Transform Estimation of Parameters for Stage-Frequency Data*

Jennifer A. Hoeting† and Richard L. Tweedie ‡
Colorado State University and University of Minnesota
February 14, 2001

Abstract

We consider multi-stage development models that occur in the maturation of biological organisms, disease progressions and industrial processes. The situations we address are distinguished by the essentially destructive sampling required to assess the stage reached by each individual. We develop robust estimators of stage-dependent maturation rates and overall death rates using methods of moments based on Laplace transforms, for which we develop variance estimates under different sampling schemes. We apply these methods to examples of cattle parasite and grasshopper data, and show they compare well with maximum likelihood methods.

KEYWORDS: Growth models, survival models, Laplace transform methods, grasshoppers, parasites

1 Introduction

Multi-stage models can be used to describe many common situations: disease progression, the life-cycle of organisms which undergo a series of distinct life stages, or system reliability where there are several stages to failure. The simplest example is a one-stage model that describes, say, the time until an egg of an organism hatches. Multiple stages are common also, describing for example maturation as an organism goes from egg to larva to adult form.

Single and multi-stage models of this type have been studied in various biological contexts, as descriptions for the life-cycle of grasshoppers (Read and Ashford, 1968; Ashford et al.,

---

*Research supported in part by National Science Foundation grants DMS 9806243 and DMS 9803882
†Postal address: Department of Statistics, Colorado State University, Fort Collins, CO 80523, USA; email address: jah@stat.colostate.edu
‡Postal address: Division of Biostatistics, University of Minnesota, Minneapolis, MN 55455, USA; email address: tweedie@biostat.umn.edu
cattle-parasitic nematodes (Young et al., 1980a; Young et al., 1980b), worm infections
in mice (Hopkins et al., 1972), sperm (Morgan, 1982) and other insect development models
(van Straalen, 1982; van Straalen, 1985; Manly, 1987). The key quantities typically addressed
in these studies are the average duration in each stage, the probability of being in a given
stage at a given time, and the death rates in each stage.

In this paper we will describe models in terms of “organisms”, but the results can be
applied to all of the contexts above.

For a given organism, we assume there are \((I + 1)\) stages where the \((I + 1)\)th stage is the
final (say, mature adult) stage. We assume the “stage time” or duration \(S_i\) that an organism
spends in stage \(i\) is random, and that for the terminal stage \(S_{I+1} = \infty\). The stage time in
stage \(i\) is described by a density \(g_i\) depending on a parameter \(\theta_i\) (which may be a vector).
We assume that these stage times are independent for each organism and that each organism
progresses through the stages independently of other organisms.

For many models it is also desirable to incorporate a death rate \(\mu_i\), which may or may not
depend on the current stage of the organism. In other examples it may even be necessary
to consider the existence of a stage zero in which the numbers of organisms may not be
observed (Read and Ashford, 1968; Manly, 1987): this stage may be, for example, the egg
stage and only after hatching may organisms be able to be sampled. After considering the
simpler model without these variations we will show that we can incorporate many of these
other aspects into our general formulation.

Our main goal is to develop methods for estimating the stage-dependent maturation para-

ermeters \(\theta = \{\theta_1, \ldots, \theta_I\}\) of the stage time distributions, and death rates or initial numbers
if required, and then to use these to describe the numbers of organisms in each stage at
given times using (1) below. These predictions are usually the most important output of the
models in practice. In the cattle parasite problem, for example, one is interested in the time
at which organisms reach the infective stage, since one can remove cattle from fields that are
known to be infected.

To motivate the type of problem we are considering, we give two data sets we shall
study in Figure 1 and Figure 2, which we shall refer to as the parasite and grasshopper
data respectively. The first of these illustrates the multi-stage context with data taken from
Young et al. (1980b). Here the data are numbers of cattle nematodes in different stages
of their life-cycle. The fitted lines are from the analysis in Section 6.1, using Erlangian(2)
stage time distributions. The second is multi-stage grasshopper data, taken from Table 2
of Read, Ashford and Vickers (1970). These grasshoppers lay eggs in late summer which
Figure 1: Method of moments estimated curves for parasite data. Observed proportions: Stage 1 = ●, stage 2=△, stage 3=+, stage 4=x. Estimated curves: Stage 1 = solid line, stage 2= dotted line, stage 3= dashed line, stage 4 = long dashed line.

hatch the following spring. In this study this first stage was not observed. After hatching, the grasshoppers pass through four stages, called instars, before reaching the final adult state. Grasshoppers may die at any stage due to predation or other causes. Here the fit, described in Section 6.2, uses Erlangian(3) distributions, and we use estimators not just for the maturation rates but also for an overall death-rate, for missing data on the initial number of organisms, and for the maturation rate in the missing first stage.

The paper is organized as follows. In Section 2 we describe the models that we shall use. In Section 3 we develop estimators of the maturation rates, and their variances. The estimators are based on the Laplace transform or weighted area approach (Feigin et al., 1983; Schuh and Tweedie, 1979; Laurence and Morgan, 1987), but the estimators of variances for the parameter estimates are new and are one of the major contributions of this paper.

In Section 4 we add an overall death rate to the models and show the methods can still be implemented. In both cases we also consider the further complications that arise when we do not know initial populations or their maturation rates, a common phenomenon in
real situations. Simulations, given in Section 5 show the methods to be accurate, and the asymptotic variance calculations to be reasonable on quite small data sets. The methods are finally applied to the parasite and grasshopper data in Section 6.

2 Models and sampling schemes

2.1 Models of multistage growth

Suppose that an organism begins to develop at time zero. If \( p_i(t) \) is the probability that the organism is in stage \( i \) at time \( t \), then in terms of the stage length densities \( g_i \), we have

\[
p_i(t) = g_1 * \ldots * g_{i-1} * G_i(t), \quad i = 1, \ldots, I + 1
\] (1)

where \(*\) denotes convolution, \( G_i(t) = \int_0^t g_i(x) dx \), and \( G_{I+1}(t) \equiv 1 \).

In standard lifetime modeling contexts it is common to assume the \( i \)th stage time distribution \((g_i)\) follows a simple exponential distribution, with \( g_i(x) = \theta_i \exp(-\theta_i x) \). However, in many of the biological examples above a more sinusoidal growth curve is visible and this is conveniently modeled by an Erlangian \((a_i)\) distribution (a gamma distribution with shape \(= a_i \) and scale \(= \theta_i \)) given by (Read and Ashford, 1968; Ashford et al., 1970; Young et al., 1980a; Young et al., 1980b)

\[
g_i(x) = \theta_i^{a_i} x^{a_i-1} e^{-\theta_i x}/(a_i - 1)!, \quad x > 0.
\] (2)
Typical choices in the studies above are to take the shape parameters as $a = 2$ or $a = 3$, although this model also covers the simple exponential when $a = 1$. This class of models not only suits the Laplace transform methods below, but also gives flexible growth curves which permit interpretation in terms of maturation rates in each stage. We develop methods to estimate $\theta_i$ in Section 3, assuming that the shape parameters are fixed and known.

These models assume there is no death rate present in the population, so organisms progress from stage to stage, with (perhaps) stage $I + 1$ denoting a dead organism. If there is a death rate $\mu_i$ in the $i$th stage then we have

$$h_i(t) = g_i(t)e^{-\mu_i t}$$

for the density of a maturation through stage $i$ without dying in stage $i$ (Schuh and Tweedie, 1979); and then, using $p^*_i(t)$ to denote the probability that the organism is in stage $i$ at time $t$, we have that

$$p^*_i(t) = h_1 \ast \ldots \ast h_{i-1} \ast H_i(t), \quad i = 1, \ldots, I + 1$$

(3)

where $H_i(t) = e^{-\mu_i t} \int_t^\infty g_i(x)dx$, and $H_{I+1}(t) = e^{-\mu_I t}$. We develop methods to estimate an overall death rate, and the $\theta_i$ in the presence of such a death rate, in Section 4.

