Theory and Application of the Logistic Probability Density Function as a Population Growth Model

J. H. Matis[1] and M. J. Al-Muhammed[2]

[1] Department of Statistics, Texas A&M University, USA
[2] Department of Mathematics, Damascus University, Syria

Received 15/09/2009
Accepted 29/03/2010

ABSTRACT

The well-known Verhulst-Pearl model in ecology, $N' = (\lambda - \delta \cdot N(t)) \cdot N(t)$ where N(t) denotes current population size, has a solution which may be written in the form of a logistic cumulative distribution function (cdf). It is widely used to describe population growth curves (Renshaw, 1991). An alternative model $N' = (\lambda - \delta \cdot F(t)) \cdot N(t)$ where F(t) is the integral of N(s) from 0 to t, was recently proposed by Kindlmann (1985) and solved analytically by Prajneshu (1998). The solution to this model was written in the form of a logistic probability density function (pdf) by Matis *et al.* (2009). The model has been previously fitted only to aphid data. We illustrate this pdf solution by fitting it to data on the gypsy moth (*Lymantria dispar*), a harmful insect which defoliates forests, from Latakia, Syria. The logistic pdf solution fits this gypsy moth data very well, which provides a mechanism for the statistical analysis of moth count data. Consequently, effective control strategies for gypsy moths can be developed with the objective of avoiding forest defoliation in Latakia. This successful fitting also suggests investigations using the model to describe population growth curves of other insect species.

Keywords: Logistic Growth Model, Cumulative Growth Model, Logistic Probability Density Function, Gypsy Moth.

نظرية دالة التوزيع الاحتمالي اللوجستي وتطبيقها كنموذج لوصف تكاثر مجتمع

جيمس هنري ماتيس[1] و محمد جاسم المحمد[2]

[1] قسم الاحصاء ــ جامعة A&M في تكساس ــ الولايات المتحدة الامريكية
[2] قسم الرياضيات ــ كلية العلوم ــ جامعة دمشق ــ سورية.

تاريـــخ الإيداع 2009/09/15
قبل للنشر في 2010/03/29

الملخص

نماذج النمو اللوجستي هي تقنيات رياضية وإحصائية تستخدم لوصف نمو المجتمعات. بشكل خـــاص، تستخدم هذه النماذج لوصف تكاثر مجتمعات الحشرات ودورة حياتها في حالات يكون فيها هذا التكاثر مضطرداً، ويبلغ قيمة توازن تسمى "قدرة الاستيعاب carrying capacity" (أي العدد الأعظمي من الحشرات التي يمكن أن تعيش معاً في المكان نفسه في لحظةٍ ما). يوفر النموذج اللوجستي تقنيات مهمة جداً للعاملين في مجال حماية النباتات للتنبؤ بعدد الحشرات في أية لحظة زمنية، ومن ثمَ تحضير وسائل مكافحة فعالة للقضاء على هذه الحشرات التي تدمر أنواعاً مختلفة من المحاصيل والنباتات. على الرغم من فعالية النماذج اللوجستية في وصف نمو المجتمعات فإن لها عيوبها. بالتحديد هذه النماذج غير قادرة على إظهار القيم الشاذة و غير المعتادة في نمو المجتمع. وهذه القيم الشاذة قد يكون لها أهمية علمية جديرة بالملاحظة.

