Estimating the variance of multiplicative noise

Dafydd Evans

School of Computer Science, Cardiff University, 5 The Parade, Cardiff CF24 3AA, Wales, UK

Abstract. When constructing non-parametric models from noisy data, it is useful to have information regarding the statistical properties of the noise distribution. In many cases, such information is not explicitly available, and must be estimated directly from the data. Under the hypothesis of *additive* noise, algorithms for estimating the variance of the noise distribution have appeared in the literature. In this paper we present a novel algorithm for estimating the noise variance under a *multiplicative* hypothesis.

Keywords: nonlinear modelling, multiplicative noise, near-neighbours
PACS: 02.50.Cw, 02.60.Gf, 02.70.Rr, 05.45.Tp

INTRODUCTION

Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a smooth function, mapping an input $x \in \mathbb{R}^m$ to an output $y \in \mathbb{R}^n$. The goal of statistical modelling is to estimate the function f from a finite set of observations $S = \{(x_1, y_1), \ldots, (x_M, y_M)\}$, a task made considerably more difficult by the presence of noise in the data. In this paper, *noise* is defined to be that part of the output which cannot be accounted for by any smooth transformation of the input. Noise is usually represented by a *random variable*, and simply knowing its *variance* σ^2 is important for many function estimation techniques. For example, σ^2 is used to define a stopping criterion for neural network training [1], and also to determine a threshold for wavelet de-noising [2]. Noise is invariably assumed to be *additive*:

$$y = f(x) + r \qquad \text{where} \qquad E(r) = 0 \tag{1}$$

The Gamma test [3] is a non-parametric algorithm for estimating the variance of additive noise, using only the available data (x_i, y_i). A useful overview of the method can be found in [4]. In [5], the estimate computed by the algorithm is shown to be (weakly) consistent as the number of data points increases. In this paper, we present a related algorithm for estimating the variance of *multiplicative* noise:

$$y = f(x)r \qquad \text{where} \qquad E(r) = 1 \tag{2}$$

NOISE ESTIMATION

To illustrate the ideas underpinning our algorithm, we first describe a new algorithm for the estimating the variance of additive noise. Our algorithms exploit the *near-neigbour* structure of the input points $x_1, \ldots, x_M$, which may be computed in time $O(M \log M)$ using *kd*-trees [6]. For any input point x, choose x' and x'' to be any two points from among its first p nearest neighbours in the set $\{x_1, \ldots, x_M\}$; let y' and y'' be the outputs corresponding to x' and x'', and let r' and r'' represent the noise measured on the outputs y' and y'' respectively. We assume that the noise values r, r' and r'' are independent and identically distributed, and also independent of the associated input points x, x' and x''; the input points are also assumed to be independent and identically distributed.

Our algorithm for additive noise is based on the product of *differences* $(y-y')(y-y'')$. Starting with (1) and applying the above conditions, it is easily shown that

$$E\left((y-y')(y-y'')\right) = \sigma^2 + E\left((f(x)-f(x'))(f(x)-f(x''))\right) \tag{3}$$

where σ^2 is the variance of the noise distribution, and the expected value on the right is taken with respect to the distribution of the input points. We think of $E\left((y-y')(y-y'')\right)$ as an estimate of σ^2, while $E\left((f(x)-f(x'))(f(x)-f(x''))\right)$ quantifies the associated estimation error. Because f is approximately locally linear[1] at x, this error term satisfies

$$E\left((f(x)-f(x'))(f(x)-f(x''))\right) \approx G_f E\left(|x-x'||x-x''|\right) \tag{4}$$

where G_f is related to the expected gradient of f (with respect to the input distribution).

