Visualizing rate of change: application to age-specific fertility

Han Lin Shang
Australian National University

Abstract

First and second derivatives of a continuous function measure the rate of change. Illustrated by the age-specific Australian fertility rates from 1921 to 2009, we introduce the phase-plane plot to the field of demography as a means of visualizing the rate of change.

Keywords: B-spline basis; Data visualization; Derivatives; Functional data; Velocity; Acceleration.
1 Introduction

Methods for analyzing age-specific demographic data can be divided into two categories: those considering age either as a discrete variable in real space $\mathbb{R}^p$ ($p$ represents the total number of ages) or as a continuous variable in function space bounded within a finite interval, denoted as $\mathcal{I}$. In the discrete case, Bell (1992), Lee and Carter (1992) and Lee (1993) have done extensive work on age-specific fertility and mortality, where the demographic rates at different ages are considered as discrete data points. In the continuous case, Hyndman and Ullah (2007), D’Amato et al. (2011) and Hyndman et al. (2013) introduced functional data analysis to model and forecast age-specific demographic rates as a continuous function, which can be converted to discrete ages at any sampling interval.

Since we cannot observe data continuously, how is a continuous age-specific function obtained? The continuous function is often approximated by a set of basis functions. Basis functions can be collections of fixed functions, such as polynomial basis functions (which are constructed from the monomials $1, t, t^2, \ldots; t \in \mathcal{I}$), Bernstein polynomial basis functions (which are constructed from $1, 1 - t, t, (1 - t)^2, 2t(1 - t), t^2, \ldots$), Fourier basis functions (which are constructed from (1, sin($wt$), cos($wt$), sin(2$wt$), cos(2$wt$), $\ldots$) or spline basis functions; or they may be data-driven orthogonal basis functions, such as principal component basis functions. Each set of basis functions is tailored to a specific data structure: for instance, Fourier basis functions are ideal for periodic data, while spline basis functions are suitable for non-periodic data due to the simplicity in computing derivatives (see appendix for a brief overview). Hyndman and Ullah (2007) emphasized that the main advantage of functional data analysis is that the smoothing technique can be incorporated into the modeling and forecasting of functions, so missing or noisy data can be naturally dealt with.
The ability to consider derivatives, a by-product of conceiving of the data as functions, is also of great advantage. The first derivative, also known as velocity in physics, measures the rate at which the value of a function $f(x)$ changes with respect to the change of a variable $x$. The second derivative, also known as acceleration, measures the rate at which the rate of change alters with respect to the change of variable. For example, Ramsay and Silverman (2002, Section 6) considered human growth curve. When a child is growing more quickly, the velocity increases rapidly, decreasing when growth slows. When a child starts to grow, the absolute acceleration is at its maximum, before gradually decreasing during the rapid growing period. Then, the absolute acceleration starts to increase at the slow growing period, before stabilizes when a child stops growing. Thus, analyzing velocity and acceleration may reveal patterns that are not easily seen from the original functions themselves. Illustrated by Australian age-specific fertility rates described in Section 2, we introduce the phase-plane plot to the field of demography for visualizing the rate of change in Section 3. Section 4 concludes.

2 Data set

We consider annual age-specific Australian fertility data from 1921 to 2006. The data set has been obtained from the Australian Bureau of Statistics (Cat. No. 3105.0.65.001, Table 38) and is available in the rainbow package (Shang and Hyndman, 2013) in R (R Core Team, 2014). The data consist of annual fertility rates by age of mother in single years from ages 15 to 49. Graphical displays of the data are given in Figure 1. From the rainbow plot in Figure 1a, we see the phenomenon of fertility postponement in the most recent years. From the contour plot in Figure 1b, we see the increases in fertility between ages 20 and 30 from
1940 to 1980, this reflects the baby boom period.

Figure 1: Plots of observed age-specific fertility rates for Australia from 1921 to 2006. Source: Australian Bureau of Statistics (Cat. No. 3105.0.65.001, Table 38).

