Biological Applications of the Lagrangian Poisson Distribution
Author(s): Konanur G. Janardan, Harold W. Kerster and David J. Schaeffer
Source: *BioScience*, Vol. 29, No. 10 (Oct., 1979), pp. 599-602
Published by: Oxford University Press on behalf of the American Institute of Biological Sciences
Stable URL: http://www.jstor.org/stable/1307766
Accessed: 03-09-2017 19:05 UTC

Biological Applications of the Lagrangian Poisson Distribution

Konanur G. Janardan, Harold W. Kerster, and David J. Schaeffer

Biology has frequently benefited from the introduction of simple mathematical models. Conspicuous examples are the Hardy-Weinberg law, logistic growth equation, the fundamental theorem of natural selection, and Cole's demonstration of the power of the simple (Cole 1946a) and compound (Cole 1946b) Poisson distributions. This paper describes the utility of the Lagrangian Poisson distribution as a tool for understanding biological processes.

The Poisson distribution has been used in a wide variety of situations to describe the behavior of biological systems. Common examples are the distribution of hemocytometer counts, decay of radioactive isotopes, and the distributions of plankton, animals under cover, and pests on leaves. The Poisson distribution is generated by processes in which a large number of intervals (hemocytometer squares, minutes of radioactivity counting time, etc.) are hit by a relatively small number of events (red blood cells, particles from nuclear decay). The occurrence or non-occurrence of a hit in an interval has no effect on the further occurrence or non-occurrence of hits in that interval; e.g., a counter minute with lots of counts is as likely to get another count as is a minute with fewer counts. Much of the value of the Poisson distribution arises from the powerful insights it provides into the sophistication of biological systems. For example, the failure of pest eggs on a plant leaf to be distributed in a Poisson frequency suggests purposeful behavior by the pest to concentrate or disperse, as the case may be, its eggs to its advantage.

There are many biologically important cases where the simple Poisson assumptions are not satisfactory. For example, if the number of plant leaves available to an ovipositing pest is severely limited, then the assumption of a large number of intervals is not satisfied and the investigator lacking tools more sophisticated than the simple Poisson is denied analytical power. The Lagrangian Poisson is a member of a class of discrete distributions; it is nearly as easy to use and is vastly more powerful than the Poisson itself. This paper presents procedures for employment of these more powerful techniques. We work through a number of examples and show further examples more briefly to demonstrate the great breadth of application of these techniques to biological research.

SIMPLE POISSON DISTRIBUTION

Cole's Arthropoda Data

We begin our demonstration with a portion of the classic set of data given by Cole (1946a). Cole scattered boards on a forest floor and periodically counted the arthropod forms under them. We may ask: "If 102 spiders (C) are to be distributed at random among 240 pieces of cover (B), how many of the pieces of cover would we expect to find empty, how many with one spider, how many with two, . . . *n* spiders?" The Poisson analysis of this question proceeds as follows.

Assume the boards are equally attractive trap sites (or intervals) for spiders. If spiders arrive at and stay under boards without respect to the number (if any) of spiders already under a board, then the spiders can be said to be behaving Poisson random. A simple expression generates the expected numbers of boards housing different numbers of spiders:

$$N_k = Ne^{-g}g^k/k!\ (k = 0, 1, 2 \ldots)$$

where e is the base of natural logs, k is the frequency class, N is the total number of spiders, and g is the average number of spiders per board. The number of boards housing no spiders (N_0) is estimated by the equation:

$$\hat{N}_0 = (240)(2.718)^{-(102/240)}\ (102/240)^0/0!$$
$$= (240)(2.718)^{-.425} = 157.2$$

and three spiders by:

$$\hat{N}_3 = Ne^{-g}g^3/3!$$
$$= (240)(0.425)(2.718)^{-0.425}/((3)(2)(1)) = 2.0$$

where g has been replaced by its estimate $\hat{g} = 0.425$.