If one can observe the times of transition from one stage to the next for each organism, and their times of death, then estimation of $\theta$ is straightforward from the independence assumptions we have made.

The context that we are interested in, however, does not allow us to view these transition times, due to the sampling used to assess the progression of organisms. Typically, for the parasite and the grasshopper data, and in many other situations, sampling is destructive and so the data are severely censored.

The convolution form of (1) or (3) leads to methods of moments estimators of closed form, based on empirical Laplace transforms. These have been developed by Feigin, Tweedie and Belyea (1983), Schuh and Tweedie (1979), Laurence and Morgan (1987), and others cited therein. In this paper we extend these methods. Our main goal is to find asymptotic variance estimates for the multistage case in the sampling situations below, but we also show that the transform methods can be used when considerable amounts of data on initial values for the populations are missing.

2.2 Multinomial and Poisson sampling schemes

There are two different sampling schemes which we shall examine, and for which the properties of the estimators vary somewhat.
In each, we assume that are a number of sampling time points \( t_k, \ k = 1, \ldots, K \). At the sampling time \( t_k \), a separate (kth) population of organisms is assessed to see in which stage each organism lies. The observed data are then just

\[ N_i(t_k) = \text{the number of organisms alive and in the i}^{\text{th}} \text{ stage at time } t_k \]

for \( i = 1, \ldots, I + 1 \) and \( k = 1, \ldots, K \), and each organism is observed only once (see Mann, Schaffer and Singapurwallah, 1974, Sections 5.1.1 and 6.2.1). This occurs frequently, as in the biological contexts described above, and similar destructive testing can occur in industrial applications (Denby et al., 1975).

The two different sampling models in the literature which we shall examine are the “multinomial model” and the “Poisson model”.

In the multinomial model we assume there are fixed (and usually known, although we address the question later) initial numbers \( N_k \) of organisms in the population observed at each time point, so (in the absence of any deaths) \( N_k = \sum_{i=1}^{I+1} N_i(t_k) \). It follows that the \( N_1(t_k), \ldots, N_{I+1}(t_k) \) are multinomially distributed with parameters \( N_k \) and \( p_1(t_k), \ldots, p_{I+1}(t_k) \), respectively. This model applies to cattle parasite data studied by Young et al. (1980a; 1980b) and to other destructive testing situations where there is a controlled starting point.

In the Poisson model, we assume there is a large external population \( N \) and the sampling at each time point is from a small random subset of this population. The \( N_i(t_k) \) are then independent Poisson random variables, with

\[ \text{E}[N_i(t_k)] = \text{Var}[N_i(t_k)] = N\pi_k p_i(t_k) \]

where \( \pi_k \) is the sampling fraction at time \( k \) and \( p_i(t_k) \) is defined in (1). Note that in this model one might start with either knowledge of \( N\pi_k \) or of \( N_k \), the actual starting number in the \( k \)th population, which is itself Poisson with mean \( N\pi_k \). The Poisson model is considered by Read and Ashford (1968) and Ashford Read and Vickers (1970): in their case \( \pi_k \) describes the physical proportion of an overall grasshopper habitat sampled at time \( t_k \).

Without knowledge of the individual stage time or death time data, the multinomial and Poisson likelihoods are not simple to maximize, even for choices of \( g_i \) such as the exponential or the Erlangian distribution (2), since the parameters do not appear in simple closed form in \( p_i(t) \) in (1). Read and Ashford (1968) used a maximum likelihood approach, and showed that it is complex for Erlangian densities; Laurence and Morgan (1987) give equations for the maximum likelihood estimate (MLE) for the simplest one stage models but note that methods giving more explicit estimators would be clearly attractive. With current computing power
the problems of maximizing complex likelihoods are easier numerically than when Read and Ashford carried out their work, but as \( g_i \) becomes more complicated, numerical solutions are not always simple or well-conditioned, and the methods we now describe seem to give robust and computationally attractive options.

3 Estimation with no death rate present

3.1 Empirical Laplace transforms and parameter estimators

We first consider the case where there is no death rate, so evolution is governed by (1). In this case, the Laplace estimation method developed in Feigin et al. (1983) and Schuh and Tweedie (1979) can be extended slightly to take the following form. Suppose that \( c_k \) are cut-points between the sampling times: for example, \( c_0 = 0, \ldots, c_k = (t_k + t_{k+1})/2, \ldots, t_K \leq c_K = A \leq \infty \). For arbitrary \( s > 0 \) construct the exponential weights

\[
w_k(s) = [e^{-sc_k} - e^{-sc_k}]/[1 - e^{-sc_K}].
\]

Now for each stage \((i = 1, \ldots, I + 1)\) construct the weighted sums

\[
\psi_i(s) = \sum_{k=1}^{K} w_k(s)N_i(t_k)/N_k,
\]

and from these define

\[
L_i(s) = \sum_{j=i}^{I+1} \psi_j(s), \quad \Delta_i(s) = L_{i+1}(s)/L_i(s).
\]

We shall assume that \( c_K \) is large enough that \( L_1(s) = \sum_{k=1}^{K} w_k(s) \approx 1 \), at least to the order of accuracy of our other approximations.

The Laplace transform method utilizes the fact that \( \psi_i(s) \) is essentially an empirical Laplace transform approximation. To see this, we denote the transform of \( G_i \) as

\[
\beta_i(s) = \int e^{-su}g_i(u)du;
\]

then from (1), we have

\[
E[\psi_i(s)] = \sum_{k=1}^{K} w_kp_i(t_k)
\approx s \int_{0}^{\infty} e^{-su}p_i(u)du
\]

\[
= s\beta_1(s) \ldots \beta_{i-1}(s)[1 - \beta_i(s)]/s.
\]
We now equate the expected and observed values of $\psi_i(s)$. For $i = 1$ we find $1 - \beta_1(s) \approx \psi_1(s)$ so that we may take

$$
\beta_1(s) = 1 - \sum_{k=1}^{K} w_k(s)N_1(t_k)/N_k
$$

$$
= \sum_{k=1}^{K} [w_k(s)/N_k][N_k - N_1(t_k)] = L_2(s) = \Delta_1(s).
$$

(7)

Inductively take $\beta_{j-1}(s) = \Delta_{j-1}(s)$; substituting in (6) we have

$$
\beta_{j}(s) = \frac{\Delta_1(s) \ldots \Delta_{j-1}(s) - \psi_j(s)}{\Delta_1(s) \ldots \Delta_{j-1}(s)} = \frac{L_j(s) - \psi_j(s)}{L_j(s)} = \frac{L_{j+1}(s)}{L_j(s)}.
$$

Thus we have essentially deconvolved the overall data to get an estimator of the individual stage transform

$$
\beta_j(s) = \Delta_j(s).
$$

(8)

These estimators assume we have knowledge of $N_k$. In the Poisson situation, we typically know $N$ and the sampling proportions $\pi_k$, but not the observed $N_k$. In this case the point estimators are similar but in the estimators above we replace $\psi_i(s), L_i(s), \Delta_i(s)$ with

$$
\psi^p_i(s) = \sum_{k=1}^{K} w_k(s)N_i(t_k)/N\pi_k
$$

$$
L^p_i(s) = \sum_{j=i}^{I+1} \psi^p_j(s), \quad \Delta^p_i(s) = L^p_{i+1}(s)/L^p_i(s).
$$

(9)

(10)

Note that $N$ factors out of the form of the $\beta_j(s)$, which is a benefit in contexts where we know the sampling fraction but not the original population.

