The Shorth Plot

John H. J. EINMAHL, Maria GANTNER, and Günther SAWITZKI

The shorth plot is a tool to investigate probability mass concentration. It is a graphical representation of the length of the shorth, the shortest interval covering a certain fraction of the distribution, localized by forcing the intervals considered to contain a given point x. It is easy to compute, avoids bandwidth selection problems, and allows scanning for local as well as for global features of the probability distribution. The good performance of the shorth plot is demonstrated through simulations and real data examples. These data as well as an R-package for computation of the shorth plot are available online.

Key Words: Data analysis; Distribution diagnostics; Probability mass concentration.

1. INTRODUCTION

Using exploratory diagnostics is one of the first steps in data analysis. Here graphical displays are essential tools. For detecting specific features, specialized displays may be available. For example, if there is a model distribution F which needs to be compared with, from a mathematical point of view the empirical distribution F_n is a key instrument, and its graphical representations by means of *PP*-plots

$$x \mapsto (F(x), F_n(x))$$

or *QQ*-plots

$$\alpha \mapsto (F^{-1}(\alpha), F_n^{-1}(\alpha))$$

are tools of first choice. If we consider the overall scale and location, box-and-whisker plots are a valuable tool. The limitation of box-and-whisker plots is that they give a global view which ignores any local structure. In particular, they are not an appropriate tool for analyzing the modality of a distribution. In this case, more specialized tools are needed, such as for example the silhouette and the excess density plot, both being introduced in Müller and Sawitzki (1991).

John Einmahl is Professor of Statistics, Department of Econometrics & OR and CentER, Tilburg University, 5000 LE Tilburg, The Netherlands (E-mail: *j.h.j.einmahl@uvt.nl*). Maria Gantner is Ph.D. Candidate of Statistics, Department of Econometrics & OR and CentER, Tilburg University, 5000 LE Tilburg, The Netherlands (E-mail: *m.gantner@uvt.nl*). Günther Sawitzki is Academic Director, StatLab, Institute for Applied Mathematics, Im Neuenheimer Feld 294, D 69120 Heidelberg, Germany (E-mail: *gs@statlab.uni-heidelberg.de*).

Journal of Computational and Graphical Statistics, Volume 19, Number 1, Pages 62–73
DOI: 10.1198/jcgs.2009.08020

Whereas we have some instruments for specific tasks, the situation is not satisfactory if it comes to general purpose tools. *PP*-plots and *QQ*-plots need considerable training to be used as diagnostic tools, as they do not highlight the qualitative features of the data.

The classical statistical approach is to focus on the density in contrast to the distribution function, which leads to density estimators and their visual representations, such as histograms and kernel density plots. These, however, introduce another complexity, such as the choice of cutpoints or bandwidth choice. The qualitative features revealed or suggested by density estimators may critically depend on bandwidth choice. A large bandwidth tends to lead to oversmoothing and hides features, and a small bandwidth is prone to undersmoothing and will produce sample artifacts. In Silverman (1981) these properties were exploited to develop a test for multimodality. The need to consider various bandwidths and cutpoints for histograms or kernel density estimators is widely recognized, but the practical application is limited, because the plots for different bandwidths overlap and do not give a possibility for a multiscale representation.

The SiZer approach in Chaudhuri and Marron (1999) tried a way out by using nominal test levels for varying bandwidths. There is considerable literature on good or optimal choice of bandwidth. However, most of this literature concentrates on kernel methods or histograms as density estimators and typically uses optimality criteria based on squared errors. In applications, however, these methods are more often used to inspect data for distribution features, and unfortunately a criterion like the mean integrated squared error does not translate easily to any feature of interest.

The shorth plot, however, discloses the features of the distribution without bandwidth selection due to its monotonicity property (see the end of this section). In contrast to density estimation, it is advised to plot several coverage levels in one picture, which then reveal the different characteristics of the data. Moreover, estimating a density is a more specific task than understanding the shape of a density. Density estimation based methods are prone to pay for bandwidth selection in terms of slow convergence or large fluctuation, or disputable choices of smoothing. Hence the shorth plot, which displays the qualitative features of the distribution without estimating the density, can have a faster rate of convergence than density estimators.

We will use the length of the shorth to analyze the qualitative shape of a distribution. Originally, the shorth is the shortest interval containing half of the distribution; more generally, the α-shorth is the shortest interval containing fraction α of the distribution. The shorth was introduced in the Princeton robustness study as a candidate for a robust location estimator, using the mean of a shorth as an estimator for a mode; see Andrews et al. (1972). As a location estimator, it performs poorly; it has an asymptotic rate of only $n^{-1/3}$, with nontrivial limiting distribution, as shown in Andrews et al. (1972, p. 50), or Shorack and Wellner (1986, p. 767). Moreover, the shorth interval is not well defined, because there may be several competing intervals. The length of the shorth, however, is a functional which is easy to estimate and it gives a graphical representation which is easy to interpret. As pointed out in Grübel (1988), the length of the shorth has a convergence rate of $n^{-1/2}$. In this article, we extend the definition of the length of the shorth to supply localization. We will vary the coverage α, and hence allow for multiscale analysis. Thus the global estimator is extended into a tool for local and global diagnostics. The shorth plot is already valuable for sample sizes of $n = 50$ (for coverage levels above $\alpha = 0.1$).

