Log SiZer

J. S. Marron
Department of Statistics
University of North Carolina
Chapel Hill, NC 27599-3260

R. Riedi
Department of Electrical and Computer Engineering
Rice University
Houston, TX 77251-1892

Abstract

A version of SiZer that flags multiplicative structure in data.

1 SiZer

Given a "true underlying curve", f (density or regression function), its "scale space representation" indexed by scale h is denoted by $f_h = f * K_h$, where $*$ denotes convolution, and where K_h is the "scale h Gaussian window function", $K_h(\cdot) = \frac{1}{h}K\left(\frac{\cdot}{h}\right)$. For each scale, f_h represents "the true curve at that level of resolution". Let $\widehat{f}_h(x)$ be a nonparametric curve estimator (either kernel density estimate, or local linear smooth), at location x, for "bandwidth", i.e. "scale" h, of $f_h(x)$. The conventional SiZer map studies "significant changes" in $f_h(x)$, i.e. SIgnificant ZERo crossings of the derivative $f_h(x)$. This is done via confidence intervals based on the derivative (with respect to x), $\widehat{f}'_h(x)$. To calculate this, the randomness in $\widehat{f}'_h(x)$ is approximated by the limiting (as $n \to \infty$) distribution

$$\widehat{f}'_h(x) - f'_h(x) \xrightarrow{d} N(0, \sigma^2_{x,h}).$$

In the case of density estimation, calculations in Appendix 1 show that this (local, in scale space) variance may be written as

$$\sigma^2_{x,h} = n^{-1}\left\{(K'_h)^2 * f - f'_h(x)^2\right\} \tag{1}$$

which is well estimated by $\widehat{\sigma^2_{x,h}}$, whose exact form is given at (18) in Appendix 1. This results in a confidence interval for $f'_h(x)$, with endpoints

$$\widehat{f}'_h(x) \pm q \cdot \widehat{\sigma_{x,h}}, \tag{2}$$

where q is an appropriate version of the Gaussian quantile, chosen to give "simultaneous inference". The Sizer map uses the color blue when this interval is wholly above 0 (i.e. where $\widehat{f}_h(x)$ is significantly increasing), red when it is wholly below 0 (i.e. where $\widehat{f}_h(x)$ is significantly decreasing), and the intermediate purple when it contains 0 (i.e. where there is significant change in $\widehat{f}_h(x)$).

2 Log SiZer

The main idea here is to replace the curve estimate $\widehat{f}_h(x)$ by its logarithm (more appropriate for handling randomness which works on a "multiplicative scale"). To study "significant changes" in $log\widehat{f}_h(x)$, consider its derivative with respect to x,

$$\frac{d}{dx} log\widehat{f}_h(x) = \frac{\frac{d}{dx}\widehat{f}_h(x)}{\widehat{f}_h(x)} = \frac{\widehat{f}'_h(x)}{\widehat{f}_h(x)}.$$

Confidence intervals of the form (2) are based on the distribution of

$$\frac{\widehat{f}'_h(x)}{\widehat{f}_h(x)} - \frac{f'_h(x)}{f_h(x)} = \frac{\widehat{f}'_h(x)f_h(x) - \widehat{f}_h(x)f'_h(x)}{\widehat{f}_h(x)f_h(x)}.$$

A crude method of approximating this distribution is to use the fact, from (11) and (16) in the Appendix, that (as $n \to \infty$, $h \to 0$),

$$\begin{aligned} \widehat{f}_h(x) - f_h(x) &= O_p\left(n^{-1}h^{-1}\right), \\ \widehat{f}'_h(x) - f'_h(x) &= O_p(n^{-1}h^{-3}), \end{aligned}$$

so "variability of $\widehat{f}_h(x)$ is negligible". This results in the approximation

$$\frac{\widehat{f}'_h(x)}{\widehat{f}_h(x)} - \frac{f'_h(x)}{f_h(x)} \approx \frac{\widehat{f}'_h(x) - f'_h(x)}{f_h(x)}. \tag{3}$$

The corresponding limiting distribution is then

$$\frac{\widehat{f}'_h(x) - f'_h(x)}{f_h(x)} \xrightarrow{d} N(0, \frac{\sigma^2_{x,h}}{f_h(x)^2}),$$

which results in approximate confidence intervals with endpoints

$$\frac{\widehat{f}'_h(x)}{\widehat{f}_h(x)} \pm q \cdot \frac{\widehat{\sigma_{x,h}}}{\widehat{f}_h(x)}.$$

Note that this confidence interval contains 0 exactly when the conventional SiZer interval in (2) does. This suggests that the approximation (3) is too crude. In particular, the *added* variability of $\widehat{f}_h(x)$ must be taken into account.