3 Visualizing rate of change: phase-plane plot

Derivatives are useful in visualizing, modeling and forecasting functional data. To visualize the rate of change, the phase-plane plot was proposed and applied to economic data by (Ramsay and Silverman, 2002, Section 3). Here, we demonstrate the use of phase-plane plot for visualizing the rate of change in age-specific fertility. The phase-plane plot is a two-dimensional plot of the acceleration against the velocity of a function. Let us consider a simple function \( \sin(2\pi x) \), which has the first derivative \( 2\pi \cos(2\pi x) \) and second derivative \(-4\pi \sin(2\pi x)\), for \( x = 0, 0.01, \ldots, 1 \). A two-dimensional plot of the acceleration against the velocity is shown in Figure 2, where the highest absolute velocity or acceleration is reached when the other variable is zero.
Figure 2: A two-dimensional plot of the acceleration against the velocity.

From the highest absolute acceleration to the highest absolute velocity (1st and 3rd quadrants), it displays an increase in rate of change. From the highest absolute velocity to the highest absolute acceleration (2nd and 4th quadrants), it shows the rate of change goes to zero. When the velocity reaches zero, it indicates the occurrence of maximum or minimum fertility. As noted by Ramsay and Silverman (2006, pp.30-33), the size of the radius can imply the magnitude of rate of change; the greater the radius is, the greater the magnitude of rate of change. If the horizontal location of the center is to the right, there is net positive velocity, which implies the speed increasing to the maximum fertility is faster than the speed decreasing to the minimum fertility. In the context of fertility, when the age-specific fertility distribution
is right-skewed, it indicates a net positive velocity; when the age-specific fertility distribution is left-skewed, it indicates a net negative velocity.

The phase-plane plot allows us to visualize changes in the shapes of the cycles from year to year, from which we aim to reveal possible social changes in age-specific fertility, including the phenomena of baby boom and postponement. In Figure 3, we display the phase-plane plot and scatter plot for age-specific fertility at the first year of data in 1921 and last year of data in 2006.

Figure 3: Phase-plane plot and scatter plot of the age-specific fertility rates for Australia in 1921 (as shown by dark gray color) and 2006 (as shown by light gray color).

From Figure 3a, the age-specific fertility rates in 1921 have an increase in velocity from ages 15 to 19, they reach the highest positive velocity at 19 before a decreasing velocity from 20 to 26. The maximum fertility occurs at 26, as shown by zero velocity. The absolute velocity increases slowly from 27 to 40 and it reaches the maximum negative velocity at 41. The
absolute velocity then decreases gradually to zero from 42 to 49. Since the horizontal location of the center is to the right, there is net positive velocity which implies the speed increasing to the maximum fertility is faster than the speed decreasing to the minimum fertility.

By contrast, the age-specific fertility rates in 2006 show an increase in velocity from 15 to 18, then a decrease from 19 to 21 before another increase in velocity from 22 to 26. This indicates the presence of bimodality in data. It reaches the maximum positive velocity at 26. The velocity decreases from 27 to 30 and the maximum fertility occurs at about 31, as shown by zero velocity. From 32 to 35, there is an increase in absolute velocity, before reaching the maximum negative velocity at 36. From 37 to 49, the absolute velocity gradually goes to zero. Since the horizontal location of the center is to the left, there is a net negative velocity which implies the speed increasing to the maximum fertility is slower than the speed decreasing to the minimum fertility.

4 Conclusion

From an aspect of functional data analysis, we consider the problem of visualization in demography. We applied the phase-plane plot to visualize the rate of change and demonstrated its usefulness by the Australian age-specific fertility. Similar to the scatter plot, the phase-plane plot identifies the age associated with the maximum fertility and it displays the skewness of age-specific fertility distribution based on the net velocity. Differing from the scatter plot, the phase-plane plot identifies the age associated with maximum positive or negative velocity and compare the magnitude of the rate of change between two years based on the size of the radius. Thus, the phase-plane plot should be considered as a complementary graphical tool to the scatter plot for exploring features in demographic data.
Appendix: $B$-spline basis functions and their derivatives