The article is organized as follows. In the remainder of this section we present the definition and elementary properties of the localized length of the shorth. In Section 2 we define the shorth plot, the central object of this article, and show its use. In Section 3 we study several real data examples. The paper is completed by a discussion section.

To be more explicit we specify our setup and notation. Let $X_1, \ldots, X_n$, $n \geq 1$, be independent random variables with common distribution function F. Let P be the probability measure corresponding to F. Let $\mathbb{I} = \{[a,b] : -\infty < a < b < \infty\}$ be the class of closed intervals and let $\mathbb{I}_x = \{[a,b] : -\infty < a < b < \infty, x \in [a,b]\}$ be the class of closed intervals that contain $x \in \mathbb{R}$. Define the empirical measure P_n on the Borel sets $\mathbb{B}$ on $\mathbb{R}$ by

$$P_n(B) = \frac{1}{n} \sum_{i=1}^{n} 1_B(X_i), \qquad B \in \mathbb{B},$$

where 1_B denotes the indicator function. Let $|\cdot|$ denote Lebesgue measure.

Definition 1: The length of the shorth at point $x \in \mathbb{R}$ for coverage level $\alpha \in (0,1)$ is

$$S_\alpha(x) = \inf\{|I| : P(I) \geq \alpha, I \in \mathbb{I}_x\}.$$

By taking $\inf_x S_{0.5}(x)$, we get the length of the shorth as originally defined. The definition in terms of a theoretical probability P has an immediate empirical counterpart, the empirical length of the shorth

$$S_{n,\alpha}(x) = \inf\{|I| : P_n(I) \geq \alpha, I \in \mathbb{I}_x\}.$$

To get a picture (see Figure 1) of the optimization problem behind the length of the shorth, we consider the bivariate function

$$(a,b) \longmapsto (|I|, P(I)) \quad \text{with } I = [a,b], \text{ where } a < b.$$

This is defined on the half-space $\{(a,b) : a < b\}$ above the diagonal. The level curves of $|I|$ are parallel to the diagonal. The level curves of $P(I)$ depend on the distribution. The α-shorth minimizes $|I|$ in the area above the level curve at level α, that is, $P_I(I) \geq \alpha$. In the empirical version, the level curves of $P(I)$ are replaced by those of $P_n(I)$. In Figure 1, the theoretical curves for the Gaussian distribution and for a Gaussian sample are shown. Localizing the α-shorth at a point x restricts optimization to the (gray) top-left quadrant anchored at (x,x).

Let the distribution function F be absolutely continuous with density f. Assume there exist $-\infty \leq x_* < x^* \leq \infty$ such that $f(x) > 0$ on $\mathcal{S} = (x_*, x^*)$ and $f(x) = 0$ outside $\mathcal{S}$; also assume that f is uniformly continuous on $\mathcal{S}$. As a consequence, F is strictly increasing on $\mathcal{S}$ and f is bounded.

We have the following elementary properties concerning $S_\alpha(x)$:

- *Minimizing intervals*: For every α and x, there exists an interval I with length $S_\alpha(x)$ such that $x \in I$ and $P(I) = \alpha$.

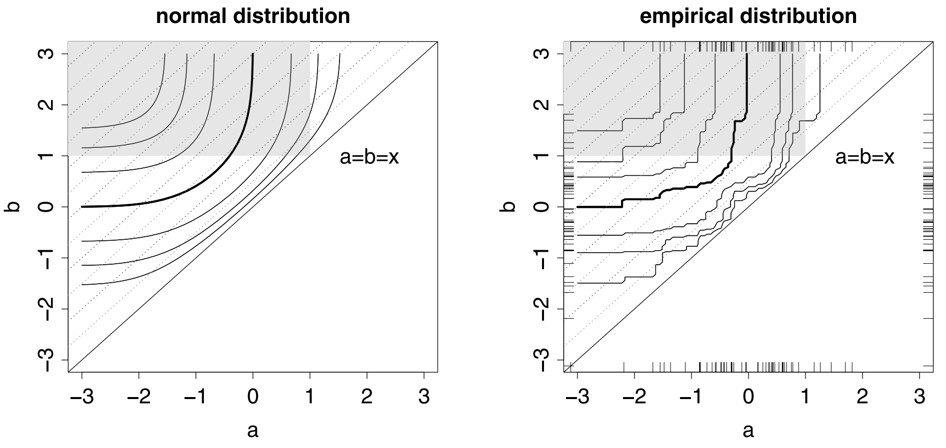

Figure 1. The length of the shorth as an optimization problem: minimize $|[a,b]|$ under the restriction $P([a,b]) \geq \alpha$. Localizing at x restricts the optimization to the quadrant top left of (x,x). The solid lines are the level curves for $P([a,b]) = const = \alpha$, for $\alpha = 0.9375$, 0.875, 0.75, 0.5, 0.25, 0.125, 0.0625 from top to bottom. The bold line depicts $\alpha = 0.5$. The dashed lines are the level curves for $|[a,b]| = const$. For the empirical distribution, a sample of size 100 from a standard normal distribution is taken.

