Published in final edited form as:

Stat Med. 2014 December 20; 33(29): 5166–5176. doi:10.1002/sim.6289.

The Estimation of Branching Curves in the Presence of Subject-Specific Random Effects

Angelo Elmi^{1,*}, Sarah J. Ratcliffe², and Wensheng Guo²

¹Department of Epidemiology and Biostatistics, George Washington University, Washington, D.C. 20052

²Department of Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104

Abstract

Branching curves are a technique for modeling curves that change trajectory at a change (branching) point. Currently, the estimation framework is limited to independent data, and smoothing splines are used for estimation. This article aims to extend the branching curve framework to the longitudinal data setting where the branching point varies by subject. If the branching point is modeled as a random effect, then the longitudinal branching curve framework is a Semiparametric Nonlinear Mixed Effects Model. Given existing issues with using random effects within a smoothing spline, we express the model as a B-spline Based Semiparametric Nonlinear Mixed Effects Model. Simple, clever smoothness constraints are enforced on the B-splines at the change point. The method is applied to Women's Health data where we model the shape of the labor curve (cervical dilation measured longitudinally) before and after treatment with oxytocin (a labor stimulant).

Keywords

Branching Curves; Mixed Effects Models; Curve Registration

1. Introduction

In many studies, exposure to a risk factor of interest may vary over time. In such scenarios, the response curve may be known to consequently change shape as exposure changes. For example, children's growth trajectory might become stunted in a subset that was exposed to an environmental hazard. Alternatively, the trajectory of dilation of a woman's cervix will increase more rapidly after treatment with a labor stimulant. In cases like these, it is of interest to model how the trajectory changes as exposure/treatment status changes. Silverman and Wood [1] used branching curves to model the shape of curves where a common nonlinear trajectory diverges after treatment or exposure status changes. Thus, to better understand how the cervix dilates before and after stimulation with oxytocin, for

example, branching curves can be used to model how the shape of the labor curve (cervical dilation measured over time) changes after treatment.

Silverman and Wood [1] proposed estimation of this class of nonparametric regression models by a set of cubic splines fitted together at the branching points. A penalty term is used to constrain the smoothness of the curves at the branching points, thus leading to the solution of a sparse system of linear equations. When the curves arise from uncorrelated observations, this approach is a straight forward modification of the traditional nonparametric regression setting. Within a correlated data setting such as longitudinal data, however, the solution is far more complex. In general, the subject-specific curves will share a common shape, but the timing of the branching point will vary by subject since exposure time to treatment will not be the same for all subjects. If the timing of the branching point is modeled as a random effect, then the resulting model is a Semiparametric Nonlinear Mixed Effects Model (SNMM) (Ke and Wang [2]).

Estimation of branching curves within a longitudinal setting can be viewed as a Curve Registration (Ramsay and Li [3], Ramsay and Silverman [4]), where the common population average curve is a branching curve. The subject specific curves are aligned with respect to the branching point for each subject via random effects. The resulting likelihood function involves a complicated convolution of random effects within an unknown nonparametric function. Using a smoothing spline to estimate the unknown function, Ke and Wang [2] proposed a backfitting algorithm which iteratively fits a smoothing spline followed by a parametric nonlinear mixed model conditional on the current smoothing spline estimate. Due to the equivalence of smoothing splines and linear mixed effects models Wang [5], however, this algorithm optimizes two separate likelihoods and is not guaranteed to converge. Derivation of a single likelihood leads to a high dimensional integral approximation to the likelihood which is computationally intractable. See Lin and Zhang [6] for a further discussion of the implications of using random effects within the smoothing spline framework developed by Ke and Wang [2]. B-splines, however, have been shown to work well in accommodating random effects within a nonparametric function (Brumback and Lindstrom [7]).

In this article, we will use B-splines to estimate the nonparametric branching curves by expressing each group as the sum of a baseline (control) curve defined over the entire interval of observation and a treatment curve which is defined only after the branching point. Elmi et. al. [8] developed an algorithm for using B-splines to estimate a single longitudinal nonlinear curve with parametric transformations within Ke and Wang's framework, but extension to estimating branching curves requires clever additional modifications to the algorithm to satisfy the necessary smoothness constraints. As such, we constrain the B-splines by placing the first two breakpoints of the treatment curve at the branching point. Thus, the sum of the two curves is constrained to at least be continuous, and further smoothness constraints such as first and second order differentiability can be imposed by varying the order of the B-spline. We assume second order differentiability in the application since the effect of oxytocin is not immediate. This simple, straight forward solution gives a generally good approximation to smooth underlying shapes pre and post treatment while also being easily and accurately estimated via Adaptive Gaussian

Quadrature (AGQ) approximation to the log-likelihood when the random effects vector is relatively low dimensional (Wu [9, chap. 2], Pinheiro, and Bates [10], Pinheiro and Chao [11], Rabe-Hesketh, Skrondal, and Pickles [12]). Additionally, each function has a clear interpretation as it enables us to explicitly isolate the effect of treatment on the nonlinear trajectory by using an ANOVA type decomposition.