للتغلب على هذه العيوب، اِقتُرح نموذج لوجستي آخر يدعى: دالة الكثافة الاحتمالية اللوجستية Logistic probability density function. هذا النموذج أفضل من سابقه كونه يستطيع إظهار، وبشكل واضح، القيم الشاذة أو غير المعتادة في نمو المجتمع. طُبّق النموذج اللوجستي الجديد كسابقه لوصف تكاثر حشرة تسمى Aphid . تعدُّ هذه الحشرة واحدة من أكثر الحشرات تدميراً للمحاصيل الزراعية وهي منتشرة في كل أنحاء العالم بما فيها سورية. ولكن هذا النموذج الجديد لم يتم تطبيقه لوصف تكاثر أي نوع آخر من الحشرات من قبل. ينفرد هذا البحث وللمرة الأولى في توسيع تطبيق النموذج الاحتمالي اللوجستي على نوع آخر من الحشرات في سورية يدعى: عثة الغجر. عثة الغجر هي الحشرة الأكثر تدميراً على المستوى العالمي، وفي سورية إذ تقوم بتجريد الأشجار من أوراقها، ومن ثم تؤدي إلى موت الأشجار وتدمير الغابات. يمكن تلخيص الاسهامات العلمية لهذا البحث بالنقاط الآتية:

1. للمرة الأولى يتم توسيع تطبيق النموذج الاحتمالي اللوجستي على نوع آخر من الحشرات و بالتحديد على عثة الغجر.
2. للمرة الأولى يجري تطبيق النموذج الاحتمالي اللوجستي لوصف تكاثر مجتمع الحشرات في سورية (بالتحديد عثة الغجر). تعدُّ هذه الإسهامة مهمة جداً لأنها تمكننا و للمرة الأولى في سورية من استخدام آلية رياضية إحصائية (النموذج الاحتمالي اللوجستي) لإجراء تحليل إحصائي لتعداد (تكاثر) عثة الغجر ومن ثم تحضير طرائق فعالة لضبط هذه الآفة الضارة ومكافحتها.

الكلمات المفتاحية: نموذج النمو اللوجستي، نموذج نمو تراكمي، دالة الكثافة الإحتمالية اللوجسية، عثة الغجر (الجادوم).

Introduction

Logistic cumulative growth models are well-known for describing population growth in ecology. Specifically, they are widely used to describe insect populations and their life cycles in cases where the population increases monotonically to some equilibrium value, called the ‘carrying capacity.’ This provides people working in plant protection with powerful techniques to predict insect counts and therefore to prepare effective control strategies to deal with problems caused by harmful insects on different types of crops. Although it has a remarkable descriptive power, this model has shortcomings. Namely, the logistic cumulative distribution function (**cdf**) model is not sensitive to outlying or unusual observations, which may be of scientific interest.

A newly proposed model is the logistic probability density function (or the **logistic pdf** for short). The model describes a common type of growth curve in which a population rises to some maximum value and then declines rapidly. The logistic pdf model has been successfully used previously to describe aphid populations. The aphid family is the leading agricultural pest in the world, and it is also a problem in **Syria**. However, the logistic pdf has never before been applied to other species. This paper shows a new successful application of the logistic pdf model to describe another insect, namely the gypsy moth (*Lymantria dispar*). The gypsy moth is the most destructive insect in forests worldwide, and it also defoliates forests in **Latakia area**, **Syria**.

To this end, the contribution of this paper is twofold. First, it shows, ***for the first time***, the application of the logistic pdf to an insect species other than the recent applications to aphids. Second, it is ***the first time*** that the logistic pdf model is used to describe an insect population in **Syria**. The latter contribution is very important because it provides a mechanism (the logistic pdf model) for analyzing gypsy moth count data and preparing a foundation for effective control strategies to deal with this insect.

We present our contributions as follows. Section 2 briefly presents the logistic cdf model. Section 3 introduces the new logistic pdf model and gives some of its useful properties. Section 4 demonstrates the application of logistic pdf model to gypsy moth data from **Latakia area**, **Syria**. Section 5 concludes the paper and gives directions for future work.