- *Continuity*: For all α, $|S_\alpha(x) - S_\alpha(y)| \leq |x - y|$. Moreover, the function

$$(x, \alpha) \mapsto S_\alpha(x)$$

  is continuous as a function of two variables.
- *Monotonicity*: For all x,

$$\alpha \mapsto S_\alpha(x)$$

  is strictly increasing in α.
- *Invariance*: For all α,

$$x \mapsto S_\alpha(x)$$

  is invariant under shift transformations and equivariant under scale transformations; that is, when we apply a transformation $u' = cu + d$ (for some constants $c > 0, d$), then the new $S'_\alpha(x')$ satisfies

$$S'_\alpha(x') = c S_\alpha(x),$$

  with $x' = cx + d$.

Denote the jth-order statistic by $X_{(j)}$; $X_{(0)} = -\infty, X_{(n+1)} = \infty$. For computing the empirical length of the shorth, observe that $S_{n,\alpha}(x)$ can be interpolated from $S_{n,\alpha}(X_{(j)})$ and $S_{n,\alpha}(X_{(j+1)})$ where j is such that $X_{(j)} \leq x < X_{(j+1)}$. Therefore we can focus on computing $S_{n,\alpha}(X_i)$. Write $k_\alpha = \lceil n\alpha \rceil - 1$, with $\lceil \cdot \rceil$ the ceiling function. Then we simply have

$$S_{n,\alpha}(X_i) = \min\{X_{(j+k_\alpha)} - X_{(j)} : 1 \leq j \leq i \leq j + k_\alpha \leq n\}.$$

We can further reduce the complexity of our problem because $S_{n,\alpha}(X_i)$ can be easily related to $S_{n,\alpha}(X_{i-1})$ through a stepwise algorithm. The resulting algorithm has a linear complexity in n.

2. THE SHORTH PLOT

Definition 2 (Sawitzki 1994): The shorth plot is the graph of the functions

$$x \mapsto S_\alpha(x), \qquad x \in \mathbb{R},$$

for (all or) a selection of coverages α.

The empirical shorth plot is the graph of

$$x \mapsto S_{n,\alpha}(x), \qquad x \in \mathbb{R}.$$

Mass concentration can be represented by the graphs of $x \mapsto S_\alpha(x)$ and $x \mapsto S_{n,\alpha}(x)$; see Figure 2. A *small* length of the shorth signals a large probability mass. In contrast, *large* values of the density indicate a large probability mass. Therefore, to make the interpretation of the shorth plot easier, we will in the sequel use a downward orientation of the vertical axis so that it is aligned with the density plot.

Figure 3 shows the empirical shorth plots for a uniform, a normal, and a lognormal distribution for sample sizes 50 and 200 and the theoretical ones. Varying the coverage level α gives a good impression of the mass concentration. Small levels give information about the local behavior, in particular near modes. Higher levels give information about skewness of the overall distribution shape. The high coverage levels show the range of the distribution. A "dyadic" scale for α, for example, 0.125, 0.25, 0.5, 0.75, 0.875, is a recommended choice. The monotonicity property (Section 1) allows the multiple scales to be displayed simultaneously without overlaps, thus giving a multiresolution image of the distribution.

Let ϕ denote the standard normal density. Then

$$f(x) = \begin{cases} 0.05 & \text{for } |x| \le 0.13 \\ 0.987\big[\phi(x) + \phi(|x| - 0.13)\big] & \text{for } 0.13 < |x| < 0.26 \\ 0.987\phi(x) & \text{for } |x| \ge 0.26, \end{cases} \tag{2.1}$$

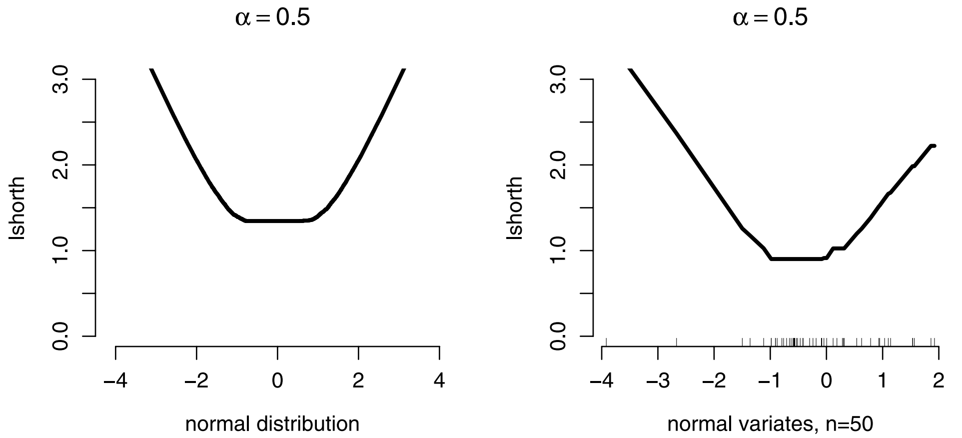

Figure 2. Short plot and empirical shorth plot for a sample of 50 standard normal random variables for $\alpha = 0.5$. Note that different scales are used.