It may be enough to do this using the variability of $\widehat{f}_h(x)$ in the numerator, and ignoring that in the denominator. The key calculation is

$$\begin{aligned} var\left(\widehat{f}_h'(x)f_h(x) - \widehat{f}_h(x)f_h'(x)\right) &= f_h(x)^2 var\left(\widehat{f}_h'(x)\right) \\ &\quad + f_h'(x)^2 var\left(\widehat{f}_h(x)\right) \\ &\quad - 2f_h(x)f_h'(x)cov\left(\widehat{f}_h'(x), \widehat{f}_h'(x)\right). \end{aligned}$$

In the case of kernel density estimation, using (16), (11) and (19) from the appendix, this can be written as:

$$\begin{aligned} var\left(\widehat{f}_h'(x)f_h(x) - \widehat{f}_h(x)f_h'(x)\right) &= f_h(x)^2\left[n^{-1}\left\{\left(K_h'\right)^2 * f - f_h'(x)^2\right\}\right] \\ &\quad + f_h'(x)^2\left[n^{-1}\left\{K_h^2 * f - f_h(x)^2\right\}\right] \\ &\quad - 2f_h(x)f_h'(x)\left[n^{-1}\left\{\left(K_h'K_h\right) * f - f_h'(x)f_h(x)\right\}\right] \end{aligned} \tag{4}$$

Sometimes it may be useful to consider the asymptotic, as $n \to \infty$ and $h \to 0$, representation

$$\begin{aligned} var\left(\widehat{f}_h'(x)f_h(x) - \widehat{f}_h(x)f_h'(x)\right) &= f_h(x)^2\left(n^{-1}h^{-3}f(x)\int K'(u)^2du + n^{-1}h^{-1}f''(x)\int \frac{u^2}{2}K'(u)^2\,du + \right. \\ &\quad + f_h'(x)^2\left(n^{-1}h^{-1}f(x)\int K(u)^2du + O(n^{-1})\right) \\ &\quad - 2f_h(x)f_h'(x)\left(O(n^{-1})\right) \\ &= n^{-1}h^{-3}f_h(x)^3\int K'(u)^2du \\ &\quad + n^{-1}h^{-1}f_h''(x)f_h(x)^2\int \frac{u^2}{2}K'(u)^2\,du \\ &\quad + n^{-1}h^{-1}f_h(x)f_h'(x)^2\int K(u)^2du \\ &\quad + O\left(n^{-1}\right), \end{aligned}$$

which uses additional approximations from the Appendix.

The expression (4) is the key to the logSiZer confidence intervals, which have endpoints:

$$\frac{\widehat{f}_h'(x)}{\widehat{f}_h(x)} \pm q \cdot \widetilde{\sigma_{x,h}}, \tag{5}$$

where in the case of density estimation

$$\begin{aligned}\widetilde{\sigma^2_{x,h}} &= \frac{var\left(\widehat{f}'_h(x)f_h(x) - \widehat{f}_h(x)f'_h(x)\right)}{f_h(x)^2}, \qquad (6)\\ &= \frac{1}{nf_h(x)^2}\left(f_h(x)^2\left[n^{-1}\left\{\left(K'_h\right)^2 * f - f'_h(x)^2\right\}\right]\right.\\ &\quad +f'_h(x)^2\left[n^{-1}\left\{K_h^2 * f - f_h(x)^2\right\}\right]\\ &\quad \left.-2f_h(x)f'_h(x)\left[n^{-1}\left\{\left(K'_hK_h\right) * f - f'_h(x)\, f_h(x)\right\}\right]\right)\end{aligned}$$

which is well estimated by $\widetilde{\widetilde{\sigma^2_{x,h}}}$ whose form is given at (21) in the Appendix.

