

5

Modeling Reaction-Time Distributions

5.1 Introduction

Hyperactivity in children has been a hot topic in recent years among educational and psychological researchers, not to mention doctors, nurses, and other health professionals. The technical term, attention deficit (hyperactive) disorder, or ADHD, captures the central issue, the difficulty these children have in focusing their attention on tasks for more than brief periods. This affliction is especially troublesome in school.

In spite of the popularity of the term, hyperactivity is actually difficult to diagnose, and may even be rather rare. Because of the frequency with which certain drugs are prescribed to calm down supposedly hyperactive children, as well as the need to test more carefully the efficacy of these drugs, it is imperative to find clearcut techniques to identify the ADHD syndrome.

One behavioral marker is the time taken to react to a visual stimulus appearing after a warning signal, but with a substantial delay. In a typical experiment, children see a warning on a computer screen that a light is about to appear, and are required to push a key as rapidly as possible when a light actually appears. When there is a delay of 10 seconds or so, ADHD individuals not only take longer to react on the average, presumably because their attention has wandered, but they also show a higher frequency of extremely long reaction times.

Figure 5.1 displays two reaction-time distributions, one for the first child in a sample of 17 ADHD children, and another for the first child in a sam-

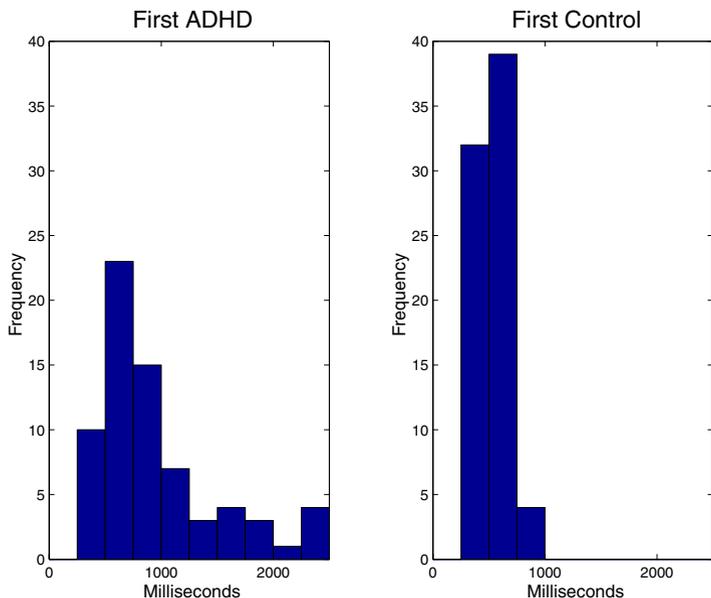


Figure 5.1. Histograms for about 70 reaction times to the onset of a signal after an eight-second delay. The left histogram is for the first of a sample of 17 attention deficit (hyperactive) disorder children, and the right for an age-matched control child.

ple of 16 age-matched control children. Each histogram is computed from about 70 reaction times. The experiment is described in Leth-Stenson, King Elbaz, and Douglas (2000). We see that the ADHD child has many reaction times beyond one second, while the control child never takes that long to respond.

A histogram, such as those in Figure 5.1, gives us only a crude impression of a distribution, and we would prefer to work with the probability density function $p(t)$, describing the reaction-time distributions. This would permit us to calculate the probability of a reaction time between two limits, t_0 and t_1 , as

$$\text{Prob}\{t_0 \leq t \leq t_1\} = \int_{t_0}^{t_1} p(t) dt .$$

But researchers who work with reaction times know that none of the standard textbook distributions capture the features of reaction-time distributions. These characteristics include

- an initial period of at least 120 milliseconds in which no reaction is possible,
- a rapid increase in the number of reaction times after this dead time,

- a strong positive skewness, and
- a very long tail with a severe tendency to outliers.

In these data, for example, we considered reaction times longer than three seconds to be outliers and did not use them, since the large majority of reaction times even for the ADHD children occurred in less time. The shortest reaction time observed was 239 milliseconds.