The Standard Logistic Growth Model

The logistic growth curve has a celebrated history, and is still in widespread use today in ecological theory as well as in math education (Renshaw, 1991). This curve was first suggested by Verhulst in 1838, and derived independently by Pearl and Reed in 1920. Let N (t) denote population size at time t, and N'(t) its derivative. The Verhulst-Pearl model is

$$N'(t) = \lambda N \cdot (1 - \frac{N}{K}) \qquad (1)$$

with parameters $\lambda>0$ and K>N(0). The simple solution to (1) may be written as

$$N(t) = \frac{K}{1 + e^{(c - \lambda t)}} \qquad (2)$$

The parameter c is related to the initial value N(0). Parameters λ, called the 'intrinsic growth rate', and K, called the 'carrying capacity', are key descriptors of population dynamics in ecology (Renshaw, 1991). For subsequent convenience, we rewrite (2) as

$$N(t) = \frac{K}{1 + e^{-\lambda (t - t_{\max})}} \qquad (3)$$

with new parameter $t_{max} = c / \lambda$.

Equations (2) and (3) with K = 1 have the form of a cumulative distribution function, or 'cdf', in statistics, as they are positive functions with N (-∞) = 0 and N (∞) = 1. This is curious, as the previous derivation has nothing to do with a random variable, but is instead the solution to a differential equation with no randomness. This distribution is called the logistic cdf in statistics. The fact that (2) and (3) have the form of a logistic cdf is very helpful, as many properties of the logistic cdf are well-known in statistics. Due to this correspondence, we call the previous logistic growth model in (2) and (3) the *logistic cdf* model.

The Verhulst-Pearl model in (1) is often reparameterized as

$$N'(t) = (\lambda - \delta N(t)) \cdot N(t) \qquad (4)$$

where $\delta = \lambda / K$. In a simple mechanistic interpretation of model (4), intrinsic growth rate λ is interpreted as the *per capita* birth rate of the population. The corresponding *per capita* death rate is δN, where δ is a death rate coefficient. The model is called 'density dependent', as the death rate is a function of current size, N. There are numerous generalizations of models (1) and (4). One in common usage is the "power-law" logistic in which N (t) in (4) is raised to a power (Banks, 1994).

The New Logistic Growth Model, the Logistic pdf

We consider now the derivation of this different logistic model. Its mechanistic formulation was first proposed in mathematical terms by Kindlmann (1985) to describe insect populations, and is given as follows:

$$N'(t) = (\lambda - \delta F(t)) \cdot N(t) \qquad (5)$$

where

$$F(t) = \int_0^t N(s)\; ds \qquad (6)$$

F(t) denotes the past population size, or 'cumulative density', since the initial time (t = 0). In simple terms, model (5) changes the *per capita* death rate in (4) from δ N (t) to δ F (t). Prajneshu (1998) derived the first analytical solution to the differential equation in (5), a nontrivial feat, in the form:

$$N(t) = \frac{ae^{-bt}}{(1+de^{-bt})^2} \qquad (7)$$

where *a, b*, and *d* are positive parameters. Matis et al. (2007) shows that the solution to (5) can also be written using stable parameters, which simplify fitting the model to data, as:

$$N(t) = \frac{4N_{\max} \cdot e^{-b(t-t_{\max})}}{(1+e^{-b(t-t_{\max})})^2} \qquad (8)$$

The parameters in this model are naturally interpretable, as parameter N_{max} denotes the maximum size of N (t) in (8) and t_{max} the time of this maximum. Parameter b is a relative rate defined subsequently. The biological meaning of the parameters in equation (8) facilitates the choice of initial parameters in iterative nonlinear estimation procedures. Of course, they also make it more user-friendly for biological investigators.

Some useful facts about model (8) follow:

1. This solution may be written more parsimoniously using the hyberbolic secant (sech) function as

$$N(t) = N_{\max} \cdot \operatorname{sech}^2\left(\frac{b(t - t_{\max})}{2}\right) \qquad (9)$$

Hyperbolic functions are defined in calculus courses, but seldom used in practice. A novelty about (9) is that we will later fit "real-world" data to this squared hyperbolic function.