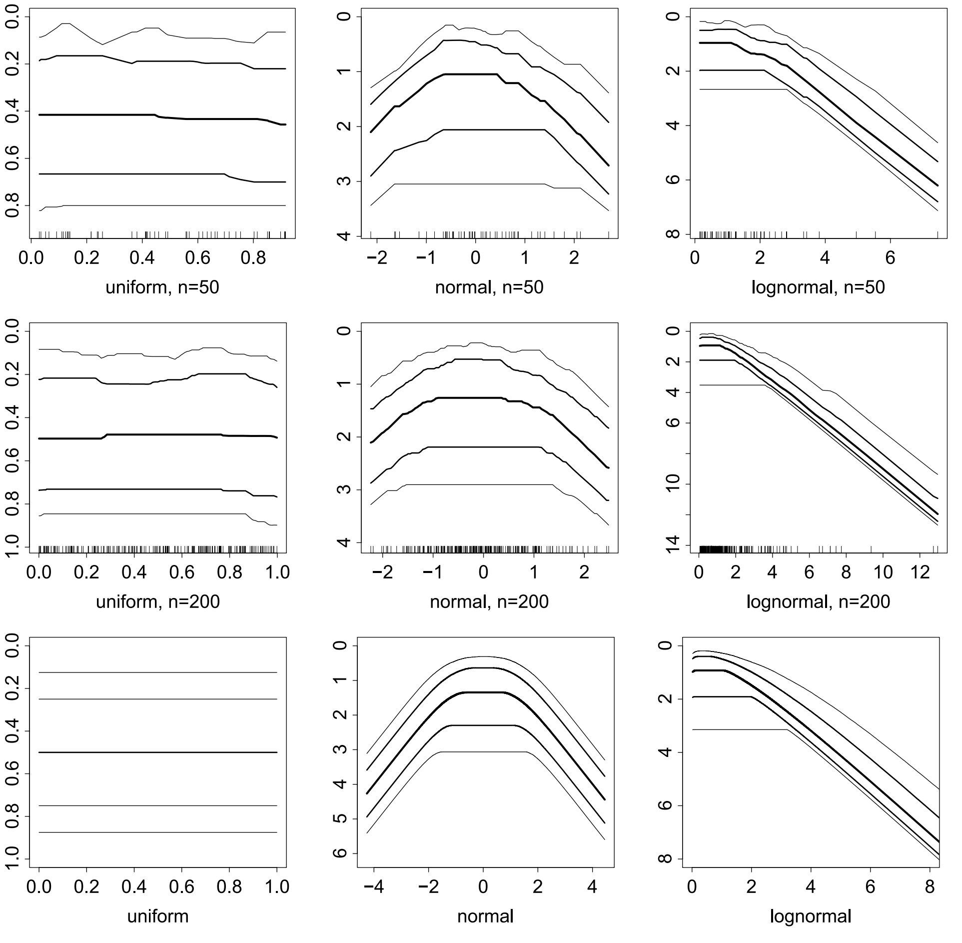

Figure 3. Shorth plots for a uniform, a normal, and a lognormal distribution for sample sizes 50 and 200 and the theoretical ones, for coverage levels (from top to bottom) $\alpha = 0.125, 0.25, 0.5, 0.75, 0.875$. Note that the vertical axes have a downward orientation and that different scales are used.

is an example of a density where the shorth plot outperforms other common methods. The plot of this density is depicted in Figure 4. We consider a random sample of size 500. The fixed bandwidth kernel density estimate with an "optimal" bandwidth [calculated with Silverman's rule of thumb; Silverman 1986, p. 48, (3.31)] suggests a density close to a standard normal; see Figure 5 (upper-left panel). Estimating the density with a varying bandwidth results in a plot even more similar to the standard normal; see the upper-middle panel of Figure 5. (This kernel density estimator is calculated according to Loader 1999, chaps. 3 and 5, using the default nearest-neighbor fraction.) Also a histogram with 17 bins (upper-right panel of Figure 5) sheds no light on the true shape of the density.