These distributional features reflect the sequence of neural activities that must precede a reaction, including passage of peripheral excitation to the brain, processing of this information to yield a decision, assembly of the excitation patterns required to generate a response, transmission of these to the neural/muscle interfaces, and delays within muscle tissues before an observable response is possible. All this is compounded by intrusions of attentional lapses, other higher priority events such as a sneeze, and so forth.

We therefore want to explore the implications of ADHD for reaction times without relying on parametric models for the reaction-time distributions. At the same time, we also want to explore the variation in reaction-time distributions from child to child within each group. Our perspective here is that we have two samples of reaction-time distributions, each distribution being identified by around 70 observations. After eliminating reaction times that exceeded three seconds, there were 1111 reaction times for the normal control group, and 1138 for the ADHD group. Our objective is to use functional principal components analysis within each sample to get some picture of the typical modes of variation. But this raises a technical issue: Density functions are by definition positive and integrate to one, but functional principal components analysis is more naturally applied to families of unconstrained functions. It is with this issue in mind that we consider methods of density estimation and modeling that avoid the constraints implicit in the definition of a density function.

5.2 Nonparametric modeling of density functions

A probability density $p(t)$ must satisfy the constraints

- $p(t) > 0$ over some interval $[t_L, t_U]$ of interest, and
- the area under this curve is one; that is,

$$\int_{t_L}^{t_U} p(t) dt = 1 .$$

Given any function $W(t)$, we can construct a probability density function $p(t)$ by

$$p(t) = C \exp W(t), \tag{5.1}$$

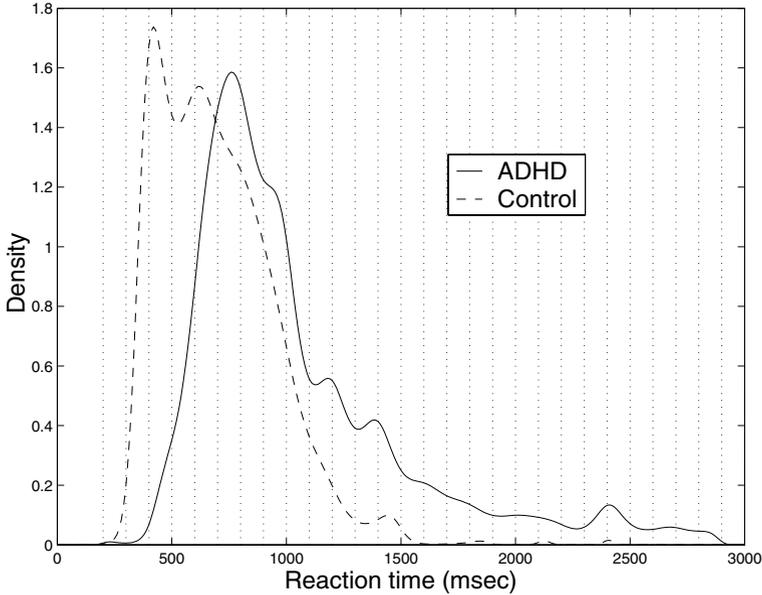


Figure 5.2. The solid curve is the density function for 1143 reaction times observed for the 17 ADHD children, and the dashed curve the density function for 1113 times for the 16 control children. The vertical dotted lines indicate the knot placement in the B-spline basis described in Section 5.6. The smoothing method requires the choice of a smoothing parameter λ , which was set to 10^6 . The density values have been multiplied by 1000.

where

$$C = \left[\int_{t_L}^{t_U} \exp W(x) dx \right]^{-1} .$$

Without any constraints on the function $W(t)$, the conditions for $p(t)$ to be a probability density function will be satisfied automatically. The function $W(t)$ and hence the density $p(t)$ can be estimated by a *penalized maximum likelihood* method, as described in Section 5.6.

Figure 5.2 displays the density functions estimated for the combined data for the two groups. We see that even the fast ADHD times are slower by around 200 milliseconds than fast times for the controls. For example, for the ADHD group, only 8% of the times are faster than 600 msec, compared with the control group which has 40% of the times. We also see that the distributions show some bimodality, and even a hint of trimodality. A distinctive feature of the ADHD times is the large shoulder and long tail on the positive side of the distribution. For example, fewer than 1% of control group times exceed 1600 msec, as compared to 12% of the ADHD times.

