

## Nonparametric and parametric estimation for a linear germination-growth model

CHIU, Sung Nok; Quine, M. P.; Stewart, M.

*Published in:*  
Biometrics

*DOI:*  
[10.1111/j.0006-341X.2000.00755.x](https://doi.org/10.1111/j.0006-341X.2000.00755.x)

Published: 01/01/2000

[Link to publication](#)

*Citation for published version (APA):*

CHIU, S. N., Quine, M. P., & Stewart, M. (2000). Nonparametric and parametric estimation for a linear germination-growth model. *Biometrics*, 56(3), 755-760. <https://doi.org/10.1111/j.0006-341X.2000.00755.x>

### General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent publication URLs

# Nonparametric and parametric estimation for a linear germination–growth model

S. N. Chiu,<sup>1,\*</sup> M. P. Quine,<sup>2</sup> and M. Stewart<sup>2</sup>

<sup>1</sup>Department of Mathematics, Hong Kong Baptist University, Kowloon, Hong Kong

<sup>2</sup>School of Mathematics and Statistics, University of Sydney, N. S. W. 2006, Australia

\**email*: snchiu@math.hkbu.edu.hk

SUMMARY. Seeds are planted on the interval  $[0, L]$  at various locations. Each seed has a location  $x$  and a potential germination time  $t \in [0, \infty)$  and it is assumed that the collection of such  $(x, t)$  pairs forms a Poisson process in  $[0, L] \times [0, \infty)$  with intensity measure  $dx d\Lambda(t)$ . From each seed which germinates an inhibiting region grows bidirectionally at rate  $2v$ . These regions inhibit germination of any seed in the region with a later potential germination time. Thus, seeds only germinate in the uninhibited part of  $[0, L]$ . We want to estimate  $\Lambda$  on the basis of one or more realisations of the process, the data being the locations and germination times of the germinated seeds. We derive the maximum likelihood estimator of  $v$  and a nonparametric estimator of  $\Lambda$  and describe methods of obtaining parametric estimates from it, illustrating these with reference to gamma densities. Simulation results are described and the methods applied to some neurobiological data. An Appendix outlines the S-PLUS code used.

KEY WORDS: Boolean model; DNA replication; Germination–growth process; Inhibition; Maximum likelihood estimation; Nucleation; Synaptic transmission.

## 1. Introduction

The estimation problem considered in this paper was motivated by applications in different biological contexts. Bennett and Robinson (1990) proposed the following model of autoinhibited release of neurotransmitters at a synapse. The terminal of a neuronal axon at the neuromuscular junction has branches consisting of strands containing many randomly scattered sites. At a synapse an action potential triggers the release of neurotransmitter at these sites. Each quantum released is assumed to cause release of an inhibitory substance which diffuses along the terminal at a constant rate preventing further releases in the inhibited region. Measurements of time and amplitude of release are possible. Vanderbei and Shepp (1988) and Cowan, Chiu and Holst (1995) considered the replication of a DNA molecule in higher animals. The replication commences at specific sites (replication origins) which are randomly scattered along a topologically linear molecule. Each site is recognised, after some random delay time, by an enzyme-complex which then binds to the site. The complex immediately initiates a bidirectional movement along the DNA. Replication takes place at the moving ‘frontier’ and enzyme-complexes continue to bind to sites at unreplicated parts of the molecule. Measurements of time and location of each initiation of replication can be taken by autoradiography. Although the releases of neurotransmitters and the initiations of replication origins are in a sense observable, it is important to know the intensities of all potential releases and potential initiations to have a better understanding of the underlying biological mechanisms.

The two biological processes described above have been modelled by the following stochastic model. Consider a Poisson process  $\Psi_L$  on  $[0, L] \times [0, \infty)$  with intensity measure  $dx d\Lambda(t)$ , where  $\Lambda(t) = 0$  for  $t < 0$ ,  $\Lambda(t) < \infty$  for  $t < \infty$  and  $\Lambda(\infty) > 0$ . Each point of  $\Psi_L$  represents the location and the potential germination time of a seed. Once the seed  $(x_i, t_i)$  is germinated at time  $t_i$ , two bidirectional *moving frontiers* commence at  $x_i$ . Each frontier moves at a con-

stant speed  $v$  until it meets an opposing one or it reaches the end of the interval. The region passed over by moving frontiers is *inhibited* such that germination of any seed in the region with a later potential germination time is not allowed. Seeds continue to germinate on uninhibited parts of  $[0, L]$  until the whole interval is inhibited. The data available are the times and the locations of germinations; neither ungerminated seeds nor inhibited regions are observable. The function  $\Lambda$  has to satisfy that  $\mu \equiv \int_0^\infty \exp \left\{ - \int_0^t 2v(t-u)d\Lambda(u) \right\} d\Lambda(t) < \infty$ . This regularity condition ensures that the number of germinated seeds in a bounded interval is almost surely finite. Such a germination–growth process was first suggested and studied by Kolmogorov (1937) to model crystal growth (for details see Chiu, 1995, 1997; Chiu and Quine, 1997; Okabe et al., 2000).

