MATHEMATICAL BIOSCIENCES
AND ENGINEERING
Volume **5**, Number **4**, October **2008**

doi:10.3934/mbe.2008.5.647

pp. **647–667**

QUANTIFYING UNCERTAINTY IN THE ESTIMATION OF PROBABILITY DISTRIBUTIONS

H.T. Banks and Jimena L. Davis

Center for Research in Scientific Computation, North Carolina State University
Raleigh, North Carolina 27695-8205

Abstract. We consider ordinary least squares parameter estimation problems where the unknown parameters to be estimated are probability distributions. A computational framework for quantification of uncertainty (e.g., standard errors) associated with the estimated parameters is given and sample numerical findings are presented.

1. **Introduction and motivation.**

The importance of estimating time and spatially dependent function parameters as coefficients in distributed parameter models has been recognized for some time [16]. This is especially true when one is trying to determine mechanistic-based terms in a model. General theoretical and computational ideas (called *function space estimation convergence* or *FSPEC* in [16]) for approximation schemes for such problems were developed some years ago and now are used somewhat routinely by practitioners. A diverse range of examples involving systems of the form

$$\frac{\partial u}{\partial t} + V \cdot \nabla u = \nabla \cdot (D \nabla u) - \mu u \tag{1}$$

for the state variables $u = u(t, x)$ is discussed in Chapter 7 of [16], where parameters to be estimated are generally vector *functions* of the form $q = (D, V, \mu)$ and are to be chosen from some set Q of admissible parameter functions. As summarized in [16], spatially dependent coefficients $D = D(x)$ are used in [18] to study the effects of bioturbation on volcanic ash records in core samples from deep sea sediments. Functional coefficients are also needed in the insect dispersal studies of [13, 14] where vegetation effects on dispersal lead to spatially dependent advection $V = V(x)$ and time-dependent emigration/immigration $\mu = \mu(t)$ terms are important in capture-mark-release flea beetle experiments (these are used to characterize "initial disturbance" effects due to the trauma from capture, handling, etc.). Similar studies involving time-dependent anemotaxis ($V = V(t)$) and emigration/immigration ($\mu = \mu(t)$) in cabbage root fly dispersal [25] are described in [15].

In these problems one uses data $\{y_k\}$ for the parameter dependent model values $u(\tau_k; q)$ (where typically $\tau_k = (t_i, x_j)$ are time/spatial covariates) to estimate

2000 *Mathematics Subject Classification.* 35L60,62E20,62F25,62G15,65M32,92D25.

Key words and phrases. approximation, asymptotic standard error theory, confidence bands, parameter estimation, probability distributions, size-structured populations.

Acknowledgements: This research was supported in part by the US Department of Energy Computational Science Graduate Fellowship to J.L. Davis under grant DE-FG02-97ER25308 and in part by the National Institute of Allergy and Infectious Disease under grant 9R01AI071915-05.

functions $q \in Q$. The data $\{y_k\}$ can be regarded as a realization of the observation process

$$Y_k = u(\tau_k; q_0) + \epsilon_k, \quad k = 1, \ldots, n, \tag{2}$$

where the ϵ_k are measurement or observation errors and q_0 are underlying "true" parameters (assumed to exist in theoretical formulations). This leads to estimates $\hat{q}$ defined by

$$\hat{q} = \arg\min_{q \in Q} \sum_{k=1}^{n} [u(\tau_k; q) - y_k]^2 \tag{3}$$

and corresponding ordinary least squares (OLS) estimator

$$q_{OLS}(Y) = \arg\min_{q \in Q} \sum_{k=1}^{n} [u(\tau_k; q) - Y_k]^2, \tag{4}$$

which is a Q-space valued random variable. The distribution of this infinite dimensional random variable (called the "sampling distribution") is a probability distribution on Q and is of great interest since knowledge of this will lead to information about the uncertainty associated with the estimates $\hat{q}$. In finite dimensional problems, there is a rather complete asymptotic theory to provide such results (see Chapter 12 of [30]). The major focus of our interest here is the development of an infinite dimensional analogue.

Another class of problems to which such an infinite dimensional theory would be immediately applicable is that involving estimation of parameters in the Fokker-Planck or forward Kolmogorov equation [1, 24] for transition probabilities $p(s, y; t, x)$ for the stochastic diffusion process $X(t)$ for a growth process

$$\frac{\partial p}{\partial t} + \frac{\partial [a(t,x)p]}{\partial x} = \frac{1}{2} \frac{\partial^2 [b(t,x)p]}{\partial x^2}. \tag{5}$$

Here $a(t, x)$, the "drift" or mean growth rate, and $b(t, x)$, the "diffusion" or second moment of the rate of increase, are the functional parameters $q = (a, b)$ to be estimated. Because the population density $u(t, x)$, where growth is assumed to be a stochastic diffusion process, also satisfies such an equation (see [28]), this model can be used as a stochastic alternative (e.g., see [19]) to the Sinko-Streifer deterministic growth model [5, 27]

$$\frac{\partial v}{\partial t} + \frac{\partial}{\partial x}(g(t,x)v) = -\mu(t,x)v. \tag{6}$$

The estimation of time-dependent mortalities in these equations is important in recent problems for sublethal effects of pesticides [2, 3] in insect populations where constant parameters μ prove inadequate in describing population life data.

In this note we consider another class of estimation problems where the functions to be estimated are actually probability distributions or densities. As explained in detail below, this class of problems arises when one assumes that a probability distribution describes the distribution of growth rates g in the model (6). Such formulations are called *Growth Rate Distribution (GRD)* models [7, 8, 11, 12]. Before introducing these models, we give a summary of the finite dimensional asymptotic distribution theory for which we seek a function space analogue.

2. **Overview of asymptotic standard error theory for finite dimensional parameters.** We briefly outline the standard statistical framework for asymptotic

distributions of finite dimensional ordinary least squares (OLS) estimators [22, 23, 26, 30]. We begin by considering the following nonlinear statistical model

$$Y_j = Y(\bar{x}_j) = f(\bar{x}_j, \theta_0) + \epsilon_j, \quad j = 1, \ldots, n, \tag{7}$$

where $\bar{x}_j$ is a vector in $\mathbb{R}^n$, $f(\bar{x}_j, \theta_0)$ represents the mathematical model, and θ_0 is a vector in the constraint set $\Theta \subset \mathbb{R}^{M+1}$ that represents the "true" parameter value. We also note the assumption that the ϵ_j are *i.i.d.* with mean 0 and constant variance σ_0^2, where $\sigma_0^2 > 0$ represents the "true" variance. Generally, θ_0 and σ_0^2 are not known but are estimated by the parameters θ and σ^2, respectively. Since ϵ_j is a random variable, Y_j is also a random variable with

$$E[Y_j] = f(\bar{x}_j, \theta_0) \quad \text{and} \quad \mathrm{Var}[Y_j] = \sigma_0^2.$$

The following OLS estimator (which is also a random variable denoted here by $\theta_{OLS} = \theta_{OLS}(Y)$) is used in the inverse problem for the estimation of θ :

$$\theta_{OLS} \equiv \arg\min_{\theta \in \Theta} \sum_{j=1}^{n} \left(Y_j - f(\bar{x}_j, \theta)\right)^2. \tag{8}$$

As $n \to \infty$, the sampling distribution for a random variable $\theta_{OLS}(Y)$ is given by the multivariate normal distribution; i.e.,

$$\theta_{OLS}(Y) \sim \mathcal{N}_{M+1}(\theta_0, \sigma_0^2[\mathcal{X}^T(\theta_0)\mathcal{X}(\theta_0)]^{-1}) \approx \mathcal{N}_{M+1}(\theta_0, \Sigma_0^n),$$

where $\mathcal{X}(\theta) = \mathcal{X}^n(\theta) = \dfrac{\partial F}{\partial \theta}(\theta) = F_\theta(\theta)$ is the $n \times (M+1)$ sensitivity matrix with elements

$$\mathcal{X}_{jk}(\theta) = \frac{\partial f(\bar{x}_j, \theta)}{\partial \theta_k},$$

and Σ_0^n is a covariance matrix approximated below in (9). As we noted, θ_0 is generally unknown; however, we can determine an estimate $\hat{\theta}$ for θ_0 using the OLS estimator. For a particular realization (data set) $\{y_j\}$ the estimates $\hat{\theta}$ minimize

$$\sum_{j=1}^{n} \left(y_j - f(\bar{x}_j, \theta)\right)^2.$$

We can also compute an estimate for σ_0^2 (which is also usually unknown) using the following estimate $\hat{\sigma}^2$:

$$\sigma_0^2 \approx \hat{\sigma}^2 = \frac{1}{n-(M+1)} \sum_{j=1}^{n} \left(y_j - f(\bar{x}_j, \hat{\theta})\right)^2.$$

