Published in final edited form as:

Sci China Ser A Math Phys Astron. 2009 June ; 52(6): 1342–1350. doi:10.1007/s11425-009-0088-5.

Modeling and prediction of children's growth data via functional principal component analysis

HU Yu¹, HE Xuming², TAO Jian¹, and SHI Ning-Zhong¹

HU Yu: huy578@nenu.edu.cn; HE Xuming: x-he@uiuc.edu; TAO Jian: taoj@nenu.edu.cn; SHI Ning-Zhong: shinz@nenu.edu.cn

¹Key Laboratory for Applied Statistics of MOE, School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, Jilin, P.R. China

²Department of Statistics, University of Illinois at Urbana-Champaign, 725 South Wright Street, Champaign, IL 61820, USA

Abstract

We use the functional principal component analysis (FPCA) to model and predict the weight growth in children. In particular, we examine how the approach can help discern growth patterns of underweight children relative to their normal counterparts, and whether a commonly used transformation to normality plays any constructive roles in a predictive model based on the FPCA. Our work supplements the conditional growth charts developed by Wei and He (2006) by constructing a predictive growth model based on a small number of principal components scores on individual's past.

Keywords

eigenfunction; functional principal component analysis; LMS method; growth curve

1 Introduction

Anthropometric methods for constructing reference growth charts date back to Quetelet in the 19th century, and have experienced vigorous subsequent growth in modern times. Almost every country has developed children's reference growth charts for key measures, such as height, weight, and head circumference. These charts are widely used by pediatricians and parents. In recent years, two new directions have emerged in the study of growth charts. First, researchers began to consider conditional growth charts by incorporating individual's past records (see [1]). Second, researchers advocated singling out subgroups (e.g., the group with low birth weight) for more careful comparisons (see [2-4], among others). The present paper follows these new directions by modeling and predicting children's weight growth by sub-groups and by utilizing individual's past records.

Weight is an important health indicator for children of all ages. To control weight gain velocity may be a useful intervention to avoid obesity (see [5]). In particular, the low birth weight group of children deserve special attention, because their weight growth is associated with their cognition development, educational attainment, and physical well-being; as made clear by [6,7].

Typical growth charts used to screen children's growth are based on the marginal percentile curves. Wei and He[1] advocated the use of past records to construct conditional growth charts. It is generally agreed upon that a child's past growth pattern is a good indicator for his or her future growth. In this paper, we move away from screening abnormality relative to a reference population. Instead, we focus on modeling and predicting individual growth paths. To do so, we approach the problem by the functional data analysis. As shown in [8], the functional data analysis has wide applications, and the functional principal component analysis is a powerful tool for data summary.

Since the weight measurements of children are not taken at regular intervals, we follow the idea of [9] and use a framework of nonparametric functional principal component analysis (FPCA). We show that taking the first two principal components from FPCA would greatly improve the prediction of the individual growth path of weight given prior measurements. We also show that regardless of whether the analysis is performed on the raw data or their z-scores the prediction errors are mostly comparable.

The rest of the paper is organized as follows. In Section 2, we briefly review the LMS transformations and show how they work for the children's weight data we analyze in the paper. We also introduce the functional principal component analysis in this section. In Section 3 we examine the first two eigen-functions of weight growth for four subgroups of children. Section 4 examines model fitting, and discusses the number of principal components needed for modeling the weight data. In Section 5, we evaluate the predictive performance of the model for the weight between 2 and 3 years based on the prior measurements in the first 2 years. We show that there is a very noticeable improvement in predicting individual growth from using the first one or two principal components, but using additional components adds little value. Section 6 concludes the paper by discussing possible directions for future work.

2 Methodology

2.1 LMS transformation

To construct growth charts with a series of percentile curves, it often helps to transform the data to fit a convenient parametric family. The most popular transformation is the LMS method first proposed by Cole[10], which generalizes the power transformation of Box and Cox[11] as an efficient tool for reducing skewness. In this paper we shall also consider modeling the LMS-transformed data, and investigate whether a transformation to normality has any benefits to individual prediction.