The shorth plot, however, detects as well the spikes of the density at $0.13 < |x| < 0.26$ as the dip at $|x| \leq 0.13$, and clearly outperforms both density estimators. This behavior is revealed by the small coverage levels; to make it more visible, we plot in the lower-middle

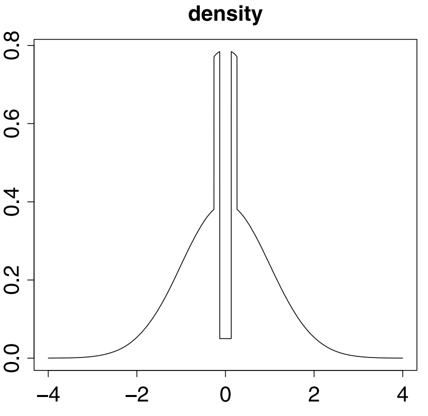

Figure 4. Plot of the density f in (2.1).

panel of Figure 5 only the empirical shorth plot for coverage level $\alpha = 0.125$ on a finer scale. For comparison, the theoretical shorth plot for $\alpha = 0.125$ is given in the lower-right panel of Figure 5. Increasing coverage levels in the shorth plot show the symmetry and the range of the distribution, respectively.

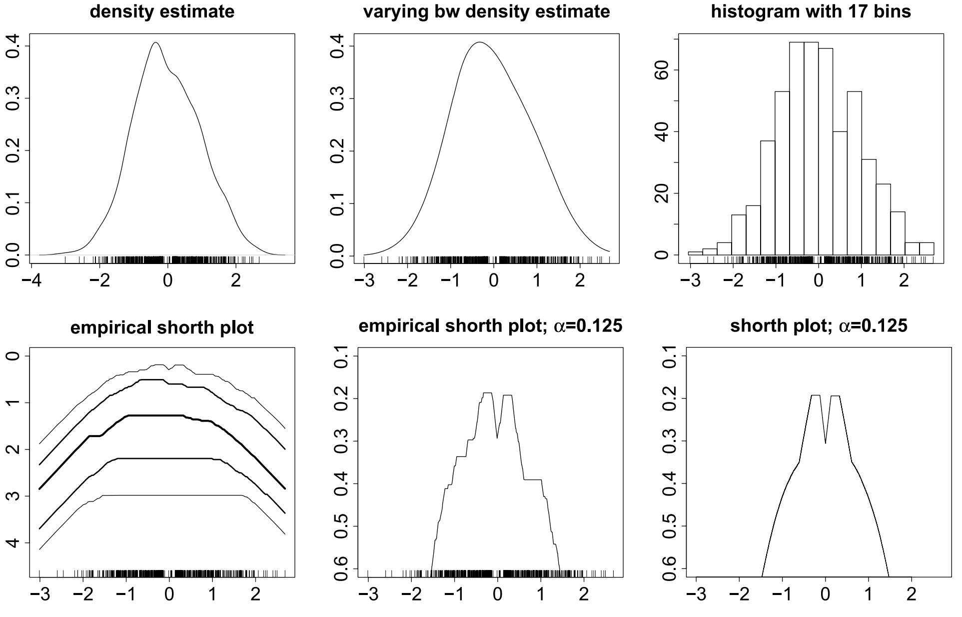

Figure 5. Density estimators (upper panels) and shorth plot (with estimators; lower panels) of f in (2.1); sample size $n = 500$. *Upper panels*: Kernel density estimate of f with "optimal" bandwidth $h = 0.2524$; kernel density estimate of f with varying bandwidth; histogram with 17 bins. *Lower panels*: Empirical shorth plot of the sample with coverage levels (from top to bottom) $\alpha = 0.125, 0.25, 0.5, 0.75, 0.875$; empirical shorth plot of the sample for $\alpha = 0.125$ (note the fine scale); shorth plot of f for $\alpha = 0.125$.

3. EXAMPLES

In this section, we use the shorth plot to depict several real data examples and explain how various features of the distribution can be “read” from this plot.

3.1 ANNUAL MAXIMUM RIVER DISCHARGES OF THE MEUSE RIVER

The annual maximum discharges of the Meuse river from 1911 through 1995 at Borgharen in The Netherlands are an example of right-skewed unimodal data. The data are extensively studied in Beirlant et al. (2004), using extreme value theory. The shorth plot shows clearly the unimodality and the right-skewness. The mode (about 1300) is indicated by small values of the shorth for the small coverage levels; see Figure 6 (right). In Beirlant et al. (2004) it was shown that the Gumbel distribution fits the data quite well and hence supports its common use in hydrology for annual river discharge maxima. For comparison, we picture the shorth plot of the fitted Gumbel distribution in the left panel of Figure 6. The right-skewness is in both plots clearly visualized.

3.2 OLD FAITHFUL GEYSER

As a second example, we use the eruption durations of the Old Faithful geyser. The data are just one component of a bivariate time series dataset. Looking at a one-dimensional marginal distribution ignores the process structure. However, these data have been used repeatedly to illustrate smoothing algorithms like (fixed bandwidth) kernel density estimators (Figure 7, left), and we reuse them to illustrate our approach (Figure 7, right). This is a good-natured dataset showing two distinct nodes with sizeable observation counts, and some overall skewness. The high coverage levels of the shorth plot ($\alpha = 0.75, 0.875$) just show the overall range of the data. The 0.5 level indicates a pronounced skewness, whereas the small levels reveal two modes of approximately the same height (which the density estimate fails to show). The multiscale property of the shorth plot allows us to combine these aspects in one picture. The histogram in Figure 8 confirms that the modes have about the same height.