The estimates $\hat{\theta}$ and $\hat{\sigma}^2$ are used in computing an estimate of the covariance matrix Σ_0^n :

$$\Sigma_0^n \approx \hat{\Sigma} = \hat{\sigma}^2[\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})]^{-1}. \tag{9}$$

We then determine the standard errors for the estimates $\hat{\theta}$ by computing

$$SE(\hat{\theta}_k) = \sqrt{\hat{\Sigma}_{kk}}, \quad k = 0, \ldots, M.$$

Confidence intervals for the finite dimensional parameter $\hat{\theta}$ are constructed using the standard errors. The endpoints of the confidence intervals are given by

$$\hat{\theta}_k \pm t_{1-\alpha/2} SE(\hat{\theta}_k), \quad k = 0, \ldots, M,$$

where $t_{1-\alpha/2}$ is a distribution value that depends on the level of significance that is chosen [21]. After the level of significance is chosen, we determine the corresponding

$t_{1-\alpha/2}$ value from a statistical table for the Student's t-distribution. The confidence intervals constructed in this manner provide us with a means of quantifying the uncertainty in the estimates obtained from the estimation procedure constructed from a realization of Y. In the following section, we will present some computational results in which we have used this asymptotic standard error theory to compute nodal confidence intervals for finite dimensional parameters.

3. Computational example: Size-structured mosquitofish population.

We next present some computational results demonstrating the construction of confidence intervals for finite dimensional parameters based on the asymptotic theory for OLS estimators discussed briefly in the previous section. These computations were carried out in MATLAB and are based on simulated data that will be described shortly. Additional results for this example along with a more detailed discussion can be found in [8].

3.1. Mathematical model.

The computational results presented in this section and the next section involve the estimation of growth-rate distributions for size-structured mosquitofish populations. We use the Growth Rate Distribution (GRD) model, a modification of the Sinko-Streifer (SS) model, to describe this population [7, 11]. The Sinko-Streifer model [31], which is used to model both age and size-structured populations, for the mosquitofish population is given by

$$\begin{aligned} \frac{\partial v}{\partial t} + \frac{\partial}{\partial x}(gv) &= -\mu v, \quad x_0 < x < x_1, \quad t > 0 \\ v(0,x) &= \Phi(x) \\ g(t,x_0)v(t,x_0) &= \int_{x_0}^{x_1} K(t,\xi)v(t,\xi)d\xi \\ g(t,x_1) &= 0. \end{aligned} \tag{10}$$

We note here that $v(t,x)$ represents the size or population density (with units of number per size class), t represents time, and x represents the size, or length, of the mosquitofish. The number of mosquitofish in the population at time t with sizes between x_0 and x_1 is

$$N(t) = \int_{x_0}^{x_1} v(t,x)dx.$$

The growth rate of the individual mosquitofish is given by $g(t,x)$, where

$$\frac{dx}{dt} = g(t,x)$$

for each individual, and the mortality rate of the mosquitofish is given by $\mu(t,x)$. The initial condition at $t = 0$ is given by the initial size density function $\Phi(x)$. The boundary condition at $x = x_0$ represents the recruitment, or birth, rate and is in terms of the fecundity kernel $K(t,x)$. At $x = x_1$ the boundary condition ensures the maximum size of the mosquitofish is x_1.

All individual mosquitofish of the same size are assumed to have the same growth rate in the SS model. However, with this assumption, solutions to (10) do not exhibit the dispersion and bifurcation of the population density observed in data collected from rice fields where mosquitofish have been used in the place of chemicals to control mosquito populations. To capture the features of dispersion and bifurcation typical of the mosquitofish population, the SS model was modified so that the individual growth rates of the mosquitofish vary across the population [7, 11, 12]. The GRD model [7, 11] is given by

$$u(t,x;P) = \int_G v(t,x;g)dP(g), \tag{11}$$

where $v(t,x,;g)$ is the solution to (10) with growth rate g, G is the collection of admissible growth rates, and P is a probability measure on G. Based on work in [7], the admissible growth rates are assumed of the form

$$g(x;b,\gamma) = \begin{cases} b(\gamma - x), & x_0 \le x \le \gamma, \\ 0, & \text{otherwise}, \end{cases}$$

where the intrinsic growth rate and maximum size of the mosquitofish is represented by b and $\gamma = x_1$, respectively. To satisfy the assumption of varying growth rates, we assume that b and γ are random variables that belong to compact sets B and Γ, respectively. The collection of admissible growth rates is then characterized as

$$G = \{g(x;b,\gamma) | b \in B, \gamma \in \Gamma\},$$

where both B and Γ are bounded closed intervals (i.e., compact sets in $\mathbb{R}^2$). In the following computational results we set $\gamma = 1$ and assume that the family of growth rates is parameterized only by the intrinsic growth rates b.

3.2. **Approximation methods.** We are interested in determining the growth-rate distribution P^* that gives the best fit of the underlying model to the data. However, this parameter estimation problem involves both an infinite dimensional state space (u) and an infinite dimensional parameter space (the space $\mathcal{P}$ of probability measures). Therefore computationally efficient approximation methods are important for this purpose. We will now briefly discuss the different approximation methods that we have previously considered in the inverse problem for the estimation of the growth-rate distributions of the mosquitofish population. A more thorough discussion of these methods can be found in [8].

In the first approach that we considered for this problem, we used the standard parametric approach based on the assumption that we have a priori knowledge about the exact form of the probability distribution on the growth rates of the mosquitofish. We will discuss later the simulated data used in our computations, which was generated with a bi-Gaussian distribution. The bimodality typically seen in the mosquitofish data has been attributed to the fact that male and female mosquitofish grow to different maximum sizes. Thus, male and female mosquitofish must grow at different rates. Previous simulations [7] demonstrated that an assumption of a bi-Gaussian distribution on the growth rates leads to both dispersion and bifurcation, qualitative features characteristic of mosquitofish data. Therefore, we choose to use the following bi-Gaussian probability density function p

$$p(b;\bar{b}_1,\sigma_{b_1}^2,\bar{b}_2,\sigma_{b_2}^2) = \frac{\exp\left\{-\frac{(b-\bar{b}_1)^2}{2\sigma_{b_1}^2}\right\}}{2\sqrt{2\pi\sigma_{b_1}^2}} + \frac{\exp\left\{-\frac{(b-\bar{b}_2)^2}{2\sigma_{b_2}^2}\right\}}{2\sqrt{2\pi\sigma_{b_2}^2}}, \tag{12}$$

where the parameters $(\bar{b}_1,\bar{b}_2)$ and $(\sigma_{b_1}^2,\sigma_{b_2}^2)$ represent the means and variances, respectively, of the bi-Gaussian distribution on the intrinsic growth rates b. (A more general formulation along with an additional Gaussian example can be found in [9].) Since P is continuous, we note that the GRD model (11) becomes

$$u(t,x;\theta) = \int_B v(t,x;g(x;b))p(b;\theta)db, \tag{13}$$

where $\theta = (\bar{b}_1, \sigma_{b_1}^2, \bar{b}_2, \sigma_{b_2}^2)$ are the parameters that are associated with the a priori probability density and distribution. We will denote this approach by PAR(M,N), where M is one less than the number of parameters in θ and N is the number of quadrature nodes used in approximating the integral above with the composite trapezoidal rule [29]. Although we will find estimates for all four parameters, we will set M= 3 so when using the asymptotic standard error theory as outlined in the previous section the correct factor will be used in our computations. The ordinary least squares problem that we wish to solve for $\hat{\theta}$ is given by

$$\hat{\theta} = \arg \min_{\theta \in \mathbb{R}_+^{M+1}} J(\theta) = \sum_{i,j} |u(t_i, x_j; \theta) - \hat{u}_{ij}|^2, \tag{14}$$

where $\{\hat{u}_{ij}\}$ is the data. After determining an optimal value for θ, we can then use this value in the bi-Gaussian probability density function (12) to determine the estimated probability density and distribution.