The methodology developed herein will be demonstrated on a data set involving a sample of women attempting a vaginal birth after cesarean (VBAC). This population was of interest due to their increased risk of uterine rupture, a potentially fatal condition. The VBAC study (Macones et. al [13]), examined the risk of uterine rupture in a multi-center cohort study. A corresponding nested case-control study gave rise to the labor curves in our example and has provided valuable information about how a woman's cervix dilates over time. Figure 1 shows a small sample of the observed labor curves categorized by the two binary factors we will consider, oxytocin and uterine rupture (both yes (1) vs. no(0)). It is readily apparent that we can expect each individual curve to follow a common shape apart from horizontal variability. The branching point, the time when oxytocin was initially administered, was not available from this data set and was treated as an additional parameter which was estimated by maximizing the profile log-likelihood over a grid search. The number and existence of branching points was and in general should be determined based on clinical knowledge. In general, one should not use this framework to test for the presence of a branching point.

2. Methodology

2.1. Review of Branching Curves

In this section, an overview of branching curves with B-splines is presented. More details on branching curves can be found in (Silverman and Wood [1]). The branching curve, $f(t)$, depends on a single covariate, t . Assume, without loss of generality, that $t \in [0, 1]$. Then $f(t)$ can be expressed as the sum of a baseline curve $h_0(t)$, defined over $[0, 1]$, and $h_j(t)$ a treatment curve which is nonzero in the j^{th} treatment group only on the subinterval $(\Delta_j, 1]$. Δ_j is the j^{th} treatment group's branching point ($j = 1, \dots, m$). Define $\lambda = j$ to indicate membership in the j^{th} treatment group. Then, letting $I(t \in [\Delta_j, 1], \lambda = j)$ be an indicator function which equals 1 if the observation was made after the branching point in the j^{th} treatment group and 0 otherwise, we can write the branching curve generally as:

$$f(t) = h_0(t) + \sum_{j=1}^m I(t \in [\Delta_j, 1], \lambda = j) h_j(t).$$

where

$$h_0(t) = \mathbf{B}_{k_0}(t) \mathbf{a}_{k_0} \text{ for } t \in [0, 1]$$

$$h_j(t) = \mathbf{B}_{k_j}(t) \mathbf{a}_{k_j} \text{ for } t \in [\Delta_j, 1] \quad j = 1, \dots, m,$$

for a v^{th} order B-spline design matrix $\mathbf{B}_{k_j}(t)$ and a set of coefficients, $\mathbf{a}_{k_j}$, which is indexed by the number of knots or breakpoints k_j .

For the untreated or baseline function, $h_0(t)$, the knots can be chosen to be quantiles of the time domain. By placing the breakpoints appropriately and using different order B-spline bases for, $h_j, j = 1, \dots, m$, smoothness constraints can be easily applied. Three more practical

cases will be described here. Extension to more complex cases is similar. At each branching point, t_j , the first two breakpoints of the B-spline are placed. The remaining breakpoints can then be placed at quantiles of the interval, $(t_j, 1]$, for $j = 1, \dots, m$. The most common types of branching points are described as follows:

TYPE A: Continuity of the branching function. Place the first two breakpoints for h_j at t_j and model h_j with a quadratic B-spline ($v=3$). This approach constrains $h_j(t_j) = 0$, and forces continuity of the function $h_0(t) + I(t \in [t_j, 1], \lambda = j)h_j(t)$.

TYPE B: Continuity of the branching function and its first derivative. Place the first two breakpoints for h_j at t_j and model h_j with a cubic B-spline ($v=4$). This approach enforces the constraints of TYPE A branch points while also forcing the first derivative, $h_j'(\Delta_j) = 0$. Thus, the function $h_0(t) + I(t \in [t_j, 1], \lambda = j)h_j(t)$ has continuity of the function and first derivative.

TYPE C: Continuity of the branching function and its first two derivatives. Place the first two breakpoints for h_j at t_j and model h_j with a quartic B-spline ($v=5$). This approach enforces the constraints of TYPE B branch points while also forcing the second derivative, $h_j''(\Delta_j) = 0$. Thus, the function $h_0(t) + I(t \in [t_j, 1], \lambda = j)h_j(t)$ has continuity of the function and first and second derivatives.

For most applications, branching functions of TYPE A, TYPE B, or TYPE C should suffice. It is not required for all branching points in the model to be of the same type. The application at hand should generally provide the basis for determining which type of branching points to use. For example, if treatment is known to abruptly affect the trajectory after application, then a type A branching point would be most appropriate, while prior knowledge of a more gradual effect might suggest usage of type B or C branching points.