Let $\lambda(t)$, $\mu(t)$ and $\sigma(t)$ denote the age-specific power, median, and scale functions, respectively. We standardize the raw weight by

$$Z(t) = \frac{(Y(t)/\mu(t))^{\lambda(t)} - 1}{\lambda(t)\sigma(t)},$$

which are assumed to be normally distributed, and will be called z-scores. The functions $\{\lambda(t), \mu(t), \sigma(t)\}$ are assumed to evolve smoothly with age, so Cole and Green[12] proposed estimating these functions by maximizing the following penalized log-likelihood,

$$L(\lambda, \mu, \sigma) = v_\lambda \int (\lambda''(t))^2 dt - v_\mu \int (\mu''(t))^2 dt - v_\sigma \int (\sigma''(t))^2 dt,$$

where $L(\lambda, \mu, \sigma)$ is the Box-Cox log-likelihood function,

$$L(\lambda, \mu, \sigma) = \sum_{i=1}^n [\lambda(t_i) \log(Y(t_i)/\mu(t_i)) - \log \sigma(t_i) - Z^2(t_i)/2],$$

and the smoothing parameters ($v_\lambda, v_\mu, v_\sigma$) are chosen to control the equivalent degrees of freedom (edf) for those functions. We use the algorithm described by Carey *et al.* [13] in our work.

Figure 1 gives the three functions, $\lambda(t)$, $\mu(t)$ and $\sigma(t)$, over the ages of 0 to 4 years controlled at edf = (7, 10, 7). The results show that the correction of skewness is more necessary as the age increases. Earlier work has shown the clear advantages of using the LMS-transformed data to estimate percentile curves, but we will later examine the impact of the transformation for model fitting and individual prediction.

2.2 Functional principal component analysis

Consider $X_i(t)$ are independent random realizations of a random function $X(t)$ with mean function $EX(t) = \mu(t)$ and covariance function

$$G(s, t) = \text{cov}(X(s), X(t)).$$

We assume that the domain of $X(t)$ is a closed interval $[0, T]$, and there exists an orthogonal representation (in the L^2 sense) with orthonormal eigenfunctions ϕ_k and corresponding eigenvalues $e_1 \geq e_2 \geq \dots \geq 0$ satisfying $\sum_{i=1}^{\infty} e_i < \infty$:

$$G(s, t) = \sum_{k=1}^{\infty} e_k \phi_k(s) \phi_k(t), \quad s, t \in [0, T].$$

By the Karhunen-Loève expansion of $X(t)$, we have

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} \xi_k \phi_k(t),$$

where ξ_k are uncorrelated random variables (principal component scores) with mean zero and variance e_k .

We now model $Y_i(t)$ as

$$Y_{ij} = X_i(t_{ij}) + \varepsilon_{ij} = \mu(t_{ij}) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t_{ij}) + \varepsilon_{ij},$$

where Y_{ij} is the j -th observation (either the raw measurement on weight or the z -score derived from the LMS transformation) of the i -th subject, ε_{ij} are i.i.d. measurement errors distributed as $N(0, \tau^2)$ with an unknown τ^2 . A useful principal component analysis uses a truncation of the *Karhunen-Loève* expansion:

$$Y_{ij} = X_i(t_{ij}) + \varepsilon_{ij} = \mu(t_{ij}) + \sum_{k=1}^M \xi_{ik} \varphi_k(t_{ij}) + \varepsilon_{ij},$$

for some relatively small integer M . We follow [14] to estimate the decomposition, including using a local polynomial smoothing method to estimate $\mu(t)$. We use the fraction of variances explained (FVE) to guide our selection of M .

3 Functional Principal Components for Each Group

We use weight data obtained from the 1991 longitudinal follow-up to the National Maternal and Infant Health Survey (NMIHS) conducted by Centers for Disease Control (CDC) in the United States[15]. Longitudinal measurements on weight are available on 5,062 children born in 1988. After removing a small proportion of the subjects with missing birth weight records, we have 2,553 boys and 2,509 girls between 0 and 4 years of age. We use month as the unit of time, but the measurement times are irregular. The median number of measurements of all children is 7, and one child has a maximum of 46 measurements.

Since gender and birth weight are often considered to be important factors for weight growth, we divide the subjects into four subgroups based on these two factors. Those children with birth weight at 5.5 lb or less are classified into the low birth weight groups. The four subgroups are labeled as Low1 (623 low birth weight boys), Low2 (675 low birth weight girls), nonLow1 (1930 normal birth weight boys) and nonLow2 (1834 normal birth weight girls).