2. Once the observed data are fitted to model (8), the estimates of N_{max}, t_{max}, and b may be used to estimate the parameters of underlying model (5). We show in Matis et al. (2007) that

$$\lambda = b \cdot \frac{d-1}{d+1}$$
$$\delta = \frac{b^2}{2N_{\max}} \qquad (10)$$
$$N(0) = \frac{4d \cdot N_{\max}}{(1+d)^2}$$

Where:

$$d = e^{b t_{\max}} \qquad (11)$$

Parameter d is typically very large, hence it follows from (10) that parameter b is an accurate approximation for the *per capita* birth rate λ.

3. It is also clear that solution (8) could be written as

$$N(t) = K \cdot p(t) \qquad (12)$$

where p(t) is the logistic probability density function, or 'pdf', defined as

$$p(t) = \frac{b \cdot e^{-b(t - t_{\max})}}{(1 + e^{-b(t - t_{\max})})^2} \qquad (13)$$

and K is the constant

$$K = 4N_{\max} / b \qquad (14)$$

One can show that K is the total area under the N (t) curve, also denoted as AUC, by integrating (12) over the full range of t, noting that p (t) is a pdf integrating to 1. The AUC is an important descriptor in ecology. The logistic pdf (also called the sech-squared pdf due to (9)) is also well-known in statistics. It is symmetric with heavier tails than a Normal pdf with the same mean and variance (Johnson and Kotz, 1970).

We call new model (8) the logistic pdf model, due to (12). It seems curious again that mechanistic model (8) also, which is devoid of any randomness, would have as its solution a scaled form of a well-known probability distribution in statistics. We note, however, that one can formulate a stochastic analog to model (5), in which the current count N(t) is a random variable. As a contrast to 'deterministic' model (8), which is the solution to a single differential equation, Matis *et al.* (2005) presents an example of an 'exact' solution to a stochastic analog of (5). This 'exact' solution is based on solving a system of over 180,000 (Kolmogorov) differential equations.

Application of the New Logistic pdf Model to Gypsy Moth Data

Both Kindlmann, the Czech ecologist who first formulated (5), and Prajneshu, the Indian mathematician who obtained solution (7), sought specifically to describe aphid population growth. Theoretical reasons why the aphid family satisfies the mechanistic assumptions in (5) are reviewed in Matis et al. (2007). This model formulation has also been verified empirically by fitting model (8) successfully to a number of aphid species, as reviewed in Matis et al. (2009). These past studies demonstrate the versatility of the model to describe observed aphid abundance curves, whether the peak count is high or low, the spread narrow or wide, or the time of maximum short or long.

To our knowledge, however, model (8) has never before been fitted in a published study to a species other than the aphid. Yet it would not be surprising if this versatile logistic pdf fitted successfully population size data from other species as well, even though their population dynamics might be completely different from those applicable to the aphid in model (8).

We illustrate the logistic pdf and at the same time explore the hypothesis that it might fit other insect species using gypsy moth data (Al alouni, 2009). Gypsy moths defoliate forests in **Syria**, and studies are conducted to control them. The observed data are given in Table 1 and are illustrated graphically in Figure 1 as small solid circles. The data consist of the average number of gypsy moths, denoted Y (t), caught in four pheromone traps over time *t* (in days) near **Latakia** (in **Slunfeh**).

Table 1: Average Moth Counts

Elapsed Time (days)	Count
0	0.4
6	10.6
11	101.2
19	467.6
26	756.8
33	309.6
40	18.4
47	24.5
55	9.6
62	4.8
68	2
74	0.8
78	0