The other two methods that we consider are nonparametric approaches that do not require any assumptions with respect to the form of the probability distribution. Based on work in [4] and [17], we are guaranteed convergence (in the Prohorov metric [6, 20]) of distributions with the families of approximating functions that we will now discuss. The first method, involving delta functions and which we denote by DEL(M), has also been discussed and used in [11] and [12]. We use M+1 delta functions in this approximation method. The form of the approximating probability distributions $\mathcal{P}^M$ placed on the growth rates is assumed to be piecewise constant on the collection of admissible growth rates G^M, where $G^M = \{g_k^M\}_{k=0}^M$. We note that $g_k^M(x; b_k^M) = b_k^M(1-x)$ for $k = 0, 1, \dots, M$. This method leads to the following approximation for $u(t,x;P)$ in (11):

$$u(t,x;\{p_k^M\}) = \sum_{k=0}^{M} v(t,x;g_k^M)p_k^M, \tag{15}$$

where $v(t,x;g_k^M)$ is the subpopulation density from (10) with growth rate g_k^M and p_k^M is the probability that an individual is in subpopulation k with growth rate g_k^M.

The second nonparametric approximation scheme involves the use of piecewise linear spline functions to approximate the density $P' = \dfrac{dP}{db} = p(b)$. Using piecewise linear splines in the place of delta functions we provide a much smoother approximation of (11) when the "true" probability distribution on the growth rates of the mosquitofish is continuous. Denoting this method as SPL(M,N), where M+1 is the number of basis elements (splines) used to approximate the distribution on the growth rates and N is the number of quadrature nodes used to approximate the integral found below in (16), $u(t,x;P)$ from (11) is approximated by

$$u(t,x;\{a_k^M\}) = \sum_{k=0}^{M} a_k^M \int_B v(t,x;g(x;b)) l_k^M(b)db, \tag{16}$$

where $g(x;b) = b(1-x)$ and $p_k^M(b) = a_k^M l_k^M(b)$ is the probability density for individuals in subpopulation k. The piecewise linear spline functions are represented by l_k^M. The composite trapezoidal rule was also used to approximate the integral in (16).

When using the two nonparametric approaches, we observe that the estimates for the growth-rate distribution are determined by solving the following least squares problem

$$\begin{aligned}\min_{P\in\mathcal{P}^M(G)} J(P) &= \sum_{i,j} |u(t_i,x_j;P)-\hat{u}_{ij}|^2 \\ &= \sum_{i,j}\left(u(t_i,x_j;P)^2 - 2u(t_i,x_j;P)\hat{u}_{ij} + (\hat{u}_{ij})^2\right),\end{aligned} \tag{17}$$

where $\{\hat{u}_{ij}\}$ is again the data and $\mathcal{P}^M(G)$ is the finite dimensional approximation to $\mathcal{P}(G)$. The finite dimensional approximation $\mathcal{P}^M(G)$ when using DEL(M) is given by

$$\mathcal{P}^M(G) = \left\{P\in\mathcal{P}^M(G) | P' = \sum_k p_k^M \delta_{b_k^M}, \sum_k p_k^M = 1\right\}, \tag{18}$$

where $\delta_{b_k^M}$ is the delta function with an atom at b_k^M. When using SPL(M,N), the finite dimensional approximation $\mathcal{P}^M(G)$ to the probability measure space $\mathcal{P}(G)$ is given by

$$\mathcal{P}^M(G) = \left\{P\in\mathcal{P}(G) | P' = \sum_k a_k^M l_k^M(b), \sum_k a_k^M \int_B l_k^M(b)db = 1\right\}. \tag{19}$$

We note that the least squares problem in (17) reduces to the constrained quadratic programming problem [11, 12]

$$F(\mathbf{p}) \equiv \mathbf{p}^T\mathbf{A}\mathbf{p} + 2\mathbf{p}^T\mathbf{b} + c, \tag{20}$$

which is minimized over $\mathcal{P}^M(G)$, where $\mathbf{p}$ is the vector containing $p_k^M, 0\le k\le M$, or $a_k^M, 0\le k\le M$ when using DEL(M) or SPL(M,N), respectively. Additional details on this formulation can be found in [8, 11, 12]. We remark that we had to include non-negativity constraints on the coefficients $\{p_k^M\}$ and $\{a_k^M\}$, as well as the last constraint in (18) and (19) in the programming of the inverse problem.

Before presenting the results from our simulations, we define the functions and variables used in the asymptotic standard error theory outlined in the previous section. We begin by noting that $\{\bar{x}_j\}_{j=1}^n$ corresponds to $(t_l, x_m), l = 1,\dots,n_t$, $m = 1,\dots,n_x$ pairs, where n_t and n_x represent the number of time and size values, respectively, used in generating the data ($n = n_t\cdot n_x$). The parameter θ, which will be estimated with each method, is finite dimensional and is given by $\theta = (\bar{b}_1, \sigma_{b_1}^2, \bar{b}_2, \sigma_{b_2}^2)$ for PAR(M,N), $\theta = \{p_k^M\}_{k=0}^M$ for DEL(M), and $\theta = \{a_k^M\}_{k=0}^M$ for SPL(M,N). The mathematical model $f(\bar{x}_j,\theta_0)$ in the statistical model is also approximated differently for each method considered here. When using PAR(M,N), we have that

$$f(\bar{x}_j,\theta_0) \approx u(\bar{x}_j;\theta) = \int_B v(\bar{x}_j;g)p(b;\theta)db.$$

However, when using DEL(M), we note that

$$f(\bar{x}_j,\theta_0) \approx u(\bar{x}_j;\{p_k^M\}) = \sum_{k=0}^M v(\bar{x}_j;g_k^M)p_k^M,$$

where $g_k^M(x;b_k^M) = b_k^M(1-x)$. When using SPL(M,N), we find that

$$f(\bar{x}_j,\theta_0) \approx u(\bar{x}_j;\{a_k^M\}) = \sum_{k=0}^{M} a_k^M \int_B v(\bar{x}_j;g) l_k^M(b)db.$$

Lastly, we remark that the entries in the sensitivity matrix $\mathcal{X}(\theta)$ are also different for the different methods that we consider here. Recall that the elements of the $n \times (M+1)$ sensitivity matrix $\mathcal{X}(\theta)$ are given by

$$\mathcal{X}_{jk}(\theta) = \frac{\partial f(\bar{x}_j,\theta)}{\partial \theta_k}.$$

When using the parameterized OLS method PAR(M,N), we find that the sensitivity elements in $\mathcal{X}(\theta)$ are given by

$$\mathcal{X}_{jk}(\theta) = \frac{\partial f(\bar{x}_j,\theta)}{\partial \theta_k} = \int_B v(\bar{x}_j;g)\frac{\partial p(b;\theta)}{\partial \theta_k}db.$$

The entries in $\mathcal{X}(\theta)$ for DEL(M) are given by

$$\mathcal{X}_{jk}(\theta) = \frac{\partial f(\bar{x}_j,\theta)}{\partial \theta_k} = v(\bar{x}_j;g_k^M),$$

where the growth rate $g_k^M(x;b_k^M) = b_k^M(1-x)$. We see that the sensitivity elements for SPL(M,N) are given by

$$\mathcal{X}_{jk}(\theta) = \frac{\partial f(\bar{x}_j,\theta)}{\partial \theta_k} = \int_B v(\bar{x}_j;g) l_k^M(b)db.$$

Using these expressions for the corresponding methods, we are able to compute estimates of the covariance matrix Σ_0^n and then compute standard errors for the estimates $\hat{\theta}_k$. We are then able to compute nodal confidence intervals for the estimated parameter $\hat{\theta}$. As noted earlier, the endpoints of the nodal confidence intervals are given by

$$\hat{\theta}_k \pm t_{1-\alpha/2} SE(\hat{\theta}_k), \quad k = 0,\ldots,M, \tag{21}$$

where $t_{1-\alpha/2}$ is a distribution value that is determined from a statistical table for Student's t-distribution based on the level of significance that is chosen [21]. For the following simulations, we chose to use $\alpha = 0.05$ for a significance level of 95%, which corresponds to $t_{1-\alpha/2} = 1.96$ when the number of degrees of freedom is large, i.e., $n \geq 30$.