The two densities in Figure 5.2, however, ignore individual differences in response times, and in particular, are likely to have more spread than

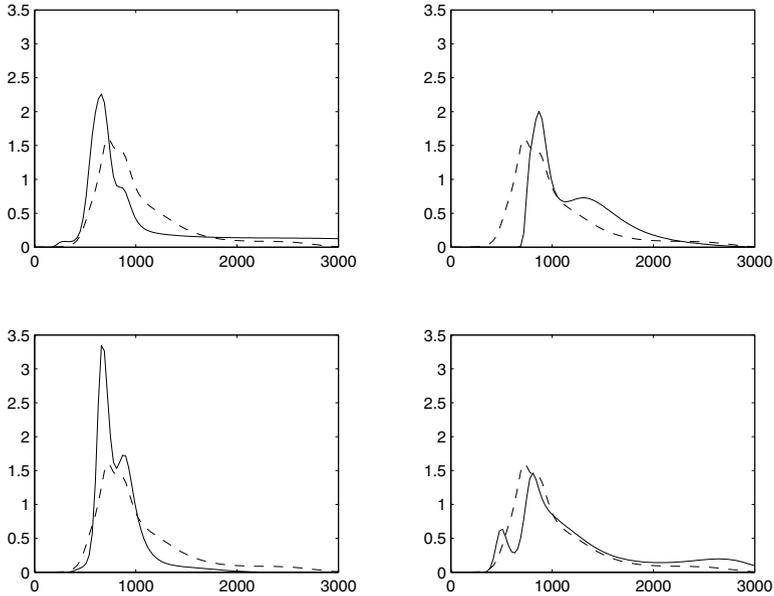


Figure 5.3. Individual reaction-time densities for four of the ADHD children (solid curve) along with the reaction-time density estimated for the entire group (dashed curve). The density values are multiplied by 1000, and the smoothing parameter was set to 10^7 .

individual distributions since, within each sample, there are children who are systematically fast and others who are systematically slow.

Figure 5.3 displays estimated ADHD density functions for selected individuals. We see that there is indeed considerable interchild variation in their shapes. The upper-left panel indicates more fast reactions than average, but a nearly uniform distribution of times beyond one second. The lower-left panel has a density more typical of control children, with no appreciably long tail. Both the right panels show a pronounced secondary mode to the right of one second, and the bottom panel even has a slight tertiary mode.

5.3 Estimating density and individual differences

The individual densities plotted in Figure 5.3 give us a visual impression that the ADHD children vary considerably among themselves in terms of their reaction-time distributions. This would be consistent, for example, with the presence of a disability that varied in severity. How can we represent the unusual shape of these reaction time distributions, and still give some indication of how these children vary?

A few critical remarks are in order about the usual way in which reaction-time data are analyzed. The nearly universal practice of using mean reaction time to represent a subject's typical performance has serious drawbacks. The first is that the mean is a much less appropriate measure of centrality when the distribution is strongly skewed than it is for nearly symmetric distributions like the normal. The long positive tail tends to pull the mean toward it and at the same time increase the variability of its estimate. The mode, by contrast, would be a better indication of a typical reaction time.

The other defect of the mean has to do with how it is modeled. Standard statistical tools such as analysis of variance and regression analysis postulate that whatever changes the typical reaction changes it *additively*. This amounts to saying that a little bit is added or subtracted to all reaction times by causal factors such as the presence of ADHD. But the results in Figure 5.2 suggest something more like a *multiplicative* impact of ADHD in which short reaction times are affected less than long reaction times, and leading, consequently, to the long positive tail being exaggerated. According to a multiplicative impact model, reaction times are affected by a percentage increase rather than a simple shift.

Let us explore, therefore, the variation from child to child and from group to group by using the following transformation for reaction time t measured in milliseconds,

$$z = \log_{10}(t - 120) . \quad (5.2)$$

The constant 120 is first subtracted because this is more like the true “zero” of reaction times, being about the fastest reaction time that is achievable. The log transformation of the shifted reaction times acknowledges that the impact of ADHD is more multiplicative than additive, and therefore that the impact on a logarithmic time scale will be more additive than it will be in the original time scale.