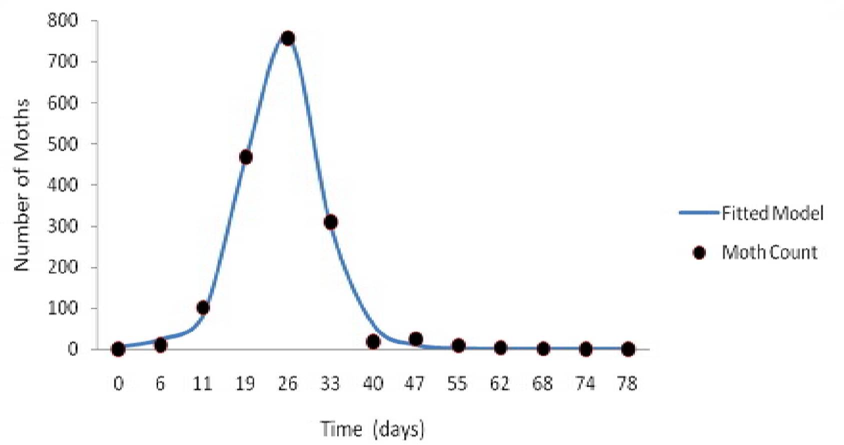

Figure 1: Illustration of Logistic pdf Curve for Moth Data

We illustrate fitting this data set using standard nonlinear least squares (Neter *et al.,* 1996;), as implemented in SPSS (2007). With Y(t) representing the observed population size, we assume regression model:

$$Y(t) = N(t) + \varepsilon \qquad (15)$$

where ε denotes an *independent* random error term. The parameters in (8), with their standard errors in parentheses, are N_{max} = 778.5

(16.4) moths, t_{max} = 24.76 (0.15) days, and b = 0.2576 (0.0067)/days. The resulting fitted curve is

$$Y(t) = \frac{(4)\cdot(778.5)\cdot e^{-0.258(t-24.8)}}{(1+e^{-0.258(t-24.8)})^2} \qquad (16)$$

which is illustrated also in Figure 1 as a smooth curve. This curve fits the moth data remarkably well.

Though a well-fitting curve is always prized for publication purposes, the fitted curve alone is not sufficient for subsequent statistical analysis. The curve is useful primarily for providing parameter estimates. The parameter estimates provide response variables, which might be used in ecological studies to detect and describe the impact of environmental factors on observed abundance curves. For example, we have shown in Matis et al. (2008) using data from a designed experiment that such response variables are helpful in explaining the effects of nitrogen and of irrigation treatments on cotton aphid abundance.

We expect that the model will be useful in the same way in relating environmental effects to moth abundance in **Syria**, which will be a useful part of any control strategy. More specifically, the data in Figure 1 come from one of four sites of varying elevation and forest composition in Al alouni (2009). One approach for analyzing these data is to relate model parameters from a site to its environmental factors. For example, letting Z_i denote a general response variable, useful response variables from the moth data in Figure 1 include:

1. $Z_1 = N_{max}$, which estimates the peak count, is of paramount interest to ecologists as a measure of maximum infestation. In this example, recall N_{max} = 778.5 moths.
2. $Z_2 = t_{max}$, which is of interest as a time to maturity for the moth in its life cycle at the site. In the present example, t_{max} = 24.76 days from the date when observations started (May 25, 2008).
3. Z_3 = b, which in the moth context may be interpreted as an initial per capita 'emergence' rate in the life cycle of the moth. This is b = 0.2576/day for these data.
4. $Z_4 = \delta$, which in the moth context is a population decline coefficient. For this example, using (10), one has $\delta = (0.2576)^2 / 1557 = 4.26\ 10^{-5}$ (moth-days^2)$^{-1}$.
5. Z_5 = K, which to an ecologist is the 'total cumulative density', and is used as a measure of the total environmental impact of the moth

infestation on its available resources. In this example, K = 3114/0.2576 = 12088 moth-days.

6. $Z_6 = \sigma$, which is the standard deviation of the logistic pdf. The standard deviation of the logistic is, from Johnson and Kotz (1970):

$$\sigma = \frac{\pi}{b\sqrt{3}} \quad (17)$$

For the present example this gives $\sigma = \pi/(0.2576*\sqrt{3}) = 7.04$ days. We use 4* σ as a measure of the duration of substantial moth infestation. Hence the duration of infestation for these data is estimated to be 28.2 days.