For Erlangian distributions we can now construct estimators of the stage parameters in a particularly nice closed form, since when $g_i$ is Erlangian($a_i$) we have

$$
\beta_i(s) = (\theta_i/(\theta_i + s))^{a_i}.
$$

Substituting in (8) the Laplace transform (LT) estimator is then defined by

$$
\hat{\theta}^L_i := s/([\Delta_i(s)]^{-1/a_i} - 1),
$$

(11)

or in the Poisson case by

$$
\hat{\theta}^L_i := s/([\Delta^p_i(s)]^{-1/a_i} - 1).
$$

(12)

This result was stated in Section 4.4 of Schuh and Tweedie (1979) for $c_K = \infty$ and also given in Feigin and Tweedie (1983) for constant fixed $N$, although the derivation in the
latter has some typographical errors. Note that if \( c_K \) is finite and \( s \) is chosen near zero, or even negative as has been done in some one-stage models (Feigin et al., 1983; Laurence and Morgan, 1987), the approximations used in (6) may not have the required accuracy to be useful.

These point estimators, although developed some time ago, have suffered from lack of associated variance estimates, even though they have been shown in simulations to exhibit good empirical properties. In the next section we develop variance estimators, rendering this approach much more effective for practical use.

The best choice of \( s \) for the Laplace transform estimators is not obvious. An iterative method is given in Schuh and Tweedie (1979), while various optimal choices have been developed for the single stage model (Feigin et al., 1983; Laurence and Morgan, 1987). We will use the approximate variance results below as a further way in which one might also choose \( s \) in this context.

### 3.2 Approximate variances of the LT estimators

We use the delta method to get approximate variances for the LT estimator in (11). These are similar in style to those that have been developed for the simple one-stage models (Feigin et al., 1983; Laurence and Morgan, 1987), and more explicitly use the sampling structure than the limiting methods recently proved for general weighted area models (Yao and Morgan, 1999).

For notational convenience we usually drop the dependence on \( s \) in the equations below. Noting that as a function of \( \theta_i \) we have \( \beta_i(s) = f(\theta_i) = (\theta_i/(\theta_i + s))^a \), we first have

\[
\text{Var} \left( \widehat{\beta}_i \right) = \text{Var} \left( \beta_i \right) \left[ \frac{\partial f(\theta)}{\partial \theta} \right]^{-2} = \text{Var} \left( \Delta_i \right) \left( \frac{(\theta + s)^{a+1}}{a\theta^{a-1}} \right)^2
\]

where we will substitute \( \theta = \widehat{\theta}_i \) to get a variance estimate. In calculating the asymptotic variance of the ratio \( \Delta_i = L_{i+1}/L_i \) we first simplify using \( \Delta_i(s) = 1 - \psi_i(s)/L_i(s) \), so that

\[
\text{Var} [\Delta_i(s)] = \text{Var} [\psi_i(s)/L_i(s)].
\]

We shall systematically use the Taylor series formula (Casella and Berger, 1990, p. 331)

\[
\text{Var} \left[ \frac{X}{Y} \right] \approx \frac{\text{E}[X]}{\text{E}[Y]}^2 \left[ \frac{\text{Var}[X]}{[\text{E}[X]]^2} + \frac{\text{Var}[Y]}{[\text{E}[Y]]^2} - 2 \frac{\text{Cov}[X,Y]}{\text{E}[X]\text{E}[Y]} \right]
\]

for the variance of ratios. We now calculate the various terms in (15) for the ratio (14), in both the multinomial and Poisson cases.
Multinomial sampling

In the multinomial case the form (14) for \( \text{Var}(\Delta_i) \) can be evaluated using the following calculations (see for example Agresti (1990, p. 423)):

\[
E[L_i] = \sum_{k=1}^{K} w_k \sum_{j=i}^{I+1} p_j(t_k); \quad E[\psi_i] = \sum_{k=1}^{K} w_k p_i(t_k);
\]

\[
\text{Var}[L_i] = \sum_{k=1}^{K} w_k^2 \text{Var} \left[ \sum_{j=i}^{I+1} N_j(t_k)/N_k \right]
\]
\[
= \sum_{k=1}^{K} w_k^2 \left[ \sum_{j=i}^{I+1} \text{Var} \left[ N_j(t_k)/N_k \right] + \sum_{l,j=i,l\neq j}^{I+1} \text{Cov} \left[ N_j(t_k)/N_k, N_l(t_k)/N_k \right] \right]
\]
\[
= \sum_{k=1}^{K} w_k^2 \left[ \sum_{j=i}^{I+1} p_j(t_k)(1 - p_j(t_k))/N_k - \sum_{l,j=i,l\neq j}^{I+1} p_j(t_k)p_l(t_k)/N_k \right]; \quad (16)
\]

\[
\text{Var}[\psi_i] = \sum_{k=1}^{K} w_k^2 \text{Var} \left[ N_i(t_k)/N_k \right]
\]
\[
= \sum_{k=1}^{K} w_k^2 [p_i(t_k)(1 - p_i(t_k))/N_k];
\]

\[
\text{Cov}[\psi_i, L_i] = \text{Cov} \left[ \psi_i, \psi_i + L_{i+1} \right]
\]
\[
= \text{Var}[\psi_i] + \sum_{k=1}^{K} w_k^2 \sum_{j=i+1}^{I+1} \text{Cov} \left[ N_i(t_k)/N_k, N_j(t_k)/N_k \right]
\]
\[
= \text{Var}[\psi_i] - \sum_{k=1}^{K} w_k^2 \sum_{j=i+1}^{I+1} p_i(t_k)p_j(t_k)/N_k.
\]

In the case \( i = 1 \), this reduces to the form given in Laurence and Morgan (1987): the situation there is much simpler since the covariance terms are absent, as \( L_1 = 1 \).

In using the estimate of variance, we need to evaluate the various estimated values of \( p_i(t_k) \) using the estimates of \( \hat{p}_i \) in (1). This is straightforward numerically although somewhat tedious, especially for larger values of \( a \).

Poisson sampling

In the case of Poisson sampling, we still have the first expression (13), but some other aspects are simplified since now the \( N_i(t_k) \) are independent. Here we evaluate (15) for the ratio \( \Delta_P^i = L_{i+1}^P/L_i^P \) using

\[
E[L_i^P] = \sum_{k=1}^{K} w_k \sum_{j=i}^{I+1} p_j(t_k); \quad E[\psi_i^P] = \sum_{k=1}^{K} w_k p_i(t_k);
\]

\[
\text{Var}[L_i^P] = \sum_{k=1}^{K} w_k^2 \left[ \sum_{j=i}^{I+1} \text{Var} \left[ N_j(t_k)/N\pi_k \right] \right] = \sum_{k=1}^{K} w_k^2 \left[ \sum_{j=i}^{I+1} p_j(t_k)/N\pi_k \right]; \quad (17)
\]
\[
\text{Var}[\psi_i^p] = \sum_{k=1}^{K} w_k^2 \text{Var}\left[ \frac{N_i(t_k)}{N \pi_k} \right] = \sum_{k=1}^{K} w_k^2 \left[ p_i(t_k) / N \pi_k \right];
\]

\[
\text{Cov}[\psi_i^p, L_i^p] = \text{Var}[\psi_i^p].
\]

### 3.3 Method of Moments Estimators

As \( s \to 0 \) the weights in the LT calculations become closer to uniform, and this leads to a related set of estimators called the method of moments (MOM) estimators, also studied by Feigin, Tweedie and Blyea (1983). We set \( w_k(0) = [c_k - c_{k-1}] \) provided \( c_K < \infty \), and put \( \psi_i(0) = \left[ \sum_{k=1}^{K} w_k(0) N_i(t_k) / N_k \right] \). In this case we have

\[
E[\psi_i(0)] \approx \int_0^{c_K} p_i(t) dt
\]

\[
= \int_0^{c_K} P\{S_1 + \ldots + S_{i-1} < t < S_1 + \ldots + S_i\} dt
\]

\[
= \int_0^{c_K} P\{S_1 + \ldots + S_i > t\} dt - \int_0^{c_K} P\{S_1 + \ldots + S_{i-1} > t\} dt
\]

\[
\approx E[S_i]
\]

This does require \( c_K \) to be large enough that \( \int_{c_K}^{\infty} p_i(t) dt \) is negligible, in which case (18) gives a completely non-parametric estimate of the mean duration of the \( i \)th stage, regardless of the underlying stage density forms.