3.3. **Simulated data and computational results.** We now describe the simulated population density data used in the inverse problem for the estimation of growth-rate distributions for the mosquitofish model. We began by first choosing a true distribution P^* on the growth rates $g(x;b)$, where again $g(x;b) = b(1-x)$ and b represents the intrinsic growth rate of the mosquitofish. Recalling the assumption of the GRD model (11), we note that the growth rates of the mosquitofish vary among the population. Therefore, we assumed that b is a random variable with distribution P^*. Using this assumption, we generated a collection of admissible growth rates $G_I = \{g_0, g_1, \ldots, g_I\}$ with a corresponding distribution P_I^*, where we took $I = 128$. For the simulations shown here, we used an "approximate" truncated bi-Gaussian distribution on the intrinsic growth rates b. The bi-Gaussian distribution was an average of two Gaussian distributions with means $\bar{b}_1 = 3.3$ and $\bar{b}_2 = 5.7$ and equal variances $\sigma_{b_1}^2 = \sigma_{b_2}^2 = 0.492$. The values for b were sampled from

$B = [\bar{b}_1 - 3\sigma_{b_1}^2, \bar{b}_2 + 3\sigma_{b_2}^2]$. We were interested only in the growth-rate distribution of the mosquitofish, so we let $\mu = K = 0$ in the SS model (10) with initial size density

$$\Phi(x) = \begin{cases} \sin^2(10\pi x), & 0 \leq x \leq 0.1, \\ 0, & 0.1 < x \leq 1. \end{cases}$$

We then created simulated data exhibiting both dispersion and bifurcation by first solving the SS model (10) for each individual $g_i \in G_I$ using the method of characteristics and then computing

$$u_d(t, x; P_I^*) = \int_{G_I} v(t, x; g) dP_I^*(g) = \int_B v(t, x; g) p_I^*(b; \bar{b}_1, \sigma_{b_1}^2, \bar{b}_2, \sigma_{b_2}^2) db,$$

where $p_I^*(b; \bar{b}_1, \sigma_{b_1}^2, \bar{b}_2, \sigma_{b_2}^2)$ is the bi-Gaussian probability density function corresponding to the true bi-Gaussian probability distribution P_I^*. The integral above is approximated via the composite trapezoidal method with 128 quadrature nodes [29]. We took 50 uniformly spaced time values in the interval [0,0.5] and 50 uniformly spaced size values from the normalized range [0,1). We then added random absolute noise to the simulated data

$$\hat{u}(t, x; P_I^*) = u_d(t, x; P_I^*) + \eta \cdot \epsilon,$$

where η represents the noise level constant and ϵ represents normally distributed random values with mean 0 and variance 1. Therefore, the simulated data used in this estimation problem was of the form discussed in the previous section.

Table 1. Estimated $\bar{b}_1$, $\bar{b}_2$, $\sigma_{b_1}^2$, and $\sigma_{b_2}^2$ and confidence intervals for bi-Gaussian example with 10% absolute error when using PAR(3,128)

Theoretical CI	Computed CI
$\bar{b}_1^* \pm 1.96SE(\bar{b}_1^*)$	3.2756 ± 0.0367
$(\sigma_{b_1}^2)^* \pm 1.96SE((\sigma_{b_1}^2)^*)$	0.5342 ± 0.1042
$\bar{b}_2^* \pm 1.96SE(\bar{b}_2^*)$	5.7057 ± 0.0252
$(\sigma_{b_2}^2)^* \pm 1.96SE((\sigma_{b_2}^2)^*)$	0.5793 ± 0.1358

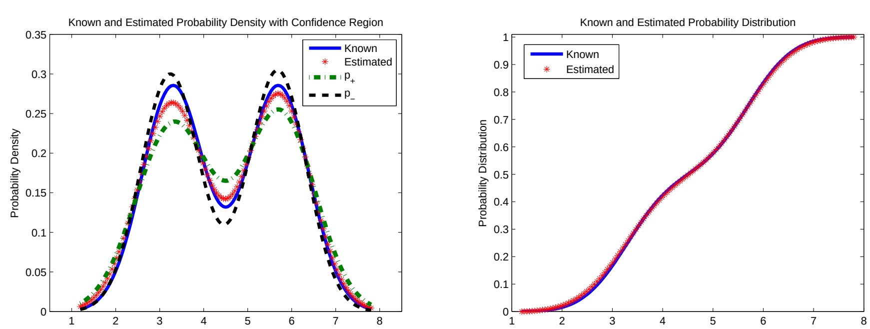

Figure 1. Estimated probability density with confidence region and probability distribution given a true bi-Gaussian distribution using PAR(3,128) to estimate the subpopulation means and variances with 10% absolute error

The first set of results shown were obtained using PAR(3,128) for the estimation of $\theta = (\bar{b}_1, \sigma_{b_1}^2, \bar{b}_2, \sigma_{b_2}^2)$ using simulated data with 10% absolute noise. Table 1 contains the optimal estimated values for θ along with the corresponding confidence intervals for each component of θ. Using this approach, the optimal cost value $J(\hat{\theta})$ is 4.1238, and the estimated variance of the system $\hat{\sigma}^2$ is 0.0017. In Figure 1, we see the known probability density and distribution used to generate the simulated data as well as the estimated probability density and distribution using the optimal estimates obtained from the inverse problem. Also shown in Figure 1 are the probability densities using the lower and upper endpoints of the confidence intervals for each of the components of θ. Based on the statistical theory outlined above, we are 95% confident that intervals constructed using PAR(3,128) would "cover" θ_0. We note that the confidence intervals are relatively small for the means in comparison to those corresponding to the variances. Based on the size of these confidence intervals in relation to the estimated parameter values, we would infer that the variances of the growth-rate distribution are more sensitive to noisy data. We feel more certain about the estimates obtained for the means from this procedure due to the smaller confidence intervals for these parameters. The confidence intervals constructed here give us an idea of the uncertainty associated with the estimated parameter θ but do not give us any indication of the uncertainty associated with the estimated probability distribution, which is the parameter of interest in our original problem.

TABLE 2. Estimated parameter values and confidence intervals for bi-Gaussian example with 10% absolute error when using DEL(8) and SPL(8,128)

p_k^M	DEL(8)	a_k^M	SPL(8,128)
p_0^8	0.1733 ± 0.0197	a_0^8	0.0818 ± 0.0302
p_1^8	0.1465 ± 0.0174	a_1^8	0.0389 ± 0.0189
p_2^8	0.1615 ± 0.0155	a_2^8	0.2378 ± 0.0161
p_3^8	0.1501 ± 0.0137	a_3^8	0.2517 ± 0.0144
p_4^8	0.1044 ± 0.0119	a_4^8	0.1132 ± 0.0128
p_5^8	0.1022 ± 0.0105	a_5^8	0.2608 ± 0.0114
p_6^8	0.1022 ± 0.0100	a_6^8	0.2434 ± 0.0101
p_7^8	0.0479 ± 0.0084	a_7^8	0.0432 ± 0.0095
p_8^8	0.0120 ± 0.0050	a_8^8	0.0045 ± 0.0115

We also used the delta function approximation method and the spline based approximation method in the inverse problem with the same data set used above with PAR(3,128). The optimal estimates along with the corresponding confidence intervals are given in Table 2 for DEL(8) and SPL(8,128). We note that the optimal cost using DEL(8) is 31.3867, while the optimal cost when using SPL(8,128) is 4.1282. The estimates of σ_0^2 for DEL(8) and SPL(8,128) are 0.0126 and 0.0017, respectively. Figure 2 shows the plots of the estimated probability densities and nodal confidence intervals for both DEL(8) and SPL(8,128). Also shown in Figure 2 are the estimated probability distributions that were constructed by using the estimates of $\{p_k^M\}$ and $\{a_k^M\}$ for DEL(8) and SPL(8,128), respectively. We point out again that these confidence intervals correspond to the finite dimensional parameters that we have estimated by solving the OLS problem. However, we are interested

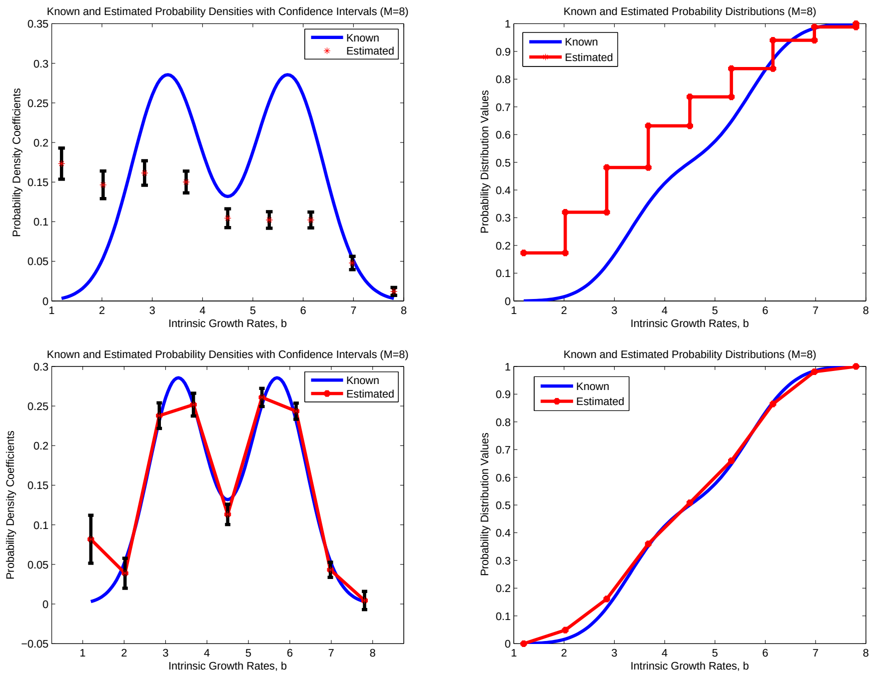

FIGURE 2. Estimated probability densities with confidence intervals and probability distributions given a true bi-Gaussian distribution using DEL(8)(top) and SPL(8,128)(bottom) with 10% absolute error

in making remarks about the uncertainty associated with the estimated probability distributions. In the following section, we will outline how to construct confidence bands for the estimated probability distributions based on the confidence intervals computed using the standard error theory for the finite dimensional parameters.

