The bbemkr Package

Han Lin Shang
Australian National University

Abstract

The multivariate kernel regression provides a flexible way to estimate possible non-linear relationship between a set of predictors and scalar-valued response. As with any type of kernel regression, it requires an optimal selection of smoothing parameter, called bandwidth. In the literature of multivariate kernel regression, bandwidth parameter is often selected by least square cross validation. In this article, we present a Bayesian bandwidth estimation method that uses the information about error density to help with the optimal selection of bandwidths in the regression function. We first describe the proposed Bayesian method in a multivariate kernel regression. Illustrated by a series of simulation studies, the Bayesian method is then implemented using a readily-available $\mathbb{R}$ add-on package.

Keywords: bandwidth selection, Bayesian model selection, Nadaraya-Watson estimator, kernel-form error density, marginal likelihood, adaptive random-walk Metropolis, simulation inefficiency factor.

1. Introduction

The aim of this article is to describe the $\mathbb{R}$ functions that are readily-available in the **bbemkr** package (Shang and Zhang 2013) for estimating bandwidth parameters in a multivariate nonparametric regression. In the literature of nonparametric regression, many developments focus on the estimation of nonparametric regression function. Some commonly used nonparametric regression estimators include: Nadaraya-Watson (NW) estimator (Bowman and Azzalini 1997), local linear estimator (Simonoff 1996), k-nearest neighbour estimator (Wand and Jones 1995), and many others. Because of simplicity and mathematical elegance, we consider the NW estimator in this paper.

In all of the aforementioned nonparametric estimators, the estimation accuracy of the conditional mean crucially depend on the optimal selection of bandwidths. Commonly, the optimal bandwidths are selected by the least-squares cross validation. Least-squares cross validation aims to minimise L_2 loss function and has the appealing feature that no estimation of error variance is required. However, since residuals affect the estimation accuracy of regression function, least-squares cross validation may select a sub-optimal bandwidth. This in turn leads to inferior estimation accuracy of regression functions. As an alternative, we present a Bayesian bandwidth estimation method that simultaneously estimates the optimal bandwidths in the regression function and kernel-form error density by minimising the generalised loss function.

This article aims to draw close connection between the accurate estimations of error density and regression function. The estimation of error density is important for assessing the goodness of fit of a specific distribution (see for examples, Akritas and Van Keilegom 2001; Cheng and Sun 2008);

the estimation of error density is also useful to test the symmetry of the residual distribution (see for examples, Ahmad and Li 1997; Neumeyer and Dette 2007); the estimation of error density is important to statistical inference, prediction and model validation (see for examples, Efromovich 2005; Muhsal and Neumeyer 2010); and the estimation of error density is also useful for the estimation of the density of the response variable (see for an example, Escanciano and Jacho-Chávez 2012). Therefore, being able to estimate the error density is as important as being able to estimate the regression function.

Before introducing the Bayesian bandwidth estimation method, we first define the problem more precisely. Let $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^\top$ be a vector of scalar responses, and $\boldsymbol{x}_i = (x_{1i}, \ldots, x_{pi})^\top$ for $i = 1, \ldots, n$ be p-dimensional real-valued predictors, where $^\top$ represents matrix transpose. We consider the nonparametric regression model given by

$$y_i = m(\boldsymbol{x}_i) + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \tag{1}$$

where $m(\boldsymbol{x}) = \mathrm{E}(y|\boldsymbol{x})$ is the conditional mean, and $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are independent and identically distributed (iid) errors with an unknown probability density function, denoted as $f(\varepsilon)$. We assume that there is no correlation between the covariates in the regression function and errors, that is

$$\mathrm{E}(\varepsilon_i|\boldsymbol{x}_i) = 0.$$

The flexibility of the nonparametric regression comes from the fact that the unknown regression function $m(\cdot)$ does not need to have a specific parametric functional form. With some smoothness properties, $m(\cdot)$ can be estimated by the kernel estimator, such as the Nadaraya-Watson estimator given by

$$\widehat{m}(\boldsymbol{x}; \boldsymbol{h}) = \frac{\sum_{i=1}^n K_{\boldsymbol{h}}(\boldsymbol{x} - \boldsymbol{x}_i) y_i}{\sum_{i=1}^n K_{\boldsymbol{h}}(\boldsymbol{x} - \boldsymbol{x}_i)},$$

where $\boldsymbol{h} = (h_1, h_2, \ldots, h_p)^\top$ represents a vector of bandwidths.