When \( g_i \) is Erlangian\((a_i)\), the MOM estimator is defined by

\[
\widehat{\theta}_i^M := a_i / \psi_i(0)
\]

and has been shown to be typically close to the optimal (in \( s \)) LT estimator (Feigin et al., 1983), at least for one-stage models.

As in the previous section, we now derive a form of the asymptotic variance of this estimator. Again using the delta method for the MOM estimator, we can write \( \psi_i(0) = f(\theta_i) = a / \theta_i \), and we have

\[
\text{Var} \left( \widehat{\theta}_i^M \right) = \text{Var}(\psi_i(0)) \frac{\theta_i^4}{a^2}.
\]

#### Multinomial sampling

When we have the multinomial structure we then have from (16)
\[ \text{Var}(\psi(t)) = \sum_{k=1}^{K} w_k^2 \text{Var} \left[ \frac{N(t_k)}{N_k} \right] = \sum_{k=1}^{K} w_k^2 \left[ \frac{p(t_k)(1 - p(t_k))}{N_k} \right]. \] (21)

**Poisson sampling**

In the case where sampling is Poisson we similarly need only

\[ \text{Var}(\psi^p(t)) = \sum_{k=1}^{K} w_k^2 \text{Var} \left[ \frac{N(t_k)}{N \pi_k} \right] = \sum_{k=1}^{K} w_k^2 \left[ \frac{p(t_k)}{N \pi_k} \right]. \] (22)

Thus the forms of both the estimator and its variance are relatively simple in the MOM case.

### 3.4 Missing initial data

It is possible that the initial numbers \( N_k \) or \( N \) are unknown, as in the grasshopper study (Read and Ashford, 1968; Ashford *et al.*, 1970). In the grasshopper data and also in the spring-tail data of van Straalen (1982) there is also an unobserved zero-stage population that feeds into the first-stage population. Thus even if there is a zero death rate one does not observe total population numbers and cannot estimate \( N_k \). This complicates analysis considerably.

One of the strengths of the Laplace transform method is that, if the sampling fraction \( \pi_k \) is known as in the Poisson case, then the form of \( \Delta_i \) enables one to factor out \( N \) and reduce \( N_k = N \pi_k \) to \( \pi_k \). In the grasshopper data the values of \( \pi_k \) are given in Ashford, Read and Vickers (1970). In fact, if the \( \pi_k \) are equal, as they often will be, then the common value also factors out given the form of the \( \Delta_i \), and does not need to be known. However, as we have seen we do need need knowledge of the initial values in order to calculate the approximate variances of the estimators.

A further positive property of these methods is that, even if the zero stage is unobserved, estimation of the parameter \( \theta_j \) using the LT or MOM methods depends only on the values observed in the stages \( j, j+1, \ldots, I+1 \). Thus we can estimate the duration of all later stages even if the numbers in the zero stage are not available. Again, we note that to estimate variances we require knowledge of the total numbers, which usually implies knowledge of the zero stage.

Perhaps surprisingly, as we now show, it is possible to estimate simultaneously both a common \( N \equiv N_k \) and the value of the parameter \( \theta_1 \) in this situation even if the initial stage data is missing. If we set

\[ \phi(s) = NL_2(s) = \sum_{k=1}^{K} w_k \sum_{j=2}^{I+1} N_j(t_k) \]
then we have from (7)

\[ E[\phi(s)] = N \beta_1(s). \]  

(23)

Suppose that the first stage duration is Erlangian(a). Choosing two values \( s_1, s_2 \) we find by equating \( \phi(s) \) with its approximate expectation as before that we have the estimates

\[ \hat{\theta}_1 = \frac{s_2[\phi(s_2)]^{1/a} - s_1[\phi(s_1)]^{1/a}}{[\phi(s_1)]^{1/a} - [\phi(s_2)]^{1/a}} \]

(24)

\[ \hat{N} = \phi(s_1) \left( \frac{\hat{\theta}_1 + s_1}{\hat{\theta}_1} \right)^a. \]

Approximate variances on these estimates are possible also, and we sketch the steps needed. Setting \( K = [\phi(s_2)/\phi(s_1)] \), we can rewrite (24) as

\[ \hat{\theta}_1 = \frac{s_2 K^{1/a} - s_1}{1 - K^{1/a}} \]

and as a function of \( \theta \) we have \( K = [(s_1 + \theta_1)/(s_2 + \theta_1)]^a \). We can then use a Taylor series expansion to write the variance of \( \hat{\theta}_1 \) in terms of the variance of \( K^{1/a} \), and this is a simple ratio of weighted sums of the \( N_j(t_k) \) which can be dealt with as before.

At first sight \( \hat{N} \) looks rather less tractable. However, substituting in the form of \( \hat{\theta}_1 \) shows that we have

\[ \hat{N}^{-1/a} = \frac{s_2 \phi(s_1)^{-1/a} - s_1 \phi(s_2)^{-1/a}}{s_2 - s_1} \]

and so repeated use of Taylor series gives an estimate of the variance of \( \hat{N} \) in terms of variances and covariances of \( \phi(s_1) \) and \( \phi(s_2) \).

In practice the most likely value of the results in this section will be in using the point estimator of \( N \) in the calculation of variances of the stage-dependent parameters, and so we omit these details.

**Remark:** The idea of using more than one value of \( s \) is developed in Leedow and Tweedie (1983). The best choice of these values of \( s \) is not obvious, and various ways have been suggested: minimizing residual sums of squares or maximizing likelihoods for the fitted values corresponding to different choices of \( s \) (Feigin et al., 1983; Leedow and Tweedie, 1983), or cross-validation approaches (Laurence and Morgan, 1987) have all been suggested. In general it has been found that choosing \( s_1 \) close to \( s_2 \) leads to good results, and so the search can be essentially one-dimensional along the diagonal of \( s \)-space: some theoretical justification for this has recently been found by Yao and Morgan (1999).
4 The effect of death rates

4.1 Estimating death rates

The work above is valid provided there is no death rate present in the population. If there is a death rate \( \mu_i \) in the \( i \)th stage then as in (3) we have that

\[
p_i^*(t) = h_1 \ast \ldots \ast h_{i-1} \ast H_i(t), \quad i = 1, \ldots, I + 1
\]  

(25)

where \( h_i(t) = \int g_i(t) e^{-\mu_t} dt, \quad H_i(t) = e^{-\mu t} \int_t^\infty g_i(x) dx, \) and \( H_{I+1}(t) = e^{-\mu t_{I+1}}. \)

If the \( \mu_i \) are not all equal then the Laplace transform methods do not seem well suited to estimation since the variables in (25) do not separate well, and other methods may be preferable (Hoeting and Tweedie, 2001). However, if \( \mu_i = \mu \) independent of stage then we may emulate the methods in Schuh and Tweedie (1979).