4. **Extension of asymptotic standard error theory to functional parameters: Computational results.** In the previous section, we demonstrated how to construct nodal confidence intervals for finite dimensional parameters (i.e., $\{p_k^M\}_{k=0}^M$, $\{a_k^M\}_{k=0}^M$, and θ) using the standard asymptotic theory for OLS estimators. The finite dimensional parameters that we determined by solving the inverse problem were at the level of the probability density. As shown in the previous section, we constructed estimates of the parameter of interest in our original problem (the probability distribution) by using the estimates of the probability density obtained from the inverse problem. While we can use the standard error theory that has already been established to quantify the uncertainty associated with the estimates of the finite dimensional parameters, we cannot apply this same theory to the estimated probability distributions, which are in an infinite dimensional setting. Since standard error theory does not exist for problems with functional parameters, we would like to develop the mathematical and asymptotic statistical theory for OLS problems where the parameter of interest is a probability distribution. In this section we again focus on a bi-Gaussian example, referring the reader to [9] for similar discussions for a Gaussian example. We will provide computational results displaying

the concept of confidence bands that will aid in quantifying the uncertainty in the estimated probability distributions.

To construct confidence bands for the estimated probability distributions, we use the confidence intervals obtained for the finite dimensional parameters. We will first discuss and present the results obtained using the standard parametric approach PAR(M,N). When using PAR(M,N), we use an a priori probability density in the GRD model (11), which we assume is continuous. After using the standard error theory to compute a confidence interval for θ, we construct a confidence band for the estimated probability distribution by using the endpoints of the confidence interval in the known probability density function (pdf). We note that the confidence region for the estimated probability density is formed by plotting

$$p_- = p(b; \hat{\theta} - 1.96SE(\hat{\theta})) \quad \text{and} \quad p_+ = p(b; \hat{\theta} + 1.96SE(\hat{\theta})),$$

where $\hat{\theta}$ represents the estimates of θ that solve the OLS problem. Then, using the fact that the probability density function p also represents the derivative of the probability distribution function P, we construct the upper confidence band for the estimated probability distribution by using the portions of p_- and p_+ that lie above the estimated probability density when this function is increasing (i.e., the slope is positive). When the estimated probability density is decreasing and the slope is negative, the portions of p_- and p_+ that lie below the estimated probability density are used to construct the upper confidence band. We use this same technique to create the lower confidence band by using the portions of p_- and p_+ that lie below (above) the estimated probability density when the slope is positive (negative). We integrate over these values and then normalize by an appropriate factor so that the confidence bands are "true" distributions (integrate to 1). Using this method, we obtained the following results with PAR(3,128) from the inverse problem using the simulated data described earlier with 20% absolute noise. A larger percentage of absolute noise was added to the simulated data for the results in this section so the reader could differentiate visually between the estimated probability distribution and the confidence bands with PAR(3,128). Results were also obtained with 10% absolute noise; however, the confidence bands were much tighter around the estimated probability distribution. The optimal cost for this set of results is 14.6131, while the estimate of $\hat{\sigma}^2$ is 0.0059. We also computed the condition number of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$, which is used in computing the standard errors for the finite dimensional parameter θ, and obtained a value of 35.0084. Table 3 contains the optimal estimate of θ as well as the corresponding confidence intervals. In Figure 3 the plots of the known and estimated probability densities along with p_- and p_+ are shown on the left as well as the plots of the known and estimated probability distributions and corresponding confidence bands on the right. We note that the estimated probability distribution lies within the confidence bands constructed using the technique that we have just outlined.

When using the nonparametric approaches, DEL(M) and SPL(M,N), the confidence intervals computed using the standard error theory correspond to the weights, $\{p_k^M\}_{k=0}^M$ and $\{a_k^M\}_{k=0}^M$, used in the approximations. In some cases, the lower confidence endpoints for these estimated weights may be negative, which violates the non-negativity condition required of probability densities (see results for SPL(8,128) in Figure 2). Thus, before constructing the confidence band for the estimated probability distribution, we first truncate any negative values to zero in order to have a "true" density. We then note if the estimated probability density is monotone

TABLE 3. Estimated $\bar{b}_1$, $\bar{b}_2$, $\sigma^2_{b_1}$, and $\sigma^2_{b_2}$ and confidence intervals for bi-Gaussian example with 20% absolute error when using PAR(3,128)

Theoretical CI	Computed CI
$b_1^* \pm 1.96SE(b_1^*)$	3.0979 ± 0.0711
$(\sigma^2_{b_1})^* \pm 1.96SE((\sigma^2_{b_1})^*)$	0.5464 ± 0.1935
$\bar{b}_2^* \pm 1.96SE(\bar{b}_2^*)$	5.6611 ± 0.0480
$(\sigma^2_{b_2})^* \pm 1.96SE((\sigma^2_{b_2})^*)$	0.5926 ± 0.2575

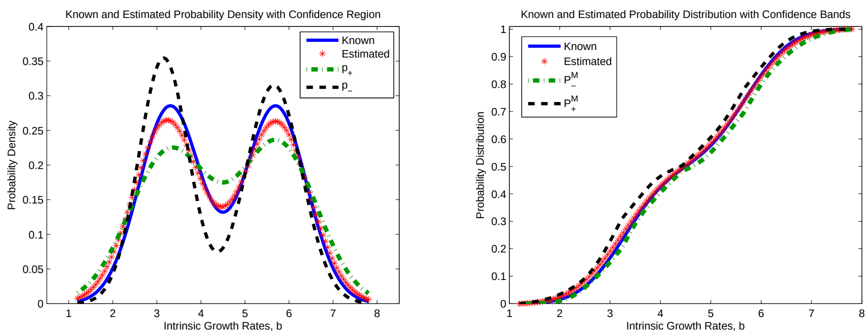

FIGURE 3. Estimated probability density and probability distribution with confidence region and confidence bands given a true bi-Gaussian distribution using PAR(3,128) to estimate the subpopulation means and variances with 20% absolute error

and increasing, the upper (lower) confidence band for the estimated distribution is constructed by integrating over the upper (lower) confidence interval endpoints and normalizing by an appropriate factor so that the confidence band is a "true" probability distribution. In the case that the estimated probability density is not monotone (which is the case in the examples shown here), the construction of the confidence bands using DEL(M) and SPL(M,N) again depends on the slope of the estimated probability density. The technique employed in these cases mimics that described when using PAR(M,N). The upper (lower) confidence band is created by integrating over the upper (lower) confidence interval endpoints when the slope of the estimated probability density is positive and the lower (upper) confidence interval endpoints when the slope is negative. We again normalize by an appropriate factor so that the confidence bands for the estimated probability distribution are also "true" distributions.