This article proceeds as follows. The Bayesian bandwidth estimation method is first described and its estimation accuracy is then compared based on the idea of marginal likelihood. Through a series of simulation studies, the sampling algorithm is demonstrated using the $\mathbb{R}$ functions in the **bbemkr** package. Conclusions will then be presented.

2. Bayesian bandwidth estimation

2.1. Estimation of error density

The unknown error density $f(\varepsilon)$ can be approximated by a location-mixture of Gaussian densities given by

$$f(\varepsilon; b) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{b} \phi\left(\frac{\varepsilon - \varepsilon_j}{b}\right), \tag{2}$$

where $\phi(\cdot)$ is the probability density function of the standard Gaussian distribution and the component Gaussian densities have means at ε_j, for $j = 1, 2, \ldots, n$ and a common standard deviation b. Equation (2) is simply a univariate kernel density estimator with Gaussian kernel and

bandwidth b. Although error ε_j is unknown, it can be estimated by the NW estimator. Thus, the density of y_i is approximated by the estimated error density $\widehat{f}(\varepsilon; b)$, expressed as

$$\widehat{f}(\varepsilon; b) = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{b}\phi\left(\frac{\varepsilon - \widehat{\varepsilon}_j}{b}\right),$$

where b represents residual bandwidth. In order to avoid the possible selection of $b = 0$, a leave-one-out version of the kernel likelihood is often used, given by

$$\widehat{f}(\widehat{\varepsilon}_i; b) = \frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{1}{b}\phi\left(\frac{\widehat{\varepsilon}_i - \widehat{\varepsilon}_j}{b}\right),$$

where $\widehat{\varepsilon}_i = y_i - \widehat{m}(\boldsymbol{x}_i; h)$ is the ith residual for $i = 1, 2, \ldots, n$, in the multivariate nonparametric regression. Given (h, b) and iid assumption of the errors, the leave-one-out version of the kernel likelihood of $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^\top$ can be approximated by

$$\widehat{L}(\boldsymbol{y}|h, b) = \prod_{i=1}^{n}\left[\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{1}{b}\phi\left(\frac{\widehat{\varepsilon}_i - \widehat{\varepsilon}_j}{b}\right)\right].$$

2.2. Prior density

We now discuss the issue of prior density for the bandwidths. Let $\pi(h^2)$ and $\pi(b^2)$ be the independent prior of squared bandwidths h and b. Since h^2 and b^2 play the role of variance parameters in the Gaussian densities, we assume that the priors of h^2 and b^2 are inverse Gamma density, denoted by $\text{IG}(\alpha_h, \beta_h)$ and $\text{IG}(\alpha_b, \beta_b)$, respectively. Thus, the prior densities of h^2 and b^2 are given by

$$\pi(h^2) = \frac{(\beta_h)^{\alpha_h}}{\Gamma(\alpha_h)}\left(\frac{1}{h^2}\right)^{\alpha_h+1}\exp\left(-\frac{\beta_h}{h^2}\right),$$
$$\pi(b^2) = \frac{(\beta_b)^{\alpha_b}}{\Gamma(\alpha_b)}\left(\frac{1}{b^2}\right)^{\alpha_b+1}\exp\left(-\frac{\beta_b}{b^2}\right),$$

where $\alpha_h = \alpha_b = 1.0$ and $\beta_h = \beta_b = 0.05$ are hyper-parameters. Sensitivity results studied in Zhang, King, and Shang (2011) show that the choices of hyper-parameters and inverse Gamma densities do not influence the estimation of posterior density.