In order to estimate an overall death rate, note that we can ignore the stage in which any organism lies and count the total number of live organisms at time \( t_k \); and that this is binomial or Poisson with mean \( N_k \exp(-\mu t_k) \) or \( N \pi_k \exp(-\mu t_k) \), depending on the sampling scheme. Using our previous notation, we get

\[
E[L_1(s)] = \sum_{k=1}^{K} w_k(s) E[\sum_j N_j(t_k)/N_k] = \sum_{k=1}^{K} w_k(s) e^{-\mu t_k} \approx \int_0^\infty se^{-st} e^{-\mu t} dt = s/(s + \mu), \tag{26}
\]

where again we assume \( c_K \) is large enough that the tail of the integral is negligible. Hence in the multinomial case we may define

\[
\hat{\mu}^L := s[L_1(s)^{-1} - 1]
\]  

(27)

as a Laplace transform estimator of the death rate. In the Poisson case the same estimator can be used with \( L_1^P \) replacing \( L_1 \).

We follow the previous pattern to get an asymptotic variance expression for \( \hat{\mu}^L \). Since \( L_1 = s/(s + \mu) \), and \( \partial L_1/\partial \mu = -s/(s + \mu)^2 \),

\[
\text{Var}[\hat{\mu}^L] = \left[ \frac{(s + \hat{\mu})^4}{s^2} \right] \text{Var}[L_1]
\]  

(28)

where in the multinomial model

\[
\text{Var}[L_1] = \sum_k w_k^2 \text{Var} \left[ \sum_{j=1}^{I+1} N_j(t_k)/N_k \right] = \sum_k w_k^2 \left[ e^{-\hat{\mu} t_k} (1 - e^{-\hat{\mu} t_k})/N_k \right]. \tag{29}
\]

In the Poisson model we similarly have

\[
\text{Var}[\hat{\mu}^L] = \left[ \frac{(s + \hat{\mu})^4}{s^2} \right] \text{Var}[L_1^P] \quad \tag{30}
\]
and in this case

\[ \text{Var}[L^0_i] = \sum_k w_k^2 \left[ e^{-\hat{\mu} t_k} / N \pi_k \right]. \]  

(31)

### 4.2 Estimating stage parameters

Given an estimator for \( \mu \) we can now iteratively construct explicit estimators for the \( \theta_i \).

From (25) we have

\[ \int s e^{-su} p_i^*(u) du = s \beta_1(s + \mu) \ldots \beta_{i-1}(s + \mu) [1 - \beta_i(s + \mu)] / (s + \mu). \]  

(32)

Using (26) we have the estimating transform equation

\[ \beta_1(s + \mu) = 1 - \left( \frac{s + \mu}{s} \right) \int s e^{-su} p_i^*(u) du \]

\approx 1 - [1/L_1(s)] \psi_1(s) \]  

\approx L_2(s)/L_1(s); \]

(33)

inductively we then get as in the \( \mu = 0 \) case

\[ \hat{\beta}_i(s + \mu) = \Delta_i(s). \]  

(34)

This result is derived in Schuh and Tweedie (1979), although there is a typographical error in their equation.

When \( g_i \) is Erlangian(\( a_i \)) we know \( \beta_i(s + \mu) = (\theta_i / (s + \mu + \theta_i))^{a_i} \). Thus in this case, using (34) the Laplace transform estimator of \( \theta_i \) is now defined by

\[ \hat{\theta}_i^L := (s + \hat{\mu}) / ([\Delta_i(s)]^{-1/a_i} - 1). \]  

(35)

**Variance: Multinomial case**

To derive the variance of \( \hat{\theta}_i^L \), we write (35) as

\[ \hat{\theta}_i^L = (s + \hat{\mu}) / (s / ([\Delta_i(s)]^{-1/a_i} - 1)), \]

which from (26) and (11) is just

\[ \hat{\theta}_i^L = \hat{\theta}_i^0 / L_1, \]

where we use \( \hat{\theta}_i^0 \) to denote the form of the zero death rate estimator in (11). To handle this ratio in the multinomial case we need
(i) the means
\[ \mathbb{E}[\widehat{\theta}_i^0] = \frac{s}{s + \mu} \theta_i; \quad \mathbb{E}[L_1] = \frac{s}{s + \mu}; \]

(ii) \( \text{Var}[L_1] \) given by (29);

(iii) \( \text{Var}[\theta_i^0] \) which we have already calculated in (16) and (13) (with \( p_i^*(t) \) in place of \( p_i(t) \) throughout); and

(iv) \( \text{Cov}[\theta_i^0, L_1] \) which is given by
\[
\text{Cov} \left[ \frac{L_{i+1}}{L_i}, L_1 \right] = \frac{(\theta_i + s)^{a+1}}{a \theta_i^{a-1}},
\]
using a Taylor expansion. Now
\[
\text{Cov} \left[ \frac{L_{i+1}}{L_i}, L_1 \right] = \mathbb{E} \left[ \frac{L_{i+1}}{L_i} L_1 \right] - \left( \frac{\theta_i}{\theta_i + s + \mu} \right)^a \left( \frac{s}{s + \mu} \right)
\]
and this can be estimated by
\[
\mathbb{E} \left[ \frac{L_{i+1}}{L_i} L_1 \right] = \mathbb{E}[L_1 - \psi_i L_1 / L_i]
\]
\[
= \left( \frac{s}{s + \mu} \right) - \mathbb{E}[\psi_i (1 + \sum_{j<i} \psi_j / L_i)] \quad (37)
\]
\[
= \left( \frac{s}{s + \mu} \right) - \sum_k w_k p_i^*(t_k)
\]
\[
- \sum_{j<i} \left[ \sum_k w_k^2 p_i^*(t_k) p_j^*(t_k) [1 - 1/N_k] + \sum_k \sum_{m \neq k} w_k w_m p_i^*(t_k) p_j^*(t_m) \right] / \sum_{j \geq i} \sum_k w_k p_j^*(t_k)
\]

Variance: Poisson case

In the Poisson sampling case the calculations are similar. The means (i) are identical, the variance of \( L_i^p \) is given by (31), and \( \text{Var}[\theta_i^0] \) is as in (17), again with \( p_i^*(t) \) in place of \( p_i(t) \) throughout. However, in (iv) we need to change the final calculation to reflect the Poisson sampling, using the form
\[
\mathbb{E} \left[ \frac{L_{i+1}^p}{L_i^p} L_1 \right] = \left( \frac{s}{s + \mu} \right) - \sum_k w_k p_i^*(t_k)
\]
\[
- \frac{\sum_{j<i} \left[ \sum_k w_k p_i^*(t_k) \sum_m w_m p_j^*(t_m) \right]}{\sum_{j \geq i} \sum_k w_k p_j^*(t_k)}
\]

4.3 Missing initial data

Again there will be situations in which the initial data \( N \) or \( N_k \), and even the first stage numbers, are missing. Given the form of the estimators, it is clear that in principle the
comments at the end of Section 3.1 are also valid in this case. However, the existence of the death rate renders the computations somewhat more complex.

Consider first the situation where neither the initial (common) number \( N \) nor the death rate \( \mu \) is known. Setting, in this case, \( \Psi(s) = \sum_k w_k(s)[\sum_j N_j(t_k)] \), we have now

\[
E[\Psi(s)] = Ns/(s + \mu);
\]

choosing two values \( s_1, s_2 \) leads to the estimators

\[
\hat{\mu} = \frac{s_1 s_2 [\Psi(s_1) - \Psi(s_2)]}{s_1 \Psi(s_2) - s_2 \Psi(s_1)},
\]

\[
\hat{\mu} = \frac{s_1 s_2 [\Psi(s_1) - \Psi(s_2)]}{s_1 \Psi(s_2) - s_2 \Psi(s_1)};
\]

\[
\hat{\mu} = \Psi(s_1)(s_1 + \hat{\mu})/s_1.
\]

The variance of \( \hat{\mu} \) in this context can be estimated by noting that \( \hat{\mu} \) has the form

\[
\hat{\mu} = \sum_k w'_k(\sum_j N_j(t_k))/\sum_k w''_k(\sum_j N_j(t_k)),
\]

for suitably defined weights \( w'_k, w''_k \). Now in the multinomial case

\[
\text{Var}(\sum_k w'_k(\sum_j N_j(t_k))) = \sum_k (w'_k)^2 N e^{-\mu t_k}(1 - e^{-\mu t_k})
\]

and in the Poisson case

\[
\text{Var}(\sum_k w'_k(\sum_j N_j(t_k))) = \sum_k (w'_k)^2 N \pi k e^{-\mu t_k}
\]

These are comparatively straightforward since we do not have to worry about the stages of the organisms, and so the variance of \( \hat{\mu} \) in (39) has a relatively simple form using (15).