The behavior of real spiders (Table 1, row 2) compared with the mythical random spiders in row 3 indicates that spiders seem indifferent to each other's presence. We draw this conclusion because their mass behavior, as observed, is well-described by a mathematical model (Poisson distribution), which is based on an assumption of indifference of spiders to each other.

Cole also reported the distribution of sowbugs under boards (Table 2, row 2). This distribution shows an inflation of the "zero" class (and a shortage in the "one" class) relative to the Poisson expectation of row 3. The biological interpretation of these findings is that one does not find lone sowbugs; they are social. The inflation of the "zero" class is brought about by sowbugs leaving "one" class boards for more social conditions. This removes boards from the "one" class, and places them in the "zero" class, simultaneously inflating the social classes (hypodispersion or contagious distribution).

In this example, the Poisson distribution very poorly fits the behavior of real sowbugs. However, the poor fit reveals that sowbugs are social, at least in com-

Janardan is associate professor of mathematics and chairman, Math Systems, Sangamon State University, Springfield, IL 62708; Kerster is associate professor of environmental studies, Environmental Studies Center, California State University, Sacramento, CA 95819; and Schaeffer is science adviser, Division of Water Pollution Control, Illinois Environmental Protection Agency, 2200 Churchill Road, Springfield, IL 62706.

parison to the spiders or hypothetical Poisson sowbugs shown in Table 2. Thus, the astute biologist can gain valuable insights into the behavior of biological systems from statistical models that fail.

Mitchell's Weevil Experiments

Having examined a nonsocial case (spider) and a social case (sowbug), we turn our attention to an antisocial case (bean weevil). Mitchell (1975) studied oviposition tactics in laboratory cultures of bean weevils. Female weevils deposit single eggs on mung beans, which thus accumulate zero to three eggs; larval survival decreases as egg number increases. Beans vary in size, and large beans offer a greater certainty of larval success than do smaller ones.

Egg numbers on beans (Table 3, row 2) are not Poisson distributed (row 3). Instead, the eggs are hyperdispersed, and Mitchell showed that the hyperdispersion enhances larval survival. Although the mechanisms are unclear, increasing the number of eggs on beans drastically reduces the survival chances for all of the eggs on a bean.

Such behavior suggests that an ideal weevil would distribute eggs in a fashion such that no bean has more than one egg deposited on it. However, each weevil is physically capable of examining only a portion of the total beans available, and if there are many weevils all seeking suitable sites, there is competition between them for such sites. In the early stages of oviposition, weevils finding an otherwise suitable site already occupied by an egg might seek an un-egged site that is also suitable. As suitable sites become contaminated with eggs, the competition among the weevils would require their choosing between sites with already deposited eggs and any un-egged, less desirable, beans that remain. This competition results in overdispersion of eggs relative to a Poisson process (Table 3).

A slight increase in the complexity of our model permits a quantitative description of the behavior of weevils (and sowbugs) and improves biological insight. At this point, all we know about them is that they do not distribute Poisson. Cole (1946b) approached this problem by suggesting a model in which sowbugs move 1, 2, 3 . . . n at a time. If the movements of these groups are each independent Poissons, their sum is a compound Poisson. (For a lucid explanation of this model and a treatment of the sowbug case, see Cole 1946b. The compound Poisson is suitable when event sizes vary.) Another approach with great analytical power is the Lagrangian Poisson, which Janardan and Schaeffer (1977) developed as a model to explain the statistical thermodynamics of the induction and restitution processes associated with the production of chromosome aberrations by inducing agents.

LAGRANGIAN POISSON DISTRIBUTION

Mathematical Model

The Poisson process which we considered first is that governing the dispersal of weevils to beans as if there is no difference between the attractiveness of egged and un-egged beans. However, because beans with eggs may be more or less attractive than beans without eggs, this competition can be modeled as another Poisson. Then the overall process, obtained by Janardan and Schaeffer (1977) as the Lagrangian transformation of two Poissons, leads to the Lagrangian Poisson distribution (LPD):

$$N_k = N\, g_1 (g_1 + g_2 k)^{k-1} e^{-(g_1 + g_2 k)}/k!$$

When $g_2 = 0$, the Lagrangian Poisson is the same as the simple Poisson. The terms in this expression are defined as they are used to work the examples below.