Each of these variables defines some naturally interpretable characteristic of a growth curve, and each is useful in describing environmental effects. An experimenter might be able to obtain some reasonably accurate, subjective estimate of N_{max} and t_{max} without a model such as (8). Accurate estimates of b, δ, K, and σ, however, would be very difficult to obtain without such a model.

These response variables could be calculated for the other sites as well, to determine statistically whether specific environmental factors are correlated with one or more of the Z_1 to Z_6 variables. As an illustration, published studies have shown that moth abundance in North America is related to elevation and to forest composition (Sharov et al., 1997). Studies are currently in progress to determine, for example, whether the above population measures for the gypsy moth in **Syria** are related to elevation and to the proportion of oak trees in a forest. We expect that such studies will lead to abundance predictions, and hopefully ultimately to moth control strategies.

Conclusions

The paper shows that the logistic pdf model may be used to describe the population size of insects other than aphids. This fact is demonstrated, for the first time in the published literature to our knowledge, by the successful fitting of the model to gypsy moth data from **Syria**. The model parameters will be correlated with ambient environmental variables, thereby providing a new mathematical tool for entomologists to predict moth outbreaks, and hopefully ultimately to develop effective moth control strategies. We expect that the pdf model will become widely used in practice.

REFERENCES

Al alouni, U. 2009. Biological study of Gypsy moth *L. dispar* in Syrian forests (Fronloq, Abo kbies) and its control methods. Master's thesis, Department of Plant Protection, Damascus University.

Banks, R. B. 1994. Growth and Diffusion Phenomena. Springer-Verlag, Berlin

Johnson, N. L. and S. Kotz. 1970. Continuous Univariate Models – 2. Wiley, NY.

Kindlmann, P. 1985 A model of aphid population with age structure. In: Capasso, V., Grosso, E., Paveri-Fontana, S. L. (Eds.) Mathematics in Biology and Medicine, Proceedings, Bari, 1983: Lecture Notes in Biomathematics, Springer, Berlin, pp. 72-77.

Matis, J. H., T. R. Kiffe, T. I. Matis, and D. E. Stevenson. 2005. Nonlinear stochastic modeling of aphid population growth. Mathematical Biosciences 198: 148-168.

Matis, J. H., T. R. Kiffe, T. I. Matis, J. A. Jackman, and H. Singh. 2007. Population size models based on cumulative size, with application to aphids. Ecological Modelling 205: 81-92.

Matis, T. I., M. N. Parajulee, J. H. Matis, and R. B. Shrestra. 2008. A mechanistic model based analysis of cotton aphid population dynamics data. Agricultural and Forest Entomology, 10: 355-362.

Matis, J. H., T. R. Kiffe, W. van der Werf, A C. Costamagna, T. I. Matis and W. E. Grant. 2009. Population dynamics models based on cumulative density dependent feedback: A link to the logistic growth curve and a test for symmetry using aphid data. Ecological Modelling 220: 1745-1751.

Neter, J., M. H. Kuttner, C. J. Nachtsheim and W. Wasserman. 1996. Applied Linear Statistical Models , 4th ed. Irwin, Chicago, Il.

Prajneshu. 1998. A nonlinear statistical model for aphid population growth. J Indian Society of Agricultural Research 51: 73-78.

Renshaw, E. 1991. Modeling Biological Populations in Space and Time. Cambridge University Press, New York.

Sharov, A. A., A. M. Liebhold and E. A. Roberts. 1997. Correlation of counts of gypsy moths (Lepidoptera: Lymantriidae) in pheromone traps with landscape characteristics. Forest Science 43: 483-490.

SPSS. 2007. SPSS 16.0 for Windows. SPSS Inc., Chicago, IL.