The form of \( \hat{\mu} \) can also be written using appropriate weights as

\[
\hat{\mu} = \frac{\sum_k w'_k(\sum_j N_j(t_k)) \sum_k w''_k(\sum_j N_j(t_k))}{\sum_k w''_k(\sum_j N_j(t_k))};
\]

approximating this variance is again possible and involves third moments of the Binomial(\( N, e^{-\mu t_k} \)) or Poisson(\( N \pi k e^{-\mu t_k} \)) variables involved. We omit the details.

The situation is considerably more complicated if one is also missing data on the first stage. However, using \( \phi(s) \) as in (23) we have

\[
E[\phi(s)] \approx \frac{N s}{s + \mu} \beta_1(s + \mu).
\]
Using three values \(s_1, s_2, s_3\) we find, if we assume that the first stage is Erlangian(a), the estimators

\[
\tilde{\theta}_1 = \frac{(s_2 + \bar{\mu})[\phi(s_2)(s_2 + \bar{\mu})/s_2]^{1/a} - (s_1 + \bar{\mu})[\phi(s_1)(s_1 + \bar{\mu})/s_1]^{1/a}}{[\phi(s_1)(s_1 + \mu)/s_1]^{1/a} - [\phi(s_2)(s_2 + \bar{\mu})/s_2]^{1/a}};
\]

\[
\bar{N} = \phi(s_1) \frac{s + \bar{\mu}}{s} \left[ \tilde{\theta}_1 + s + \bar{\mu} \right]^{1/a};
\]

an estimator of \(\mu\) is then given by solving the non-linear equation

\[
\frac{(s_2 + \mu)[\phi(s_2)(s_2 + \mu)/s_2]^{1/a} - (s_1 + \mu)[\phi(s_1)(s_1 + \mu)/s_1]^{1/a}}{(s_3 + \mu)[\phi(s_3)(s_3 + \mu)/s_3]^{1/a} - (s_1 + \mu)[\phi(s_1)(s_1 + \mu)/s_1]^{1/a}}
\]

\[
= \frac{[\phi(s_1)(s_1 + \mu)/s_1]^{1/a} - [\phi(s_2)(s_2 + \mu)/s_2]^{1/a}}{[\phi(s_1)(s_1 + \mu)/s_1]^{1/a} - [\phi(s_3)(s_3 + \mu)/s_3]^{1/a}}.
\]

The existence of a solution to this equation is not obvious. A simpler, more informal, approach is to take the data only from some point \(c_T\) where the first stage organisms have essentially all passed to stage two (or died). The data will usually indicate when such a point has been reached. Let

\[
\phi_T(s) = \sum_{k=T}^{K} w_k(s) \sum_{j=2}^{I+1} N_j(t_k).
\]

We then have, as in (38),

\[
E[\phi_T(s)] \approx \int_{c_T}^\infty Nse^{-st}e^{-\mu t}dt
\]

\[
= \frac{Ns}{s + \mu} e^{-(s + \mu)c_T}.
\]

Again choose two values \(s_1, s_2\). From (45) we have

\[
\frac{\phi_T(s_1)}{\phi_T(s_2)} = \frac{(s_2 + \mu)s_1}{(s_1 + \mu)s_2} e^{-(s_1 - s_2)c_T}
\]

so that as in (39) we can take

\[
\tilde{\mu}_T = \frac{s_1s_2[\phi_T(s_1)e^{s_1c_T} - \phi_T(s_2)e^{s_2c_T}]}{s_1\phi_T(s_2)e^{s_2c_T} - s_2\phi_T(s_1)e^{s_1c_T}};
\]

we can then estimate \(\theta_1\) from (42), and \(N\) by either (42) or

\[
\bar{N}_T = e^{(s_1 + \tilde{\mu}_T)c_T} \phi_T(s_1)(s_1 + \tilde{\mu}_T)/s_1.
\]

The variance of \(\tilde{\mu}_T\) can be approximated as a ratio of linear combinations of binomials or Poissons, as in (40). The form \(\bar{N}_T\) seems harder to handle, as does the form (42) for \(\theta_1\), and we do not pursue this here.
Table 1: Simulation data. The number of organisms \( N(t_k) \) in each stage.

<table>
<thead>
<tr>
<th>time</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
<th>Stage 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.52</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.94</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1.36</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1.79</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2.21</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2.63</td>
<td>0</td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>3.05</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3.47</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>3.89</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>4.31</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>4.74</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>5.16</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>5.58</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>6.00</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
</tr>
</tbody>
</table>

5 Evaluation on simulated data

5.1 Parameter estimates

In this section we evaluate both the accuracy of the estimates of the parameters produced by the LT and MOM approaches and the asymptotic variances of the estimators using simulated data.

To simulate data of a realistic nature, we generated 10 stage times from an Erlangian(2) distribution with \( \theta_i = 1.5 \) for each of 3 stages, at 15 equally spaced sampling time points between 0.1 and 6. Based on cumulative sums of the generated hatch times, we counted the number of observations in each stage at each sampling time, and these are given in Table 1.

Parameter estimates are given in Table 2 for various LT and MOM methods, using the count data in Table 1. We used equally spaced cut points about the observation times except that \( c_0 = 0, c_{15} = \infty \) for the LT estimates, and \( c_{15} = 6.21 \) for the MOM estimate. The LT estimate depends upon an unknown parameter \( s \). For the “iterated LT” estimate we used an iterative approach suggested in the last section of Schuh and Tweedie (1979, page 63) to choose \( s \) values of 0.68, 0.65, and 0.78 for stages 1–3, respectively. For the “minimum variance LT” estimate we found the \( s \) values that minimized the sum of the asymptotic
Table 2: Parameter estimates for simulated data. The MLE estimates are based on the simulated stage time data which are usually unobserved. Other estimates are based on the count data in Table 1. The estimated standard error is given in parentheses.

<table>
<thead>
<tr>
<th>Method</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>MLE</td>
<td>1.48 (0.09)</td>
<td>1.50 (0.09)</td>
<td>1.40 (0.08)</td>
</tr>
<tr>
<td>MOM</td>
<td>1.38 (0.17)</td>
<td>1.35 (0.23)</td>
<td>1.69 (0.21)</td>
</tr>
<tr>
<td>LT (iterated)</td>
<td>1.37 (0.19)</td>
<td>1.29 (0.25)</td>
<td>1.57 (0.42)</td>
</tr>
<tr>
<td>LT (min. var.)</td>
<td>1.37 (0.16)</td>
<td>1.32 (0.22)</td>
<td>1.54 (0.34)</td>
</tr>
</tbody>
</table>

Table 3: Residual sum of squares for curves in Figure 3.