3 Appendix 1: detailed calculations for density estimation

Details are given here, for the case of kernel density estimation. This setting assumes data $X_1, ..., X_n$ independent and identically distributed from the probability density f, assumed to have a bounded 3rd derivative. The kernel density estimator of $f(x)$, using bandwidth h, is

$$\widehat{f}_h(x) = n^{-1}\sum_{i=1}^{n} K_h\left(x - X_i\right),$$

where K_h denotes the rescaling of the kernel function K, $K_h(\cdot) = \frac{1}{h}K\left(\frac{\cdot}{h}\right)$, where K is say the Gaussian density. See e.g. Scott (1992) and Wand and Jones (1995) for further references and for discussion of many important properties and aspects of this estimator. See Fan and Marron (1994) for discussion of a fast binned implementation of the local linear smoother.

Note that

$$\begin{aligned}E\widehat{f}_h(x) &= En^{-1}\sum_{i=1}^{n} K_h\left(x - X_i\right) \qquad (7)\\ &= EK_h\left(x - X\right)\\ &= \int K_h\left(x - t\right) f(t)dt\\ &= K_h * f\\ &= f_h\left(x\right) \qquad (8)\\ &= \int K(u) f\left(x - hu\right) du\\ &= f(x) + O\left(h^2\right),\end{aligned}$$

as $h \to 0$, where X is a random variable with the same density f, where $\int$ is understood to mean definite integration over the full real line, and where the

substitution $u = (x-t)/h$ was used. Also note that

$$\begin{aligned} var\left(\widehat{f}_h(x)\right) &= var\left(n^{-1}\sum_{i=1}^{n} K_h\left(x - X_i\right)\right) \\ &= cov\left(n^{-1}\sum_{i=1}^{n} K_h\left(x - X_i\right), n^{-1}\sum_{i'=1}^{n} K_h\left(x - X_{i'}\right)\right) \\ &= n^{-2}\sum_{i=1}^{n} var\left(K_h\left(x - X\right)\right) + 0 \\ &= n^{-1} var\left(K_h\left(x - X\right)\right), \end{aligned}$$

where the independence of the X_i and thus of the $K_h\left(x - X_i\right)$ has been used. But

$$\begin{aligned} E\left(K_h\left(x - X\right)^2\right) &= \int K_h\left(x - t\right)^2 f(t) dt \qquad (9) \\ &= K_h^2 * f \qquad (10) \\ &= \frac{1}{h}\int K(u)^2 f\left(x - hu\right) du \\ &= h^{-1} f(x)\int K(u)^2 du + O(h), \end{aligned}$$

from which it follows that,

$$\begin{aligned} var\left(K_h\left(x - X\right)\right) &= E\left(K_h\left(x - X\right)^2\right) - \left\{E K_h\left(x - X\right)\right\}^2 \\ &= K_h^2 * f - f_h\left(x\right)^2 \\ &= h^{-1} f(x)\int K(u)^2 du + O(1), \end{aligned}$$

and so

$$\begin{aligned} var\left(\widehat{f}_h(x)\right) &= n^{-1}\left\{K_h^2 * f - f_h\left(x\right)^2\right\} \qquad (11) \\ &= n^{-1} h^{-1} f(x)\int K(u)^2 du + O(n^{-1}). \qquad (12) \end{aligned}$$