Table 1 lists the FVEs for the first two principal components in each of the four subgroups based on the LMS-transformed data. If we require the FVE to reach 0.85, the minimum number of the principle components, M , is 2 for two subgroups but 7 and 10 for the other two subgroups. Even for the two subgroups with $\hat{M} = 7$ or 10, the first two principal components already have the FVE value of over 0.75. As a comparison, the FVEs for the first two principal components are also reported in Table 1. Obviously, the FVEs based on the two types of data are very similar.

We examine the first two eigenfunctions for each sub-group. We can see from Figure 2 that the first eigenfunction captures the trend of the long-term variation of the body weight over the four years. Interestingly, the first eigenfunction curve of the Low1 group rises more rapidly in the initial months than those of the other groups. After 3 months, a similar trend is observed for the first eigenfunction in all subgroups, even though the boy's birth weight of the Low1 group is obviously lower than the nonLow1 group. The same observation can be made for the eigenfunctions based on the raw data or the LMS-transformed data. The second eigenfunction shown in Figure 3 explains the remaining residual variance that is not related to the mean function. The high residual variance localized at the ages of 3.5 to 4 years is mainly due to the sparsity of the observed data in that region.

4 Fitting Individual Growth Curves

Even though it takes $M = 10$ in the nonLow2 subgroup to explain 85% of the variance, a model with two principal components fits most of the individual growth curves well. For illustration, we examine one subject (ID 12501798). The fitted growth curves and the observed

measurements for this subject, based on 0, 1, 2 and 10 principal components, are given in Figure 4. The upper left panel is the fitted curve obtained by 10 components, and the upper right panel is just the mean function with 0 principal components used. Clearly the mean function alone cannot predict accurately the weight growth of this individual. The two lower panels show the fitted curves obtained from 1 and 2 principal components, respectively. We note that the fitted curves from $M = 1$ or 2 are nearly as good as that from $M = 10$. We note again that the LMS transformation had little impact on the quality of the fitted curves.

Based on our empirical results, we recommend fitting the weight growth of each subgroup with 2 principal components. As pointed out by James, Hastie, and Sugar[16], a large M risks overfitting. This risk is not sufficiently compensated by the improvement in the overall fitting.

5 Predicting Individual Growth Curves

For each group, we are often most interested in making predictions for future growth paths. To evaluate the accuracy of prediction, we consider using the growth records in the first 2 years to predict the measurements in years 2 to 3. We include subjects for whom there is at least one observed point before two years of age and at least one more point between 2 and 3 years of age. We divide them into two subsets: a training sample (80 percents of subjects in each subgroup) and a testing sample. We use the training sample (with all the measurements) to estimate the unknown principal components and then predict the growth path (over ages 2 to 3) of subjects in the testing sample based on their measurements prior to age 2.

For further information, we construct $100(1 - \alpha)\%$ asymptotic pointwise prediction intervals for the subjects in the testing sample. We shall describe how this can be done for the LMS-transformed data.

Suppose that we want to predict the growth pattern of subject i . Let Y_i^* denote the z-scores of

subject i prior to age 2, $\widehat{\Omega}_M = \widehat{\Lambda} - \widehat{H} \widehat{\Sigma}_{Y_i^*}^{-1} \widehat{H}^T$, where $\widehat{\Lambda} = \text{diag}\{\widehat{e}_1, \dots, \widehat{e}_M\}$, $\widehat{H} = (\widehat{e}_1 \widehat{\phi}_{i1}, \dots, \widehat{e}_M \widehat{\phi}_{iM})^T$, M is the number of components used to make the prediction, and $\widehat{\Sigma}_{Y_i^*}$ is the covariance