2.3. Posterior density

Let $\boldsymbol{\theta} = (h^2, b^2)$ be the parameter vector and $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^\top$ be the data. According to the Bayes theorem, the posterior of $\boldsymbol{\theta}$ is written by

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{\widehat{L}(\boldsymbol{y}|\boldsymbol{\theta})}{L(\boldsymbol{y})}, \tag{3}$$

where $\widehat{L}(\boldsymbol{y}|\boldsymbol{\theta})$ is the approximated likelihood function with squared bandwidths and $L(\boldsymbol{y})$ is the marginal likelihood, which can be expressed as

$$\int \widehat{L}(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}.$$

In practice, the posterior in (3) can be approximated by (up to a normalising constant):

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \widehat{L}(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

We use the adaptive random-walk Metropolis algorithm to sample $\boldsymbol{\theta}$ (see Garthwaite, Fan, and Sisson 2010, for details). In order to assess the convergence of the Markov chain Monte Carlo (MCMC) algorithm, we use the notion of simulation inefficiency factor (Meyer and Yu 2000). This is a measure of autocorrelation among iterations and provides an indication of how many iterations are required to have the iid draws from the posterior distributions. It is noteworthy that a full range of diagnostic tools in the coda package (Plummer, Best, Cowles, and Vines 2006) can also be applied to check the convergence of MCMC.

2.4. Adaptive estimation of error density

In kernel density estimator, it has been noted that the leave-one-out estimator may be heavily affected by extreme observations in the data sample (see for example, Bowman 1984). Because of the use of a global bandwidth, the leave-one-out kernel error density estimator is likely to overestimate the tails of the density. To overcome this deficiency, it is possible to use localised bandwidths by assigning small bandwidths to the residuals in the high density region and large bandwidths to the residuals in the low density region. The localised error density estimator can be given by

$$\widehat{f}(\widehat{\varepsilon}_i; b, \tau_\varepsilon) = \frac{1}{n-1}\sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{1}{b(1+\tau_\varepsilon|\widehat{\varepsilon}_j|)}\phi\left(\frac{\widehat{\varepsilon}_i - \widehat{\varepsilon}_j}{b(1+\tau_\varepsilon|\widehat{\varepsilon}_j|)}\right),$$

where $b(1+\tau_\varepsilon|\widehat{\varepsilon}_j|)$ is the bandwidth assigned to residual $\widehat{\varepsilon}_j$ and the vector of parameter is now (h, b, τ_ε). Again, the adaptive random-walk Metropolis algorithm can be used to sample these parameters, where the prior density of $\tau_\varepsilon \sim U(0,1)$.

2.5. Sampling algorithm

We use the adaptive random-walk Metropolis algorithm of Garthwaite *et al.* (2010) to sample (h^2, b^2), the sampling algorithm is briefly described below. For simplicity of notation, I shall let $\boldsymbol{\theta} = (h^2, b^2)$ to represent a vector of squared bandwidths.

Step 0 Specify a Gaussian proposal distribution, with an arbitrary starting point h^2 and b^2. The starting points can be drawn from a uniform distribution $U(0,1)$.

Step 1 At the kth iteration, the current state $b^2_{(k)}$ is updated as $b^2_{(k)} = b^2_{(k-1)} + \tau_{(k-1)}\varepsilon$, where $\varepsilon \sim N(0,1)$, and $\tau_{(k-1)}$ is an adaptive tuning parameter with an arbitrary initial value $\tau_{(0)}$.

Step 2 The updated $b^2_{(k)}$ is accepted with probability $\min\left\{\frac{\pi\left(b^2_{(k)},h^2_{(k-1)}|\boldsymbol{y}\right)}{\pi\left(b^2_{(k-1)},h^2_{(k-1)}|\boldsymbol{y}\right)},1\right\}$, where π represents the posterior density.

Step 3 By using the stochastic search algorithm of Robbins and Monro (1951), the tuning parameter is

$$\tau_{(k)} = \begin{cases} \tau_{(k-1)} + c(1-p)/k & \text{if } b^2_{(k)} \text{ is accepted;} \\ \tau_{(k-1)} - cp/k & \text{if } b^2_{(k)} \text{ is rejected.} \end{cases}$$

where $c = \frac{\tau_{(k-1)}}{p(1-p)}$ is a varying constant, and $p = 0.44$ is the optimal acceptance probability for drawing one parameter while $p = 0.234$ is the optimal acceptance probability for drawing multiple parameters (Roberts and Rosenthal 2009).

Step 4 Repeat Steps 1-3 for $h^2_{(k)}$, conditional on $b^2_{(k)}$ and $\boldsymbol{y}$.