For $k \in \mathbb{N}$, let x_{i_k} denote the kth nearest neighbour of x_i among the input points $x_1,\dots,x_M$, and let y_{i_k} denote the output corresponding to x_{i_k}. For every $k \in \{1,\dots,p\}$, we estimate the expected values $E\left((y-y')(y-y'')\right)$ and $E\left(|x-x'||x-x''|\right)$ by the sample means

$$\Gamma_k = \frac{1}{M}\sum_{i=1}^{M}(y_i-y_{i_k})(y_i-y_{i_{k+1}}) \quad \text{and} \quad \Delta_k = \frac{1}{M}\sum_{i=1}^{M}|x_i-x_{i_k}||x_i-x_{i_{k+1}}| \tag{5}$$

respectively. In [5], is shown that sample means such as Γ_k and Δ_k satisfy a weak law of large numbers, in the sense that they converge to their expected values in probability as the number of data points increases. In view of this, we substitute the sample means of (5) for the distribution means of (3) and (4), which yields

$$\Gamma_k \approx \sigma^2 + G_f\Delta_k \tag{6}$$

where we think of G_f as a (finite) *constant*, in the sense that it is independent of any particular realisation of the sample data $\{(x_i,y_i)\}$. Finally, we exploit this (approximate) linear relation between Γ_k and Δ_k, by performing linear regression on the pairs (Δ_k,Γ_k) to estimate the value of Γ_k in the limit as $\Delta_k \to 0$, which provides an estimate for σ^2.

THE ALGORITHM

To estimate the variance of multiplicative noise, rather than look at the product of differences $(y-y')(y-y'')$ we now consider the product of *ratios* $(y/y')(y/y'')$, or

[1] The unknown function f is assumed to be smooth, and is therefore approximately linear in sufficiently small regions around x. Here, the region of interest is the *p-nearest neighbour ball*, centred at x and having the pth nearest neighbour of x on its boundary. Whether this ball is 'sufficiently small' to ensure local linearity depends on the density of the input distribution near x.

equivalently $y^2/y'y''$. Because the noise values r, r' and r'' are independent of the inputs x, x' and x'', starting from (2) it easily follows that

$$\frac{E(y^2)}{E(y'y'')} = (\sigma^2+1)\left(\frac{E(f(x)^2)}{E(f(x')f(x''))}\right) \tag{7}$$

where we have used the fact that $E(r'r'') = E(r')E(r'') = 1$ and $E(r^2) = \sigma^2 - 1$. Furthermore, because f is smooth, by Taylor's theorem

$$f(x)^2 \approx f(x')f(x'') - \big((x-x')f(x'') + (x-x'')f(x')\big)\nabla f(x) \tag{8}$$

Thus by (7), and using the fact that x, x' and x'' are identically distributed,

$$\frac{E(y^2)}{E(y'y'')} \approx (\sigma^2+1) + G_f E\big((x-x') + (x-x'')\big) \tag{9}$$

where

$$G_f = -(\sigma^2+1)E\big(f(x)\nabla f(x)\big)/E(f(x')f(x'')) \tag{10}$$

Because f is smooth, G_f is bounded provided the noise variance σ^2 is finite, and also provided f is bounded (and not identically zero) over the set of possible inputs. Following the discussion leading to (5), for every $k \in \{1,\ldots,p\}$ we estimate the ratio $E(y^2)/E(y'y'')$ and the expected value $E\big((x-x')+(x-x'')\big)$ by the empirical values

$$\Gamma_k = \frac{\sum_{i=1}^M y_i^2}{\sum_{i=1}^M y_{i_k} y_{i_{k+1}}} \quad \text{and} \quad \Delta_k = \frac{1}{M}\sum_{i=1}^M \big((x_i - x_{i_k}) + (x_i - x_{i_{k+1}})\big) \tag{11}$$

respectively. As in (6), we then replace the expected values of (9) by the empirical values of (11), leading to

$$\Gamma_k \approx (\sigma^2+1) + G_f \Delta_k \tag{12}$$

Finally, we compute our estimate for σ^2 by exploiting this approximate linear relation between Γ_k and Δ_k, using simple linear regression to estimate the value of Γ_k in the limit as $\Delta_k \to 0$ (note that because the intercept estimates σ^2+1, we must subtract one from this to get the final estimate). It is interesting to note that the gradient of the regression line can be interpreted as an estimate of G_f, and might therefore represent some useful information regarding the unknown function f.

Our algorithm is explicitly stated as Algorithm 1.