We first present some of the results obtained using DEL(M) for various values of M in the estimation problem using the same data set with 20% absolute noise. The optimal cost values, estimates of $\hat{\sigma}^2$, and condition numbers κ of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ can be found in Table 4 for M$= 8, 12, 16, 24, 32, 48,$ and 64. Figures 4 through 10 show the estimated probability densities and confidence intervals as well as the estimated probability distributions and confidence bands. As the value of M is increased, we observe the optimal cost and estimate of σ_0^2 decrease, which we expect because we are allowing more degrees of freedom. The estimated probability distribution converges to the known distribution as M is increased. However, we also note

TABLE 4. Optimal cost values, $\hat{\sigma}^2$, and condition number of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ for bi-Gaussian example with 20% absolute error when using DEL(M)

M	J^*	$\hat{\sigma}^2$	$\kappa(\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta}))$
8	40.6264	0.0163	15.6702
12	31.9494	0.0128	16.2206
16	25.8704	0.0104	16.9203
24	20.0946	0.0081	19.0560
32	15.9340	0.0065	22.1191
48	14.3649	0.0059	53.5280
64	14.2212	0.0058	105.5634

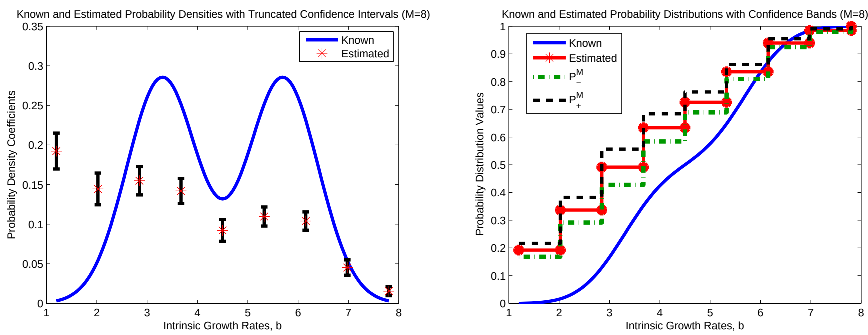

FIGURE 4. Estimated probability density and probability distribution with confidence intervals and bands given a true bi-Gaussian distribution using DEL(8) with 20% absolute error

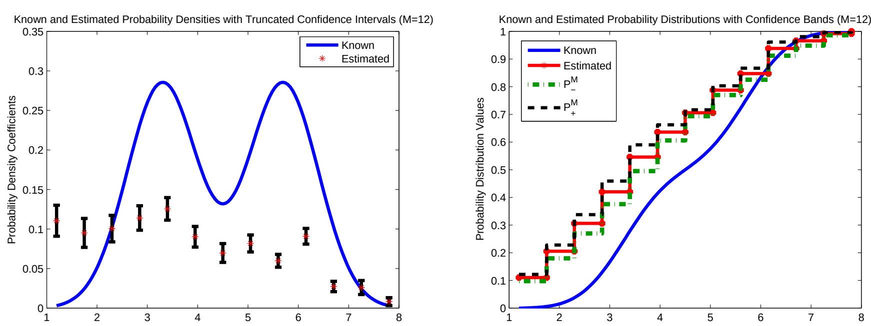

FIGURE 5. Estimated probability density and probability distribution with confidence intervals and bands given a true bi-Gaussian distribution using DEL(12) with 20% absolute error

from Table 4 that as M is increased, $\kappa(\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta}))$ increases. Once M becomes too large, the problem becomes over-parametrized and ill-conditioned (exhibited by the larger condition numbers of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$), and we observe the confidence bands become larger. As M is increased from 8 to 32, the confidence bands appear to

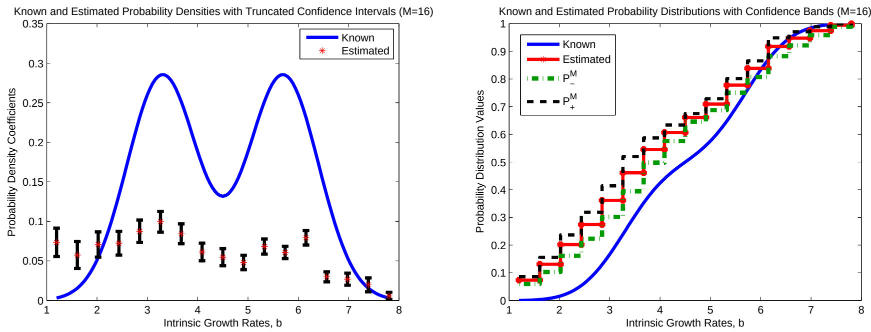

FIGURE 6. Estimated probability density and probability distribution with confidence intervals and bands given a true bi-Gaussian distribution using DEL(16) with 20% absolute error

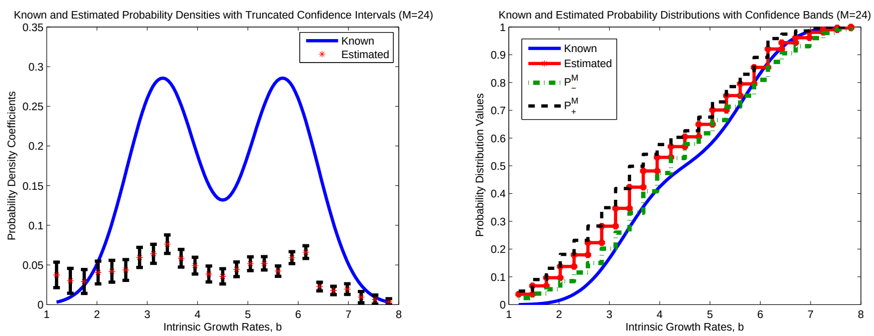

FIGURE 7. Estimated probability density and probability distribution with confidence intervals and bands given a true bi-Gaussian distribution using DEL(24) with 20% absolute error

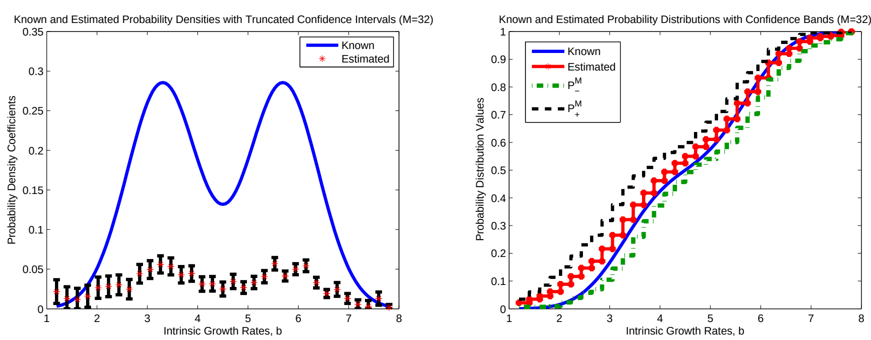

FIGURE 8. Estimated probability density and probability distribution with confidence intervals and bands given a true bi-Gaussian distribution using DEL(32) with 20% absolute error

be converging nicely; however, when M is increased from 32 to 48 and from 48 to 64, we no longer observe nice convergence of the confidence bands. However,

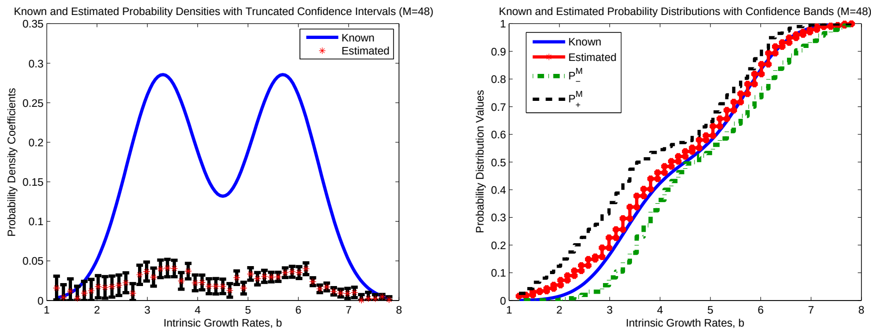

FIGURE 9. Estimated probability density and probability distribution with confidence intervals and bands given a true bi-Gaussian distribution using DEL(48) with 20% absolute error

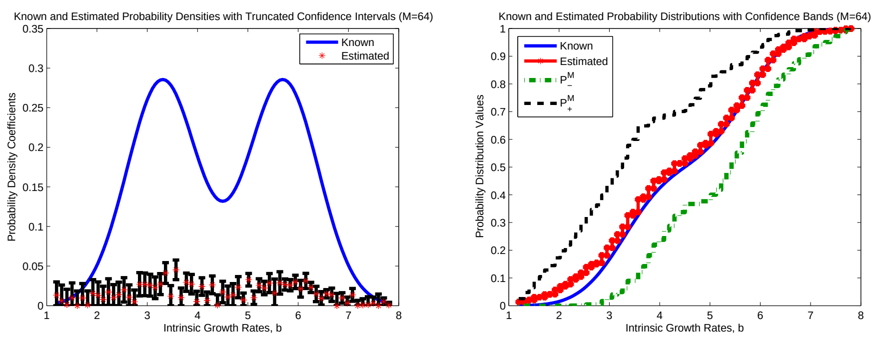

FIGURE 10. Estimated probability density and probability distribution with confidence intervals and bands given a true bi-Gaussian distribution using DEL(64) with 20% absolute error

by examining the condition number of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$, we can better understand the behavior of the confidence bands, which appear to converge nicely until the problem becomes over-parametrized (beyond $M = 32$).