Step 5 Repeat Steps 1-4 for $M+N$ times, discard $\left(h^2_{(0)},b^2_{(0)}\right),\left(h^2_{(1)},b^2_{(1)}\right),\ldots,\left(h^2_{(M)},b^2_{(M)}\right)$ for burn-in in order to let the effects of the transients wear off, estimate $\widehat{h}^2 = \frac{\sum_{k=M+1}^{M+N} h^2_{(k)}}{N}$ and $\widehat{b}^2 = \frac{\sum_{k=M+1}^{M+N} b^2_{(k)}}{N}$. The burn-in period is taken to be $M = 1,000$ iterations, and the number of iterations after burn-in period is $N = 10,000$ iterations. The analytical form of the kernel-form error density can be derived based on h^2 and b^2. It is noteworthy that a similar error density result can be obtained by taking the average of the kernel-form error densities computed at all iterations, but at the cost of much slower computational speed.

2.6. Bayesian model selection

How could the Bayesian model selection be useful in multivariate kernel regression? The answer lies in the comparison of different error-density assumptions. Under the normal and student-t error densities, marginal likelihood can be used to compare the kernel-form error density and assumed Gaussian error density.

In Bayesian inference, model selection or averaging is calculated through the Bayes factor of the model of interest against a competing model. The Bayes factor reflects a summary of evidence provided by the data supporting the model as opposed to its competing model. The Bayes factor is defined as the expectation of likelihood with respect to the prior of parameters. It is seldom computed as the integral of the product of the likelihood and prior of parameters, but instead is often computed numerically (Gelfand and Dey 1994; Newton and Raftery 1994; Chib 1995; Kass and Raftery 1995; Geweke 1999).

Chib (1995) showed that the marginal likelihood under error-density assumption A is expressed as

$$L_{\mathrm{A}}(\boldsymbol{y}) = \frac{\widehat{L}_{\mathrm{A}}(\boldsymbol{y}|\boldsymbol{\theta})\pi_{\mathrm{A}}(\boldsymbol{\theta})}{\pi_{\mathrm{A}}(\boldsymbol{\theta}|\boldsymbol{y})},$$

where $\widehat{L}_{\mathrm{A}}(\boldsymbol{y}|\boldsymbol{\theta})$, $\pi_{\mathrm{A}}(\boldsymbol{\theta})$ and $\pi_{\mathrm{A}}(\boldsymbol{\theta}|\boldsymbol{y})$ denote the kernel likelihood, prior and posterior under error-density assumption A, respectively. $L_{\mathrm{A}}(\boldsymbol{y})$ is often computed at the posterior estimate of $\boldsymbol{\theta}$. The numerator has a closed form and can be computed analytically, but the denominator can be

approximated by the MCMC posterior draws. However, for relatively small number of parameters, the denominator can be estimated by its kernel density estimator based on the simulated chain of $\boldsymbol{\theta}$ through a posterior sampler.

The Bayes factor of error-density assumption A against error-density assumption B is defined as

$$\frac{\mathrm{L_A}(\boldsymbol{y})}{\mathrm{L_B}(\boldsymbol{y})}.$$

Based on the Bayes factor, we can determine which model is more superior than a competing model with different levels of evidence (see Kass and Raftery 1995, for more details).

3. Simulation study

Consider the relationship between y and $\boldsymbol{x} = (x_1, x_2, x_3)^\top$ given by

$$y_i = \sin(2\pi x_{1,i}) + 4(1 - x_{2,i})(1 + x_{2,i}) + \frac{2x_{3,i}}{1 + 0.8x_{3,i}^2} + \varepsilon_i, \tag{4}$$

for $i = 1, 2, \ldots, n$. A sample of 100 observations was generated by drawing $x_{1,i}, x_{2,i}$ and $x_{3,i}$ independently from $U(0, 1)$, and ε_i from either a normal distribution $N\left(0, 0.9^2\right)$ or the student distribution with four degrees of freedom t_4.