For the neurobiology application described above, Quine and Robinson (1992) discussed the estimation of the speed  $v$  and the intensity  $\Lambda$ , assuming that  $\Lambda(t) = \lambda t$  for some positive finite  $\lambda$ , i.e. the underlying Poisson process is homogeneous. The estimation uses only the times of the first germination and of the second germination, if there is one. However, this homogeneous model was oversimplified for the application and shown to be inadequate by data. A more realistic  $\Lambda$  suggested in the literature (e.g. Thomson et al., 1995) is an unnormalised gamma density  $d\Lambda(t) = \{\lambda\gamma^k/\Gamma(k)\}t^{k-1}e^{-\gamma t}dt$ , where  $\lambda$ ,  $\gamma$  and  $k$  are all positive and finite.

For the molecular biology application mentioned, Vanderbei and Shepp (1988) suggested that random sites and times are starting points and times of the DNA molecule replication, which corresponds to  $\Lambda(t) = \lambda t$  for a positive finite constant  $\lambda$ . Cowan et al. (1995) took the local stereochemistry of the DNA into consideration and suggested that replication takes places at random sites and the starting times are exponentially distributed, which means  $d\Lambda(t) = \lambda\gamma e^{-\gamma t}dt$ , where  $\lambda$  and  $\gamma$  are both positive and finite.

Both the synaptic transmission and the DNA replication have two models which differ in

$\Lambda$ . Thus, estimation of nonparametrised  $\Lambda$  would be useful in choosing a more appropriate model.

In this paper we study the maximum likelihood estimation of  $v$  and a nonparametric estimation of  $\Lambda$ , based on the time and location of each germination, i.e. the time and the location of each release of neurotransmitter or each initiation of replication of a DNA molecule. The nonparametric estimation of  $\Lambda$  yields an empirical function which can then be used to estimate the parameters of parametrised  $\Lambda$ . We apply these methods on simulated and real data.

## 2. Maximum likelihood estimation of the growth rate

Consider the two adjacent germinated seeds  $p_1 = (x_1, t_1)$  and  $p_2 = (x_2, t_2)$ , where  $x_1 < x_2$ . The growth rate  $v$  is at most  $v_o = (x_2 - x_1)/(t_2 - t_1)$ . If  $v > v_o$ , then the seed  $p_2$  cannot germinate, since at time  $t_2$  the position  $x_2$  has already been inhibited. Let  $A_v$  represent the meeting-time and position of the two opposing frontiers, each with speed  $v$ , from  $x_1$  to  $x_2$  and from  $x_2$  to  $x_1$ . For any  $v \leq v_o$ , denote by  $\Delta_v$  the triangle with vertices  $p_1$ ,  $p_2$  and  $A_v$ . Note that  $\Delta_{v_1} \subset \Delta_{v_2}$  if  $v_1 > v_2$ . The probability that there is no seed germinated inside  $\Delta_v$  is  $\exp \left\{ - \iint_{\Delta_v} d\Lambda(t) dx \right\}$ .

Maximising the likelihood is equivalent to maximising the product of probabilities in the form  $\exp \left\{ - \iint_{\Delta_v} d\Lambda(t) dx \right\}$ . This can be achieved by maximising the growth rate  $v$ , since all  $\Delta_v$  shrink as  $v$  increases. Therefore, the maximum likelihood estimator of  $v$  based on one experimental result is simply the maximum possible growth rate.

Each experiment producing two or more germinated seeds yields one maximum likelihood estimator of  $v$ . Suppose the experiments are repeated  $n$  times. The maximum likelihood estimator of  $v$ , denoted by  $\hat{v}_n$ , is the minimum of all these estimators of  $v$ . Approximately,  $\hat{v}_n$  can be regarded as a single-experiment maximum likelihood estimator where the length of strand of the experiment is the sum of the lengths of the strands used in the experiments

producing two or more germinated seeds. Thus, in the estimation of  $v$ , repeating the experiment a large number of times is practically the same as increasing the length of the strand used in the experiment. However, by reducing  $L$ , we lose some of the dependence structure on which the analysis depends.

### 3. Nonparametric estimation of $\Lambda$

At each time point  $t$ ,  $L\Lambda(t)$  is the mean number of points in the Poisson process  $\Psi_L = \{(x_i, t_i)\}$  such that  $t_i \leq t$ . For each  $i$ , if at time  $t_i$  the location  $x_i$  is already inhibited, the point  $(x_i, t_i)$  will not germinate. Thus, the total number of germinations at or before time  $t$  in an experiment with a strand of length  $L$  has an expected value less than  $L\Lambda(t)$ ; it should not be an estimator of  $L\Lambda(t)$ .