We may now propose the following additive model to describe the log transformed reaction time z_{ijk} of child i on trial j in group k :

$$z_{ijk} = \mu_k + \alpha_{i|k} + U_{ijk} . \quad (5.3)$$

The parameter μ_k quantifies the typical performance of children in group k and the parameter $\alpha_{i|k}$, read “child i within group k ,” quantifies the individual typical performance of this child. As is usual in ANOVA models, we fix the relative sizes of these effects by imposing the restriction

$$\sum_{i=1}^{N_k} \alpha_{i|k} = 0,$$

where N_k is the number of children in group k .

The residual term U_{ijk} expresses the lack of fit of the model for a specific reaction time, and it is the variation of the values of U_{ijk} that we see in

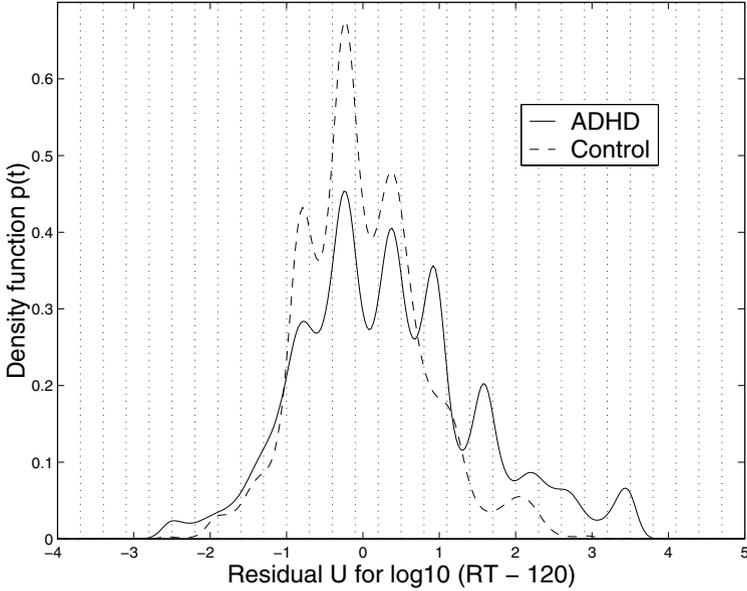


Figure 5.4. The densities of log-shifted reaction-time residuals U_{ijk} for 17 ADHD or hyperactive children (solid line) and 16 normal children (dashed line). Mean effects for individual control children have been removed, so that this group's density is centered on 0. The ADHD density is centered on 0.33 (corresponding to 122 msec) in order to emphasize the coherence of the modes. The vertical dotted lines indicate the knot placement in the B-spline basis described in Section 5.6.

the distribution of transformed reaction times for a specific child. We are assuming that this variable has a mean of zero. If we want to express what model (5.3) means for reaction time itself, then we can reverse the transformation (5.2) to get

$$\tau_i = 120 + 10^{\mu_k + \alpha_{i|k} + U_{ijk}} .$$

But can we be so sure that the distribution of the residuals U_{ijk} is normal? Not at all. We will want to preserve the idea of nonparametric estimation of the density function $p_k(u)$, where the subscript k indicates that we allow the distribution to be different for the two groups. Our technique for estimating these densities starts from that used to estimate the densities in Figures 5.2 and 5.3, but adds the capacity to estimate the model components μ_k and $\alpha_{i|k}$ in addition.

The estimated densities for the residuals for the two groups in this experiment are displayed in Figure 5.4. We see that the hyperactive children show greater variability in residuals U_{ijk} , even after the shifted log transformation, and we also see that the transformed times remain somewhat positively skewed. The mean μ_1 of the ADHD group was 2.92, correspond-

ing to 952 msec, and the control mean μ_2 was 2.72, or 645 msec. As expected, this difference was highly significant ($t = 4.8$).

The pattern of modes in the two densities is striking. We can see this level of detail in the group densities because individual effects have been removed by estimating the individual shift parameters $\alpha_{i|k}$. In the plot, the center of the ADHD density has been shifted by what is equivalent to about 120 msec to show how well lined up the modes are. Initially, there was a suggestion that this multimodal behavior pointed to a substantive feature of brain function. On reflection, however, the experimenters realized that it was an artifact of the instrumentation, which gave some preference to times on particular cycles. Although this conclusion is not as exciting as the neurophysiological one, it illustrates the way in which statistical analyses can be important in drawing experimenters' attention to aspects they had previously overlooked.