function at the observed time points t_i^* corresponding to Y_i^* . Obviously, $\widehat{\Sigma}_{Y_i^*}^{-1}$ and $\widehat{H}$ depend on Y_i^* . For $t \in [2, 3]$, let $\widehat{\phi}_{M,t} = (\widehat{\phi}_1(t), \dots, \widehat{\phi}_M(t))^T$, and $\widehat{X}_i^M(t) = \widehat{\mu}(t) + \widehat{\varphi}_{M,t}^T \widehat{\xi}_i^M$, where $\widehat{\xi}_i^M = (\widehat{\xi}_{i,1}, \dots, \widehat{\xi}_{i,M})^T$ is determined by Y_i^* and the M principal components, $\widehat{\xi}_{j,i} = \sum_{k=1}^{n_i^*} (Y_{i,k}^* - \widehat{\mu}(T_{i,k})) \widehat{\varphi}_j(T_{i,k})(T_{i,k} - T_{i,k-1})$, $0 = T_{i,0} \leq T_{i,1} \leq \dots \leq T_{i,n_i^*} \leq 2$, $\widehat{\phi}_j$, similarly to their multivariate counterparts, is obtained by orthogonal decomposition of empirical covariance function $\widehat{G}$ estimated by training sample, $j = 1, \dots, M$. Now we construct the $100(1 - \alpha)\%$ approximate pointwise prediction interval for $X_i(t)$, the measurement at time $t \in (2, 3)$ as follows:

$$\widehat{X}_i^M(t) \pm \Phi^{-1}(1 - \alpha/2) \sqrt{\widehat{\varphi}_{M,t}^T \widehat{\Sigma}_M \widehat{\varphi}_{M,t}},$$

where $\Phi(\cdot)$ is the standard Gaussian c.d.f. For simplicity, we use $[LB^*, UB^*]$ to denote this interval on the z-scores. When evaluating the accuracy of predictions, we transform them back to the raw scale, and the resulting intervals are denoted as $[LB, UB]$.

We evaluate the performance of prediction intervals using both coverage and average prediction error on the raw weight scale. The results are summarized in Table 2, where the FPCA is used for both the raw data and the LMS-transformed data. We make two important observations.

First, the coverage probabilities are below the nominal level of 95%. This indicates that the asymptotic prediction intervals are too liberal at the sample sizes of our study. Better finite-sample prediction intervals need to be developed. Second, the LMS-transformation appears to lead to better prediction intervals (smaller errors and higher coverage) for boys with normal birth weight, but the benefit is not clear for the other subgroups.

Most importantly, we note that the use of $M = 2$ principal components offers a substantial benefit over any prediction based on the group means, but going beyond $M = 2$ has little added value.

6 Conclusion and Further Work

By analyzing the weight growth data from the four subgroups of children, we have made a convincing case for using a small number of functional principal components in modeling or predicting individualized growth paths based on their prior records. We also find that transformation to normality, which is often valuable in constructing centile curves of the same measurements, may not provide much benefit to individualized prediction.

Two future research problems arise from our study. One is how to use the functional principal component models to adjust for additional covariates (such as parental information). The other is how to obtain prediction intervals that are better than those based on the asymptotic normality so that the coverage probabilities are well preserved. We hope that our article adds further credential to [1] for the value of conditional growth analysis and encourages additional research in this important area.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 10828102) and a Changjiang Visiting Professorship. It was also supported by the Training Fund of NENU'S Scientific Innovation Project (NENU-STC07002) and the NIH Grant R01GM080503-01A1 (USA).

References

1. Wei Y, He X. Discussion paper: Conditional growth charts. *Ann Stat* 2006;34:2069–2097.
2. Ogden CL, Troiano RP, Briefel RR, Kuczmarski RJ, Flegal KM, Johnson CL. Prevalence of overweight among preschool children in the United States, 1971 through 1994. *Pediatrics* 1997;99:1–7. [PubMed: 8989329]
3. Stettler N, Zemel BS, Kumanyika S, Stallings VA. Infant weight gain and childhood overweight status in a multi-center cohort study. *Pediatrics* 2002;109:194–188. [PubMed: 11826195]
4. Waring ME, Lapane KL. Overweight in children and adolescents in relation to attention-deficit/hyperactivity disorder: results from a national sample. *Pediatrics* 2008;122:1–7. [PubMed: 18595979]
5. Belfort MB, Rifas-Shiman SL, Rich-Edwards JW, Kleinman KP, Oken E, Gilman MW. Infant growth and children cognition at 3 years of age. *Pediatrics* 2008;122:689–695.
6. Schimidt LA, Miskovic V, Boyle MH, Saigal S. Shyness and timidity in young adults who were born at extremely low birth weight. *Pediatrics* 2008;122:181–187.
7. Cole CR, Hansen NI, Higgins RD, Ziegler TR, Stoll BJ. Very low birth weight preterm infants with surgical short bowel syndrome: incidence, morbidity and mortality, and growth outcomes, at 18 to 22 months. *Pediatrics* 2008;122:573–582.
8. Ramsay, JO.; Silverman, BW. *Functional Data Analysis*. 2nd. New York: Springer; 2005.