Similar calculations apply to the derivative estimator

$$\widehat{f}_h'(x) = n^{-1}\sum_{i=1}^{n} K_h'\left(x - X_i\right).$$

In particular

$$\begin{aligned} E\widehat{f}_h'(x) &= En^{-1}\sum_{i=1}^{n} K_h'(x-X_i) && (13)\\ &= EK_h'(x-X) \\ &= \int K_h'(x-t)\, f(t)dt \\ &= K_h' * f && (14)\\ &= \int K_h(x-t)\, f'(t)dt \\ &= K_h * f' \\ &= f_h'(x) && (15)\\ &= \int K(u) f'(x-hu)\, du \\ &= f'(x) + O(h^2), \end{aligned}$$

where integration by parts has been used. And also

$$\begin{aligned} var\left(\widehat{f}_h'(x)\right) &= var\left(n^{-1}\sum_{i=1}^{n} K_h'(x-X_i)\right) \\ &= cov\left(n^{-1}\sum_{i=1}^{n} K_h'(x-X_i), n^{-1}\sum_{i'=1}^{n} K_h'(x-X_{i'})\right) \\ &= n^{-2}\sum_{i=1}^{n} var\left(K_h'(x-X)\right) + 0 \\ &= n^{-1} var\left(K_h'(x-X)\right). \end{aligned}$$

But

$$\begin{aligned} E\left(K_h'(x-X)^2\right) &= \int K_h'(x-t)^2 f(t)dt \\ &= \left(K_h'\right)^2 * f \\ &= h^{-3}\int K'(u)^2 f(x-hu)\, du \\ &= h^{-3}\int K'(u)^2 \left\{ f(x) - huf'(x) + \frac{(hu)^2}{2} f''(x) + O\left(h^3\right)\right\} du \\ &= h^{-3} f(x)\int K'(u)^2 du + h^{-1} f''(x)\int u^2 K'(u)^2\, du + O(1), \end{aligned}$$

since $K_h'(x) = \frac{d}{dx}\left(\frac{1}{h}K\left(\frac{x}{h}\right)\right) = \frac{1}{h^2}K'\left(\frac{x}{h}\right)$, from which it follows that (as $h \to 0$),

$$\begin{aligned} var\left(K_h'(x-X)\right) &= E\left(K_h'(x-X)^2\right) - \left\{EK_h'(x-X)\right\}^2 \\ &= \left(K_h'\right)^2 * f - f_h'(x)^2 \\ &= h^{-3}f(x)\int K'(u)^2 du + h^{-1}f''(x)\int \frac{u^2}{2}K'(u)^2\,du + O(1), \end{aligned}$$

and so

$$\begin{aligned} var\left(\widehat{f}_h'(x)\right) &= n^{-1}\left\{\left(K_h'\right)^2 * f - f_h'(x)^2\right\} \qquad (16) \\ &= n^{-1}h^{-3}f(x)\int K'(u)^2 du + n^{-1}h^{-1}f''(x)\int \frac{u^2}{2}K'(u)^2\,du + O(n^{-1}). \qquad (17) \end{aligned}$$

These calculations show that it is natural to estimate the variance $\sigma_{x,h}^2$ in (1), using

$$\widehat{\sigma_{x,h}^2} = n^{-1}\left\{n^{-1}\sum_{i=1}^{n}K_h'(x-X_i)^2 - \widehat{f}_h'(x)^2\right\}. \qquad (18)$$

A useful observation is that $\widehat{f}_h'(x)$ and $\widehat{f}_h(x)$ are "nearly independent", which comes from:

$$\begin{aligned} cov\left(\widehat{f}_h'(x), \widehat{f}_h(x)\right) &= cov\left(n^{-1}\sum_{i=1}^{n}K_h'(x-X_i), n^{-1}\sum_{i'=1}^{n}K_h'(x-X_{i'})\right) \\ &= n^{-2}\sum_{i=1}^{n}cov\left(K_h'(x-X), K_h(x-X)\right) + 0 \\ &= n^{-1}cov\left(K_h'(x-X), K_h(x-X)\right). \end{aligned}$$