<table>
<thead>
<tr>
<th>Method</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
<th>Stage 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.0516</td>
<td>0.1080</td>
<td>0.1641</td>
<td>0.2299</td>
</tr>
<tr>
<td>MLE</td>
<td>0.0489</td>
<td>0.1073</td>
<td>0.1625</td>
<td>0.2105</td>
</tr>
<tr>
<td>MOM</td>
<td>0.0398</td>
<td>0.0936</td>
<td>0.1771</td>
<td>0.2119</td>
</tr>
<tr>
<td>LT (iterated)</td>
<td>0.0395</td>
<td>0.0957</td>
<td>0.1695</td>
<td>0.1990</td>
</tr>
<tr>
<td>LT (min. var.)</td>
<td>0.0396</td>
<td>0.0941</td>
<td>0.1660</td>
<td>0.1990</td>
</tr>
</tbody>
</table>

variance for the three parameters to choose the $s$ values of 0.20, 0.34, and 0.46 for stages 1–3, respectively. Note that these $s$ values are rather smaller than the $s$ values in the ad-hoc iterative scheme, but both approaches lead to very similar parameter estimates.

The maximum likelihood estimates (MLE) given in Table 2 are based on the actual organism-by-organism stage time data: these are usually unobserved, and the difference in magnitude of the standard errors indicates the loss of efficiency in moving to the censored data.

The MOM and LT estimates are all within one standard error of the true value, and generally appear to produce similar estimates of the inverse mean hatch time $2/\theta_i$ in each stage. MOM and LT overestimate the parameter for the final stage. The estimated standard errors for the MOM and LT estimates are, as expected, larger than the standard errors for the MLE, reflecting the additional uncertainty when the stage times are unobserved.

For the MOM and LT approaches, the estimated probability that an organism is in each stage across time is shown in Figure 3 along with the true probability curves (equation 1 evaluated at $\theta = 1.5$). These approaches produce good estimates of the true proportion in
stage 1 but seem to underestimate the highest observed proportions in stages 2–4.

Table 3 gives the residual sum of squares (RSS) for the curves in Figure 3. The RSS from the true probability curves reflect the random error in the simulation. All of the RSS results are quite similar, but it is noteworthy that the transform methods typically give smaller residuals than the MLE methods. This might be expected given that the transform methods, in matching areas under curves, are working with quantities closely related to the RSS.

5.2 Accuracy of the variance approximations

To investigate the accuracy of the variance approximations, we generated 100 data sets with the structure of Table 1. We first compare the empirical standard errors of the MOM estimates with their asymptotic theoretical values \( \hat{\sigma}(\hat{\theta}) \), since these are not dependent upon
Table 4: Accuracy of the MOM estimates. For 100 simulations, the mean and the standard deviation of the parameter estimates and the mean of the asymptotic estimates of the standard error (13).

<table>
<thead>
<tr>
<th>Stage</th>
<th>mean of ( \hat{\theta} )</th>
<th>s.d. of ( \hat{\theta} )</th>
<th>mean of ( \hat{\sigma}(\hat{\theta}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.53</td>
<td>0.20</td>
<td>0.17</td>
</tr>
<tr>
<td>2</td>
<td>1.53</td>
<td>0.22</td>
<td>0.23</td>
</tr>
<tr>
<td>3</td>
<td>1.71</td>
<td>0.30</td>
<td>0.28</td>
</tr>
</tbody>
</table>

so a comparison is simpler. The asymptotic estimates of standard error seem acceptably close to the empirical standard deviation of the 100 LT estimates, as shown in Table 4, although there is an indication of upward bias in the MOM estimates of \( \theta_3 \).

The top row of plots in Figure 4 shows the LT estimate from (11) averaged across the 100 simulations. For the first stage, estimates of \( \theta_1 \) are fairly accurate for a wide range of \( s \) values. For the other two stages, the parameter estimates are much more sensitive to \( s \), and appear unstable for \( s \) near 0. For this simulation context, choosing the \( s \) value that minimizes the asymptotic variance results in a very accurate estimate of \( \theta \) for all three stages. The iterated method is less accurate, but it still produces reasonable estimates.

The bottom row of plots in Figure 4 compares the standard deviation of the 100 LT estimates to the mean of the asymptotic standard errors from the 100 simulations for a range of \( s \) values. For very small \( s \) values the asymptotic estimate of the variance is poor, but it produces a good estimate of the variance for a moderate range of \( s \) values, which in this example covers the \( s \)-values that would be used for either the minimum variance or the iterative methods.
Figure 4: LT estimates from 100 simulations. The top row shows the mean of the LT estimates from 100 simulations. The horizontal line in the top row of plots corresponds to the true value of $\theta = 1.5$. The dashed line shows the average $s$ value that minimized the asymptotic variance. The dotted line shows the $s$ value where the iterated LT estimate would converge if $\hat{\theta} = 1.5$. The bottom row of plots shows the s.d. of the LT estimates from 100 simulations (solid line) and the mean of the asymptotic standard error (the square root of equation 13) computed for each simulation (dotted line).

6 Multi-stage Growth Examples

6.1 Parasitic nematode data

The life cycle for the cattle parasite *Ostertagia ostertagi* includes several stages of eggs and larva. Young *et al.* (1980a) estimate the parameters for the stage time distributions for this parasite using the data shown in Figure 1. We shall consider these data, given in Table 5,
Table 5: Data for parasite example. The proportion of observations in each stage of *Ostertagia ostertagi*, expressed as a percentage. Stage 1 = eggs (unembryonated and embryonated), stage 2 = 1st stage larvae, stage 3 = 2nd stage larvae, stage 4 = 3rd stage larvae. Time is in hours.

<table>
<thead>
<tr>
<th>Time</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
<th>Stage 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>87</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>12</td>
<td>81</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>65</td>
<td>11</td>
<td>13</td>
<td>74</td>
<td>2</td>
</tr>
<tr>
<td>90</td>
<td>6</td>
<td>7</td>
<td>65</td>
<td>14</td>
</tr>
<tr>
<td>115</td>
<td>14</td>
<td>2</td>
<td>54</td>
<td>21</td>
</tr>
<tr>
<td>140</td>
<td>10</td>
<td>2</td>
<td>24</td>
<td>61</td>
</tr>
<tr>
<td>160</td>
<td>0</td>
<td>0</td>
<td>17</td>
<td>67</td>
</tr>
<tr>
<td>185</td>
<td>0</td>
<td>0</td>
<td>33</td>
<td>51</td>
</tr>
<tr>
<td>210</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td>260</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>70</td>
</tr>
</tbody>
</table>

Table 6: Parameter estimates for parasite example. The estimated standard error is given in parentheses.

<table>
<thead>
<tr>
<th>Method</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
<th>Stage 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOM</td>
<td>0.044 (0.002)</td>
<td>0.078 (0.007)</td>
<td>0.028 (0.001)</td>
<td></td>
</tr>
<tr>
<td>LT (iterated)</td>
<td>0.047 (0.004)</td>
<td>0.075 (0.011)</td>
<td>0.021 (0.001)</td>
<td></td>
</tr>
<tr>
<td>LT (min. var.)</td>
<td>0.042 (0.002)</td>
<td>0.068 (0.006)</td>
<td>0.021 (0.001)</td>
<td></td>
</tr>
</tbody>
</table>

Table 7: Residual sum of squares for estimated curves in Figure 5.

<table>
<thead>
<tr>
<th>Method</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
<th>Stage 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOM</td>
<td>0.1477</td>
<td>0.3614</td>
<td>0.2579</td>
<td>0.0827</td>
</tr>
<tr>
<td>LT (iterated)</td>
<td>0.1365</td>
<td>0.3493</td>
<td>0.2858</td>
<td>0.1143</td>
</tr>
<tr>
<td>LT (min. var.)</td>
<td>0.1665</td>
<td>0.3696</td>
<td>0.3489</td>
<td>0.1691</td>
</tr>
</tbody>
</table>

which were read from Figure 2B of Young et al. (1980a). The data, which were adjusted to have zero death rate, are the proportion of organisms in each stage out of approximately 606 eggs in the initial population per time period.