It should also be noted that placement of the first two breakpoints at t_j for $h_j(t)$, $j = 1, \dots, m$, creates a notational redundancy in the use of the domain indicator, $I(t \in [t_j, 1])$, since the B-spline basis expansion for $h_j(t)$ is by definition equal to zero for $t < t_j$ under this formulation. However, we choose to include the indicator in our notation as it makes explicit the domain for $h_j(t)$ and does not create any computational inconveniences. Further details regarding computation based on the B-spline recursion formulas can be found in DeBoor [14, chap. 10].

2.2. The Model and Assumptions

The assumptions of our model are similar to the Semiparametric Nonlinear Mixed Effects Model (Ke and Wang [2]) where the population average shape function is a branching curve (Silverman and Wood [1]). It can be written as:

$$\begin{aligned} \mathbf{y}_i &= \eta(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, f, \mathbf{t}_i) + \boldsymbol{\varepsilon}_i \quad i=1, \dots, N \\ \mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}), \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i), \mathbf{b}_i \perp \boldsymbol{\varepsilon}_i \\ f(t) &= h_0(t) + \sum_{j=1}^m I(t \in [\Delta_j, 1], \lambda_i = j)h_j(t). \end{aligned} \quad (1)$$

For $i = 1, \dots, N$, the n_i dimensional response vector, $\mathbf{y}_i$, is expressed as the sum of a known link function η and an n_i dimensional vector of errors, $\boldsymbol{\varepsilon}_i$. η depends on the branching function, $f(t)$, a p -dimensional set of fixed effects, $\boldsymbol{\beta}$, and a q -dimensional set of random effects, $\mathbf{b}_i$, which are structured by the $n_i \times p$ and $n_i \times q$ dimensional design matrices, $\mathbf{A}_i$ and $\mathbf{B}_i$. $\mathbf{b}_i$ and $\boldsymbol{\varepsilon}_i$ are assumed to be mutually independent and follow covariance matrices $\mathbf{D}$ and $\mathbf{R}_i$ respectively that are parameterized by a small number of variance components and correlation coefficients. Note that if random effects are included within $f(t)$, as is the case for our application, then this model can be viewed as an extension of Curve Registration (Ramsay and Li [3]) to the branching curve setting. Conditional on the random effects, each subject can be seen to follow a branching curve which depends on their treatment status λ_i .

2.3. The Log-Likelihood

For any of the three types of branching functions (TYPE A, TYPE B, TYPE C), the log-likelihood depends on the parameter vector $\boldsymbol{\theta}$, where

$$\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \mathbf{a}_{k_0}^T, \dots, \mathbf{a}_{k_m}^T, \boldsymbol{\sigma}^T, \boldsymbol{\rho}^T)^T.$$

$\boldsymbol{\sigma}^T$ and $\boldsymbol{\rho}^T$ represent the vectors of variance and correlation parameters which comprise $\mathbf{D}$ and $\mathbf{R} = \text{diag}(\mathbf{R}_1, \dots, \mathbf{R}_N)$. Letting $\boldsymbol{\lambda}$ be an N -dimensional vector of treatment group membership for each subject, the multivariate normality assumptions from (1) suggest that the log-likelihood for $\boldsymbol{\theta}$ is formed from integration with respect to the random effects of the product of the conditional density of $\mathbf{y}_i|\mathbf{b}_i$ and the prior density of $\mathbf{b}_i$ as follows:

$$\begin{aligned} \ell(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\lambda}) &= \sum_{i=1}^N \log [\int \text{Pr}(\mathbf{y}_i, \mathbf{b}_i|\boldsymbol{\theta}, \lambda_i) d\mathbf{b}_i] \\ &= \sum_{i=1}^N \log \left\{ \int \frac{|\mathbf{R}_i|^{-\frac{1}{2}} |\mathbf{D}|^{-\frac{1}{2}}}{(2\pi)^{\frac{n_i+q}{2}}} \exp \left[-\frac{1}{2} g(\mathbf{b}_i, \mathbf{t}_i, \boldsymbol{\theta}) \right] d\mathbf{b}_i \right\} \quad (2) \end{aligned}$$

where

$$g(\mathbf{b}_i, \mathbf{t}_i, \boldsymbol{\theta}) = [\mathbf{y}_i - \eta(\mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i, f, \mathbf{t}_i)]^T \mathbf{R}_i^{-1} [\mathbf{y}_i - \eta(\mathbf{A}_i \boldsymbol{\beta} + \mathbf{B}_i \mathbf{b}_i, f, \mathbf{t}_i)] + \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i.$$

Direct computation of this log-likelihood is not possible; however, numerical integration techniques such as Adaptive Gaussian Quadrature (AGQ) have been shown to yield good approximations for nonlinear mixed models (Pinheiro and Bates [15]). Numerical maximization is then used on the approximation to obtain the estimate $\hat{\boldsymbol{\theta}}$. Maximization over a constrained space can also be used to enforce constraints such as monotonicity or positivity on the fitted function $f(t)$ (Schumaker [16, chap. 4.9]). For example, we restrict the first derivative, $f'(t) \geq 0$, in our application since labor curves are known to be nondecreasing.