At each time  $t$  the inhibited region on  $[0, L]$  is a random closed set. More precisely it is a Boolean model (see e.g. Stoyan, Kendall and Mecke, 1995, p. 59 for the definition), the intensity of which is to be estimated. The estimation of the intensity of a stationary Boolean model in  $\mathbb{R}^d$  has been discussed in the literature (see e.g. Stoyan et al., 1995, pp. 89-95; Molchanov, 1995, 1997; Molchanov and Stoyan, 1994). However, in the current context we are estimating the intensity measure of a nonstationary Boolean model in  $[0, L] \times [0, \infty)$ . Nevertheless, since the Boolean model is stationary with respect to one coordinate, estimation methods for stationary Boolean model can be borrowed to yield pointwise estimates for the intensity measure, and this will be described as follows.

Suppose  $0 \leq t < L/(2v)$ . Consider the inhibited region  $\Xi_t = \bigcup_{i: t_i \leq t} [x_i - v(t - t_i), x_i + v(t - t_i)] \cap [0, L]$  on  $[0, L]$  at time  $t$ . The intersection  $\Xi_t \cap [vt, L - vt]$  is a stationary Boolean model observed on  $[vt, L - vt]$ . Denote by  $p_t$  the mean fraction of the total length of the inhibited region  $\Xi_t$  in the interval  $[vt, L - vt]$ , i.e.  $p_t = E(\|(\Xi_t \cap [vt, L - vt])\|) / (L - 2vt)$ , where  $\|\cdot\|$  denotes the total length. Because of the stationarity of  $\Xi_t$  on  $[vt, L - vt]$ ,  $p_t$  is equal to the probability that an arbitrary but fixed point  $o$  in  $[vt, L - vt]$  is contained in  $\Xi_t$

(see e.g. Stoyan et al., 1995, p. 201).

Consider the point process  $\Phi_t = \{y_i : y_i = x_i + v(t - t_i), (x_i, t_i) \in \Psi_L, t_i \leq t\}$ . This point process is obtained by shifting the location of each (germinated or ungerminated) seed from  $x_i$  to the position at time  $t$  of the right-hand side frontier of the (actual or potential) bidirectional movement initiated by the (actual or potential) germination of the seed  $(x_i, t_i)$ . Thus, a germinated and ungerminated seed will be shifted to the right end point and the interior, respectively, of an inhibited interval. The shift of each Poisson point  $x_i$  is independent of the other points. Therefore,  $\Phi_t$  is a stationary Poisson process on the interval  $[vt, L - vt]$  with intensity  $\Lambda(t)$  for  $t < L/v$ .

Let  $N_t$  be the number of points of  $\Phi_t$  in  $[vt, L - vt]$  which are also the right end points of the *separate* inhibited intervals in  $\Xi_t$ .

Slivnyak's (1962) theorem implies that for each point  $y_j$  in a stationary Poisson process, if all the points of the Poisson process are shifted such that  $y_j$  is shifted to an arbitrary but fixed point  $o$ , the collection of all shifted points except  $o = y_j$  itself forms a stationary Poisson process with the same intensity as before. This leads to the conclusion that for each  $y_j \in \Phi_t \cap [vt, L - vt]$ , the probability that  $y_j$  is not the right end point of any one of the separate intervals of  $\Xi_t$  is  $p_t$ . It is because if  $y_j$  is not a right end point, then  $y_j$  is contained in some interval  $[y_i - 2v(t - t_i), y_i]$ , where  $i \neq j$ . Since  $t_i$  and  $t_j$  are independent and identically distributed, by Slivnyak's theorem,  $\Pr(y_j \in \cup_{i: t_i \leq t, i \neq j} [y_i - 2v(t - t_i), y_i]) = \Pr(o \in \Xi_t) = p_t$  for some  $o$  in  $[vt, L - vt]$ .

The argument above leads to  $E(N_t) = (1 - p_t)(L - 2vt)\Lambda(t)$ . Thus, if  $\hat{v}$  is an estimator of  $v$ , by the method of moment,  $\Lambda(t)$  for  $0 \leq t < L/(2\hat{v})$  can be estimated by

$$\hat{\Lambda}(t) = \frac{\hat{N}_t}{(L - 2\hat{v}t)(1 - \hat{p}_t)},$$

where

$$\widehat{N}_t = \begin{cases} \# \text{ separate intervals in } \widehat{\Xi}_t \cap [\widehat{v}t, L - \widehat{v}t] & \text{if } L - \widehat{v}t \notin \widehat{\Xi}_t \\ \# \text{ separate intervals in } \widehat{\Xi}_t \cap [\widehat{v}t, L - \widehat{v}t] - 1 & \text{if } L - \widehat{v}t \in \widehat{\Xi}_t, \end{cases}$$

$$\widehat{p}_t = \frac{\|(\widehat{\Xi}_t \cap [\widehat{v}t, L - \widehat{v}t])\|}{L - 2\widehat{v}t}, \text{ and } \widehat{\Xi}_t = \bigcup_{i: t_i \leq t} [x_i - \widehat{v}(t - t_i), x_i + \widehat{v}(t - t_i)] \cap [0, L].$$

For each  $t$ , this estimator is very similar to the intensity estimator of an observable stationary Boolean model by the method of moment (see e.g. Stoyan et al., 1995, p. 89). However, since the Boolean model, i.e. the inhibited region, is actually not observable in our context, we reconstruct the Boolean model by using the estimated speed  $\widehat{v}$  and then use the reconstructed Boolean model to estimate  $p_t$ . There are other methods of estimating the intensity of an observable stationary Boolean model. These methods can also be modified in order to estimate  $\Lambda(t)$ . However, simulation shows that the method of moment is the most precise method in estimating the intensity of a Boolean model (Stoyan et al., 1995, p. 95). Thus, we do not consider these other methods here.