The smooth and non-linear relationship between y_i and $\boldsymbol{x}_i$ can be modelled by the nonparametric regression, where bandwidths are estimated through the Bayesian sampling algorithm under the two error-density assumptions. When the true error density is Gaussian, we find that the Bayesian sampling algorithm with the Gaussian error density performs better than the one with kernel-form error density, as measured by the marginal likelihood using Geweke's method. This can be obtained as follows:

```
# install and load the R package
R> install.packages("bbemkr")
R> require(bbemkr)

# set random seed
R> set.seed(123456)

# initial bandwidths obtained from the normal reference rule

x = log(nrr(data_x = data_x, logband = FALSE)^2)

# Initial cost function

inicost = cost_gaussian(x = x, data_x = data_x, data_y = data_ynorm, prior_p = 2, prior_st = 1)

# Burn-in period (error density is Gaussian)

warmup_res = warmup_gaussian(x = x, inicost = inicost, mutsizp = 1.0, data_x = data_x,
                             data_y = data_ynorm, warm = 1000)
```

```
# MCMC recording period (error density is Gaussian)

mcmc_res = mcmcrecord_gaussian(x = warmup_res$x, inicost = warmup_res$cost,
                               mutsizp = warmup_res$mutsizplast, data_x = data_x,
                               data_y = data_ynorm, xm = xm, warm = 1000, M = 1000)

# initial bandwidths obtained from the normal reference rule

x = c(log(nrr(data_x = data_x, logband = FALSE)^2),2)

# Initial cost function

nicost = cost_admkr(x = x, data_x = data_x, data_y = data_ynorm)

# Burn-in period (error density is the kernel-form)

warmup_res_admkr = warmup_admkr(x = x, inicost = inicost, mutsizp = 1.0, errorsizp = 1.0,
                                data_x = data_x, data_y = data_ynorm, warm=1000)

# MCMC recording period (error density is the kernel-form)

mcmc_res_admkr = mcmcrecord_admkr(x = warmup_res_admkr$x, inicost = warmup_res_admkr$cost,
                  mutsizp = warmup_res_admkr$mutsizp, errorsizp = warmup_res_admkr$errorsizp,
                  data_x = data_x, data_y = data_ynorm, xm = xm, warm = 1000, M = 1000)

# marginal likelihoods for both error-density assumptions

round(mcmc_res$marginalike, 2)
round(mcmc_res_admkr$marginalike, 2)
```

When the error density is t_4, we find that the Bayesian sampling algorithm with the kernel-from error density performs better than the one with Gaussian error density, as measured by the marginal likelihoods using Chib's and Geweke's methods. This can be obtained as follows:

```
# initial bandwidths obtained from the normal reference rule

x = log(nrr(data_x = data_x, logband = FALSE)^2)

# Initial cost function

inicost = cost_gaussian(x = x, data_x = data_x, data_y = data_yt, prior_p = 2, prior_st = 1)

# set random seed

set.seed(123456)

# Burn-in period (error density is Gaussian)

warmup_res = warmup_gaussian(x = x, inicost = inicost, mutsizp = 1.0, data_x = data_x,
                             data_y = data_yt, warm = 1000)
```

```
# MCMC recording period (error density is Gaussian)

mcmc_res = mcmcrecord_gaussian(x = warmup_res$x, inicost = warmup_res$cost,
                      mutsizp = warmup_res$mutsizplast, data_x = data_x,
                      data_y = data_yt, xm = xm, warm = 1000, M = 1000)

# initial bandwidths obtained from the normal reference rule

x = c(log(nrr(data_x = data_x, logband = FALSE)^2), 2)

# Initial cost function

inicost = cost_admkr(x = x, data_x = data_x, data_y = data_yt)

# set random seed

set.seed(123456)

# Burn-in period (error density is the kernel-form)

warmup_res_admkr = warmup_admkr(x = x, inicost = inicost, mutsizp = 1.0, errorsizp = 1.0,
                                data_x = data_x, data_y = data_yt, warm = 1000)

# MCMC recording period (error density is the kernel-form)

mcmc_res_admkr = mcmcrecord_admkr(x = warmup_res_admkr$x, inicost = warmup_res_admkr$cost,
                        mutsizp = warmup_res_admkr$mutsizp,
                        errorsizp = warmup_res_admkr$errorsizp,
                        data_x = data_x, data_y = data_yt, xm = xm, warm = 1000, M = 1000)

# marginal likelihoods for both error-density assumptions

round(mcmc_res$marginalike, 2)
round(mcmc_res_admkr$marginalike, 2)
```

4. Conclusion

This article describes the Bayesian bandwidth estimation method in a multivariate nonparametric regression, using the $\mathbb{R}$ functions that are readily-available in the **bbemkr** package. The method allows us to simultaneously estimate optimal bandwidths in the regression function approximated by the NW estimator and kernel-form error density. Illustrated by a series of simulation studies, we found that when the error density is correctly specified, the Bayesian method with kernel-form error density is sub-optimal; when the error density is wrongly specified, the proposed method performs the best. In practice, given the error density is often unknown, the proposed method provides a robust approach towards bandwidth estimation.