Algorithm 1

---

1. Compute the p-nearest neighbour structure of the input points $\{x_1,\ldots,x_M\}$.
for $k \in \{1,\ldots,p\}$ **do**
    2. Compute the pair (Δ_k, Γ_k) as defined in (11).
end for
3. Perform linear regression on the pairs $\{(\Delta_k,\Gamma_k) : k = 1,\ldots,p\}$.
4. Return the intercept of the regression line with the $\Delta_k = 0$ axis (minus one).

EXPERIMENTAL RESULTS

We generated a set of 2000 points x_i, each selected uniformly at random from the unit interval $[0,1]$, and a set of 2000 noise values r_i, each selected according to a Gaussian distribution of unit mean and variance 0.2. The output points y_i were then constructed according to the rule $y_i = f(x_i)r_i$ where $f(x) = 3\sin(8\pi x) + \cos(23\pi x)$.

The left-hand plot of Figure 1 shows the output points y_i plotted against the input points x_i. The multiplicative nature of the noise is evident here – the noise becomes more pronounced as the distance between the underlying curve $f(x)$ and the $y = 0$ axis increases.

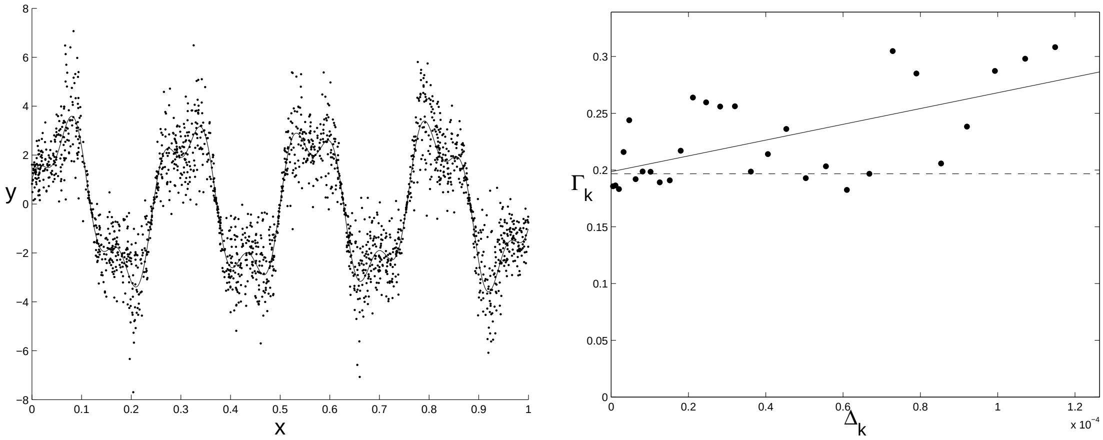

FIGURE 1. The noisy data (x_i, y_i) and the regression plot produced by the algorithm.

The data set $\{(x_i, y_i)\}$ was processed according to Algorithm 1, with $p = 30$. The resulting regression plot is shown on the right of Figure 1, where it can be seen that the intercept of the regression line (solid line) is in close agreement with the variance of the noise (dashed line). It can also be seen that the point estimates Γ_k become increasingly inaccurate as Δ_k increases. Experimental evidence suggests that our algorithm can successfully estimate the variance of multiplicative noise, provided there are sufficient data available. We have observed that more data points are required as the compexity of the function f increases, a fact that will be addressed in a future study.

ACKNOWLEDGMENTS

The author acknowledges the support of the Royal Society (URF 516002.K5736).

REFERENCES

1. S. Haykin, *Neural Networks*, Prentice Hall, 1998.
2. D. L. Donoho, and I. M. Johnstone, *Biometrika*, **81**, 425–455 (1994).
3. A. Stefánsson, N. Končar, and A. J. Jones, *Neural. Comput. Appl.*, **5**, 131–133 (1997).
4. A. J. Jones, *Comp. Manage. Sci.*, **1**, 109–149 (2004).
5. D. Evans, and A. J. Jones, *Proc. R. Soc. Lond. A*, **458**, 2759–2799 (2002).
6. J. L. Bentley, *Comm. ACM*, **18**, 309–517 (1975).