The empirical function  $\widehat{\Lambda}$  is a single-experiment estimator of  $\Lambda$ . Consider that the experiment is repeated  $n$  times, and the maximum likelihood estimator  $\widehat{v}_n$  is used to estimate  $v$ . Suppose in the  $i$ th experiment the strand used is of length  $L_i$  and the estimated inhibited region at time  $t$  is  $\widehat{\Xi}_t^{(i)}$ . Denote by  $\widehat{N}_t^{(i)}$  the empirical value of  $N_t$  in the  $i$ th experiment and let

$$\widehat{p}_t = \frac{\sum_{i=1}^n \|(\widehat{\Xi}_t^{(i)} \cap [\widehat{v}_n t, L_i - \widehat{v}_n t])\|}{\sum_{i=1}^n L_i - 2n\widehat{v}_n t},$$

with the convention that  $[\widehat{v}_n t, L_i - \widehat{v}_n t] = \emptyset$  if  $t \geq L_i/(2\widehat{v}_n)$ . Then  $\Lambda(t)$  for  $0 \leq t < \max\{L_i : i = 1, \dots, n\}/(2\widehat{v}_n)$  can be estimated by

$$\widehat{\Lambda}_n(t) = \frac{\sum_{i=1}^n \widehat{N}_t^{(i)}}{(\sum_{i=1}^n L_i - 2n\widehat{v}_n t)(1 - \widehat{p}_t)}.$$

Thus, similar to the estimation of  $v$ , in the estimation of  $\Lambda$ , repeating the experiment is essentially the same as increasing the length of the strand used in the experiment.



Further developments of estimation for nonstationary observable and unobservable Boolean models can be found in Molchanov and Chiu (2000).

#### 4. Parametric estimation of $\Lambda$

Suppose that  $d\Lambda(t) = \{\lambda\gamma^k/\Gamma(k)\}t^{k-1}e^{-\gamma t}dt$ , which was used in Cowan et al. (1995) (with  $k = 1$ ) and Thomson et al. (1995). Then the empirical function  $\widehat{\Lambda}_n$  can be used to estimate  $\lambda$ ,  $\gamma$  and  $k$ . The methods discussed below can also be used to estimate the parameters of other forms of  $\Lambda$ .

There are two ways to estimate the parameters. The first one is as follows. The estimators  $\widehat{\lambda}_1$ ,  $\widehat{\gamma}_1$  and  $\widehat{k}_1$  are obtained by minimising the sum of the absolute value of deviations between the empirical function  $\widehat{\Lambda}_n$  and the theoretical  $\Lambda$  with parameters  $\widehat{\lambda}_1$ ,  $\widehat{\gamma}_1$  and  $\widehat{k}_1$ . The minimisation procedure requires starting values and upper and lower bounds for the three parameters. For  $\lambda$ , we used the initial value  $\lambda_o = \sup_t \widehat{\Lambda}_n(t)$  with arbitrary bounds  $\lambda_o \pm \lambda_o^{1/2}$ . We chose this initial value because  $\Lambda(t) \rightarrow \lambda$  as  $t \rightarrow \infty$ . It was not clear how to deal with  $\gamma$  and  $k$ , so we assigned lower bounds of zero, and chose arbitrary starting values and upper bounds. In an application there may well be some *a priori* information about all three parameters.

Since  $\Lambda(\infty) = \lambda < \infty$ , the probability that there is no seed germinated in an experiment with a strand of length  $L$  is not zero but  $\exp(-\lambda L)$ . If there are  $m$  such experiments,  $1 \leq m < n$  and the lengths of the strand used in the experiments are all equal to  $L$ , then  $\lambda$  can also be estimated by the method of moment:  $\widehat{\lambda}_2 = (\log n - \log m)/L$ . This  $\widehat{\lambda}_2$  leads to another pair of estimators of  $\gamma$  and  $k$ :  $\widehat{\gamma}_2$  and  $\widehat{k}_2$ , respectively. They are obtained by minimising with respect to  $\gamma$  and  $k$  the absolute value of deviations between  $\widehat{\Lambda}_n$  and the  $\Lambda$  with parameters  $\widehat{\lambda}_2$ ,  $\gamma$  and  $k$ .