We also obtained computational results for the inverse problem using SPL(M,128) using the data set with 20% absolute noise for various values of M. For M$= 8, 12, 16, 24,$ and 32, we report the optimal cost values, the estimates $\hat{\sigma}^2$, and the conditions numbers of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ in Table 5. The figures displaying the estimated probability densities with the nodal confidence intervals as well as the estimated probability distributions with the functional confidence bands for these values of M are shown in Figures 11 through 15. Since we added absolute noise instead of relative noise to the simulated data used in the inverse problem calculations shown here, the tails of the estimated probability density functions (pdfs) are very poor. We observe the same type of behavior in the confidence bands here as noted when using DEL(M).

As M is increased, there is a small decrease in the optimal cost. The decrease in the estimate of the variance of the system $\hat{\sigma}^2$ is so small that it is not noticeable when reported to only four significant digits. We also note the increase in $\kappa(\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta}))$ as M is increased, and again, we can use this to explain the behavior we observe in the confidence bands constructed for these values of M. The confidence

Table 5. Optimal cost values, $\hat{\sigma}^2$, and condition number of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ for bi-Gaussian example with 20% absolute error when using SPL(M,128)

M	J^*	$\hat{\sigma}^2$	$\kappa(\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta}))$
8	14.5734	0.0059	22.4873
12	14.5058	0.0058	31.3596
16	14.4767	0.0058	40.2284
24	14.4577	0.0058	63.5889
32	14.3953	0.0058	91.0741

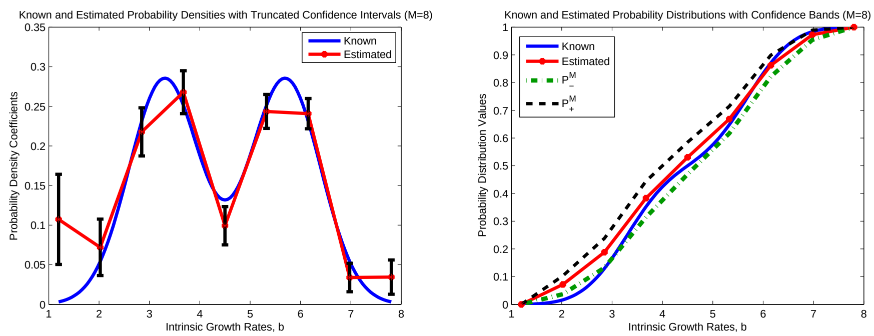

Figure 11. Estimated probability density and probability distribution with confidence intervals and bands given a true bi-Gaussian distribution using SPL(8,128) with 20% absolute error

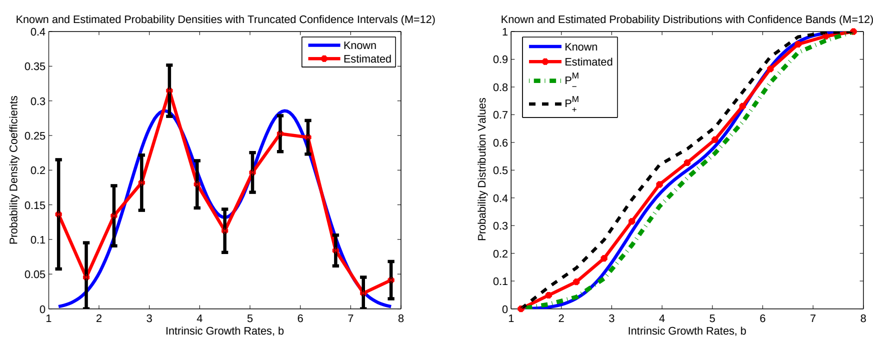

Figure 12. Estimated probability density and probability distribution with confidence intervals and bands given a true bi-Gaussian distribution using SPL(12,128) with 20% absolute error

bands appear to be converging nicely as M is increased from 8 to 16. However, the confidence bands begin to grow larger as M is increased beyond 16, which is also accompanied by a much larger increase in the condition number of $\mathcal{X}^T(\hat{\theta})\mathcal{X}(\hat{\theta})$ for the values of M above 16. Over-parametrization of the inverse problem does not only affect the estimates obtained but the confidence bands as well. However, for

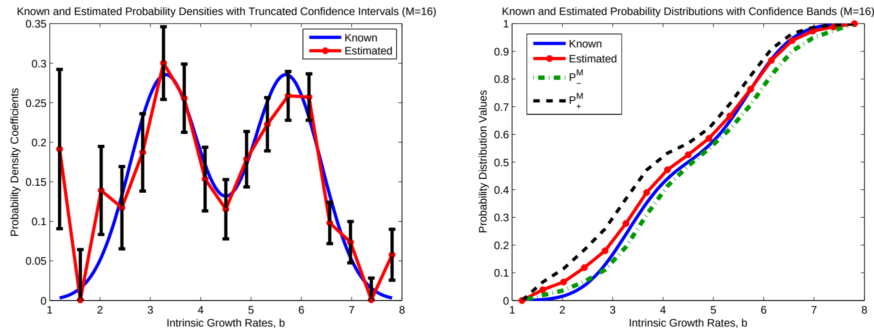

FIGURE 13. Estimated probability density and probability distribution with confidence intervals and bands given a true bi-Gaussian distribution using SPL(16,128) with 20% absolute error

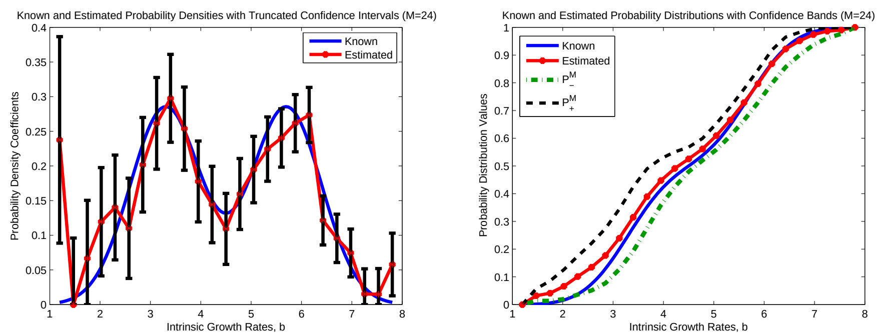

FIGURE 14. Estimated probability density and probability distribution with confidence intervals and bands given a true bi-Gaussian distribution using SPL(24,128) with 20% absolute error

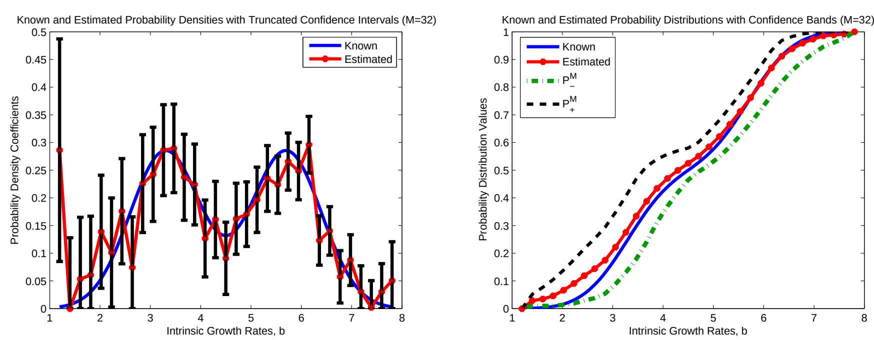

FIGURE 15. Estimated probability density and probability distribution with confidence intervals and bands given a true bi-Gaussian distribution using SPL(32,128) with 20% absolute error

appropriately chosen values of M, we observe very nice convergence of the confidence bands constructed using the technique outlined above for the approximation methods DEL(M) and SPL(M,N).

Other examples illustrating this behavior can be found in [9]. Moreover, the ideas outlined here can also be adopted to treat many other (non probability density) functional parameter estimation problems including those described in the Introduction.

5. **Summary and concluding remarks.** In this note we presented computational and statistical results for both parametric and nonparametric versions of the inverse problem for the estimation of growth-rate distributions in size-structured mosquitofish populations. The results discussed here demonstrate some of the strengths and weaknesses associated with each method. When the form of the probability distribution is known a priori, the parametric approach PAR(M,N) is the better method to use because the number of parameters to be estimated is typically small so computations are usually not very expensive. We also observed much tighter confidence bands around the estimated probability distributions in this example; hence, we are fairly confident about the reliability of estimates obtained with this method. However, the accuracy of the parameter estimates relies heavily on one's ability to correctly specify the form of the probability distribution a priori.