Bean Weevil Experiment

If we reconsider the weevils, g_1 is the rate of the Poisson process affecting movement of weevils to beans. Similarly, g_2 is a complex function of the competition rate, which is easy to estimate. In this process, g_1 (in the Poisson process, g) is expressed as events per unit time, i.e., organism arrivals per board, eggs per bean, etc. In the present case, g_1 and g_2 can be estimated from the mean and variance of the data. The sample

TABLE 1. The distribution of 102 spiders under 240 boards.

Calculation	Number of spiders per board			
1. Assumed	0	1	2	3
2. Observed frequency	159	64	13	4
3. Poisson expected	157.2	66.5	14.2	2.0

TABLE 2. The distribution of sowbugs (*Trachelipus rathkei*).*

Calculation	Number per board								
1. Assumed	0	1	2	3	4	5	6	7	8
2. Observed frequency	28	28	14	11	8	11	2	3	3
3. Simple Poisson	5.7	17.5	26.8	27.3	20.9	12.8	6.5	2.8	1.1
4. Lagrangian Poisson	26.1	23.6	18.1	13.4	9.9	7.3	5.5	4.1	3.1
1. Assumed	9	10	11	12	13	14	15	16	17
2. Observed frequency	3	3	2	0	1	2	1	0	2
3. Simple Poisson	0.6	0.2	0.1	0	0	0	0	0	0
4. Lagrangian Poisson	2.4	1.8	1.4	1.1	0.9	0.7	0.5	0.4	0.3

*N = 122, $\bar{X}$ = 3.2951, g_1 = 1.5416, g_2 = 0.5321, b = 2.9.

TABLE 3. The distribution of eggs per bean for oviposition of the bean weevil *Callosobrachus maculatus* on mungbeans.

Calculation	Number of eggs per bean			
1. Assumed	0	1	2	3
CASE A*				
2. Observed frequency	138	46	—	—
3. Simple Poisson	143.3	35.8	4.5	0.4
4. Lagrangian Poisson	138	46	—	—
CASE B†				
2. Observed frequency	26	117	35	1
3. Simple Poisson	61.9	65.7	34.9	12.3
4. Lagrangian Poisson	29	107.8	43.1	7.8
CASE C‡				
2. Observed frequency	5	68	88	32
3. Simple Poisson	33.1	58.4	51.4	30.2
4. Lagrangian Poisson	8.7	57.6	97.4	29.5

*$\bar{X}$ = 0.25 eggs per bean N = 184, g_1 = 0.2878, g_2 = −0.1515, b = −1.9.
†$\bar{X}$ = 1.06 eggs per bean N = 179, g_1 = 1.8191, g_2 = −0.7138, b = −2.55.
‡$\bar{X}$ = 1.76 eggs per bean N = 193, g_1 = 3.1027, g_2 = −0.7612, b = −3.1.

variance, which is the square of the sample standard deviation, is calculated from the sums of the observations (ΣX) and the sums of the squares of the observations (ΣX^2), as follows:

$$\text{Variance} = (\Sigma X^2 - (\Sigma X)^2/N)/(N - 1)$$

Then, the estimates of g_1 and g_2 are (Janardan and Schaeffer 1977):

$$\hat{g}_2 = 1 - \hat{D}^{-0.5}$$
$$\hat{g}_1 = (\text{mean})\ (1 - \hat{g}_2)$$

where $\hat{D}$, the estimated dispersion, is given by the ratio of the sample variance to the sample mean.

In terms of the Lagrangian Poisson, the rate of the initial (Poisson) process—attraction of weevils to beans—is given by g_1, and the net effect of dispersion—competition of weevils for more attractive sites—is given by g_2. The first, g_1, has units of eggs deposited per bean, and g_2 is dimensionless. Table 3 (row 4) shows that the Lagrangian Poisson is a useful model of this process in the first two data sets, where both attraction and repulsion are important.