When  $t$  gets large,  $[0, L]$  is almost covered by intervals, and so  $\widehat{p}_t$  tends to 1 and  $\widehat{\Lambda}_n(t)$  goes to infinity. Thus, it would be better not to use the whole range of  $t$  in estimating

the parameters. We introduce an estimation parameter  $\alpha$ . Suppose there is a  $t_o$  such that  $1 - p_{t_o} < \alpha$ , then we do not use  $\widehat{\Lambda}_n$  on  $[t_o, L/(2v)]$  in estimating the parameters.

## 5. Simulation

The main objective of the simulation was to assess the quality of the estimators. First consider  $\widehat{v}$ . We take the target value of  $v = 0.2$ , which is encountered in practice (see e.g. Holst, Quine and Robinson, 1996, p. 921). From its definition it is clear that  $\widehat{v}$  will always overestimate  $v$ , but the bias is extremely small for reasonably large values of  $L$  and  $n$ , e.g. the estimated biases are  $2.7 \times 10^{-4}$  and  $2.9 \times 10^{-6}$  for  $L = n = 10$  and  $50$ , respectively. The precision of  $\widehat{v}$  can be seen in Table 1.

To assess the quality of the nonparametric estimator  $\widehat{\Lambda}_n(t)$  and at the same time that of the parametric estimator based on  $\widehat{\lambda}_1$ ,  $\widehat{\gamma}_1$  and  $\widehat{k}_1$  we give in Figure 1 the results of four simulations with different values of  $n$  and  $L$ . In each case  $\lambda = 5$ ,  $\gamma = 2$ ,  $k = 4$ ,  $v = 0.2$ , and  $\alpha = 0.05$ .

[ Figure 1 about here.]

It can be seen that even for these rather modest sample sizes, both the parametric and nonparametric curves provide a good fit at least for small values of  $t$ . For large  $t$ , the estimates sometimes lie above the true intensity curve and sometimes below it. Increasing either  $n$  or  $L$  improves the quality of the fit.

The effect of the choice of the sampling fraction  $\alpha$  on the parametric estimators  $\widehat{\lambda}_1$ ,  $\widehat{\gamma}_1$  and  $\widehat{k}_1$  can be gauged from Table 1 which gives means and standard deviations of these estimates and of  $\widehat{v}$ , from 50 iterations with  $n = 10$ ,  $L = 25$ ,  $\lambda = 5$ ,  $\gamma = 2$ ,  $k = 4$  and  $v = 0.2$ . The figures suggest that for these values of  $n$  and  $L$ ,  $\alpha$  should be chosen in the range 0.1–0.2.

[Table 1 about here.]

The case when some runs produce no germinations at all is of particular interest in practice. There are three options available to us: (1) discard these runs and use the estimates

$(\hat{\lambda}_1, \hat{\gamma}_1, \hat{k}_1)$  calculated from the remaining runs; (2) estimate  $\lambda$  by  $\hat{\lambda}_2$  and use this as a starting point for the trivariate minimisation; (3) estimate  $\lambda$  by  $\hat{\lambda}_2$  and then do a bivariate minimisation on the other two variables. The first option is less efficient than the others in the sense that it ignores an independent estimator of  $\lambda$ . However in simulations all three options gave very similar output; in fact the intensity graphs are visually indistinguishable.

Further simulation results related to the current paper can be obtained from

<http://www.maths.usyd.edu.au/u/malcolmq/>

## 6. A neurobiological application

The data, provided by Professor M.R. Bennett of the Neurobiology Research centre at the University of Sydney, consists of the results of 800 experiments in which the times of release and the amplitudes were measured. The number of releases ranges from 0 to 4. The frequencies of 0's, 1's, ... is 101, 387, 237, 66, 9. These figures are very similar to the results of another series of 800 experiments given as Table 2 of Quine and Robinson (1992). In an attempt to exploit the bivariate nature of the data, we take a transformation of amplitude to be the distance surrogate: we assume that there is some inverse relation between amplitude and distance from the measuring device. Of all the inverse power transformations we tried,  $\text{distance} = 1/\sqrt{\text{amplitude}}$  gave the most symmetric histogram and the closest to uniform, so we used this. There were 50 experiments with two identical amplitude readings. Since in our model it is impossible to have two germinated seeds at the same  $x$ -location, we ignored these experiments. Of the transformed data, there are four outliers above 0.2 (all with value 0.2234) so we deleted these experiments as well, leaving us with 101 experiments with no germinations and 645 with at least one germination, the frequencies of 1's, ..., 4's now being 387, 210, 45, 3. The 'location values' were multiplied by 5 so they were roughly uniform on  $[0, 1]$ ; that is, we take  $L = 1$  in our model. The positive time values ranged up to 1000

so we divided them by 500 so as to get roughly the range of  $t$ -values we encountered in the simulations.