But

$$\begin{aligned} E\left(K_h'(x-X)\cdot K_h(x-X)\right) &= \int K_h'(x-t)\,K_h(x-t)\,f(t)dt \\ &= \left(K_h'K_h\right) * f \\ &= h^{-2}\int K'(u)K(u)f(x-hu)\,du \\ &= h^{-2}\left[f(x)\int K'(u)K(u)du + O\left(h^2\right)\right] \\ &= O(1), \end{aligned}$$

since $\int K'(u)K(u)du = \left[\frac{1}{2}K^2(u)\right]_{-\infty}^{\infty} = 0$, from which it follows that (as $h \to 0$),

$$\begin{aligned} cov\left(K_h'(x-X), K_h(x-X)\right) &= \left(K_h'K_h\right) * f - f_h'(x)\,f_h(x) \\ &= E\left\{K_h'(x-X)\cdot K_h(x-X)\right\} - \left\{EK_h'(x-X)\cdot EK_h(x-X)\right\} \\ &= O(1), \end{aligned}$$

and so

$$cov\left(\widehat{f}_h'(x), \widehat{f}_h(x)\right) = n^{-1}\left\{(K_h' K_h) * f - f_h'(x)\, f_h(x)\right\} \quad (19)$$

$$= O(n^{-1}). \quad (20)$$

Finally all of the above can be combined to give the estimate of $\widetilde{\sigma^2_{x,h}}$ from (6),

$$\widehat{\widetilde{\sigma^2_{x,h}}} = \frac{1}{n f_h(x)^2}\left[f_h(x)^2\left\{n^{-1}\sum_{i=1}^n K_h'(x-X_i)^2 * f - f_h'(x)^2\right\}\right. \quad (21)$$

$$+f_h'(x)^2\left\{n^{-1}\sum_{i=1}^n K_h(x-X_i)^2 * f - f_h(x)^2\right\}$$

$$\left.-2f_h(x)f_h'(x)\left\{\left(n^{-1}\sum_{i=1}^n K_h'(x-X_i)\,K_h(x-X_i)\right) * f - f_h'(x)\, f_h(x)\right\}\right]$$

4 Appendix 2: detailed calculations for regression

Details are given here, for the case of local linear regression. This setting assumes pairs of data $(X_1, Y_1), ..., (X_n, Y_n)$ independent and identically distributed from some joint probability density, with regression function $f(x) = E(Y_i | X_i = x)$ assumed to have a bounded 3rd derivative. The local linear estimator of $f(x)$, using bandwidth h, is

$$\widehat{f}_h(x) = \arg\min_a \sum_{i=1}^n \left[Y_i - (a + b(X_i - x))\right]^2 K_h(x - X_i), \quad (22)$$

where $\arg\min_a$ is interpreted to mean "minimize jointly over a and b, but use the a value". The main idea is that for each x, a line is fitted to the data, using K_h - weighted least squares, i.e. the data "in a Gaussian window." See e.g. Wand and Jones (1995) and Fan and Gijbels (1996) for further references and for useful discussion of many properties of this smoothing method. See Fan and Marron (1994) for discussion of a fast binned implementation of the local linear smoother.

The overall structure of the log SiZer variance calculation is similar, but the details are quite different. The first important difference is that the estimate of the derivative is no longer the derivative of the estimate, because quotient rule differentiation gets very messy. Instead it is more natural to use the slope of the local linear fit

$$\widehat{f}_h'(x) = \arg\min_b \sum_{i=1}^n \left[Y_i - (a + b(X_i - x))\right]^2 K_h(x - X_i).$$

which is very similar to (22), except the slope is kept instead of the intercept.

The second important difference is that the previous variances are now viewed as conditional on $X_1, ..., X_n$. In particular,

$$\begin{aligned}\sigma^2_{x,h} &= var\left(\widehat{f}'_h(x)|X_1, ..., X_n\right) \\ &= var\left(\sum_{i=1}^n W'_h(x, X_i)Y_i|X_1, ..., X_n\right) \\ &= \sum_{i=1}^n \sigma^2\left(Y_i|X_i\right)\left(W'_h(x, X_i)\right)^2,\end{aligned}$$

using the fact that the local linear smoother can be written as a local average

$$\widehat{f}'_h(x) = \sum_{i=1}^n W'_h(x, X_i)Y_i,$$

where the symbol $'$ appearing in the notation of the weights $W'_h(x, X_i)$ is only intended to convey that these are the weights needed to estimate the local slope,

$$W'_h(x, X_i) = \frac{\left\{S_{0,h}(x)\left(X_i - x\right) - S_{1,h}(x)\right\}K_h\left(X_i - x\right)}{S_{2,h}(x)S_{0,h}(x) - S_{1,h}(x)^2},$$

using the notation, for $j = 0, 1, ...$

$$S_{1,h}(x) = \sum_{i=1}^n \left(X_i - x\right)^j K_h\left(X_i - x\right).$$

A natural estimate of $\sigma^2\left(Y_i|X_i\right)$ is based on a simple smooth of the residuals, e.g.