Following Young et al. (1980a), we used Erlangian(2) distributions for the stage time
distribution for all three stages. We used equally spaced cut points about the observation times with $c_0 = 0$, $c_{10} = \infty$ for LT, and $c_{10} = 285$ for MOM. As in Section 5.1 we used both the iterative approach and the minimum variance approach to choose the unknown parameter $s$. For the iterative approach, the $s$ values for stages 1–3 were 0.0233, 0.0377, and 0.0107, respectively; for the minimum variance approach, the $s$ values for stages 1–3 were 0.0029, 0.0034, and 0.0091, respectively.

The parameter estimates and corresponding estimated curves for the MOM and LT estimates are shown in Table 6 and Figure 5, respectively. All three methods underestimate the observed rather sudden high proportion of the population in stage 2 at 45 hours, and it is clear that this particular Erlangian distribution is probably inappropriate for stage 2. This lack of fit for stage 2 is also reflected in the RSS (Table 7). The minimum variance approach has the highest RSS for all stages, suggesting that this approach for selecting $s$ values may
be more influenced by outlying values than is the iterated approach. The MOM estimated curves for all four stages are shown on Figure 1, and although the data are sparser in this example, the fit curves do describe the data as well as one might hope.

6.2 Grasshopper Data

Our final example describes stages of grasshopper development shown in Figure 2. The goal is to model the life cycle of the grasshopper, *Chorthippus parallelus*. The life cycle for the grasshopper starts at the egg stage and progresses through four interim stages, called instars, before reaching the final adult stage. A grasshopper may die due to various causes at any stage of its development. The data are taken from Ashford, Read and Vickers (1970) (hereafter, ARV), and in this case no information about initial numbers or maturation or death rates of unhatched eggs is available. ARV describe results for 1964 and 1965, and we focus here on the 1965 data set. Additional analysis of the 1965 data is given in Read and Ashford (1968).

The data were collected on 37 days between May 14 and September 27, 1965 in Devon, England, over a 2500 m$^2$ area where conditions were considered to be reasonably homogeneous. On each sampling date, all the live grasshopper were removed from a number of small areas which were chosen at random by reference to a rectangular grid laid out over the area. On the first 29 sampling dates 0.0028 of the area was sampled and 0.004 of the area was sampled thereafter. Results given here are adjusted so all dates use the 0.0028 sampling fraction. More details on the sampling are given in ARV.

Our model is similar to ARV's model where the probability that the organism is in stage $i$ at time $t$ is described in (3), $i = 1, \ldots, 6$, and the $i$th stage time distribution (2) follows an Erlangian(3) distribution. We use the Poisson model described in Section 2.2 to model these data where $\pi_k=0.0028$, $k = 1, \ldots, 37$.

In their analysis ARV ignore the mortality of eggs (i.e., they assume zero death rate in the egg stage) and assume a constant death rate after the egg stage. In contrast, we assume a constant death rate for all stages. As described in Section 4.3, we can estimate the parameters for the first stage even though we did not observe any data for this stage. We adopted the ARV estimate of May 10 as the time when the season begins for the 1965 data set.

In analyzing this model, our estimate for $\theta_1$, the unobserved hatch stage parameter, is from (42); the estimates for $\theta_2, \ldots, \theta_5$ are based on (35); and the estimates for $\mu$ and $N$ are based on (46) and (47), respectively. The value of $s_1 = .03$ was chosen based on the value
Table 8: Comparison of parameter estimates for grasshopper example. The “hatch” and “death rate” parameters are not equivalent as ARV assumed no death rate for the “hatch” stage. The notation under “parameters” follows this paper. The estimated standard error is given in parentheses where available.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>$N$</th>
<th>6690</th>
<th>(969)</th>
<th>17224</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hatch</td>
<td>$\theta_1$</td>
<td>0.13</td>
<td>(0.015)</td>
<td>0.08</td>
</tr>
<tr>
<td>Instar 1</td>
<td>$\theta_2$</td>
<td>0.17</td>
<td>(0.010)</td>
<td>0.23</td>
</tr>
<tr>
<td>Instar 2</td>
<td>$\theta_3$</td>
<td>0.22</td>
<td>(0.030)</td>
<td>0.21</td>
</tr>
<tr>
<td>Instar 3</td>
<td>$\theta_4$</td>
<td>0.18</td>
<td>(0.015)</td>
<td>0.16</td>
</tr>
<tr>
<td>Instar 4</td>
<td>$\theta_5$</td>
<td>0.16</td>
<td>(0.020)</td>
<td>0.17</td>
</tr>
<tr>
<td>Adult</td>
<td></td>
<td>0.03</td>
<td>(&gt;0.36)</td>
<td></td>
</tr>
<tr>
<td>Death rate</td>
<td>$\mu$</td>
<td>0.021</td>
<td>(0.004)</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Table 9: Residual sum of squares for estimated curves in Figure 6.

<table>
<thead>
<tr>
<th>Method</th>
<th>Instar 1</th>
<th>Instar 2</th>
<th>Instar 3</th>
<th>Instar 4</th>
<th>Adult</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT</td>
<td>75</td>
<td>34</td>
<td>25</td>
<td>22</td>
<td>36</td>
<td>192</td>
</tr>
<tr>
<td>MLE</td>
<td>104</td>
<td>36</td>
<td>28</td>
<td>20</td>
<td>35</td>
<td>223</td>
</tr>
</tbody>
</table>

that gave the smallest total residual sum of squares, with $s_2 = 1.1 \times s_1$: note we do not have all variances available due to the complexity of the missing data, so we cannot minimize the variances to get $s$. We compare our Laplace transform estimates of the parameters to the MLE estimates of ARV, who derive these using complex closed form expressions of (3) derived as part of the Ph.D. dissertation of one of the authors. In more complicated situations, this approach is not feasible.

The parameter estimates from the two methods are given in Table 8. The two approaches yield very different estimates of the starting population ($N$), which is largely explained by the fact that ARV assumed a zero death rate for the hatch stage. The differences between the hatch parameter estimates will also be affected by differences in the modeling of the death rate parameters.

The overall death rate estimates are relatively close. The parameter estimates for instar 1 are rather different, as discussed further below, but the parameter estimates for the remaining stages are similar. The estimated standard errors for instars 1-4 from the LT method are
Figure 6: Observed number of grasshoppers and estimated curves for grasshopper data for stages.

considerably larger than the estimated standard errors from the MLE approach. While this may be due in part to the approximations in the LT approach, the estimated curves in Figure 6 suggest that the LT standard error estimates may better reflect reality. Simulations performed under the conditions for the 1965 data set also indicate that the LT parameter estimates presented here are reasonable.

The estimated curves for all five stages are shown in Figure 6 and the corresponding
residual sum of squares (RSS) are given in Table 9. The curves and RSS results are quite similar for all but the instar 1 stage. This stage is difficult to fit as there are three small observations around 20 days after May 10 and two large observations around 40 days after May 10. The MLE approach models the large observations better than the LT approach, but is penalized in the RSS for over-estimating the small values. The estimated curves from both methods underestimate the maximum of the observed grasshoppers in each stage. This suggests that this particular Erlangian distribution may not reflect the sharpness of the maturation process well, and this is born out by attempts to fit the same model to the 1964 data, where the fit was worse.

We conclude that the RSS criterion for deciding on s-values for fitting the data appears to mesh well with the concept of equating weighted areas under the maturation curves. Overall, the LT method seems to give results which fit at least as well to the data as the MLE method, and with considerably less work. Its extension to more general choice of a is straightforward, and for particular data sets this would be a fruitful approach to pick up more rapidly spiking population movements.

References