2.4. Approximation of the Log-Likelihood and Estimation

The log-likelihood in (2) depends on an integral with respect to the random effects which does not take an analytical form, in general. The multivariate normality assumptions in (1) for the densities of $\mathbf{y}_i|\mathbf{b}_i$ and $\mathbf{b}_i$, however, suggest that AGQ would give a good approximation of the log-likelihood in (2) since the kernel of the normal density involves a quadratic form. Pinheiro and Bates [10] have shown its equivalence with a deterministic form of importance sampling. Here the integrand is approximated by a normal distribution that is located and scaled by the posterior mode, $\hat{\mathbf{b}}_i$, and posterior variance, $\mathbf{H}_i^{-1}$, to capture the region where it carries most of its mass. Expressions for $\hat{\mathbf{b}}_i$ and $\mathbf{H}_i^{-1}$ are:

$$\begin{aligned}\hat{\mathbf{b}}_i &= \arg \min g(\mathbf{b}_i, \mathbf{t}_i, \boldsymbol{\theta}) \\ &= \arg \min [\mathbf{y}_i - \boldsymbol{\eta}(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, f, \mathbf{t}_i)]^T \mathbf{R}_i^{-1} [\mathbf{y}_i - \boldsymbol{\eta}(\mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, f, \mathbf{t}_i)] + \mathbf{b}_i^T \mathbf{D}^{-1} \mathbf{b}_i,\end{aligned}\quad (3)$$

and

$$\mathbf{H}_i = \frac{d^2 g}{d\mathbf{b}_i^2} \Big|_{\mathbf{b}_i = \hat{\mathbf{b}}_i}. \quad (4)$$

Letting z_j and w_j , $j = 1, \dots, N_{AGQ}$, represent the abscissas and weights respectively for a Gauss-Hermite Quadrature rule with N_{AGQ} points based on a $N(0, 1)$ kernel, an approximation to the log-likelihood can be written as:

$$\begin{aligned}\ell(\boldsymbol{\theta}|\mathbf{y}, \boldsymbol{\lambda}) &= \sum_{i=1}^N \log [\int Pr(\mathbf{y}_i, \mathbf{b}_i|\boldsymbol{\theta}) d\mathbf{b}_i] \\ &= \sum_{i=1}^N \log \left\{ \int \frac{|\mathbf{R}_i|^{-\frac{1}{2}} |\mathbf{D}|^{-\frac{1}{2}}}{(2\pi)^{\frac{n_i+q}{2}}} \exp[-\frac{1}{2}g(\mathbf{b}_i, \mathbf{t}_i, \boldsymbol{\theta})] d\mathbf{b}_i \right\} \\ &\approx -\sum_{i=1}^N \frac{n_i+Nq}{2} \log(2\pi) - \frac{N}{2} \log|\mathbf{D}| - \frac{1}{2} \sum_{i=1}^N (\log|\mathbf{R}_i| + \log|\mathbf{H}_i|) \\ &\quad + \sum_{i=1}^N \log \left\{ \sum_{j_1=1}^{N_{AGQ}} \cdots \sum_{j_q=1}^{N_{AGQ}} \prod_{n=1}^q w_{j_n} \exp \left[-\frac{1}{2}g(\hat{\mathbf{b}}_i + \mathbf{H}_i^{-1/2} \mathbf{z}_j, \mathbf{t}_i, \boldsymbol{\theta}) + \frac{1}{2} \mathbf{z}_j^T \mathbf{z}_j \right] \right\} \\ &\quad \text{where } \mathbf{z}_j = (z_{j_1}, \dots, z_{j_q})^T.\end{aligned}\quad (5)$$

Simultaneous optimization of (5) over all of the components of $\boldsymbol{\theta}$ is a computationally intensive and difficult task. Rather, we obtain $\hat{\boldsymbol{\theta}}$ by iteratively updating each component of $\boldsymbol{\theta}$ until convergence. Standard errors for parameter estimates can be based on the numerically differentiated second derivative matrix evaluated at the final estimate, $\hat{\boldsymbol{\theta}}$. Alternatively, standard errors could also be estimated via the bootstrap (Efron and Tibshirani [17], Wu [9, chap. 2]).