In contrast, the nonparametric approaches are a better choice when one cannot (correctly) identify the form of the probability distribution. The delta function approximation method DEL(M) is very easy to implement and computationally inexpensive; however, a large number of elements is usually necessary for convergence of the estimated probability distributions. The underlying theory [4] guarantees convergence of distributions, not densities, in the Prohorov metric. Therefore, estimates of the probability densities are very misleading in terms of accuracy of corresponding estimated probability distributions. While the spline-based method SPL(M,N) provides a much smoother approximation of probability distributions in comparison to DEL(M), it is more computationally expensive. With appropriate choices for the weights $\{a_k^M\}$, convergence of distributions in the Prohorov metric (with significantly fewer elements than DEL(M)) is guaranteed as well as convergence of the approximating densities in L^2 [17]. Lastly, over-parametrization of the inverse problem with the nonparametric approaches can not only result in oscillations in the estimated densities but larger confidence bands as well. As a result, we feel less certain about the reliability of the estimated probability distributions obtained with these nonparametric methods.

The computational results shown here demonstrate how to construct "functional" confidence bands for estimated probability distributions in size-structured mosquitofish populations in both parametric and nonparametric settings. However, one would like to fully develop the mathematical and asymptotic statistical theory for OLS problems with functional parameters, such as the probability distributions studied here and the time- and spatial-dependent functional parameters discussed in the Introduction. One would also like to determine if the confidence bands constructed from the approximation methods DEL(M) and SPL(M,N) are converging to some "true" smooth confidence bands. Following the work of [30], we note that this will require the sensitivity of the system being studied with respect to the probability distribution, which is actually a directional derivative [10]. We are currently working on the development of this fundamental theory in an alternate weak L^2 setting for densities.

Acknowledgments. This paper was written on the occasion and in honor of Tom Hallam's 70th Birthday. We are happy to dedicate this work to Tom in recognition of and in gratitude for his many contributions (scientific and personal) to mathematical biology and ecology. The authors are also grateful to two referees whose comments and suggestions led to improved clarity in this manuscript.

REFERENCES

[1] Linda J. S. Allen, *An Introduction to Stochastic Processes with Applications to Biology*, Pearson Education, Inc., Upper Saddle River, New Jersey, 2003.

[2] H. T. Banks, J. E. Banks, L. K. Dick and J. D. Stark, Estimation of dynamic rate parameters in insect populations undergoing sublethal exposure to pesticides, CRSC-TR05-22, NCSU, May, 2005; *Bull. Math. Biol.*, **69** (2007), 2139–2180.

[3] H. T. Banks, J. E. Banks, L. K. Dick and J. D. Stark, Time-varying vital rates in ecotoxicology: selective pesticides and aphid population dynamics, *Ecological Modelling*, **210** (2008), 155–160.

[4] H. T. Banks and K. L. Bihari, Modelling and estimating uncertainty in parameter estimation, *Inverse Problems*, **17** (2001), 95–111.

[5] H. T. Banks, V. A. Bokil, S. Hu, F. C. T. Allnutt, R. Bullis, A. K. Dhar and C. L. Browdy, Shrimp biomass and viral infection for production of biological countermeasures, CRSC-TR05-45, NCSU, December, 2005; *Math. Biosci. and Engr.*, **3** (2006), 635–660.

[6] H. T. Banks, D. M. Bortz, G. A. Pinter and L. K. Potter, Modeling and imaging techniques with potential for application in bioterrorism, CRSC-TR03-02, NCSU, January, 2003; Chapter 6 in *Bioterrorism: Mathematical Modeling Applications in Homeland Security,* (H. T. Banks and C. Castillo-Chavez, eds.) Frontiers in Applied Math, **FR28,** SIAM (2003), Philadelphia, PA, 129–154.

[7] H. T. Banks, L. W. Botsford, F. Kappel, and C. Wang, Modeling and estimation in size structured population models, LCDS-CCS Report 87-13, Brown University; *Proceedings 2nd Course on Mathematical Ecology*, (Trieste, December 8-12, 1986) World Press (1988), Singapore, 521–541.

[8] H. T. Banks and J. L. Davis, A comparison of approximation methods for the estimation of probability distributions on parameters, CRSC-TR05-38, NCSU, October, 2005; *Applied Numerical Mathematics,* **57** (2007), 753–777.

[9] H. T. Banks and J. L. Davis, Quantifying uncertainty in the estimation of probability distributions with confidence bands, CRSC-TR07-21, NCSU, December, 2007.

[10] H. T. Banks, S. Dediu and H. K. Nguyen, Sensitivity of dynamical systems to parameters in a convex subset of a topological vector space, CRSC-TR06-25, NCSU, September, 2006; *Math. Biosci. and Engr.,* **4** (2007), 403–430.

[11] H. T. Banks and B. G. Fitzpatrick, Estimation of growth rate distributions in size-structured population models, CAMS Tech. Rep. 90-2, University of Southern California, January, 1990; *Quarterly of Applied Mathematics,* **49** (1991), 215–235.

[12] H. T. Banks, B. G. Fitzpatrick, L. K. Potter and Y. Zhang, Estimation of probability distributions for individual parameters using aggregate population data, CRSC-TR98-6, NCSU, January, 1998; In *Stochastic Analysis, Control, Optimization and Applications,* (Edited by W. McEneaney, G. Yin and Q. Zhang), Birkhaüser, Boston, 1989.

[13] H. T. Banks and P. Kareiva, Parameter estimation techniques for transport equations with application to population dispersal and tissue bulk flow models, LCDS Report #82-13, July, 1982; *J. Math. Biol.*, **17** (1983), 253–273.

[14] H. T. Banks, P. M. Kareiva and P. K. Lamm, Modeling insect dispersal and estimating parameters when mark-release techniques may cause initial disturbances, *J. Math. Biol.*, **22** (1985), 259–277.

[15] H. T. Banks, P. Kareiva and L. Zia, Analyzing field studies of insect dispersal using two-dimensional transport equations, LCDS/CCS Rep. 86-48, November, 1986; *Environmental Entomology*, **17** (1988), 815–820.

[16] H. T. Banks and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*, Birkhaüser Boston, 1989.

[17] H. T. Banks and G. A. Pinter, A probabilistic multiscale approach to hysteresis in shear wave propagation in biotissue, CRSC-TR04-03, NCSU, January, 2004; *SIAM J. Multiscale Modeling and Simulation,* **2** (2005), 395–412.
[18] H. T. Banks and I. G. Rosen, Numerical schemes for the estimation of functional parameters in distributed models for mixing mechanisms in lake and sea sediment cores, LCDS Technical Report #85-27, October, 1985; *Inverse Problems*, **3** (1987), 1–23.
[19] H. T. Banks, H. T. Tran and D. E. Woodward, Estimation of variable coefficients in Fokker Planck equations using moving finite elements, CAMS Tech. Rep. 90-9, August, 1990, University of Southern California; *SIAM J. Num. Anal.*, **30** (1993), 1574–1602.
[20] P. Billingsley, *Convergence of Probability Measures,* Wiley, New York, 1968.
[21] G. Casella and R. L. Berger, *Statistical Inference,* Duxbury, California, 2002.
[22] M. Davidian and A. R. Gallant, *Nonlinear Models for Repeated Measurement Data,* Chapman & Hall, London, 1995.
[23] A. R. Gallant, *Nonlinear Statistical Models,* J. Wiley & Sons, New York, 1987.
[24] T. C. Gard, *Introduction to Stochastic Differential Equations*, Marcel Dekker, New York, 1988.
[25] C. Hawkes, The estimation of the dispersal rate of the adult cabbage root fly (*Erioischia brassicae*) in the presence of (*Brassicae*) crop, *J. Appl. Ecol.*, **9** (1972), 617–632.
[26] R. I. Jennrich, Asymptotic properties of non-linear least squares estimators, *The Annals of Mathematical Statistics,* **40** (1969), 633–643.
[27] M. Kot, *Elements of Mathematical Ecology*, Cambridge University Press, Cambridge, 2001.
[28] A. Okubo, *Diffusion and Ecological Problems: Mathematical Models*, Biomathematics, Vol. **10**, Springer-Verlag, Berlin, 1980.
[29] A. Quarteroni, R. Sacco and F. Saleri, *Numerical Mathematics,* Springer, New York, 2000.
[30] G. A. F. Seber and C. J. Wild, *Nonlinear Regression,* John Wiley & Sons, New York, 1989.
[31] J. W. Sinko and W. Streifer, A new model for age-size structure for a population, *Ecology,* **48** (1967), 910–918.

Received December 12, 2007. Accepted on March 6, 2008.

E-mail address: `htbanks@ncsu.edu (corresponding author)`
E-mail address: `jldavis9@ncsu.edu`