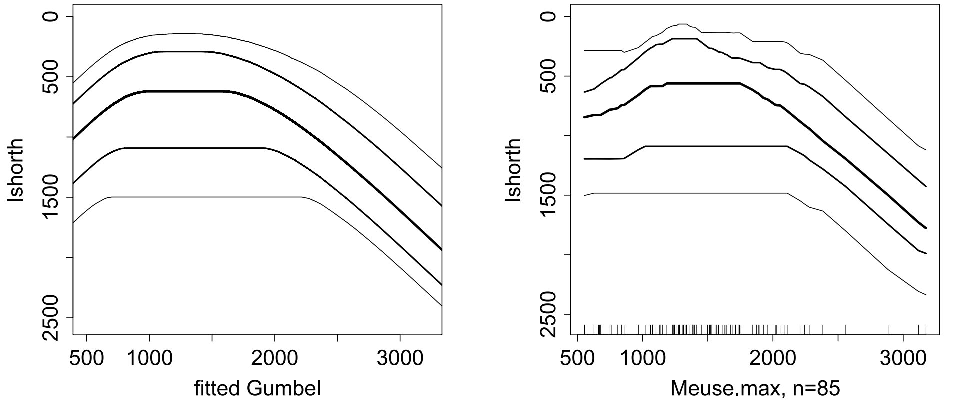

Figure 6. Shorth plots with coverage levels (from top to bottom) $\alpha = 0.125$, 0.25, 0.5, 0.75, and 0.875. *Left*: Fitted Gumbel distribution. *Right*: Annual maximum discharges of the Meuse river from 1911 through 1995.

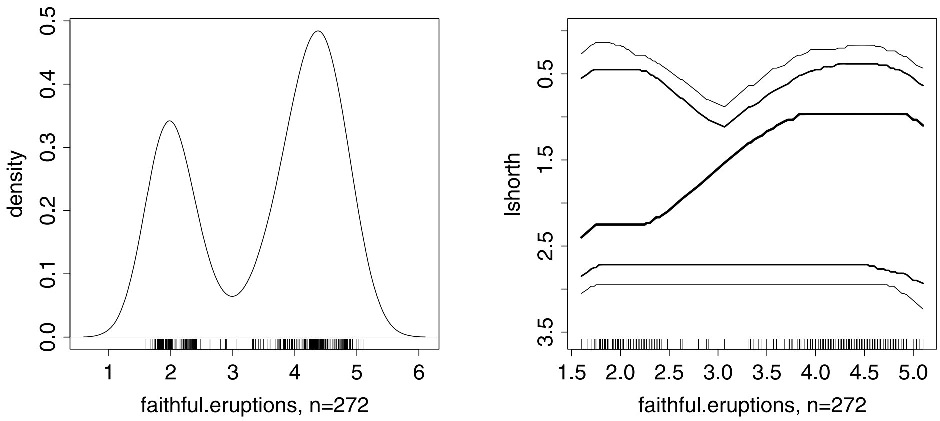

Figure 7. Eruption durations of the Old Faithful geyser. *Left*: Kernel density estimate ("optimal" bandwidth $h = 0.3348$). *Right*: Shorth plot with coverage levels (from top to bottom) $\alpha = 0.125$, 0.25, 0.5, 0.75, and 0.875.

3.3 MELBOURNE TEMPERATURE DATA

In Hyndman, Bashtannyk, and Grunwald (1996) the bifurcation to bimodality in the Melbourne temperature dataset was pointed out. We use an extended version of the dataset (Melbourne temperature data 1955–2007, provided by the Bureau of Meteorology, Victorian Climate Services Centre, Melbourne) and analyze the day-by-day difference in temperature at 3:00 p.m., conditioned on today's temperature. The shorth plot in Figure 9 clearly indicates bimodality (and some skewness) when conditioning on high temperatures, and unimodality when conditioning on the lower temperatures.

3.4 FAMILY INCOMES IN THE UNITED KINGDOM

This example was used in Chaudhuri and Marron (1999) to point out the difficulty of selecting the "optimal" bandwidth in kernel density estimation. The data are $n = 7203$ net family incomes (rescaled to mean 1) from the Family Expenditure Survey in the United

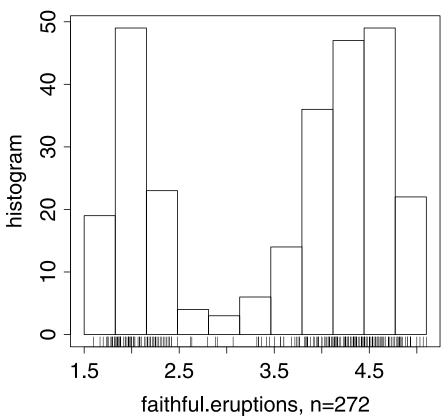

Figure 8. Histogram with 11 bins of the eruption durations of the Old Faithful geyser.