The approximation, (5), depends on the dimension of the spline coefficient vectors $\mathbf{a}_{k_j}$, $j = 0, 1, \dots, m$. In general, a B-spline of order v_j with k_j interior knots will need $k_j + v_j$ coefficients. Akaike Information Criterion (AIC) or Cross Validation (CV) techniques have been proposed for choosing the dimensions of B-splines in a nonparametric mixed model setting by minimizing one of the following criteria:

$$\text{CV: } CV(k_0, k_1, \dots, k_m) = -2 \sum_{i=1}^N \log \left[\int Pr(\mathbf{y}_i, \mathbf{b}_i | \hat{\boldsymbol{\theta}}^{(-i)}, \boldsymbol{\lambda}) d\mathbf{b}_i \right]$$

$$\text{AIC: } AIC(k_0, k_1, \dots, k_m) = -2\ell(\boldsymbol{\theta} | \mathbf{y}, \boldsymbol{\lambda}) + 2 \sum_{j=1}^m (k_j + v_j)$$

over ranges of $k_0, k_1, \dots, k_m$ (Rice and Wu [18]). For the CV criterion, $\hat{\boldsymbol{\theta}}^{(-i)}$ is the vector of parameter estimates with the i^{th} subject removed. Although CV is preferable theoretically, it may become computationally infeasible depending on the number of branching points, the number of random effects, or the number of subjects. Since simulation results from Elmi et. al [8] have shown that CV and AIC tend to agree about 50–70% of the time and that $\hat{f}(t)$ is not likely to be appreciably affected if they don't, AIC may be preferred to CV due to its considerably lower computational demand. Either criterion can be adapted in a straight forward way to also accommodate varying fixed effects structures; however as discussed in section 1, it is not recommended to use the model selection criteria to test for the existence of a branching point.

2.5. An Outline of the Algorithm

As is generally the case for complex estimation algorithms, starting values are quite important. Computation using the statistical package R is described here. The B-spline design matrix can be obtained from `splineDesign` contained within the `splines` library. `lm`, `constrOptim` or matrix manipulations can then be used to estimate the B-spline coefficients for $\hat{f}(t)$. Conditional on the initial estimate of $f(t)$, starting values for $\boldsymbol{\beta}$, $\boldsymbol{\sigma}$, and $\boldsymbol{\rho}$ can be obtained from the package `nlme` (Pinheiro and Bates [10]) by fitting the nonlinear mixed model of Lindstrom and Bates [19]. In our experience, this procedure tends to give good starting values, since it is in essence fitting our model using a less accurate approximation to the marginal likelihood based on Penalized Quasi Likelihood (Wolfinger [20]).

Given starting values, the general estimation algorithm proceeds in the following steps:

1. Find $\hat{\mathbf{b}}_i$ and $\mathbf{H}_i$ as in (3) and (4) for $i = 1, \dots, N$. This can be implemented in R through the general purpose optimization function `nlm` which implements a Quasi-Newton algorithm (Dennis and Schnabel [21]).
2. Update the estimate of the branching function, $f(t)$, by using `nlm` to update, one-by-one, each set of spline coefficients $\mathbf{a}_{k_j}$.
3. Update the estimate of $\boldsymbol{\beta}$ by applying `nlm`.
4. Update the estimates of $\mathbf{D}$ and $\mathbf{R}$ by applying `nlm`.
5. Repeat 1–4 until convergence.

Convergence can be assessed using the relative change in the objective function or its gradient, similar to `nlme` (Pinheiro and Bates [10]) and SAS `proc nlmixed` (Wolfinger [22]). The abscissas and weights used in the AGQ approximation, (5), can be obtained from the R library `statmod`.

3. Simulation Study

A simulation study was performed to assess the adequacy of the fitting algorithm. 500 replicates were generated from TYPE A, TYPE B, and TYPE C population average branching B-spline curves similar to our data application. To show how our algorithm works under settings similar to the application, we chose curves similar in shape to Figure 2.a. To simulate the data:

1. For $i = 1, \dots, 50$ subjects, n_i was generated from a Poisson(5) distribution, and the n_i time points were generated from a $U(0, 1)$ distribution.
2. Treatment group assignment, λ_i , was randomly generated from a Bernoulli($\frac{1}{2}$) distribution for $i = 1, \dots, 50$.
3. The response vector $\mathbf{y}_i$ was generated from the following model:

$$\begin{aligned} \mathbf{y}_i &= \mathbf{b}_{i2} + f(t_i - \mathbf{b}_{i1}) + \boldsymbol{\varepsilon}_i \\ \mathbf{b}_i &\sim N(\mathbf{0}, \mathbf{D}), \boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{R}_i), \mathbf{b}_i \perp \boldsymbol{\varepsilon}_i \\ \text{where } \mathbf{R}_i &= \sigma^2 \mathbf{I}_{n_i}, \mathbf{D} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}, \quad (6) \\ \text{and } f(t) &= h_0(t) + I(t \in [.6, 1], \lambda_i=1)h_1(t). \end{aligned}$$

$\mathbf{I}_{n_i}$ represents an $n_i \times n_i$ dimensional identity matrix; σ_j for $j = 1, 2$ represent variance components for the random effects b_{i1} and b_{i2} ; ρ is the correlation between the random effects. In the simulation, we set these parameters to $\sigma_1 = .1$, $\sigma_2 = 1$, $\rho = .5$, and $\sigma = 1$.