5.4 Exploring variation across subjects with PCA

For each child, the work described in Section 5.3 yields an estimated density function for the log-shifted reaction times z for that child. This density function can be regarded as a functional datum for that child. In this section, we explore the use of functional principal components analysis to get a sense of how the density functions vary from child to child, and how many substantial components of variation there are. In Section 5.2, we only looked at an elementary aspect of this variation, namely a variation only in the center of the distributions. As we have seen in previous chapters, PCA offers the possibility of uncovering modes of variation that are more complex. As in Section 5.2, we work with the density functions $p_i(z)$ for log-shifted reaction times z defined in (5.3). We look only at the ADHD group.

Principal components analysis is not well adapted to describing variation in constrained functions. This is because principal components analysis provides an expansion of the data in terms of empirically defined basis functions, namely the principal components weight functions or *harmonics*. Thus there is no convenient way to ensure that the approximation of a density based on these harmonics will remain nonnegative. Instead of analyzing the densities directly, therefore, we study the variation in the *derivatives* of the functions $W_i(z)$ defined in (5.1), that is, the log-density derivative functions

$$w_i(z) = \frac{d}{dz} W_i(z) = \frac{d}{dz} \log p_i(z) .$$

One feature that makes these functions interesting is that, for the normal distribution, $w_i(z)$ is a straight line with negative slope. We can, therefore,

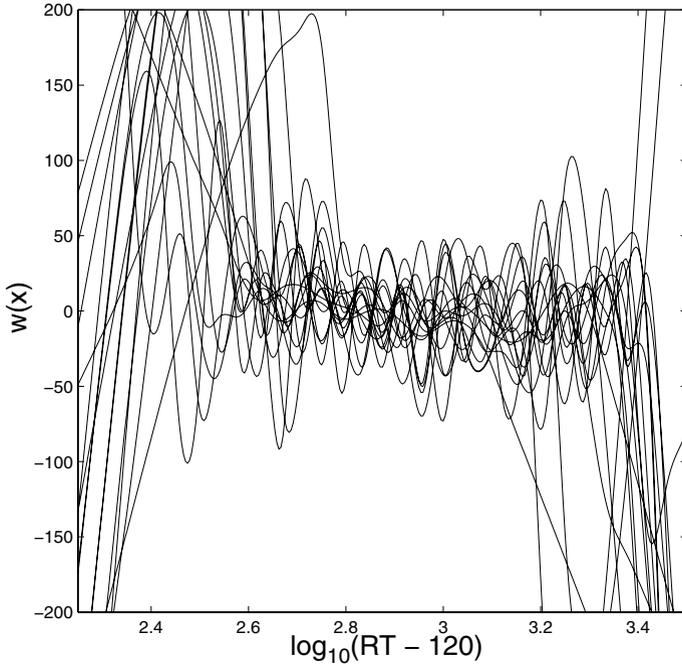


Figure 5.5. Log-density derivatives $w_i(z) = (d/dz) \log p_i(z)$ for individual log-shifted reaction-time densities for the 17 ADHD children.

investigate departures from normality such as multimodality by seeing how different these functions are from linear.

Figure 5.5 shows what these functions look like for the ADHD children. We confess that at first glance they do not look promising. But note that between about $z = 2.75$ and $z = 3.25$, there is something of a linear trend. Outside this central region, however, there is little if any structure visible. However, all the densities themselves are near zero outside the region $[2.75, 3.25]$, and we are not particularly interested in what the functions $w_i(z)$ are up to over values of z that are extremely unlikely to occur. Therefore we use a weighted version of PCA, with weight the average density $\bar{p}(z)$ for the sample. This choice of weight diminishes the role of variation in $w_i(z)$ in defining the harmonics when the density itself is small. The weighted PCA proceeds by applying a standard PCA to the functions $\bar{p}(z)^{1/2} w_i(z)$. Once the harmonics $\eta_m(z)$ are identified for this analysis, we then back-transform to compute the weighted-PCA harmonics $\xi_m(z) = \bar{p}(z)^{-1/2} \eta_m(z)$ for the original log density derivative functions $w_i(z)$.