Using the method introduced in previous sections, we obtained the following nonparametric estimate of the intensity  $\Lambda$ :

```
[1] 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
[7] 0.00000000 0.00000000 0.01292028 0.31839059 0.97645342 1.33376761
[13] 1.49481937 1.54849759 1.56754307 1.57657268 1.59285925 1.59334251
[19] 1.58637623 1.58820836 1.58240330 1.59335701 1.59874864 1.60808411
[25] 1.60959817
```

with the corresponding vector of time points (`t.vec`) being

```
[1] 0.000000 0.140525 0.281050 0.421575 0.562100 0.702625 0.843150 0.983675
[9] 1.124200 1.264725 1.405250 1.545775 1.686300 1.826825 1.967350 2.107875
[17] 2.248400 2.388925 2.529450 2.669975 2.810500 2.951025 3.091550 3.232075
[25] 3.372600
```

On inspection the vector of the estimate of  $\Lambda$  does not go positive until the ninth element, which corresponds to a time of somewhere between 0.983675 and 1.124200, i.e. there is a delay of approximately one time unit. Discarding the first 8 elements of these vectors and subtracting 1 from the  $t$ -values, then using Option 2 of the three minimisation procedures gave  $\hat{\lambda} = 1.590$ ,  $\hat{\gamma} = 13.296$ ,  $\hat{k} = 5.100$ , and  $\hat{v} = 0.018$ . The fit of the parametric curve is shown together with the nonparametric curve in Figure 2.

[Figure 2 about here.]

Of these estimates,  $\hat{\gamma}$  and  $\hat{v}$  are scale-dependent, but  $\hat{\lambda}$  and  $\hat{k}$  may be compared to the corresponding values of Thomson (see Holst et al., 1996, pp. 920–921) who suggested on biological grounds that  $k = 5$  and hence obtained an estimate of 2.04 for  $\lambda$ .

## ACKNOWLEDGEMENTS

We thank the editor and the referees for helpful suggestions. This research was supported by an FRG grant of the Hong Kong Baptist University and by the Australian Research Council.

## REFERENCES

- Bennett, M. R. and Robinson, J. (1990). Probabilistic secretion of quanta from nerve terminals at synaptic sites on muscle cells: nonuniformity, autoinhibition and the binomial hypothesis. *Proceedings of the Royal Society of London, Series B* **239**, 329-358.
- Chiu, S. N. (1995). Limit theorem for the time of completion of Johnson–Mehl tessellations. *Advances in Applied Probability* **27**, 889-910.
- Chiu, S. N. (1997). A central limit theorem for linear Kolmogorov's birth–growth models. *Stochastic Processes and their Applications* **66**, 97-106.
- Chiu, S. N. and Quine, M. P. (1997). Central limit theory for the number of seeds in a growth model in  $\mathbb{R}^d$  with inhomogeneous Poisson arrivals. *Annals of Applied Probability* **7**, 802-814.
- Cowan, R., Chiu, S. N. and Holst, L. (1995). A limit theorem for the replication time of a DNA molecule. *Journal of Applied Probability* **32**, 296-303.
- Holst, L., Quine, M. P. and Robinson, J. (1996). A general stochastic model for nucleation and linear growth. *Annals of Applied Probability* **6**, 903-921.
- Kolmogorov, A. N. (1937). On statistical theory of metal crystallisation. *Izvestia Academy of Science, USSR, series Mathematics* **3**, 355-360 (in Russian).
- Molchanov, I. S. (1995). Statistics of the Boolean model: From the estimation of means to the estimation of distributions. *Advances in Applied Probability* **27**, 63-86.
- Molchanov, I. S. (1997). *Statistics of the Boolean Model for Practitioners and Mathematicians*. Chichester: Wiley.

- Molchanov, I. S. and Chiu, S. N. (2000) Smoothing techniques and estimation methods for nonstationary Boolean models with applications to coverage processes. *Biometrika*, to appear.
- Molchanov, I. S. and Stoyan, D. (1994). Asymptotic properties of estimators for parameters of the Boolean model. *Advances in Applied Probability* **26**, 301-323.
- Okabe, A., Boots, B, Sugihara, K. and Chiu, S. N. (2000). *Spatial Tessellations. Concepts and Applications of Voronoi Diagrams*, 2nd edition. Chichester: Wiley.
- Quine, M. P. and Robinson, J. (1992). Estimation for a linear growth model. *Statistics & Probability Letters* **15**, 293-297.
- Slivnyak, I. M. (1962). Some properties of stationary flows of homogeneous random events. *Theory of Probability and its Applications* **7**, 336-341.
- Stoyan, D., Kendall, W. S. and Mecke, J. (1995) *Stochastic Geometry and its Applications*, 2nd edition. Chichester: Wiley.
- Thomson, P. C., Lavidis, N. A., Robinson, J. and Bennett, M. R. (1995). Probabilistic secretion of quanta at somatic motor-nerve terminals: the fusion pore model, quantal detection and autoinhibition. *Philosophical Transactions of the Royal Society of London. Series B* **349**, 197-214.
- Vanderbei, R. J. and Shepp, L. A. (1988) A probabilistic model for the time to unravel a strand of DNA. *Communications in Statistics. Stochastic Models* **4**, 299-314.