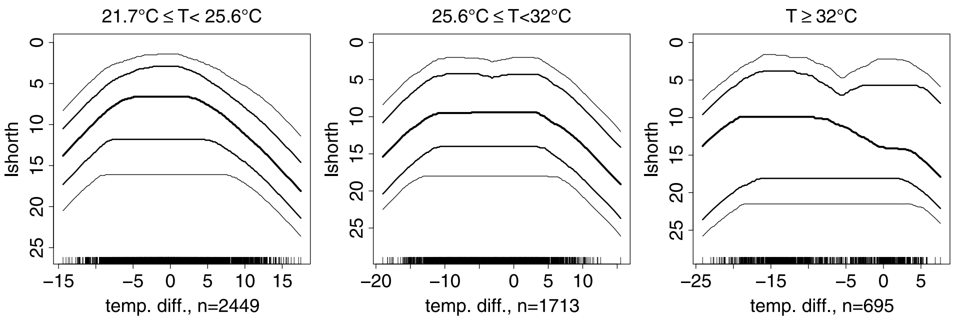

Figure 9. Melbourne day-by-day temperature difference at 3:00 p.m. conditioned on today's temperature; coverage levels (from top to bottom) $\alpha = 0.125$, 0.25, 0.5, 0.75, and 0.875.

Kingdom for the year 1975. For detailed discussion and analysis of the data see Schmitz and Marron (1992). In comparison to parametric approaches which typically lead to a unimodal representation of the data, in Schmitz and Marron (1992) the bimodality of the data was revealed by applying kernel density estimation. The bimodality can be explained by the different densities of pensioner households and nonpensioner households. The SiZer approach confirms this bimodality.

With these data, we want to compare the shorth plot and the SiZer map; see Figure 10. For the latter one, applied to the family income data, the reader is referred to Chaudhuri and Marron (1999) for a more detailed discussion. There also a comparison to a plot of kernel density estimators with various bandwidths can be found.

The first manifest characteristic revealed by the shorth plot is the right-skewness of the data. The top two lines ($\alpha = 0.0625$, 0.125) show the bimodality of the data: there is a high but short first mode and a nearly as high but broader second mode. The third coverage level ($\alpha = 0.25$) underlines this difference of mass concentration.

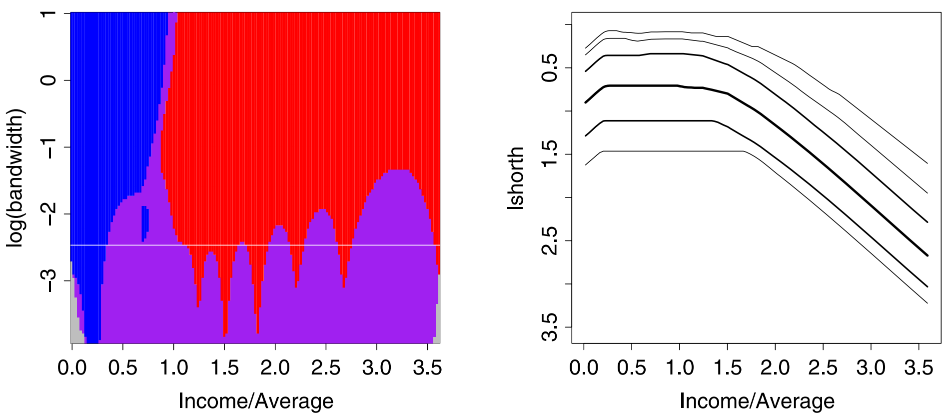

Figure 10. Family Expenditure Survey 1975: rescaled net household income. *Left*: SiZer map; "optimal" bandwidth $h = 0.0943$ marked by a white line. *Right*: Shorth plot with coverage levels (from top to bottom) $\alpha = 0.0625$, 0.125, 0.25, 0.5, 0.75, and 0.875.

Some features are also revealed by the SiZer map. (Recall that the color scheme is blue (red) in locations where the curve is significantly increasing (decreasing), purple is used for a zero-derivative, and gray indicates regions where the data are too sparse to make statements about significance.) Right-skewness and one clear mode are apparent at all bandwidths, whereas the second mode can only be detected when looking at the "good" bandwidths. (As an indication, the "optimal" bandwidth given by Silverman's (1986) rule of thumb is marked with a white line in the SiZer map.) The SiZer map, however, offers no information about the height of the modes.

4. DISCUSSION

The α-shorth is a well-defined concept. Its length was studied in detail in Grübel (1988). The length of the shorth can be extended to higher dimensions by replacing the class of intervals by a class of sets (e.g., all ellipsoids) and length by "volume"; see Einmahl and Mason (1992) for the asymptotic behavior of these minimal volumes. In higher dimensions, however, there is no canonical class of sets, like the intervals in dimension 1.