In the simulation, $\hat{h}_j(t)$ ($j = 0, 1$) was estimated using 3 or 4 knots and 2, 3, or 4 knots respectively with selection performed via the AIC criterion in section 2.4. Table 1 displays results of the parameter estimates in terms of mean and standard deviation of parameter estimates and integrated mean squared error for the curves. The parameter estimates overall were relatively unbiased, and the integrated mean squared errors were small on average for the B-spline curve estimates. The true values of the parameters and the small average number of observations per subject ($n_i = 5$) were chosen to represent a similar scenario to our application. With $N = 50$, we obtained relatively accurate parameter estimates, on average. Prior experience fitting these models suggests that other choices of the parameters would yield accurate estimates as long as there is an adequate signal to noise ratio. Since the model parameters are estimated, in essence, using a parametric Maximum Likelihood algorithm conditional on k , we expect that increasing both n_i or N would yield similarly accurate estimates. Vonesh [23] has shown Maximum Likelihood Estimates for the nonlinear mixed effects model to be consistent when using a Laplace Approximation as n_i and N increase. Adaptive Gaussian Quadrature can be thought of as a higher order Laplace Approximation (Liu and Pierce [24]). Although the Laplace Approximation can yield biased parameter estimates in Generalized Linear Mixed Models with small n_i (Breslow and Clayton [25]), Adaptive Gaussian Quadrature has been found to have good statistical properties for nonlinear mixed effects models (Pinheiro and Bates [15]). The average amount of time per simulation replicate needed to fit all of the models with varying knot

specifications to be selected by AIC was about 50 minutes for our setting. With larger N , n_i , and a large number of random effects, it may take several hours to fit the model.

4. Application: The VBAC Data

The VBAC (Vaginal Birth After Cesarean) study (Macones et. al [13]) was a large, multicenter retrospective cohort study of women who attempted to give birth vaginally after a prior cesarean between 1996 and 2000. A nested case-control study for uterine rupture was used to provide the variables of interest including the raw cervical dilation measurements, the time prior to delivery, case/control status, and the cumulative dose of oxytocin. The study was NICHD funded and conducted at 17 participating hospitals.

The primary objective is to develop a model to describe the effect of oxytocin on the shape of the average labor curve. As such, the labor curve is modeled as a branching curve. A random shift parameter was used to model the point where each individual woman's labor curve branches. It is only known a-priori that the typical labor curve is monotonically increasing with a similar shape for each woman. Since rupture status is associated with labor speed, and to demonstrate our model's ability to accommodate parametric effects, we included a vertical scaling factor that depends on the rupture indicator.

Of the original 629 subjects in the case-control study, there were 510 controls and 119 cases. Women who deliver prematurely versus at full term (gestational age ≥ 36 weeks) were excluded due to a fundamental lack of comparability to at term pregnancies. Although explicit constraints on n_i are not required, we included only subjects with at least three measurements of the response to ensure reasonable estimates of the subject specific curves. Two subjects were removed with outlying responses (extreme on the y-axis), while nine subjects with outlying time points were removed because there was not enough data in these regions of the time axis to obtain a good estimate of their labor curves. The final analysis, therefore, contained 404 subjects. Of these subjects, 238 received oxytocin while 166 did not. There were 82 cases of rupture and 322 subjects who completed labor successfully.

Time was originally recorded in minutes prior to delivery where 0 represents the time of delivery and a number such as -120 represents two hours prior to delivery. The earliest observation included in the analysis occurred at 800 minutes prior to delivery. For the purposes of more stable computations in the B-spline recursion formulas that are implemented in R, time was re-scaled to $0 = -800$ min., $1 = 0$ min. (time of natural delivery) during the fitting of the model. In choosing the branching point, a grid search was performed over $t = .3, .4, .5, .6$. $t = .5$ (-400 minutes prior to delivery) was found to have the highest value of the profiled log-likelihood, and thus was used in the analysis. To increase the practical interpretability of the plots in Figures 2 and 3, the original abscissa is used. A Box-Cox transformation with $\lambda = 2$ (Box and Cox [26]) was used to help remove any potential heteroscedasticity since the assumptions of the model depend upon normality and homogeneity of variance over time. The transformed variable was also uniformly divided by 12 to preserve the original 0–10 cm scale of cervical dilation.

The analysis consists of two factors, oxytocin _{i} , an indicator of whether or not the i^{th} woman was given oxytocin during the course of labor to stimulate cervical dilation, and case _{i} , an

indicator of whether or not the i^{th} woman had rupture. These factors specify unique curves for the four combinations of $case_i$ and $oxytocin_i$. In terms of the formulation in section 2, $oxytocin_i$ corresponds to the treatment variable for subject i , λ_i . The model is therefore written as:

$$y_{ij} = b_{i2} + \exp(\beta case_i) f(t_{ij} - b_{i1}) + \varepsilon_{ij}$$

$$\mathbf{b}_i \sim N(0, \mathbf{D}), \varepsilon \sim N(0, \mathbf{R}_i), \mathbf{b}_i \perp \varepsilon_i$$

where $\mathbf{R}_i = \sigma^2 \mathbf{I}_{n_i}$, $\mathbf{D} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}$, and $f(t) = h_0(t) + I(t \in [.5, 1], \lambda_i = 1) h_1(t)$.