Dierckx (1995) and de Boor (2001) presented the authoritative account on $B$-spline basis functions. Here, we present the procedure for computing derivatives using $B$-spline basis functions. Denote $u$ as a set of $m + 1$ non-decreasing sequences $u_0 \leq u_1 \leq u_2 \leq \cdots \leq u_m$. These $u_i$s are known as knots and the finite interval $[u_i, u_{i+1})$ is known as the $i^{th}$ knot span. Since $B$-spline basis functions are a type of piecewise polynomials, the method requires the specification of degree $p$. The $i^{th}$ $B$-spline basis function of degree $p$, written as $N_{i,p}(u)$ can be defined recursively as follows:

$$N_{i,0}(u) = \begin{cases} 
1 & \text{if } u_i \leq u \leq u_{i+1}; \\
0 & \text{if otherwise.}
\end{cases}$$

$$N_{i,p}(u) = \frac{u - u_i}{(u_{i+p} - u_i)} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u)$$

The above recursion is referred to as the Cox-de Boor recursion formula (see Cox, 1972; de Boor, 1972). When $p = 0$, it corresponds to the step function, assigning 1 if $u$ is in the $i^{th}$ knot span $[u_i, u_{i+1})$. When $p = 1$, $N_{i,1}(u)$ spans two intervals and is piecewise linear, $N_{i,2}(u)$ spans three intervals that is piecewise quadratic, $N_{i,3}(u)$ spans four intervals that is piecewise cubic. The cubic $B$-spline is the common choice, as it provides sufficient flexibility for approximating most functions.

A set of $B$-spline basis functions are given by

$$\sum_{i=0}^{m+1} c_i N_{i,p}(u), \quad (1)$$
where $c_i$ represents the control points (also known as $B$-spline coefficients). An advantage of the $B$-spline is the easy computation of derivatives,

$$D \left( \sum_{i=0}^{m+1} c_i N_{i,p}(u) \right) = (p-1) \sum_{i=0}^{m+1} \frac{c_i - c_{i-1}}{u_{i+p-1} - u_i} N_{i,p-1}(u),$$

where $D$ represents the first derivative with respect to $u$. As noted by de Boor (2001), the first derivative of a spline function can be represented by differencing the $B$-spline coefficients and a spline of one order lower. The higher-order derivatives can be calculated iteratively.
References

Bell, W. (1992). ARIMA and principal components models in forecasting age-specific fertility. In Keilman, N. and Cruijsen, H., editors, National Population Forecasting in Industrialized Countries, pages 177–200. Swets & Zeitlinger, Amsterdam.

Cox, M. G. (1972). The numerical evaluation of B-splines. Journal of the Institute of Mathematics and its Applications, 10(2):134–149.

D’Amato, V., Piscopo, G., and Russolillo, M. (2011). The mortality of the Italian population: smoothing technique on the Lee-Carter model. The Annals of Applied Statistics, 5(2A):705–724.

de Boor, C. (1972). On calculating with B-splines. Journal of Approximation Theory, 6(1):50–62.

de Boor, C. (2001). A Practical Guide to Splines, volume 27 of Applied Mathematical Sciences. Springer, New York, revised edition.

Dierckx, P. (1995). Curve and Surface Fitting with Splines. Monographs on numerical analysis. Oxford Science Publications, New York.

Hyndman, R. J., Booth, H., and Yasmeen, F. (2013). Coherent mortality forecasting: the product-ratio method with functional time series models. Demography, 50(1):261–283.

Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. Computational Statistics & Data Analysis, 51(10):4942–4956.

Lee, R. D. (1993). Modeling and forecasting the time series of US fertility: age distribution, range and ultimate level. International Journal of Forecasting, 9(2):187–202.
Lee, R. D. and Carter, L. R. (1992). Modelling and forecasting U.S. mortality. *Journal of the American Statistical Association*, 87(419):659–671.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis*. Springer-Verlag, New York.

Ramsay, J. O. and Silverman, B. W. (2006). *Functional Data Analysis*. Springer Science+Business Media LLC, New York, 2nd edition.

Shang, H. L. and Hyndman, R. J. (2013). *rainbow: Rainbow plots, bagplots and boxplots for functional data*. R package version 3.2. URL: http://cran.r-project.org/web/packages/rainbow.