman, E. R. Pike, and S. Swain, *J. Phys. A* **4**:517 (1971).
12. V. Degiorgio and J. B. Lastovka, *Phys. Rev. A* **4**:2033 (1971).
13. B. E. A. Saleh and M. F. Cardoso, *J. Phys. A* **6**:1897 (1973).
14. D. E. Koppel, *Phys. Rev. A* **10**:1938 (1974).
15. M. H. Gail and C. W. Boone, *Biophys. J.* **10**:980 (1970).
16. M. J. Box, *Biometrika* **58**:149 (1971).
17. E. D. Becker, J. A. Ferretti, R. K. Gupta, and G. H. Weiss, *J. Mag. Res.* (to appear).