#### APPENDIX: S-PLUS PROGRAMS

The simulations were run using S-PLUS. The nonparametric estimation used seven functions which we will now outline.

`NP.n.sim` has 8 arguments with default values as follows:

```
nsim=100,L=25,lam=5,gam=2,sh=4,v=.2,
```

`t.vec=seq(0,5,len=25), alpha=0.05`

Here `nsim` is the sample size ( $n$ , say), `L` the interval length, `lam` is  $\lambda$ , `gam` is  $\gamma$ , `sh` is  $k$ , `t.vec` is the initial vector of  $t$ -values at which the formulae for  $\hat{\Xi}_t$ ,  $\hat{N}_t$  and  $\hat{p}_t$  are evaluated. This vector has as default length 25; increasing this length would increase precision at the expense of computer running time but we have not quantified this aspect of the algorithm. The maximum value (default 5) is modified subsequently. Finally, `alpha` is  $\alpha$ .

`NP.n.sim` first calls the function `Psi.L` to generate the points  $(x_i, t_i)$ . This involves calls to the S-PLUS functions `rpois` to get  $N$ , the number of points, `runif`, to get the  $x_i$ 's, and `rgamma`, to get the  $t_i$ 's. Note that if  $W$  represents the rv corresponding to a value generated by `rgamma(sh=k)` then  $W/\gamma$  has pdf  $\gamma^k t^{k-1} e^{-kt} / \Gamma(k)$ ,  $t > 0$ . Next, the function `Upsil` is called to thin the points according to the specified growth rate,  $v$ .

The final call `NP.n.sim` makes is to the function `Lam.n.hat.t`. This function first makes calls to the function `What` in order to compute an estimate  $\hat{v}_n$  of  $v$  based on the combined samples, and modifies the vector `t.vec` in the light of the maximum time over all samples and the constraint  $0 \leq t \leq \max L_i / (2\hat{v}_n)$ . It then computes a preliminary version of  $\hat{p}_t$ , which involves calls to `Xi.hat.t`, which computes the  $\hat{\Xi}_t^{(i)}$ 's, and then to `Lam.hat.t`, which computes the  $\|(\Xi_t^{(i)} \cap [\hat{v}_n t, L - \hat{v}_n t])\|$ 's. If  $\hat{p}_t$  exceeds  $1 - \alpha$  for any members of `t.vec` then `t.vec` is modified again and a final version of  $\hat{p}_t$  is computed, which, together with the  $\hat{N}_t^{(i)}$ 's, which are also calculated by `Lam.hat.t`, leads to the estimate  $\hat{\Lambda}_n(t)$ .

The parametric estimators  $\hat{\lambda}_i$ ,  $\hat{\gamma}_i$  and  $\hat{k}_i$ ,  $i = 1, 2$  are found by first running the non-parametric estimation algorithm and then the functions `Pobj` and `Pmin`. The first of these sets up the objective function, being the sum of absolute differences between  $\hat{\Lambda}_n(t)$  and the corresponding parametric function, for given values of  $\lambda$ ,  $\gamma$  and  $k$ , the sum being taken over the elements of `t.vec`. The second function, `Pmin`, then calls on the S-PLUS nonlinear minimization function `nlminb`. Minimization is with respect to  $(\lambda, \gamma, k)$  in the first case and

$(\gamma, k)$  in the second.

To analyse the neurobiological data studied in Section 6 by the S-PLUS programs, we needed firstly to get the data, by now in two  $645 \times 4$  matrices, one for time, one for ‘location’, into an S-PLUS list `neuroUp` with 645 elements, each element containing a vector of  $x$ -values, a vector of  $t$ -values, and in addition to conform with our existing programs, the element  $L = 1$ . So for instance the first element of the list is

```
> neuroUp[[1]]  
  
$x:  
[1] 0.5050763 0.5330018 0.6509446  
  
$t:  
[1] 1.218 1.372 1.526  
  
$L:  
[1] 1
```

Details of the code and the dataset are available on

<http://www.maths.usyd.edu.au/u/malcolmq/>

or from the authors.



## Figure captions

Figure 1: Estimates of the intensity with various  $n$  and  $L$ .

Figure 2: Nonparametric and parametric estimates of the intensity of the neurobiological data.

**Table caption**

Table 1: Means and standard deviations (in parentheses) of 50 estimates of  $\lambda = 5$ ,  $\gamma = 2$ ,  
 $k = 4$ ,  $v = 0.2$  with  $n = 10$  and  $L = 25$ .