In future, the Bayesian method described may be extended to other nonparametric estimators for estimating regression function, such as local linear estimator (Simonoff 1996).

References

Ahmad I, Li Q (1997). "Testing symmetry of an unknown density function by kernel method." *Journal of Nonparametric Statistics*, **7**(3), 279–293.

Akritas MG, Van Keilegom I (2001). "Non-parametric estimation of the residual distribution." *Scandinavian Journal of Statistics*, **28**(3), 549–567.

Bowman AW (1984). "An alternative method of cross-validation for the smoothing of density estimates." *Biometrika*, **71**(2), 353–360.

Bowman AW, Azzalini A (1997). *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-PLUS Illustrations*. Oxford University Press, New York.

Cheng F, Sun S (2008). "A goodness-of-fit test of the errors in nonlinear autoregressive time series models." *Statistics and Probability Letters*, **78**(1), 50–59.

Chib S (1995). "Marginal likelihood from the Gibbs output." *Journal of the American Statistical Association*, **90**(432), 1313–1321.

Efromovich S (2005). "Estimation of the density of regression errors." *The Annals of Statistics*, **33**(5), 2194–2227.

Escanciano JC, Jacho-Chávez DT (2012). "$\sqrt{n}$ uniformly consistent density estimation in nonparametric regression models." *Journal of Econometrics*, **167**(2), 305–316.

Garthwaite PH, Fan Y, Sisson SA (2010). "Adaptive optimal scaling of Metropolis-Hastings algorithms using the Robbins-Monro process." *Working paper*, University of New South Wales. URL http://arxiv.org/pdf/1006.3690v1.pdf.

Gelfand AE, Dey DK (1994). "Bayesian model choice: Asymptotics and exact calculations." *Journal of the Royal Statistical Society, Series B*, **56**(3), 501–514.

Geweke JF (1999). "Using simulation methods for Bayesian econometric models: Inference, development and communication." *Econometric Review*, **18**(1), 1–73.

Kass RE, Raftery AE (1995). "Bayes factors." *Journal of the American Statistical Association*, **90**(430), 773–795.

Meyer R, Yu J (2000). "BUGS for a Bayesian analysis of stochastic volatility models." *Econometrics Journal*, **3**(2), 198–215.

Muhsal B, Neumeyer N (2010). "A note on residual-based empirical likelihood kernel density estimation." *Electronic Journal of Statistics*, **4**, 1386–1401. ISSN 1935-7524.

Neumeyer N, Dette H (2007). "Testing for symmetric error distribution in nonparametric regression models." *Statistica Sinica*, **17**(2), 775–795.

Newton MA, Raftery AE (1994). "Approximate Bayesian inference with the weighted likelihood bootstrap." *Journal of the Royal Statistical Society, Series B*, **56**(1), 3–48.

Plummer M, Best N, Cowles K, Vines K (2006). "CODA: Convergence diagnosis and output analysis for MCMC." *R News*, **6**(1), 7–11.

Robbins H, Monro S (1951). "A stochastic approximation method." *The Annals of Mathematical Statistics*, **22**(3), 327–495.

Roberts GO, Rosenthal JS (2009). "Examples of adaptive MCMC." *Journal of Computational and Graphical Statistics*, **18**(2), 349–367.

Shang HL, Zhang X (2013). *bbemkr: Bayesian bandwidth estimation for multivariate kernel regression with Gaussian error*. R package version 1.6, URL http://CRAN.R-project.org/package=bbemkr.

Simonoff J (1996). *Smoothing Methods in Statistics*. Springer, New York.

Wand MP, Jones MC (1995). *Kernel Smoothing*. Chapman & Hall, London.

Zhang X, King ML, Shang HL (2011). "Bayesian estimation of bandwidths for a nonparametric regression model with a flexible error density." *Working paper 10*, Monash University. URL http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2011/wp10-11.pdf.

Affiliation:

Han Lin Shang
Research School of Finance, Actuarial Studies and Applied Statistics
Australian National University
Canberra ACT 0200, Australia
E-mail: hanlin.shang@anu.edu.au
URL: https://sites.google.com/site/hanlinshangswebsite/