Given the biological assumptions used in setting up this model, what biological insights derive from it? We used competing Poisson processes to develop the Lagrangian Poisson. The rate of the initial process—the dispersion and the interaction of weevils with beans—was given by a Poisson process with a rate g_1. The competing process, due to the differing attractiveness of egged and un-egged beans as sites suitable for oviposition, was similarly specified, with a rate, b. This new constant, b, then, is the true parameter describing the competing process, and g_2 is then given by the ratio of g_1/b.

The examples so far discussed reveal that a negative value of g_2 has the effect of increasing the dispersion of the Poisson process. A positive value of g_2 has the opposite effect, which implies that the competing process is acting in the opposite direction (retarding the dispersion) to the Poisson process. A consideration of the Poisson process shows that g_1 must always be positive, but that b, the rate of the competing process, can either be negative or positive. This leads to hypo- or hyperdispersion, respectively. Thus, negative values of g_2 can arise only from negative values of b, that is, from processes that act in the same direction as the initial Poisson dispersal mechanism. We can infer, therefore, that the competition of weevils for good oviposition sites results in the dispersion of eggs to a greater extent than would occur by a Poisson process operating alone.

On the other hand, the value of g_1 is a measure of the desirability of un-egged beans to weevils, given an initial (hypothetical) situation in which the weevils are allowed to deposit eggs freely on a collection of beans none of which contains eggs at the start of the process. (This situation corresponds to the way in which this experiment was actually run.) Then, since g_2 is the ratio of g_1 to b—that is, of the "attractiveness" of un-egged beans relative to egged beans—the magnitude of g_2 is a measure of the relative effect of these types of sites on the dispersal process.

The effect of egged sites is to make un-egged ones more attractive to weevils, relative to the attractiveness of un-egged sites at the start of the experiment when no sites were egged. This conclusion, drawn from the statistical model and its results alone, is consistent with Mitchell's (1975) conclusion that ". . . marked departures from the Poisson will occur only if the beetles search out and oviposit on beans with no eggs on them and reject beans that already carry eggs." This analysis provides a quantitative estimate (g_2) of the relative importance of random and purposive factors in weevil egg distribution.

Fertilization of Sea Urchin Eggs

Cases in which initial occupation of an interval reduces the attractiveness of the interval are widespread in biology. Fertilization is a good example. After some small number of sperm (often one) have entered an egg, the egg becomes resistant to the entrance of further sperm. Table 4 shows that sea urchin eggs (Morgan 1975) resist penetration by more than one sperm; the Poisson fit is very poor.

A new biological insight is provided by quantification of the resistance to penetration of sperm following the first. Because g_2 is a measure of the process competing with the Poisson distribution of sperm in eggs, here (as in the weevil case) g_2 is negative—an indication that the form of the competition process is a race, i.e., both processes tend to disperse sperm on eggs.

The Case D (Table 4) outcome of 180 seconds of exposure of eggs to sperm is very poorly fitted by both the simple and Lagrangian Poisson. The sperm/egg system is saturated; only two eggs remain in the most receptive class. Failure to fit occurs because the initial competition between processes ended with the exhaustion of the class (empty eggs) on which the Poisson process depended. Sometime before 180 seconds the overall process has become simple, not Lagrangian, but starting with eggs "salted" with sperm to an unknown extent. Such saturation can occur only when g_2 is negative.

A fortunate consequence of the 180-second analysis is the knowledge that process competition ended before data were recorded. Values of g_2 from shorter experiments should form a series becoming increasingly negative with time. We obtain an asymptotic approach to the equilibrium balance between competing processes ($g_2 \geq 0$) when the trend line (Fig. 1A) becomes nearly horizontal. The asymptote occurs at roughly 8 seconds ($g_2 = 0$).

TABLE 4. Distribution of sperm per egg, sea urchins.