$h_0(t)$ and $h_1(t)$ are constrained to be monotonically increasing. Five Gauss-Hermite quadrature points were used to approximate the likelihood, and a bootstrap distribution was generated from 500 bootstrap samples. Convergence was determined as less than a .6% change in the log-likelihood at each iteration. The algorithm converged in 5 iterations.

The response variable, cervical dilation, is modeled with a branching B-spline of type C from section 2.1. The untreated curve (no oxytocin), $h_0(t)$, is assumed to follow a monotone cubic B-spline basis expansion ($\nu = 4$). The treated curve deviates from the untreated at the branching point and is twice differentiable to capture the gradual effect of oxytocin. The fixed effect, β , controls the speed of labor in cases compared with controls. The random effect b_{i1} is a subject-level time-shift parameter which aligns curves with respect to the time of first oxytocin administration. b_{i1} , therefore, shifts each curve to approximate their unknown time of oxytocin administration. The random effect b_{i2} controls varying initial dilation levels.

The plots in figure 2 demonstrate the fit of the average branching curves. Figure 2.a shows how the curves branch from one another overall, and Figure 2.b isolates the estimated treatment curve, $\hat{h}_1(t)$. Figure 2.c and 2.d show the treated (oxytocin) and untreated curves separately with 95 % pointwise bootstrap confidence intervals. From Figure 2.a we can see an overall monotonically increasing trend for all subjects. In the untreated group, labor initially progressed slowly followed by a sharp upturn around four hours prior to delivery. In the treated group, there was a small increase in the dilation rate at the branching point (−400 minutes) followed by a continuing monotonically increasing trend that consistently remained slightly higher post-treatment. From panel Figure 2.b, the estimated treatment curve $\hat{h}_1(t)$, we can see the initial accelerative effect that occurs in the minutes after the drug is given. This is consistent with how clinicians expect oxytocin to affect dilation. After the initial post-treatment acceleration in dilation, the shape of the trend stays similar to the untreated subjects; however, the dilation level does remain consistently slightly higher in the treated subjects. The overall conclusion from this analysis, therefore, is that oxytocin can be expected to have some short term benefits in terms of speeding up labor; however, over the course of an entire delivery it may need to be used multiple times since the effects may wear out after a certain period of time has elapsed. The capacity of our data set to study this in more detail, however, is limited by the unavailability of detailed timing records of when the drug was given to each woman.

Figure 3 shows sample plots for eight subjects with corresponding pointwise 95% confidence intervals. Here, the branching points occur for each subject at different time points reflecting the effect of aligning curves via the time-shift parameter, although the estimated curves for each subject show a similar shape overall. The estimated fixed effect was $\hat{\beta} = -0.12$. This parameter suggested a slightly slower dilation rate for cases of rupture, but it did not achieve statistical significance based on a percentile based bootstrap 95 % confidence interval of its sampling distribution $(-0.3699, 0.1204)$. The effect of rupture was not the primary focus of our analysis, and the parameter for rupture was primarily included in the model to demonstrate the ability to incorporate parametric effects without much added complexity.

Computational time required to fit the model to the VBAC data was approximately three hours. However, when fitting multiple models with different knot selection criteria, additional time would be required. In general, a useful strategy for fitting complex, computationally intensive models of this nature is to use the computationally simpler Laplace Approximation to get an idea about fixed effects structures, random effects structures, and the shapes of the branching curves. Once a good idea of the model structure has been formulated, the more accurate Adaptive Gaussian Quadrature approximation can be used.

5. Discussion

This article proposes an estimation framework for modeling longitudinally observed branching curves where random effects are used for both curve registration and to account for within subject correlation. Demonstration of the methodology on our labor curve example confirms the expectation of clinicians regarding the effects of oxytocin on the typical labor trajectory, but further investigation is warranted to obtain a more detailed picture. The use of B-splines has allowed a computationally straight forward development which could not be attained under alternative approaches such as with smoothing splines. Additionally, expression in terms of an ANOVA type decomposition allows for direct interpretation of the branching functions in terms of the difference between treatment groups over time. Further work will focus on studying inferential properties and covariate effects in more detail. Development of an inferential procedure for testing the existence of a branching point is of interest as well. In the case of relatively smooth underlying curves, a parametric warping function is usually sufficient to represent the time transformation; however, if the underlying curve is not as smooth, a higher dimensional model for the warping function may be warranted. In such a situation, the framework developed here based on Adaptive Gaussian Quadrature may become computationally infeasible due to the large dimensionality of the random effects vector. The VBAC data analysis on labor curves, however, provides a clinically relevant example which motivates branching curves in this framework, and there is the potential for applicability across a variety of scientific disciplines.