The first three harmonics account for 63% of the variation in this weighted PCA. This seems reasonable, considering the amount of variability that we see in Figure 5.5. Figure 5.6 indicates that the first three log eigenvalues are noticeably larger than the linear trend in the remainder.

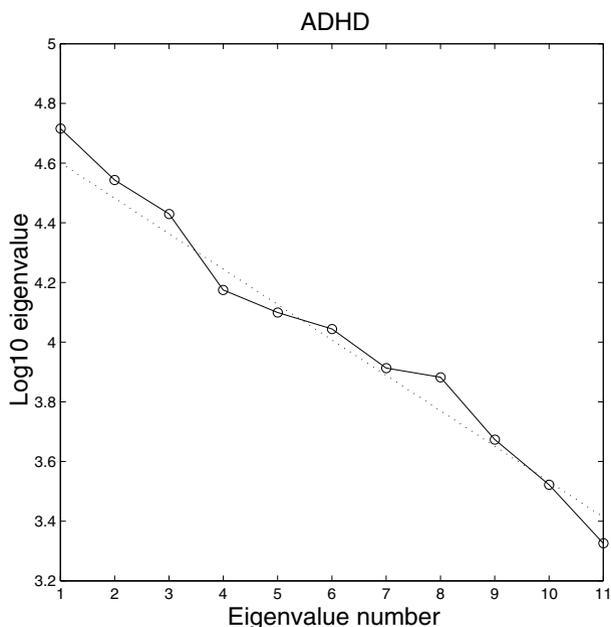


Figure 5.6. The logarithms of the eigenvalues for the weighted principal components analysis of the log-density derivative functions $w_i(z)$ for the ADHD children. The dotted line shows the linear trend for the log eigenvalues from 4 to 11.

Because of the density estimation context, we display the principal components or harmonics as effects on the mean density for the group by adding a multiple of the harmonic to the mean log density, and then converting this perturbed function to a density. The results for the first three principal components for the ADHD sample after varimax rotation are given in Figure 5.7. In each panel the density corresponding to the mean log-density derivative function $\bar{w}(z)$ is plotted as a dashed line for reference purposes.

The first harmonic mainly affects the height of the central peak of the distribution at the expense of moderate deviations from the peak. The second harmonic adds weight in the part of the distribution corresponding to very fast reaction times. The third harmonic corresponds to a density very much like the mean, but with the isolation of the three modes more sharply defined. This harmonic quantifies the strength of the quasiperiodic effect induced by the instrumentation in the experiment. These three harmonics all account for nearly equal amounts of variation.

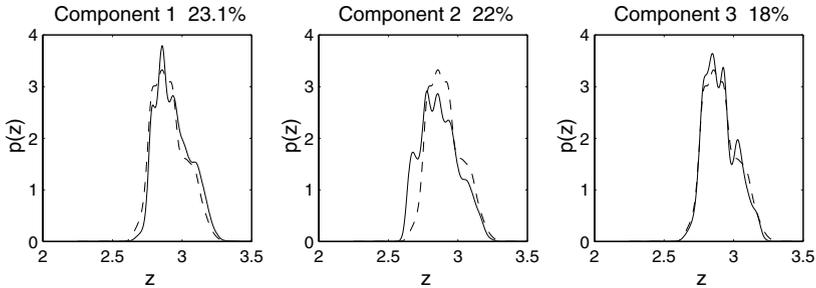


Figure 5.7. The effects of the first three varimax-rotated harmonics for the weighted principal components analysis of the log-density derivative functions $w_i(z)$ for the ADHD children. The solid line in each panel is the density resulting from adding a multiple of a harmonic to the mean function $\bar{w}(z)$ for the entire sample, and the dashed line is the density corresponding to the mean function $\bar{w}(z)$ itself.