Calculation	Number of sperm per egg				
1. Assumed	0	1	2	3	4
CASE A*					
2. Observed frequency	89	11	—	—	—
3. Simple Poisson	89.6	9.9	0.5	—	—
4. Lagrangian Poisson	89	10.9	—	—	—
CASE B†					
2. Observed frequency	42	36	6	—	—
3. Simple Poisson	47.4	27.1	7.7	1.5	0.4
4. Lagrangian Poisson	42.2	35.7	6.1	—	—
CASE C‡					
2. Observed frequency	28	44	7	1	0
3. Simple Poisson	37.3	28.5	10.8	2.8	1.1
4. Lagrangian Poisson	29.2	40.6	10.2	—	—
CASE D§					
2. Observed frequency	2	81	15	1	1
3. Simple Poisson	30.7	36.3	15.4	8.4	5
4. Lagrangian Poisson	8.5	62.3	26.5	59.1	1

*After 5 seconds N = 100, $\bar{X}$ = 0.11, g = 0.116, g_2 = −0.0546, b = −2.12.
†After 15 seconds N = 84, $\bar{X}$ = 0.5714, g_1 = 0.6895, g_2 = −0.2067, b = −3.34.
‡After 40 seconds N = 80, $\bar{X}$ = 0.7625, g_1 = 1.0077, g_2 = −0.3216, b = −3.13.
§After 180 seconds N = 180, $\bar{X}$ = 1.18, g_1 = 2.4654, g_2 = −1.0893, b = −2.26.

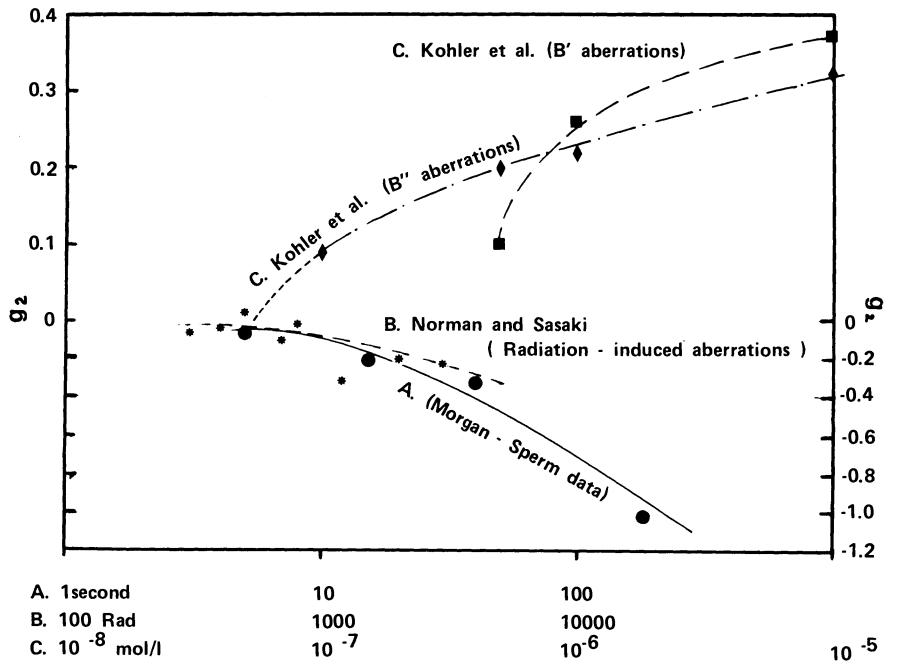

Fig. 1. Dose-response relationships.

Producing Chromosome Aberrations

The Lagrangian Poisson was originally developed to explain the production of chromosome aberrations by a variety of induction mechanisms. In generating this model, Janardan and Schaeffer (1977) considered the induction process as a group of chromosomes lined up in a Poisson queue waiting to undergo transformation. The aberrated chromosomes were also considered to be Poisson queued while awaiting restitution. At the end of the experiment, what the observer saw was the net production of aberrations from the competition between the two queues. The induction process arose from the radiation or chemical insult used to initiate the chromosome damage process.