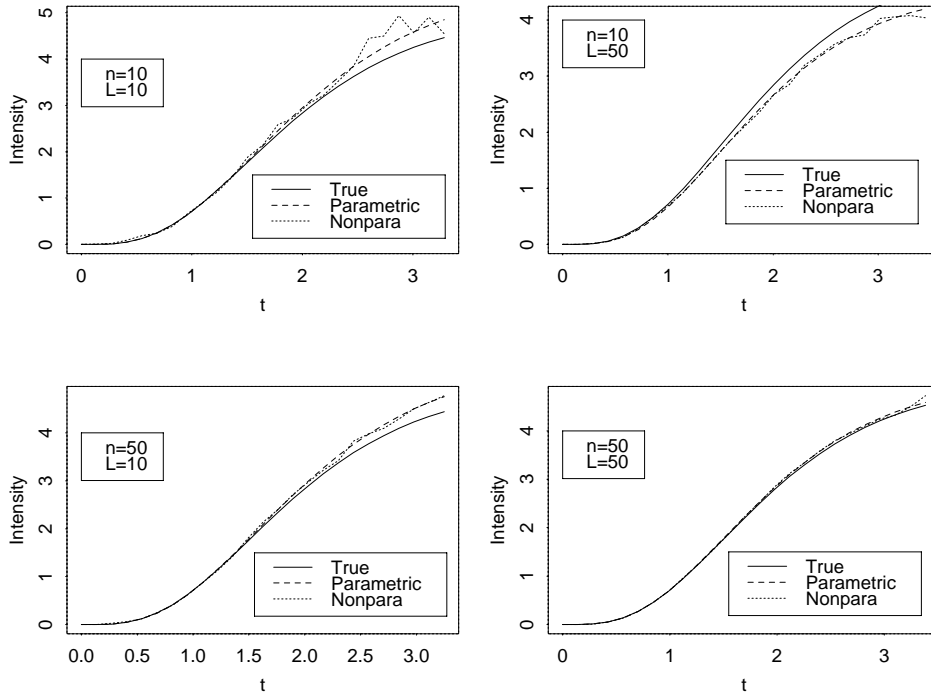


Figure 1: Estimates of the intensity with various  $n$  and  $L$ .

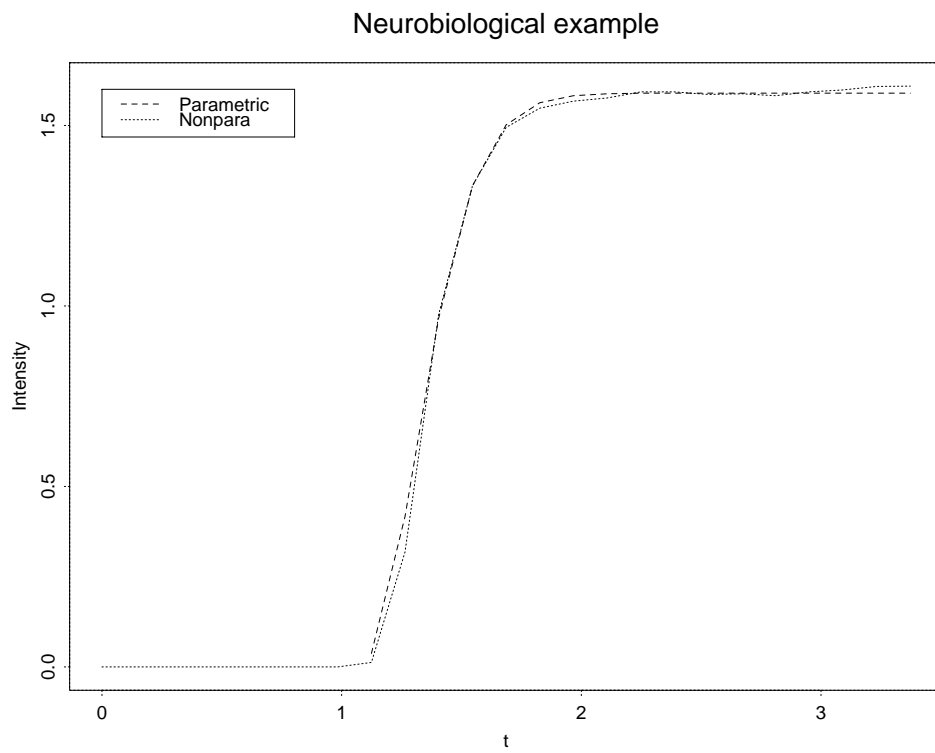


Figure 2: Nonparametric and parametric estimates of the intensity of the neurobiological data.

Table 1: Means and standard deviations (in parentheses) of 50 estimates of  $\lambda = 5$ ,  $\gamma = 2$ ,  $k = 4$ ,  $v = 0.2$  with  $n = 10$  and  $L = 25$ .

$\alpha$	$\hat{\lambda}_1$	$\hat{\gamma}_1$	$\hat{k}_1$	$\hat{v}$
0	5.821 (1.689)	1.888 (0.903)	3.884 (1.102)	0.201 (0.001)
0.05	5.140 (0.789)	2.023 (0.477)	4.012 (0.521)	0.201 (0.001)
0.1	5.082 (0.675)	2.017 (0.446)	4.019 (0.514)	0.201 (0.001)
0.25	5.007 (0.523)	2.085 (0.370)	4.105 (0.396)	0.201 (0.001)
0.5	4.027 (0.292)	2.728 (0.343)	4.715 (0.440)	0.201 (0.001)