Acknowledgments

Contract/grant sponsor: National Institutes of Health Grant R01 GM104470

References

1. Silverman B, Wood J. The nonparametric estimation of branching curves. *Journal of the American Statistical Association*. 1987; 82:551–558.
2. Ke C, Wang Y. Semiparametric nonlinear mixed effects models and their applications. *Journal of the American Statistical Association*. 2001; 96:1272–1298.
3. Ramsay J, Li X. Curve registration. *Journal of the Royal Statistical Society, Ser B*. 1998; 60:350–361.
4. Ramsay, J.; Silverman, B. *Functional Data Analysis*. 2. Springer; New York: 2005.
5. Wang Y. Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*. 1998b; 93:341–348.
6. Lin X, Zhang D. Discussion of semiparametric nonlinear mixed effects models and their applications. *Journal of the American Statistical Association*. 2001; 96:1288–1291.
7. Brumback L, Lindstrom M. Self-modeling with flexible random time transformations. *Biometrics*. 2004; 60:461–470. [PubMed: 15180672]
8. Elmi A, Ratcliffe S, Parry S, Guo W. A b-spline based semiparametric nonlinear mixed effects model. *Journal of Computational and Graphical Statistics*. 2011; 63:342–350.
9. Wu, L. *Mixed Effects Models for Complex Data*. Chapman and Hall; Florida: 2010.
10. Pinheiro, J.; Bates, D. *Mixed Effects Models in S and S-PLUS*. Springer; 2000.
11. Pinheiro J, Chao E. Efficient laplacian and adaptive gaussian quadrature algorithms for multilevel generalized linear mixed models. *Journal of Computational and Graphical Statistics*. 2006; 15(1): 58–81.
12. Rabe-Hesketh SAS, Pickles A. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The STATA Journal*. 2002; 2(1):1–21.
13. Macones G, Peipert J, Nelson D, Odibo A, Stevens E, Stamilio D, Pare E, Elovitz M, Sciscione A, Sammel M, et al. Maternal complications with vaginal birth after cesarean delivery: A multicenter study. *American Journal of Obstetrics and Gynecology*. 2005; 193:1656–1662. [PubMed: 16260206]
14. deBoor, C. *A Practical Guide to Splines*. Springer; Berlin: 2001.
15. Pinheiro J, Bates D. Approximations to the log-likelihood in nonlinear mixed effects models. *Journal of Computational and Graphical Statistics*. 1995; 4:12–35.
16. Schumaker, L. *Spline Functions: Basic Theory*, Third Edition. Cambridge University Press; New York: 2007.
17. Efron, B.; Tibshirani, R. *An Introduction to the Bootstrap*. Chapman Hall/CRC; New York: 1994.
18. Rice J, Wu C. Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*. 2001; 57:253–259. [PubMed: 11252607]
19. Lindstrom M, Bates D. Nonlinear mixed effects models for repeated measures data. *Biometrics*. 1990; 46:673–687. [PubMed: 2242409]
20. Wolfinger R. Laplace's approximation for nonlinear mixed models. *Biometrika*. 1993; 8:791–795.
21. Dennis, J.; Schnabel, R. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice Hall; Englewood Cliffs, NJ: 1983.
22. Wolfinger, R. Fitting nonlinear mixed models with the new nlmixed procedure. *Proceedings of the 24th Annual SAS Users Group International Conference (SUGI 24)*; 1999.
23. Vonesh E. A note on the use of laplace's approximation for nonlinear mixed-effects models. *Biometrika*. 1996; 83(2):447–452.
24. Liu Q, Pierce D. A note on gauss-hermite quadrature. *Biometrika*. 1994; 81:624–629.
25. Breslow N, Clayton D. Approximate inference in generalized linear mixed effects models. *Journal of the American Statistical Association*. 1993; 88:9–25.
26. Box G, Cox D. An analysis of transformations. *Journal of the Royal Statistical Society, Series B*. 1964; 26:211–252.

Figure 1.
Spaghetti Plots of Raw Data

Figure 2.
Population-Level Branching Curves

(a) Non-Rupture, oxytocin

(b) Rupture, oxytocin

(c) Non-Rupture, No oxytocin

(d) Rupture, No oxytocin

Figure 3.
Subject-Specific Plots

Table 1

Results of Simulation Study (SD = Standard Deviation)

Parameter	Mean TYPE A	SD TYPE A	Mean TYPE B	SD TYPE B	Mean TYPE C	SD TYPE C
$\sigma_1=1$	.097	.014	.098	.015	.098	.014
$\sigma_2=1$	.970	.147	.979	.153	.975	.146
$\sigma=1$	1.012	.061	1.01	.063	1.011	.057
$\rho=.5$	.491	.164	.495	.165	.498	.162
$f(\hat{h}_0-h_0)^2$	.212	-	.212	-	.219	-
$f(\hat{h}_1-h_1)^2$	.232	-	.222	-	.259	-