5.5 What have we seen?

The effects of a disorder such as ADHD on a marker variable such as reaction time can be complex. These may go beyond a simple change of their central tendency to change the shape of the distribution itself. If we only use distributions that can change in simple ways, such as the normal which can change in location and scale only, we may miss some of these important distributional shape changes, and may at the same time get a distorted picture of simple shifts in distribution. In this case, we see that ADHD seems to create a long positive tail in addition to shifting the mode. Indeed, the strength of this tail seems to be an important component of variation, suggesting that perhaps the upper tail is the true marker for the severity of the ADHD condition.

An additional feature of our analysis was its ability to highlight the quasiperiodic behavior caused by the instrumentation; not only was this visible in the mean curves for the two populations, but one of the principal components was able to quantify the strength of the effect.

The statistical technology that makes our analyses possible is the non-parametric estimation of a density function, whether $p(t)$ for the reaction times, $p(z)$ for the log-shifted reaction times, or $p(u)$ for the residuals in model (5.3). Our method is not the only one available, and kernel density estimation is an alternative approach that is better known. However, our method of estimating the log density leads naturally into using the derivatives of the log densities as functional data for further analysis.

5.6 Technical details

When studying a density function like $p(t)$, we expand the function $W(t) = \text{constant} + \log p(t)$ in a B-spline basis, as described in Section 2.5, to give the expansion

$$W(x) = \sum_{k=1}^K c_k B_k(x). \quad (5.4)$$

There is no restriction on the values of the coefficients c_k . In the work described in this chapter, we used 34 B-spline basis functions of order 5, with equally spaced knots. Splines of order 5 were used so we would be able to define roughness penalties based on high derivatives, and to ensure that the derivative of $W(x)$ was itself smooth.

Given a sample t_1, \dots, t_N modeled by the density function $p(t)$, the density is estimated using a penalized maximum likelihood method proposed by Silverman (1982). The method applies a penalty on the roughness of $W(t)$ by maximizing the penalized log likelihood criterion

$$\text{PENMLE} = \sum_i \ln p(t_i) + \lambda \int_{t_L}^{t_U} W'''(u)^2 du. \quad (5.5)$$

There are two reasons for penalizing the integrated squared third derivative of the function $W(t)$. We use the derivative $w(t) = W'(t)$ for further analysis, and the penalty expressed in terms of w is the more familiar integrated squared second derivative. In addition, the penalty will be zero if and only if $W(t)$ is a quadratic function, which corresponds to $p(t)$ being a normal density (truncated over the interval of interest). Thus, if the smoothing λ increased without limit, it would force $W(t)$ to be quadratic and consequently $p(t)$ to be the normal density, which is the standard “parametric” density estimate.

To carry out the procedure numerically, the function $W(t)$ is expanded in terms of coefficients c_k as in (5.4), and the log likelihood,

$$\ln L = \sum_i \ln p(t_i)$$

and its first two derivatives are expressed in terms of the function $W(t)$ as

$$\begin{aligned} \ln L &= \sum_{i=1}^N W(t_i) - N \ln \int \exp[W(u)] du \\ D_c \ln L &= \sum_{i=1}^N D_c W(t_i) - N E[D_c W] \\ D_c^2 \ln L &= \sum_{i=1}^N D_c^2 W(t_i) - N E[D_c^2 W] - N \text{Var}[D_c W], \end{aligned}$$

where the notations D_c and D_c^2 mean taking the first and second partial derivatives with respect to c , respectively. Also, $E[W] = \int W(u)g(u) du$, and similarly for $E[D_c W]$ and $E[D_c^2 W]$. The values of the integrals in these expressions were approximated using numerical methods rather than analytically.

We use the method of scoring, which is defined by replacing the second derivative matrix in the Newton–Raphson method by $-N \text{Var}[D_c W(t)]$. Convergence is rapid and stable in our experience. The computation is made simpler if $W(t_L)$ is zero, a condition that is easily assured if we fix the coefficient c_1 to zero for the first B-spline basis function, which is the only basis function that is nonzero at t_L .

When applying the method, the smoothing parameter values were chosen subjectively. Where the data are pooled across children, as in Figure 5.2, we used the value $\lambda = 10^6$. Where individual children are considered, and the sample size is smaller, the variability is larger and so a larger value of the smoothing parameter is appropriate. For example, in Figure 5.3, the value was $\lambda = 10^7$.

Software and further details are available from the Web page corresponding to this chapter.