9. Yao F, Muller HG, Wang JL. Functional data analysis for sparse longitudinal data. *J Amer Statist Assoc* 2005;100:577–590.
10. Cole TJ. Fitting smoothed centile curves to reference data. *J R Statist Soc A* 1988;151:385–418.
11. Box GEP, Cox DR. An analysis of transformations. *J R Statist Soc B* 1964;26:211–252.
12. Cole TJ, Green PJ. Smoothing reference centile curves: The LMS method and penalized likelihood. *Stat Med* 1992;11:1305–1319. [PubMed: 1518992]
13. Caery VJ, Yong FH, Frenkel LM, McKinney RM. Growth velocity assessment in paediatric AIDS: smoothing, penalized quantile regression and the definition of growth failure. *Stat Med* 2004;23:509–526. [PubMed: 14748042]
14. Yao F, Muller HG, Clifford AJ, Dueker SR, Follett J, Lin Y, Buchholz BA, Vogel JS. Shrinkage estimation for functional principal component scores with application to the population kinetics of plasma folate. *Biometrics* 2003;59:676–685. [PubMed: 14601769]
15. Sanderson M, Scott C, Gonzalez JF. National maternal and infant health survey: methods and response characteristics, National Center for Health Statistics. *Vital and Health Statistics, Series 2, Data evaluation and methods research* 1988;125:1–39.
16. James GM, Hastie TJ, Sugar CA. Principal component models for sparse functional data. *Biometrika* 2001;87:587–602.

Figure 1.
LMS transformations with $\text{edf} = (7, 10, 7)$.

Figure 2.
The first eigenfunction for all groups.

Figure 3.
The second eigenfunction for all groups.

Figure 4.
Fitted curves for one subject with different number of principal components.

Table 1

FVE for the first two principal components (LMS v.s. Raw)

	Low1	Low2	nonLow1	nonLow2
FVE 1	0.7377 : 0.7041	0.6717 : 0.6598	0.7593 : 0.7100	0.7169 : 0.7671
2	0.7671 : 0.7503	0.8539 : 0.8502	0.8551 : 0.8539	0.7778 : 0.7934
$\hat{M}$	7 : 6	2 : 2	2 : 2	10 : 9

Note: Row 1 for FVE is for the first principal component; Row 2 for the first two principal components. The last row for $\hat{M}$ is the first M for FVE to exceed 0.85.

Table 2

Coverage and average prediction error (LMS v.s. Raw)

<i>M</i>		Low1	Low2	nonLow1	nonLow2
# {obs}		96	99	283	284
Coverage					
0		0 : 0	0 : 0	0 : 0	0 : 0
1		0.7188 : 0.8021	0.7172 : 0.5455	0.7562 : 0.6652	0.7394 : 0.7077
2		0.7396 : 0.7604	0.7071 : 0.7475	0.7527 : 0.7088	0.8521 : 0.7077
all		0.7812 : 0.8438	0.7071 : 0.7475	0.7527 : 0.7088	0.8908 : 0.8521
Average prediction error					
0		7.9000 : 7.9874	6.5021 : 6.6222	5.8219 : 5.8118	7.2530 : 7.1719
1		4.8727 : 4.7662	3.4761 : 3.7983	3.5837 : 4.9416	3.5891 : 3.5252
2		4.7229 : 4.5987	3.6528 : 4.4299	3.9381 : 4.6231	3.5800 : 3.5219
all		4.2560 : 4.2163	3.6528 : 4.4299	3.9381 : 4.6231	3.7473 : 3.5674
FVE					
1		0.7029 : 0.6590	0.6773 : 0.6944	0.7617 : 0.7107	0.7069 : 0.7255
2		0.7493 : 0.7761	0.8556 : 0.8593	0.8570 : 0.8699	0.7539 : 0.7979