The shorth plot was introduced in Sawitzki (1994), but no further analysis or theory was provided. The asymptotic theory for the shorth plot, with convergence rate $n^{-1/2}$, can be found in Einmahl, Gantner, and Sawitzki (2009). A closely related idea is the balloonogram in Tukey and Tukey (1981). That multivariate procedure reduces in dimension 1 to considering the shortest interval, centered at a data point, that contains a certain number of data. In contrast to the balloonogram, the shorth plot avoids centering, thus reducing random fluctuation. No theory is provided, however, and also only one coverage level is used at a time.

The shorth plot is based on the concept of mass concentration, an idea shared with the excess density plot and the silhouette plot (Müller and Sawitzki 1991). Excess density and silhouette plots are designed to detect the modes of a density. They use a global approach: there is no localization in x, like in the shorth plot. In Hyndman (1996), so-called highest density regions boxplots are introduced. These boxplots use mass concentration in a regression context.

Kernel density estimators with varying bandwidths are widely studied and somewhat related to our approach. The coverage α of the shorth plot bears some similarity to the bandwidth chosen for kernel estimation. The SiZer (Chaudhuri and Marron 1999) is a kernel-based approach which studies simultaneously a wide range of bandwidths. Another approach that combines kernel estimation explicitly with detecting modes is that of the mode trees (Minnotte and Scott 1993). Here the mode locations are plotted against the bandwidth of the density estimator with those modes. In the recent paper by Dümbgen and Walther (2008), increases and decreases of a density were investigated through multiscale test statistics. Mass concentration is a local concept, but not, like a density, an infinitesimal concept. Therefore the shorth plot avoids the smoothing step and can be based directly on the empirical measure.

SUPPLEMENTAL MATERIALS

R-code and data: An R package (lshorth) for computing the shorth plot and the data used in this article are available online. (supplements.zip)

ACKNOWLEDGMENTS

We thank the editor, an associate editor, and a referee for valuable comments that led to improvements of the article.

[Received February 2008. Revised April 2009.]

REFERENCES

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972), *Robust Estimation of Location: Survey and Advances*, Princeton: Princeton University Press. [63]

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004), *Statistics of Extremes: Theory and Applications*, New York: Wiley. [69]

Chaudhuri, P., and Marron, J. S. (1999), "SiZer for Exploration of Structures in Curves," *Journal of the American Statistical Association*, 94, 807–823. [63,70-72]

Dümbgen, L., and Walther, G. (2008), "Multiscale Inference About a Density," *The Annals of Statistics*, 36, 1758–1785. [72]

Einmahl, J. H. J., and Mason, D. M. (1992), "Generalized Quantile Processes," *The Annals of Statistics*, 20, 1062–1078. [72]

Einmahl, J. H. J., Gantner, M., and Sawitzki, G. (2009), "Asymptotics of the Shorth Plot," preprint. [72]

Grübel, R. (1988), "The Length of the Shorth," *The Annals of Statistics*, 16, 619–628. [63,72]

Hyndman, R. J. (1996), "Computing and Graphing Highest Density Regions," *The American Statistician*, 50, 120–126. [72]

Hyndman, R. J., Bashtannyk, D. M., and Grunwald, G. K. (1996), "Estimating and Visualizing Conditional Densities," *Journal of Computational and Graphical Statistics*, 5, 315–336. [70]

Loader, C. (1999), *Local Regression and Likelihood*, New York: Springer. [67]

Minnotte, M. C., and Scott, D. W. (1993), "The Mode Tree: A Tool for Visualization of Nonparametric Density Features," *Journal of Computational and Graphical Statistics*, 2, 51–68. [72]

Müller, D. W., and Sawitzki, G. (1991), "Excess Mass Estimates and Tests for Multimodality," *Journal of the American Statistical Association*, 86, 738–746. [62,72]

Sawitzki, G. (1994), "Diagnostic Plots for One-Dimensional Data," in *Computational Statistics, 25th Conference on Statistical Computing at Schloss Reisensburg*, eds. R. Ostermann and P. Dirschedl, Heidelberg: Physica-Verlag/Springer, pp. 237–258. [66,72]

Schmitz, H. P., and Marron, J. S. (1992), "Simultaneous Estimation of Several Size Distributions of Income," *Econometric Theory*, 8, 476–488. [71]

Shorack, G. R., and Wellner, J. A. (1986), *Empirical Processes With Applications to Statistics*, New York: Wiley. [63]

Silverman, B. W. (1981), "Using Kernel Density Estimates to Investigate Multimodality," *Journal of the Royal Statistical Society*, 43, 97–99. [63]

Silverman, B. W. (1986), *Density Estimation*, London: Chapman & Hall. [67,72]

Tukey, P. A., and Tukey, J. W. (1981), "Data-Driven View Selection; Agglomeration and Sharpening," in *Interpreting Multivariate Data*, ed. V. Barnett, New York: Wiley. [72]