The restitution process also has a clear biological meaning as the outcome of the molecular processes of excision, insertion, and fusion (the DNA repair processes). The meaning of g_2 is clear: At equilibrium, the rate at which new aberrations are produced is exactly balanced by the rate at which aberrations are repaired. In this situation, the final counts are the net between the production and repair of aberrations. Prior to equilibrium, g_2 can be either positive or negative. A negative value can occur when the energy of the induction process or the rate of induction is high, so that the repair process actually has the effect of producing more broken chromosomes; that is, the repair process has been converted into an ancillary production process.

The radiation-induced aberrations (in Table 2 of Janardan and Schaeffer 1977) show this phenomenon. This is seen more clearly in Fig. 1B, where Norman and Sasaki's (1966) data on the production of dicentrics (as measured by g_2) are plotted against radiation dose. If the assumption of ancillary induction through the repair mechanism is right, g_2 would become increasingly negative as the radiation dose increases—a trend shown in Fig. 1.

Of far greater moment to biologists, especially those concerned with damage to biological systems, are the Kohler et al. (1976) data presented in Fig. 1C. The plots of concentration of the inductant are curvilinear against positive values of g_2 for these types of chromosome breaks. The positive values of g_2 signify that the repair process is acting to restore breaks, as opposed to the previous example. The difference in signs of g_2 is explainable in terms of the greater energy available from radiation than from chemical induction. The former may represent nonequilibrium thermodynamic processes affecting the chromosome structure. The chemically induced processes may have occurred at or near equilibrium.

We account for the curvilinear nature of the data in Fig. 1C by inferring a threshold value of the mutagen at a level of 10^{-7} molar; the rate of production at lower levels of the inducing agent will be so low as to make the effective value of g_2 zero. That is, the expected rate becomes that of a Poisson process in the absence of any (or of a significant) inducing agent. The Lagrangian model has thus led us to the middle of the cancer/mutation threshold question!

Referring to Fig. 1C for the production of B" aberrations (isochromatid breaks), Kohler et al's. data for a concentration of inductant of 10^{-6} molar result in a distribution of aberrations for which the calculated g_2 is 0.24 after one hour and 0.22 after 24 hours. These numbers are essentially the same, which suggests that the ratio of the rates of induction to repair has stabilized by the end of the first hour. The system appears to be in equilibrium (this is the first example of equilibrium). Without our going into details of the thermodynamics, the standard free energy of a reaction at equilibrium can be calculated from the equilibrium constant (here given by g_2) as $-RT\ln g_2$. Inserting the average value, 0.23, we estimate that the free energy at 37.5° C for this process is about +900 cal/mole/aberration (4.5 kcal/mole). The positive sign of this estimate indicates that the production of aberrations is not spontaneous.

Our belief in this estimate, and indeed in the process of model development and induction that led to the estimate, coincides with the biological knowledge that the production of aberrations is an uncommon event in the absence of induction. To the best of our knowledge, this constitutes the first estimate of the net energy required for chromosome rupture.

REFERENCES CITED

Cole, L. C. 1946a. A study of the cryptozoa of an Illinois woodland. *Ecol. Monogr.* 16(a): 50-86.

———. 1946b. A theory for analyzing contagiously distributed populations. *Ecology* 27(4): 329-341.

Janardan, K. G., and D. J. Schaeffer. 1977. Models for the analysis of chromosomal aberrations in human leukocytes. *Biomet. J.* 19(8): 599-612.

Kohler, W., V. Loeschcke, and G. Obe, 1976. Analysis of intercellular distributions of chromatid aberrations. *Mutat. Res.* 34: 427-436.

Mitchell, R. 1975. The evolution and oviposition tactics in the bean weevil, *Callobruchus maculatus* (F.). *Ecology* 56(6): 696-702.

Morgan, R. W. 1975. Some stochastic models to describe the fertilization of an egg. *Appl. Stat.* 24(1): 137-139.

Norman, A., and M. S. Sasaki. 1966. Chromosome exchange aberrations in human lymphocytes. *Int. J. Rad. Biol.* 11: 321-328.