$$\widehat{\sigma}^2(Y|X = x) = \frac{\sum_{i=1}^n \widehat{e}_i^2 K_h(x - X_i)}{\sum_{i=1}^n K_h(x - X_i)}.$$

where $\widehat{e}_i = Y_i - \widehat{f}_h(X_i)$. To give a fast binned implementation, $\widehat{\sigma}^2(Y|X = x)$ is calculated as

$$\widehat{\sigma}^2(Y|X = x) \approx \left(1 - \widehat{\rho}(x)^2\right)\widehat{\sigma}(x)^2,$$

where

$$\begin{aligned}\widehat{\sigma}(x)^2 &= \tfrac{U_{0,h}(x)}{S_{0,h}(x)} - \left(\tfrac{T_{0,h}(x)}{S_{0,h}(x)}\right)^2, \\ \widehat{\rho}(x)^2 &= \left(\widehat{f}'_h(x)\right)^2\left(\tfrac{S_{0,h}(x)S_{2,h}(x) - S_{1,h}(x)^2}{\widehat{\sigma}(x)^2 S_{0,h}(x)^2}\right),\end{aligned} \tag{23}$$

using the notation, for $j = 0, 1, ...$

$$\begin{aligned}T_{j,h}(x) &= \sum_{i=1}^n \left(X_i - x\right)^j K_h\left(X_i - x\right) Y_i \\ U_{0.h}(x_j) &= \sum_{i=1}^n K_h\left(X_i - x\right) Y_i^2,\end{aligned}$$

For LogSiZer, we also need to work with

$$\begin{aligned} var\left(\widehat{f}_h(x)|X_1,...,X_n\right) &= var\left(\sum_{i=1}^{n} W_h(x,X_i)Y_i|X_1,...,X_n\right) \\ &= \sum_{i=1}^{n} \sigma^2\left(Y_i|X_i\right)\left(W_h(x,X_i)\right)^2, \end{aligned}$$

using the fact that the local linear smoother can be written as a local average

$$\widehat{f}_h(x) = \sum_{i=1}^{n} W_h(x,X_i)Y_i,$$

where

$$W_h(x,X_i) = \frac{\left\{S_{2,h}(x) - S_{1,h}(x)\left(X_i - x\right)\right\} K_h\left(X_i - x\right)}{S_{2,h}(x)S_{0,h}(x) - S_{1,h}(x)^2}.$$

Finally we must also consider

$$\begin{aligned} cov\left(\widehat{f}_h'(x), \widehat{f}_h(x)|X_1,...,X_n\right) &= cov\left(\sum_{i=1}^{n} W_h'(x,X_i)Y_i, \sum_{i=1}^{n} W_h(x,X_i)Y_i|X_1,...,X_n\right) \\ &= \sum_{i=1}^{n} \sigma^2\left(Y_i|X_i\right) W_h'(x,X_i)W_h(x,X_i). \end{aligned}$$

Fast binned implementations, of the type detailed in Section 3.1 of Chaudhuri and Marron (1999), are used everywhere for computational speed.

References

[1] Chaudhuri, P. and Marron, J. S. (1999) SiZer for exploration of structures in curves, *Journal of the American Statistical Association*, 94, 807-823.

[2] Fan, J. and Gijbels, I. (1996) *Local polynomial modeling and its applications*, Chapman and Hall, London.

[3] Fan, J. and Marron, J. S. (1994) Fast implementations of nonparametric curve estimators, *Journal of Computational and Graphical Statistics*, 3, 35-56.

[4] Scott, D. W. (1992) *Multivariate density estimation, theory, practice and visualization*, John Wiley: New York.

[5] Wand, M. P. and Jones, M. C. (1995) *Kernel Smoothing*, Chapman and Hall: London.